Redbooks
ibm.com/redbooks

Implementation Guide for IBM Storage FlashSystem and IBM SAN Volume Controller

Updated for IBM Storage Virtualize Version 8.6

Tiago Moreira Candelaria Bastos
Nezih Boyacioglu
Andrew Greenfield
Byron M Grossnickle
Vasfi Gucer
Jon Herd
Mert Korcum
Sergey Kubin
Youssef Largou
Carsten Larsen
Guillaume Legmar
Hartmut Lonzer
Corne Lottering
John Nycz
Denis Olshanskiy
Uygar Sahin
Uwe Schreiber
Christian Schroeder



Storage

IBM

Redbooks



IBM Redbooks

Implementation Guide for IBM Storage FlashSystem and IBM SAN Volume Controller

May 2024

SG24-8542-00

Note: Before using this information and the product it supports, read the information in “Notices” on page xlvii.

First Edition (May 2024)

This edition applies to IBM Storage Virtualize Version 8.6.

© Copyright International Business Machines Corporation 2024. All rights reserved.
Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Figures xvii

Tables xxxix

Examples xli

Notices xlvii
Trademarks xlviii

Preface xlix
Authors l
Now you can become a published author, too! liv
Comments welcome lv
Stay connected to IBM Redbooks lv

Chapter 1. Introduction and system overview 1
1.1 IBM Storage Virtualize 2
1.1.1 Storage virtualization terminology 4
1.2 IBM SAN Volume Controller architectural overview 8
1.2.1 Controller-based approach 9
1.2.2 SAN or fabric-based appliance solution 9
1.2.3 IBM SAN Volume Controller 10
1.2.4 IBM SAN Volume Controller topology 11
1.3 Changes and enhancements in version 8.6.0 13
1.3.1 IBM Storage Virtualize V8.6.0 13
1.3.2 IBM Storage Virtualize V8.5.0 15
1.3.3 IBM Storage Virtualize V8.4.2 18
1.4 IBM SAN Volume Controller family 23
1.4.1 Components 24
1.4.2 Nodes 24
1.4.3 I/O groups 25
1.4.4 System 25
1.4.5 MDisks 26
1.4.6 Cache 27
1.4.7 Quorum disk 30
1.4.8 Storage pool 32
1.4.9 Volumes 33
1.4.10 Easy Tier 35
1.4.11 Hosts 36
1.4.12 Array 37
1.4.13 Encryption 37
1.4.14 iSCSI and iSCSI Extensions over RDMA 38
1.4.15 Data reduction pools 39
1.4.16 IP replication 39
1.4.17 IBM Storage Virtualize copy services 40
1.4.18 Synchronous or asynchronous remote copy 40
1.4.19 Policy-based replication 41
1.4.20 FlashCopy and Transparent Cloud Tiering 42
1.4.21 IBM HyperSwap 43

© Copyright IBM Corp. 2024.

1.5 IBM SAN Volume Controller models 43
1.5.1 IBM SAN Volume Controller SV3 44
1.5.2 IBM SAN Volume Controller SV2 and SA2 45
1.5.3 IBM SAN Volume Controller model comparisons 47
1.6 IBM FlashSystem family 48
1.6.1 Storage Expert Care 49
1.7 IBM FlashSystem 9500 overview 53
1.7.1 IBM FlashSystem 9500 hardware components 53
1.7.2 IBM FlashSystem 9500 control enclosure 55
1.7.3 IBM FlashSystem 9000 Expansion Enclosure Models AFF and A9F 58
1.8 IBM FlashSystem 9500R Rack Solution overview 59
1.8.1 IBM FlashSystem 9500R Rack Solution diagram 60
1.8.2 FC cabling and clustering 62
1.9 IBM FlashSystem 9000 Expansion Enclosure Models AFF and A9F 62
1.10 IBM FlashSystem 7300 overview 66
1.10.1 IBM FlashSystem 7300 control enclosures 67
1.10.2 IBM FlashSystem 7000 Expansion Enclosures 68
1.11 IBM FlashSystem 5200 overview 71
1.11.1 IBM FlashSystem 5200 expansion enclosures 74
1.12 IBM FlashSystem 5000 family overview 75
1.12.1 IBM FlashSystem 5015 77
1.12.2 IBM FlashSystem 5035 78
1.12.3 IBM FlashSystem 5045 80
1.12.4 IBM Storage Virtualize for Public Cloud 84
1.13 Storage efficiency and data reduction features 84
1.13.1 IBM Easy Tier 84
1.13.2 Data reduction and UNMAP 85
1.13.3 Compression and deduplication 88
1.13.4 Enhanced data security features 89
1.14 Application integration features 90
1.14.1 Features for manageability 91
1.15 Copy services 96
1.15.1 Policy-based replication 99
1.15.2 HyperSwap 99
1.16 IBM FlashCore Module drives, NVMe SSDs, and SCM drives 100
1.17 Storage virtualization 103
1.18 Business continuity 107
1.18.1 Business continuity with the IBM SAN Volume Controller 107
1.18.2 Business continuity with stretched clusters 108
1.18.3 Business continuity with enhanced stretched cluster 109
1.18.4 Business continuity for IBM FlashSystem and IBM SAN Volume Controller . . 109
1.18.5 Business continuity with HyperSwap 109
1.18.6 Business continuity with 3-site replication 110
1.18.7 Automatic hot spare nodes (IBM SAN Volume Controller only) 111
1.19 Management and support tools 113
1.19.1 IBM Assist On-site and Remote Support Assistance 113
1.19.2 Event notifications 113
1.20 Licensing 115
1.20.1 Licensing IBM SAN Volume Controller 115
1.20.2 Licensing IBM FlashSystem 9500/R, 9200/R, 7300, 7200, 5200 and 5045 . . . 115
1.20.3 Licensing IBM FlashSystem 5035 and 5015 116

Chapter 2. Installation and configuration planning 119

2.1 General planning guidelines . . . 120
2.2 Planning for availability . . . 121
2.3 Physical installation planning . . . 122
2.4 Planning for system management . . . 123
2.4.1 User password creation options . . . 124
2.4.2 Two person integrity . . . 125
2.5 Connectivity planning . . . 126
2.6 Portsets . . . 127
2.7 Fibre Channel SAN configuration planning . . . 130
2.7.1 Physical topology . . . 130
2.7.2 Zoning . . . 131
2.7.3 N_Port ID Virtualization . . . 131
2.7.4 Inter-node zone . . . 132
2.7.5 Back-end storage zones . . . 132
2.7.6 Host zones . . . 133
2.7.7 Zoning considerations for Metro Mirror and Global Mirror . . . 134
2.7.8 Port designation recommendations . . . 135
2.8 IP SAN configuration planning . . . 137
2.8.1 iSCSI and iSER protocols . . . 138
2.8.2 NVMe over RDMA (RoCE) and NVMe over TCP protocols . . . 138
2.8.3 Priority flow control . . . 139
2.8.4 RDMA clustering . . . 140
2.8.5 iSCSI back-end storage attachment . . . 141
2.8.6 IP network host attachment . . . 142
2.8.7 Native IP replication . . . 143
2.8.8 Firewall planning . . . 144
2.9 Planning topology . . . 144
2.9.1 High availability . . . 144
2.9.2 Volume Mirroring . . . 145
2.9.3 Three-site replication . . . 146
2.10 Back-end storage configuration . . . 146
2.11 Internal storage configuration . . . 148
2.11.1 IBM FlashSystem Systems . . . 148
2.12 Storage pool configuration . . . 152
2.12.1 Child pools . . . 154
2.12.2 Storage pool and cache relationship . . . 156
2.12.3 Provisioning policies . . . 156
2.13 Volume configuration . . . 157
2.13.1 Planning for image mode volumes . . . 157
2.13.2 Planning for standard-provisioned volumes . . . 158
2.13.3 Planning for thin-provisioned volumes . . . 158
2.13.4 Planning for compressed volumes . . . 158
2.13.5 Planning for deduplicated volumes . . . 159
2.13.6 Planning for Volume groups . . . 160
2.14 Host attachment planning . . . 162
2.14.1 Queue depth . . . 162
2.14.2 SAN boot support . . . 162
2.14.3 Planning for large deployments . . . 162
2.14.4 Planning for SCSI UNMAP . . . 163
2.15 Planning copy services . . . 163
2.15.1 FlashCopy guidelines . . . 163
2.15.2 Planning for Metro Mirror and Global Mirror . . . 164
2.15.3 Policy-based replication . . . 166

2.16 Throttles . . . 168
2.17 Data migration . . . 169
2.18 Ansible automation for IBM Storage Virtualize systems . . . 171
2.19 Container Storage integration . . . 171
2.20 Safeguarded Copy . . . 172
2.21 Performance monitoring with IBM Storage Insights . . . 173
2.22 Configuration backup procedure . . . 176

Chapter 3. Initial configuration . . . 177
3.1 Prerequisites . . . 178
3.2 System initialization . . . 179
3.2.1 System initialization process . . . 182
3.3 System setup . . . 186
3.3.1 System Setup wizard . . . 186
3.3.2 Configuring clustering by using Ethernet connections . . . 205
3.3.3 Adding an enclosure in IBM FlashSystem . . . 210
3.3.4 Adding a node or hot spare node in IBM SAN Volume Controller systems . . . 212
3.3.5 Changing the system topology . . . 216
3.3.6 Configuring quorum disks or applications . . . 220
3.3.7 Configuring the local Fibre Channel port masking . . . 223
3.3.8 Automatic configuration for IBM SAN Volume Controller back-end storage . . . 226
3.3.9 Configuring management access . . . 232

Chapter 4. IBM Storage Virtualize GUI . . . 233
4.1 Performing operations by using the GUI . . . 234
4.1.1 Accessing the GUI . . . 234
4.2 Introduction to the GUI . . . 239
4.2.1 Task menu . . . 239
4.2.2 Background tasks . . . 240
4.2.3 Notification icons and help . . . 241
4.3 System Hardware - Overview window . . . 247
4.3.1 Content-based organization . . . 247
4.4 Monitoring menu . . . 250
4.4.1 System Hardware Overview window . . . 252
4.4.2 Easy Tier reports . . . 261
4.4.3 Events option . . . 262
4.4.4 Performance window . . . 264
4.4.5 Background Processes . . . 265
4.5 Using the menus . . . 265
4.5.1 Pools . . . 265
4.5.2 Volumes . . . 266
4.5.3 Hosts . . . 267
4.5.4 Copy Services . . . 267
4.5.5 Access menu . . . 268
4.6 Ownership groups . . . 268
4.6.1 Inheriting ownership . . . 270
4.6.2 Users by groups . . . 275
4.6.3 Users . . . 278
4.6.4 Audit log . . . 284
4.7 Settings . . . 285
4.7.1 Notifications . . . 286
4.7.2 Network . . . 290
4.7.3 Using the management GUI . . . 296

4.7.4 Security 306
4.7.5 System menus 319
4.7.6 Support menu 342
4.7.7 GUI Preferences menu 350
4.8 Other frequent tasks in the GUI 356
4.8.1 Renaming components 356
4.8.2 Working with enclosures 359
4.8.3 Restarting the GUI service 362

Chapter 5. Using storage pools 365
5.1 Working with storage pools 366
5.1.1 Creating storage pools 368
5.1.2 Managed disks in a storage pool 370
5.1.3 Actions on storage pools 372
5.1.4 Child pools 378
5.1.5 Encrypted storage pools 383
5.2 Working with provisioning policies 383
5.2.1 Introduction to provisioning policies 383
5.2.2 Creating and deleting provisioning policies 383
5.2.3 Assigning and unassigning provisioning policies 385
5.3 Working with internal drives and arrays 386
5.3.1 Working with drives 386
5.3.2 RAID and distributed RAID 394
5.3.3 Creating arrays 397
5.3.4 Actions on arrays 400
5.4 Working with external controllers and MDisks 407
5.4.1 External storage controllers 407
5.4.2 Actions on external storage controllers 409
5.4.3 Working with external MDisks 411
5.4.4 Actions for external MDisks 413

Chapter 6. Volumes 419
6.1 Introduction to volumes 420
6.2 Volume characteristics 420
6.2.1 Volume type 421
6.2.2 Managed mode and image mode 422
6.2.3 VSize 425
6.2.4 Performance 426
6.2.5 Volume copies 426
6.2.6 Storage efficiency 430
6.2.7 Encryption 436
6.2.8 Cache mode 436
6.2.9 I/O throttling 436
6.2.10 Volume protection 437
6.2.11 Secure data deletion 438
6.3 Volume groups 438
6.3.1 Creating volume groups 439
6.4 Virtual volumes 444
6.5 Volumes in multi-site topologies 444
6.5.1 HyperSwap topology 445
6.5.2 Stretched topology 446
6.6 Operations on volumes 447
6.6.1 Creating volumes 447

6.6.2 I/O throttling 456
6.6.3 Volume protection 463
6.6.4 Modifying a volume 463
6.6.5 Deleting a volume 471
6.6.6 Mapping a volume to a host 473
6.6.7 Modify I/O group or Nondisruptive Volume Move 477
6.6.8 Migrating a volume to another storage pool 482
6.7 Volume operations by using the CLI 488
6.7.1 Displaying volume information 488
6.7.2 Creating a volume 489
6.7.3 Creating a thin-provisioned volume 491
6.7.4 Creating a volume in image mode 492
6.7.5 Adding a volume copy 493
6.7.6 Splitting a mirrored volume 499
6.7.7 Modifying a volume 501
6.7.8 Deleting a volume 502
6.7.9 Volume protection 502
6.7.10 Expanding a volume 503
6.7.11 HyperSwap volume modification with CLI 504
6.7.12 Mapping a volume to a host 505
6.7.13 Listing volumes that are mapped to the host 507
6.7.14 Listing hosts that are mapped to the volume 507
6.7.15 Deleting a volume to host mapping 508
6.7.16 Migrating a volume 508
6.7.17 Migrating a fully managed volume to an image mode volume 509
6.7.18 Shrinking a volume 510
6.7.19 Listing volumes that use MDisks 511
6.7.20 Listing MDisks that are used by the volume 511
6.7.21 Listing volumes that are defined in the storage pool 511
6.7.22 Listing storage pools in which a volume has its extents 512
6.7.23 Tracing a volume from a host back to its physical disks 514

Chapter 7. Hosts 517
7.1 Host attachment overview 518
7.2 Host objects overview 519
7.3 NVMe over Fibre Channel 520
7.4 NVMe over Remote Direct Memory Access 521
7.5 NVMe over TCP 521
7.6 N_Port ID Virtualization support 522
7.6.1 NPIV prerequisites 524
7.6.2 Verifying the NPIV mode state for a new system installation 524
7.6.3 Enabling NPIV on a system 525
7.7 IP multi-tenancy 530
7.7.1 Limitations and restrictions 530
7.7.2 Prerequisites 531
7.7.3 Configuring the portset 533
7.7.4 Assigning an IP address to a portset 535
7.7.5 Assigning the portset to a host object 537
7.8 Fibre Channel portset 537
7.8.1 FC portset definition 537
7.8.2 FC portsets usage scenarios 538
7.8.3 FC portset behavior for established and new installations 539
7.8.4 Configuring the FC portset 539

7.8.5 Assigning an FC port to an FC portset 541
7.8.6 Modifying an FC portset for a host 543
7.8.7 Removing a port from an FC portset 548
7.8.8 Removing an FC portset 551
7.8.9 FC portset misconfiguration and resolution 554
7.9 Hosts operations by using the GUI 555
7.9.1 Creating hosts 555
7.9.2 Host clusters 574
7.9.3 Actions on hosts 577
7.9.4 Actions on host clusters 588
7.9.5 Host management views 594
7.10 Performing hosts operations by using the CLI 597
7.10.1 Creating a host by using the CLI 597
7.10.2 Host administration by using the CLI 601
7.10.3 Adding and deleting a host port by using the CLI 603
7.10.4 Host cluster operations 605
7.10.5 Adding a host or host cluster to an ownership group 608
7.11 Host attachment examples 608
7.11.1 Prerequisites 608
7.11.2 Fibre Channel host connectivity and capacity allocation 608
7.11.3 iSCSI host connectivity and capacity allocation 613
7.11.4 FC NVMe over Fabric host connectivity example 616
7.11.5 NVMe over RDMA host connectivity example 622
7.11.6 NVMe over TCP host connectivity example 627
7.12 Container Storage Interface drivers 632

Chapter 8. Storage migration 635
8.1 Introduction 636
8.2 Storage migration overview 637
8.2.1 Interoperability and compatibility 639
8.2.2 Prerequisites 639
8.3 Storage migration wizard 640
8.4 Enclosure upgrade migration 656
8.5 Migrating data between systems nondisruptively 657
8.5.1 Nondisruptive volume migration procedure 658

Chapter 9. Advanced features for storage efficiency 669
9.1 IBM Easy Tier 670
9.1.1 Easy Tier concepts 670
9.1.2 Implementing and tuning Easy Tier 677
9.1.3 Monitoring Easy Tier activity 683
9.2 Thin-provisioned volumes 690
9.2.1 Concepts 690
9.2.2 Implementation 691
9.3 UNMAP 692
9.3.1 The SCSI UNMAP command 692
9.3.2 Back-end SCSI UNMAP 692
9.3.3 Host SCSI UNMAP 693
9.3.4 Offloading I/O throttle 694
9.4 Data reduction pools 695
9.4.1 Introduction to DRP 695
9.4.2 DRP benefits 697
9.4.3 Planning for DRP 698

9.4.4 Implementing DRP with compression and deduplication 700
9.5 Saving estimations for compression and deduplication 707
9.5.1 Evaluating compression savings by using IBM Comprestimator 707
9.5.2 Evaluating compression and deduplication 709
9.6 Overprovisioning and data reduction on external storage 709
9.7 Safeguarded child pool capability: Protection from logical data corruption 714

Chapter 10. Advanced Copy Services 717
10.1 Maximum configuration updates 718
10.2 Safeguarded Copy 718
10.3 IBM FlashCopy 719
10.3.1 Business requirements for FlashCopy 720
10.3.2 FlashCopy principles and terminology 722
10.3.3 FlashCopy mapping 722
10.3.4 Consistency groups 724
10.3.5 Crash consistent copy and hosts considerations 724
10.3.6 Volume group snapshot 726
10.3.7 Grains and bitmap: I/O indirection 726
10.3.8 Interaction with cache 732
10.3.9 Background copy rate 733
10.3.10 Incremental FlashCopy 735
10.3.11 Starting FlashCopy mappings and consistency groups 736
10.3.12 Multiple target FlashCopy 738
10.3.13 Reverse FlashCopy 743
10.3.14 FlashCopy and image mode volumes 745
10.3.15 FlashCopy mapping events 746
10.3.16 Thin-provisioned FlashCopy 747
10.3.17 Serialization of I/O by FlashCopy 748
10.3.18 Event handling 749
10.3.19 Asynchronous notifications 750
10.3.20 Interoperation with Metro Mirror and Global Mirror 750
10.3.21 FlashCopy attributes and limitations 751
10.3.22 Expanding Volumes In a FlashCopy Mapping 752
10.4 Managing FlashCopy by using the GUI 752
10.4.1 FlashCopy presets 752
10.4.2 FlashCopy window 755
10.4.3 Creating a FlashCopy mapping 757
10.4.4 Single-click snapshot 767
10.4.5 Single-click clone 768
10.4.6 Single-click backup 770
10.4.7 Creating a FlashCopy consistency group 771
10.4.8 Creating FlashCopy mappings in a consistency group 772
10.4.9 Showing related volumes 776
10.4.10 Moving FlashCopy mappings across consistency groups 777
10.4.11 Removing FlashCopy mappings from consistency groups 778
10.4.12 Modifying a FlashCopy mapping 779
10.4.13 Renaming FlashCopy mappings 780
10.4.14 Deleting FlashCopy mappings 782
10.4.15 Deleting a FlashCopy consistency group 784
10.4.16 Starting FlashCopy mappings 785
10.4.17 Stopping FlashCopy mappings 786
10.4.18 Memory allocation for FlashCopy 787
10.5 Transparent Cloud Tiering 789

10.5.1 Considerations for using Transparent Cloud Tiering. 790
10.5.2 Transparent Cloud Tiering as backup solution and data migration. 790
10.5.3 Restoring data by using Transparent Cloud Tiering 791
10.5.4 Transparent Cloud Tiering restrictions 791
10.6 Implementing Transparent Cloud Tiering 792
10.6.1 Domain Name System configuration. 792
10.6.2 Enabling Transparent Cloud Tiering 794
10.6.3 Creating cloud snapshots 797
10.6.4 Managing cloud snapshots 800
10.6.5 Restoring cloud snapshots 801
10.7 Volume mirroring and migration options 804
10.8 Remote Copy 806
10.8.1 IBM SAN Volume Controller and IBM FlashSystem system layers 807
10.8.2 Multiple IBM Storage Virtualize systems replication 808
10.8.3 Importance of write ordering 811
10.8.4 Remote Copy intercluster communication 813
10.8.5 Metro Mirror overview 814
10.8.6 Synchronous Remote Copy 815
10.8.7 Metro Mirror features. 815
10.8.8 Metro Mirror attributes. 816
10.8.9 Practical use of Metro Mirror. 816
10.8.10 Global Mirror overview 817
10.8.11 Asynchronous Remote Copy 818
10.8.12 Global Mirror features 819
10.8.13 Using Global Mirror with change volumes. 822
10.8.14 Distribution of work among nodes. 823
10.8.15 Background copy performance 824
10.8.16 Thin-provisioned background copy 824
10.8.17 Methods of synchronization 825
10.8.18 Practical use of Global Mirror 825
10.8.19 IBM Storage Virtualize HyperSwap topology 825
10.8.20 Consistency Protection for Global Mirror and Metro Mirror. 826
10.8.21 Valid combinations of FlashCopy, Metro Mirror, and Global Mirror 827
10.8.22 Remote Copy configuration limits 827
10.8.23 Remote Copy states and events. 827
10.8.24 Remote Copy commands 835
10.9 Native IP replication 841
10.9.1 Enhancements 841
10.9.2 Native IP replication technology 843
10.9.3 IP partnership limitations. 844
10.9.4 IP Partnership and data compression. 846
10.9.5 VLAN support 847
10.9.6 IP partnership and terminology 848
10.9.7 States of IP partnership 849
10.9.8 Remote Copy portsets 850
10.9.9 Supported configurations 851
10.10 Managing Remote Copy by using the GUI 862
10.10.1 Creating Fibre Channel partnership 864
10.10.2 Creating Remote Copy relationships 866
10.10.3 Creating a consistency group 872
10.10.4 Renaming Remote Copy relationships 874
10.10.5 Renaming a Remote Copy consistency group 874
10.10.6 Moving standalone Remote Copy relationships to consistency group 875

10.10.7 Removing Remote Copy relationships from consistency group 876
10.10.8 Starting Remote Copy relationships 877
10.10.9 Starting a Remote Copy consistency group 878
10.10.10 Switching a relationship copy direction 878
10.10.11 Switching a consistency group direction 880
10.10.12 Stopping Remote Copy relationships 881
10.10.13 Stopping a consistency group 882
10.10.14 Deleting Remote Copy relationships 883
10.10.15 Deleting a consistency group 884
10.11 Remote Copy memory allocation 885
10.12 A real-life implementation of IP replication: Anadolu Sigorta 886
10.12.1 Customer profile 886
10.12.2 General information about the infrastructure 886
10.12.3 IP replication implementation 887
10.12.4 Implementation decisions and results 887
10.13 Troubleshooting Remote Copy 888
10.13.1 1920 error 888
10.13.2 1720 error 890
10.14 3-Site Replication 891
10.14.1 3-Site Replication Orchestrator 892
10.15 IBM Storage volume group and volume group snapshots 894
10.15.1 Volume group snapshots 896
10.15.2 Volume group thin clones 900
10.15.3 Volume group thick clones 901
10.15.4 Orphaned volume group snapshots 902
10.15.5 Volume group snapshot scheduler 902
10.16 IBM Storage policy-based replication 905
10.16.1 Understanding policy-based replication deployment 909
10.16.2 Planning for hosts, partnerships and connectivity 921
10.16.3 Configuring policy-based replication using GUI 923
10.17 Converting Global Mirror to policy-based replication 938
10.18 Monitoring options for policy-based replication 946
10.19 Troubleshooting policy-based replication 955

Chapter 11. Reliability, availability, and serviceability 963
11.1 Reliability, availability, and serviceability 964
11.1.1 Hardware information 965
11.1.2 Dense Drawer Enclosures LED 971
11.1.3 Enclosure SAS cabling 972
11.1.4 Drive modules 974
11.1.5 Power 974
11.2 Shutting down an IBM Storage Virtualize System 975
11.2.1 Shutting down and powering on a complete infrastructure 975
11.3 Removing a node or adding a node 976
11.4 Configuration backup 979
11.4.1 Backing up by using the CLI 980
11.4.2 Saving the backup by using the GUI 982
11.5 Updating software 984
11.5.1 Storage Virtualize Upgrade Planning 984
11.5.2 Precautions before the update 985
11.5.3 IBM FlashSystem upgrade test utility 985
11.5.4 Updating your IBM FlashSystem to Version 8.6.0.0 987
11.5.5 Updating the IBM FlashSystem drive code 993

11.5.6 Manually updating the system 999
11.6 Health checker feature 1000
11.7 Troubleshooting and fixing procedures 1001
11.7.1 Managing the event log 1002
11.7.2 Running a fix procedure 1004
11.7.3 Event log details 1004
11.8 Monitoring and Event Notification 1006
11.8.1 Email notifications and the Call Home function 1007
11.8.2 Remote Support Assistance 1015
11.8.3 SNMP configuration 1019
11.8.4 Syslog notifications 1021
11.9 Audit log 1022
11.10 Collecting support information by using the GUI, CLI, and USB 1024
11.10.1 Collecting information by using the GUI 1024
11.10.2 Collecting logs by using the CLI 1027
11.10.3 Collecting logs by using a USB flash drive 1029
11.10.4 Uploading files to the IBM Support Center 1030
11.11 Service Assistant Tool 1031
11.12 IBM Storage Insights monitoring 1035
11.12.1 Capacity monitoring 1036
11.12.2 Performance monitoring 1037
11.12.3 Logging support tickets by using IBM Storage Insights 1039
11.12.4 Managing existing support tickets by using IBM Storage Insights 1043

Chapter 12. Security and encryption 1045
12.1 New security features in IBM Storage Virtualize V8.6 1046
12.1.1 Support for non-superuser ability to manage the system 1046
12.2 Introduction to IBM Storage Virtualize security 1046
12.2.1 Configuring System TLS Certificates 1047
12.3 Configuring users and password policy 1052
12.3.1 Local users 1052
12.3.2 Remote authentication 1053
12.3.3 Default user 1054
12.3.4 User groups and roles 1054
12.3.5 Configuring remote authentication 1055
12.3.6 Adding locally administered users 1059
12.3.7 Defining a password policy 1060
12.3.8 Setting password expiration and account locking 1061
12.3.9 Changing the default session timeouts 1065
12.4 Multifactor authentication 1066
12.4.1 Configuring Multifactor authentication 1067
12.4.2 User authentication using Single Sign-on 1077
12.5 Ownership groups principles of operations 1083
12.5.1 Implementing ownership groups on a new system 1085
12.5.2 Migrating objects to ownership groups 1090
12.6 Two Person Integrity (TPI) 1093
12.6.1 Prerequisites 1093
12.6.2 Tasks affected by TPI 1093
12.6.3 Enabling and Configuring TPI 1094
12.6.4 Requesting a temporary Role Elevation 1095
12.7 Encryption 1098
12.7.1 Externally virtualized storage 1098
12.7.2 Serial-attached SCSI internal storage 1099

12.7.3 Non-Volatile Memory Express internal storage . . . 1099
12.7.4 Planning for encryption . . . 1099
12.7.5 Defining encryption of data at-rest . . . 1100
12.7.6 Encryption methods . . . 1100
12.7.7 Encrypted data . . . 1101
12.7.8 Data reduction and encryption . . . 1102
12.7.9 Encryption keys . . . 1102
12.7.10 Encryption licenses . . . 1103
12.8 Activating and enabling encryption . . . 1103
12.8.1 Activating encryption . . . 1103
12.8.2 Enabling encryption . . . 1112
12.8.3 Enabling encryption by using key servers . . . 1119
12.8.4 Enabling encryption by using SGKLM . . . 1120
12.8.5 Enabling encryption by using SafeNet KeySecure or Thales CipherTrustManager . . . 1126
12.8.6 Enabling encryption by using both providers . . . 1131
12.8.7 Configuring more providers . . . 1134
12.8.8 Adding USB flash drives as a second provider . . . 1137
12.8.9 Migrating between providers . . . 1137
12.8.10 Changing from encryption key server to USB flash drive provider . . . 1138
12.8.11 Migrating between different key server types . . . 1138
12.8.12 Recovering from a provider loss . . . 1140
12.9 Using encryption . . . 1141
12.9.1 Encrypted pools . . . 1141
12.9.2 Encrypted child pools . . . 1143
12.9.3 Encrypted arrays . . . 1144
12.9.4 Encrypted MDisks . . . 1145
12.9.5 Encrypted volumes . . . 1148
12.9.6 Restrictions . . . 1149
12.9.7 Rekeying an encryption-enabled system . . . 1149
12.9.8 Disabling encryption . . . 1152

Chapter 13. Automation and scripting . . . 1155
13.1 REST API on IBM Storage Virtualize . . . 1156
13.2 Scripting . . . 1164
13.2.1 Scripting principles . . . 1164
13.2.2 Secure Shell . . . 1164
13.2.3 SMI-S . . . 1168
13.2.4 HTTPS and REST API on IBM Storage Virtualize . . . 1171
13.2.5 HTTPS and REST API on IBM Storage Control . . . 1180
13.3 Automation with Red Hat Ansible . . . 1180
13.3.1 Red Hat Ansible . . . 1181
13.3.2 Red Hat Ansible editions . . . 1181
13.3.3 Requirements . . . 1181
13.3.4 Essential terminology in an Ansible environment . . . 1182
13.3.5 Automating IBM Storage with Ansible . . . 1182
13.3.6 Getting started . . . 1183
13.3.7 Securing credentials in Ansible . . . 1187
13.3.8 Creating an Ansible playbook . . . 1187
13.3.9 More automation . . . 1193

Appendix A. Command line interface setup . . . 1195
Overview . . . 1196

Basic setup on a Windows host 1197
13.4 Basic setup on a Mac, UNIX, or Linux host 1205

Appendix B. Terminology 1209
Commonly encountered terms 1210

Appendix C. List of the demonstration videos 1243

Related publications 1249
IBM Redbooks 1249
Online resources 1249
Help from IBM 1250

Figures

1-1 Block-level virtualization overview 5
1-2 IBM Storage Virtualize features by product 7
1-3 Overview of block-level virtualization architectures. 9
1-4 IBM SAN Volume Controller conceptual and topology overview 12
1-5 IBM Safeguarded Copy Data Resilience 19
1-6 STaaS Tiers comparison 22
1-7 IBM SAN Volume Controller family 24
1-8 Separation of upper and lower cache 28
1-9 Overview of an IBM SAN Volume Controller clustered system with an I/O group 33
1-10 Striped volume 34
1-11 Sequential volume 34
1-12 Image mode volume 35
1-13 Global Mirror with change volumes. 41
1-14 IBM SAN Volume Controller SV3 front view 44
1-15 IBM SAN Volume Controller SV3 rear view 44
1-16 IBM SAN Volume Controller SV3 internal hardware components 44
1-17 IBM SAN Volume Controller SV3 internal architecture 45
1-18 IBM SAN Volume Controller SV2 or SA2 front view 45
1-19 IBM SAN Volume Controller SV2 or SA2 rear view 45
1-20 Internal hardware components 46
1-21 IBM SAN Volume Controller SV2 and SA2 internal architecture 46
1-22 IBM FlashSystem family 49
1-23 Storage Expert Care tier levels for IBM FlashSystem 5015 and IBM FlashSystem 5045 51
1-24 IBM FlashSystem 9500 front and rear views 53
1-25 IBM FlashSystem 9500 internal architecture 54
1-26 FS9500 internal hardware components 55
1-27 Logical NMVe drive placement 58
1-28 NVMe drive population with numbering 58
1-29 Legend to figures in this section 61
1-30 IBM FlashSystem 9500R Rack Solution configuration in the rack 61
1-31 IBM FlashSystem 9000 Expansion Enclosure Model AFF 63
1-32 IBM FlashSystem 9000 Expansion Enclosure Model front view 63
1-33 IBM FlashSystem 9200 system that is connected to an expansion enclosure 64
1-34 IBM FlashSystem 9500 system that is connected to an expansion enclosure 65
1-35 IBM FlashSystem 7300 front and rear views 66
1-36 IBM FlashSystem 7300 internal architecture 67
1-37 Internal hardware components 67
1-38 IBM FlashSystem 7000 LFF Expansion Enclosure Model 12G 69
1-39 Front view of an IBM FlashSystem 7000 SFF expansion enclosure. 69
1-40 Rear of an IBM FlashSystem 7000 expansion enclosure 70
1-41 IBM Dense Expansion Drawer 70
1-42 Connecting FS7300 SAS cables while complying with the maximum chain weight . . 71
1-43 IBM FlashSystem 5200 control enclosure front and 3/4 ISO view 73
1-44 IBM FlashSystem 5015, 5035 and 5045 SFF control enclosure front view 75
1-45 IBM FlashSystem 5015, 5035 and 5045 LFF control enclosure front view. 76
1-46 Front view of an IBM FlashSystem 5035 80
1-47 Rear view of an IBM FlashSystem 5035. 80

© Copyright IBM Corp. 2023.

1-48 View of available connectors and LEDs on an IBM FlashSystem 5035 single canister . . . 80
1-49 Front view of an IBM FlashSystem 5045 . . . 83
1-50 Rear view of an IBM FlashSystem 5045 . . . 83
1-51 View of available connectors and LEDs on an IBM FlashSystem 5045 single canister . . . 83
1-52 Easy Tier concept . . . 85
1-53 IBM Storage Virtualize GUI dashboard . . . 92
1-54 IBM Storage Insights dashboard . . . 95
1-55 IBM FlashCore Module (NVMe) . . . 101
1-56 Storage technologies versus latency for Intel drives . . . 103
1-57 “Star” and “Cascade” modes in a three-site solution . . . 111
2-1 System zoning . . . 131
2-2 Port masking configuration example for IBM FlashSystem . . . 135
2-3 Port masking configuration example for IBM SAN Volume Controller . . . 136
2-4 Volume Mirroring Comparison . . . 146
2-5 Distributed RAID planning guidance . . . 150
2-6 IBM Storage Insights architecture . . . 174
2-7 Data flow from the storage systems to the IBM Storage Insights cloud . . . 175
3-1 Technician port FlashSystem 9500 . . . 180
3-2 Technician port FlashSystem 7300 . . . 180
3-3 Technician port FlashSystem 5200 . . . 180
3-4 Technician port FlashSystem 5045 . . . 180
3-5 Technician port FlashSystem 5015 . . . 181
3-6 Technician port IBM SAN Volume Controller 2145-SV3 . . . 181
3-7 Technician port IBM SAN Volume Controller 2145-SV2 . . . 181
3-8 Logging in to Service Assistant by way of the technician port . . . 183
3-9 System Initialization: Canister detection . . . 183
3-10 System Initialization: Initialize the first enclosure . . . 184
3-11 System Initialization: Initialize the first IBM SAN Volume Controller node . . . 184
3-12 System Initialization: Enter Management IP . . . 185
3-13 System Initialization: Web-server restart timer counting down from 5 minutes . . . 185
3-14 System Initialization completed . . . 186
3-15 Logging in for the first time . . . 187
3-16 Welcome page in System Setup . . . 188
3-17 License agreement . . . 189
3-18 Changing the superuser password . . . 190
3-19 Entering system name . . . 190
3-20 Licensed functions . . . 191
3-21 DNS server (optional) . . . 192
3-22 Data and time . . . 192
3-23 Encryption license activation . . . 193
3-24 Encryption licensed . . . 194
3-25 Call Home information window . . . 194
3-26 Call Home transmission type . . . 196
3-27 Proxy server (optional) . . . 196
3-28 System location . . . 197
3-29 Setting up contact information for Call Home . . . 198
3-30 Email servers . . . 198
3-31 Storage Insights information window . . . 199
3-32 Support assistance . . . 200
3-33 System communicating with named IBM Support servers . . . 201
3-34 Remote support access settings . . . 202

3-35 Automatic configuration for IBM SAN Volume Controller 202
3-36 Summary page 203
3-37 System Setup: Setup completed. 204
3-38 System Setup directs the user to the system GUI Base configuration 204
3-39 No ISL connectivity 205
3-40 Shared ISL connectivity 206
3-41 Dedicated ISL connectivity 207
3-42 Node IP address setup for Remote Direct Memory Access clustering 208
3-43 Node IP addresses configured 209
3-44 Setting the node discovery subnet 210
3-45 Add Enclosure button 211
3-46 Selecting the control enclosure to add 211
3-47 Add Node button 213
3-48 Adding a node 213
3-49 IBM SAN Volume Controller is adding node to the cluster 214
3-50 Node added 214
3-51 Starting the Modify System Topology wizard 217
3-52 Assigning site names 217
3-53 Specifying the system topology 218
3-54 Assigning hosts to sites 219
3-55 Setting the bandwidth between the sites 220
3-56 Download IPv4 quorum button 221
3-57 Download IP quorum application window 221
3-58 IP quorum application that is deployed and connected 222
3-59 Changing the quorum mode 223
3-60 Applying a port mask by using a GUI 224
3-61 Modify Connection dialog box 225
3-62 Automatic Configuration wizard enablement 227
3-63 Automatic configuration: Add Enclosure 227
3-64 Defining a host cluster 228
3-65 Hosts inside an IBM SAN Volume Controller host cluster. 229
3-66 Begin the automatic configuration process 229
3-67 Automatic pool configuration. 230
3-68 Pools configuration 230
3-69 Automatic configuration running commands. 231
3-70 Automatic configuration complete. 231
4-1 Login window of the GUI. 234
4-2 Login window of the storage system when it is connected to the configuration node 235
4-3 Single Sign On 235
4-4 Welcome page with the dashboard. 236
4-5 Performance statistics. 236
4-6 Capacity overview. 236
4-7 Overview of compressed volumes 237
4-8 Compression Details. 237
4-9 List that shows overprovisioned external storage 238
4-10 System Health overview window 238
4-11 Expanded System health view 238
4-12 IBM Storage System Hardware window 239
4-13 The task menu on the left side of the GUI 240
4-14 Background tasks 240
4-15 Notification area 241
4-16 System alerts 241

4-17 View all Events 242
4-18 Power Supply Unit Input Power Failed 242
4-19 Background Processes 243
4-20 Details of a running task 243
4-21 User Task button 244
4-22 Opening the Suggested Actions window 244
4-23 Suggested Actions window 244
4-24 Selecting items by using the Shift key 245
4-25 Selecting items by using the Ctrl key 245
4-26 Access help menu 245
4-27 Example of Dashboard help content 246
4-28 IBM Storage Virtualize REST API 246
4-29 System Hardware - Overview window 247
4-30 Filter search box 248
4-31 Showing filtered rows 248
4-32 Removing the filtered view 248
4-33 Adding or removing details in a table 249
4-34 Table with an added ID column 249
4-35 Table Restore Default View 250
4-36 Reorganizing table columns 250
4-37 Monitoring menu 250
4-38 System Hardware Overview for an IBM SAN Volume Controller Cluster 252
4-39 Select the IBM SAN Volume Controller node 252
4-40 Health status of components 253
4-41 Detailed properties 253
4-42 Identify node 254
4-43 Turn LED Turn off LED 254
4-44 Backward view of a SAN Volume Controller SV3 255
4-45 System Hardware Overview for a FlashSystem 7300 255
4-46 Selecting more system hardware enclosure details 256
4-47 System hardware enclosure details 256
4-48 Showing the Adapter details 257
4-49 Drive information 257
4-50 Turning on the Identify LED 258
4-51 Turning off the Identify LED 258
4-52 Viewing the internal components 259
4-53 SAS Chain View 259
4-54 Enclosure Details window 260
4-55 Top View of a 4657-92G Expansion Enclosure 260
4-56 System Overview with attached enclosures 261
4-57 Selecting Easy Tier Reports 261
4-58 Easy Tier Reports 262
4-59 Event log list 263
4-60 Performing a fix procedure 263
4-61 Performance statistics of the IBM FlashSystem storage system 264
4-62 Sample details 264
4-63 Selecting Background Processes 265
4-64 List of running tasks 265
4-65 Pools menu 266
4-66 Volumes menu 266
4-67 Hosts menu 267
4-68 Copy Services in GUI 267
4-69 Access menu 268

4-70 Start of an Ownership Group 269
4-71 Ownership group inheritance 270
4-72 Ownership by Groups 275
4-73 Accessing Users by Group 276
4-74 Defining a User Group 276
4-75 User Group 277
4-76 Registering a user account 280
4-77 Deleting a user account 280
4-78 Setting a new password 281
4-79 Locking a user. 281
4-80 Lock verification 282
4-81 Locked user 282
4-82 Unlock user 283
4-83 Unlock confirmation 283
4-84 Expire Password 283
4-85 Expire Password verification 284
4-86 Audit log with Custom Range Filter. 285
4-87 Settings menu 285
4-88 Setting the SNMP server and traps 286
4-89 Creating Syslog Server 288
4-90 Setting the syslog messages 289
4-91 Delete Syslog Server 290
4-92 Accessing network information 291
4-93 Viewing the management IP addresses 291
4-94 Viewing service IP addresses 292
4-95 Ethernet Connectivity 292
4-96 Ethernet Ports 294
4-97 Priority flow control 296
4-98 iSCSI Configuration window 297
4-99 Fibre Channel Connectivity 298
4-100 Viewing Fibre Channel Port properties 298
4-101 NVMe Connectivity window 300
4-102 DNS information 301
4-103 Internal Proxy Server 301
4-104 portsets 302
4-105 Create Portset window 305
4-106 Security menu 306
4-107 Security Settings 306
4-108 Multifactor Authentication with IBM Security Verify. 308
4-109 Multifactor Authentication with Duo Security. 308
4-110 Remote Authentication 309
4-111 Refresh Authentication 309
4-112 Configuring Remote Authentication 310
4-113 Single Sign-on. 310
4-114 Single Sign-on login 311
4-115 Configuring secure communications and updating certificates. 312
4-116 Password Policies window 314
4-117 User Access 317
4-118 Inactivity Logout 317
4-119 SSH Rules 318
4-120 Security protocol levels 319
4-121 System option 319
4-122 Date and Time window 320

4-123 Set Date and Time window 320
4-124 Set NTP Server IP Address window 321
4-125 Licensing window 322
4-126 Update System 327
4-127 Enabling VVOLs management 327
4-128 Volume Protection window 329
4-129 Resources allocation 330
4-130 IP Quorum settings 331
4-131 IP quorum application metadata 332
4-132 IP Quorum Java application 332
4-133 Quorum settings 333
4-134 I/O group port virtualization 335
4-135 Changing NPIV settings 336
4-136 Transparent Cloud Tiering settings 339
4-137 Automatic Configuration for Virtualization option 340
4-138 Remote-copy Bandwidth Limit 342
4-139 Call Home settings 343
4-140 Support assistance 349
4-141 Support package menu 350
4-142 GUI Preferences selection window 351
4-143 Enabling the login message 351
4-144 Welcome message in the GUI 352
4-145 Welcome message in CLI 352
4-146 General GUI Preferences window 352
4-147 GUI Features 354
4-148 Notification Behavior window 355
4-149 Language selection 355
4-150 Sidebar Accent Color window 356
4-151 Customized sidebar 356
4-152 Overview window 357
4-153 Renaming the system 358
4-154 Overview window 358
4-155 Entering the new name of the node 359
4-156 Newly detected expansion enclosure 359
4-157 Adding an enclosure 360
4-158 Enclosure successfully added 360
4-159 Selecting an enclosure for removal 361
4-160 Confirming the removal 361
4-161 Enclosure removed 361
4-162 Log in Service GUI 362
4-163 Identifying the configuration node on the Service Assistant 363
4-164 Node details window 363
4-165 Restarting the Tomcat web server 364
5-1 Relationship between MDisks, storage pools, and volumes 366
5-2 Accessing the Pools panel 367
5-3 Option to create a storage pool in the Pools panel 368
5-4 Create Pool dialog box 368
5-5 Advanced pool settings 369
5-6 Creating a pool with encryption enabled 369
5-7 Newly created empty pool 370
5-8 MDisks by Pools 371
5-9 Pool actions 372
5-10 Add Storage to Pool dialog 373

5-11 Edit throttle for Pool window 374
5-12 Viewing all throttles 375
5-13 List of resources in the storage pool 376
5-14 Pool properties and details 377
5-15 Creating a child pool 379
5-16 Child pool creation window 380
5-17 Resizing a child pool 381
5-18 Deleting a child pool 381
5-19 Managing the ownership group of a child pool 382
5-20 Actions menu in Volumes by Pool 382
5-21 Provisioning Policies list 384
5-22 Create provisioning policy 384
5-23 Delete or rename provisioning policy 385
5-24 Assigning policies to the standard pool 385
5-25 Assigning a policy to a DRP and removing the assigned policy 386
5-26 Internal storage panel 386
5-27 Drive use transitions 388
5-28 Actions on internal storage 388
5-29 Taking a drive offline if a spare or rebuild area is available 389
5-30 An offline drive is marked as failed 389
5-31 Taking a drive offline fails if it might result in a loss of access to data 390
5-32 Identifying an internal drive 391
5-33 Upgrading a drive or a set of drives 392
5-34 List of volumes dependent on disks 7, 8, 9 392
5-35 Drive properties 393
5-36 Distributed RAID 6 (for simplification, not all packs are shown) 395
5-37 Single drive failure with DRAID 6 (for simplification, not all packs are shown) 396
5-38 Adding storage to a pool 397
5-39 Assign drives 397
5-40 Assigning storage to a pool 398
5-41 Assign Storage to Pool 399
5-42 Actions on arrays 401
5-43 Swapping array member with another candidate or spare drive 401
5-44 Expand a distributed array 403
5-45 Array expansion progress 403
5-46 Deleting an array from a non-empty storage pool 404
5-47 Dependent volumes for MDisk mdisk0 405
5-48 List of drives in an array 406
5-49 Array properties with expanded list of details 406
5-50 External Storage panel 409
5-51 Actions for external storage 409
5-52 Modifying the site of an external controller 410
5-53 Assigning an unmanaged MDisk 411
5-54 Assign MDisk dialog box 412
5-55 Actions for MDisks 413
5-56 Modifying an external MDisk tier 414
5-57 Selecting new quorum disks 415
5-58 Removing an MDisk from a pool 415
5-59 Dependent volumes for an MDisk 416
5-60 View Provisioning Groups 417
6-1 Striped extent allocation 421
6-2 Simple view of block virtualization 423
6-3 An image mode volume versus a striped volume 424

6-4 Smallest possible volume size 425
6-5 Volume mirroring overview 427
6-6 Data flow for write I/O processing in a mirrored volume 429
6-7 Design of an enhanced stretched cluster 430
6-8 Conceptual diagram of a thin-provisioned volume 432
6-9 Creating a volume group 439
6-10 Volume Groups view 440
6-11 Create Volume Group dialog 440
6-12 Updated Volume Group view 441
6-13 Selecting existing volumes for a volume group 441
6-14 Volume selection 442
6-15 Updated Volume Group list 442
6-16 Managing volume groups by using the Volumes view 443
6-17 Add volumes to volume group 443
6-18 What makes up a HyperSwap volume 446
6-19 Volumes menu 448
6-20 Create Volumes 448
6-21 Creating volumes, advance settings mode disable 449
6-22 Define volume properties 449
6-23 Create volume, DRP (compressed and deduplicated) 450
6-24 Create volume, DRP (compressed and deduplicated) 450
6-25 Advanced settings mode enabled 451
6-26 Mirrored volume 451
6-27 Mirrored volume advanced settings 452
6-28 Multiple volumes advanced settings disabled 453
6-29 Multiple volumes advanced settings enabled 453
6-30 Multiple volumes created 454
6-31 Create HyperSwap volume, advanced settings mode disable 454
6-32 Define HyperSwap volume properties, advanced settings disabled 455
6-33 Create HyperSwap volume, advanced settings mode enable 455
6-34 Define HyperSwap volume properties, advanced settings enabled 456
6-35 Edit Throttle menu item 457
6-36 IOPS throttle on a volume 458
6-37 View All Throttles 459
6-38 View All Throttles window 460
6-39 Filtering the throttle type 460
6-40 Edit Throttle 461
6-41 Modifying a volume throttle 462
6-42 Removing a volume throttle 462
6-43 Volume Protection configuration 463
6-44 Volume Shrink menu item 464
6-45 Specifying the size of the shrunk volume 465
6-46 Confirming the volume shrinking operation 465
6-47 Shrunk volume size 466
6-48 Volume Expand menu item 467
6-49 Specifying the expanded volume size 468
6-50 Expanded volume size 468
6-51 Modify Capacity Savings menu item 468
6-52 Capacity savings options for a volume 469
6-53 Enabling deduplication on a volume 469
6-54 Modify Mirror Sync Rate menu item 470
6-55 Setting the volume mirror sync rate 470
6-56 Modifying the volume cache mode 471

6-57 Setting the volume cache mode 471
6-58 Volume Delete menu item 472
6-59 Confirming the volume deletion 473
6-60 Volume mapping menu item 473
6-61 Mapping a volume to a host 474
6-62 Mapping a volume to a host: Summary 475
6-63 Accessing the Hosts Mapping view 475
6-64 List of volume to host mappings 476
6-65 Unmap Volumes menu item 476
6-66 Volume unmap confirmation window 477
6-67 Volume mapping removed 477
6-68 Modify I/O Group menu item 478
6-69 I/O Group selection dialog box 478
6-70 Modify the I/O Group for a mapped volume wizard: Welcome 479
6-71 Modify the I/O Group for a mapped volume wizard: I/O group selection window . . . 479
6-72 Modify the I/O Group for a mapped volume wizard: First stage completed (details) 480
6-73 Modify I/O Group for a mapped volume wizard: Validation 480
6-74 Modify I/O Group for a mapped volume wizard: Second stage completes (details) 481
6-75 Modify I/O Group for a mapped volume wizard: Operation complete 481
6-76 Migrate Volume Copy: Selecting a menu item 482
6-77 Migrate Volume Copy: Selecting the volume copy 483
6-78 Monitoring the volume migration progress 483
6-79 Volume migration complete 484
6-80 Volume copy after migration 484
6-81 Add Volume Copy menu item 485
6-82 Adding a volume copy 486
6-83 Verifying that the volume copies are synchronized 486
6-84 Setting the volume copy in the target storage pool as the primary copy 487
6-85 Deleting the volume copy in the source pool 487
6-86 Confirming the deletion of a volume copy 487
6-87 Volume copy in the target storage pool 488
7-1 Allocation of NPIV virtual WWPN ports per physical port (*failover host attach port* is not active) 523
7-2 Allocation of NPIV virtual WWPN ports per physical port after a node failure 523
7-3 I/O Groups menu 526
7-4 Change NPIV Settings windows 527
7-5 Change NPIV Settings window: Selecting Enabled option 529
7-6 Selecting the portset option 533
7-7 Creating a portset 534
7-8 Creating portset_2 for node1 in ownershipgroup0 534
7-9 Selecting Ethernet ports 535
7-10 Actions available after right-clicking the suitable node and port 536
7-11 Portset_2 is assigned IP address 9.42.162.180 537
7-12 FC portsets 537
7-13 Default FC portset 539
7-14 Selecting the portset option 539
7-15 Creating FC portset 540
7-16 Fibre Channel Ports 541
7-17 Selecting Fibre Channel Port ID 541
7-18 Assigning FC portset 542
7-19 Hosts 543

7-20 Hosts and default FC portset 543
7-21 Host Properties 543
7-22 Associated FC portset 544
7-23 Editing the Host properties for FC portset 545
7-24 Portset dropdown 546
7-25 FC portset change saved 547
7-26 New FC portset assigned 547
7-27 Unassign Portset operation 548
7-28 Selecting FC portset 549
7-29 Unassign Portset 549
7-30 Removing FC Port from an FC portset 550
7-31 FC portset 551
7-32 FC portset Delete operation 552
7-33 FC portset Delete confirmation 552
7-34 FC portset deletion 553
7-35 FC portset misconfiguration event 554
7-36 Event marked as Fixed 555
7-37 Opening the Host window 556
7-38 Fibre Channel host configuration view 557
7-39 Selecting the host WWPNs 557
7-40 Host types selection 558
7-41 Adding a host to an ownership group 559
7-42 Hosts view after creating a host 559
7-43 Network: iSCSI Configuration view 560
7-44 iSCSI Configuration modification 561
7-45 Select Manage IP Address from the “Actions” menu 562
7-46 Adding, deleting, or modifying IP addresses 563
7-47 Available actions with configured ports 563
7-48 Modifying the port for host connectivity 564
7-49 VLAN settings modification interface 564
7-50 VLAN settings: Details 565
7-51 Adding the iSCSI host object to the configuration 566
7-52 Defined hosts list 566
7-53 Creating an FC-NVMe hos 567
7-54 Defining the NQN 568
7-55 NVMe host created 568
7-56 Creating an NVMe over RDMA host 570
7-57 Defining NQN for RDMA over NVMe host 571
7-58 RDMA over NVMe host created 571
7-59 Creating an NVMe over TCP host 572
7-60 Defining NQN for TCP over NVMe host 573
7-61 TCP over NVMe host created 573
7-62 Host clusters menu 574
7-63 Create Host Cluster window 574
7-64 Host Cluster details definition 575
7-65 Create Host Cluster: Summary 575
7-66 Host Clusters view 575
7-67 Host Clusters Actions menu 576
7-68 Actions on hosts 577
7-69 Renaming a host 577
7-70 Assigning a host to a cluster 578
7-71 Assign host to host cluster confirmation 578
7-72 Conflict between private and shared volume mappings 579

7-73 Removing a host from a host cluster 579
7-74 Modifying the host volume mappings 580
7-75 Adding private mappings 580
7-76 Warning that a mapping to another host exists 581
7-77 Choosing SCSI IDs manually 581
7-78 Duplicate Mappings window 582
7-79 Import volume mappings source host selection 583
7-80 Changing the host type 584
7-81 Edit Throttle for Host dialog 584
7-82 View All Throttles window 585
7-83 Confirming the number of mappings to be removed 585
7-84 Confirming the removal of the host 586
7-85 Viewing the IP login information 586
7-86 Host properties overview 587
7-87 Listing port definitions 587
7-88 Actions that are available on a host cluster object 588
7-89 View Host Cluster Members window 588
7-90 Adding a host member 589
7-91 Confirming the addition of a host to a cluster 589
7-92 Remove Host from Cluster 590
7-93 Existing shared mappings 591
7-94 Adding shared mappings 591
7-95 Assigning a SCSI ID to mapped volumes manually 591
7-96 Summary of added shared mappings 592
7-97 Setting the I/O groups for hosts 592
7-98 Warning that host throttles exist 593
7-99 Creating a host cluster throttle 593
7-100 Removing a host cluster 594
7-101 Host management views 594
7-102 Private mappings list 594
7-103 Shared mappings list 595
7-104 Removing two private mappings 595
7-105 Confirming the mapping deletion 595
7-106 Volumes by Host window 596
7-107 Host Details: Mapped volumes 610
7-108 Host Details: Port Definitions tab 614
7-109 Ethernet Ports Configuration tab 615
7-110 Checking the host object NQN on the system 618
7-111 Volume as it is seen on the system 622
8-1 Browsing to Storage Migration 641
8-2 Starting a migration 641
8-3 Error message if no external storage is detected 642
8-4 Restrictions and prerequisites confirmation 643
8-5 Preparing your environment for storage migration 644
8-6 Steps to map the LUNs to be migrated 645
8-7 Storage Migration external storage discovery detectmdisk command detail 646
8-8 Discovering mapped LUNs from external storage 647
8-9 Storage Migration image mode volume creation detail 648
8-10 List of configured hosts to which to map the imported volume 649
8-11 Creating a host during the migration process 649
8-12 Mapping volumes to hosts 650
8-13 Selecting the host to which to map the new volume 651
8-14 Manually assign a LUN SCSI ID to a mapped volume 651

8-15 Volumes mapping summary before migration. 652
8-16 Selecting the target pool for the migration of the image mode MDisk. 653
8-17 Migration is started 654
8-18 The ongoing migration is listed in the Storage Migration window 654
8-19 Finalizing a migration 655
8-20 Migration finalization confirmation. 655
8-21 External Storage MDisks window 655
8-22 Existing volume on source system 658
8-23 Establish partnership between source and target systems. 659
8-24 Creating two volumes on the target system 659
8-25 Creating a nondisruptive system migration relationship 660
8-26 Choosing a target system 660
8-27 Choosing master and auxiliary volumes 661
8-28 Adding master and auxiliary volumes 661
8-29 Final list of volumes to migrate 662
8-30 Creating a relationship 662
8-31 Creating a relationship successful completion 663
8-32 Consistent synchronized state 664
8-33 Approval state to switch relationship. 665
8-34 Approval state to switch relationship. 665
8-35 Successful completion of switch relationship for one volume 666
8-36 Completion of one volume 666
8-37 Switch relationship completion for all volumes 667
8-38 Volume migration completed with source UUID on target system 668
9-1 Easy Tier. 671
9-2 Actions on extents. 674
9-3 Choosing the tier when assigning MDisks 677
9-4 Changing the MDisk tier 678
9-5 MDisk properties 679
9-6 Pool properties 680
9-7 Modifying the pool overallocation limit 682
9-8 Easy Tier reports are not available 684
9-9 Easy Tier Data Movement report 684
9-10 Easy Tier movement description. 685
9-11 Tier Composition chart 685
9-12 Composition description 686
9-13 Easy Tier Mappings 686
9-14 Easy Tier Workload Skew Comparison. 687
9-15 Skew Comparison Description 688
9-16 Downloading an Easy Tier heat file: Download Support Package 689
9-17 Downloading Easy Tier heat data file: dpa_heat files 689
9-18 Volume parameters for thin provisioning 691
9-19 Setting an offload throttle 694
9-20 System Offload Throttle settings. 695
9-21 Creating compressed volumes 700
9-22 DRP capacity overview 701
9-23 Capacity reporting in a DRP 701
9-24 Volumes in a DRP. 702
9-25 Add Volume Copy dialog 705
9-26 Synchronization progress 705
9-27 Estimate Compression Savings 708
9-28 Dashboard button for overprovisioned storage monitoring 712
9-29 View overprovisioned controllers and mdisks 712

9-30 Thin-provisioned MDisk properties 713
10-1 FlashCopy terminology 722
10-2 A simplified representation of grains and bitmap 727
10-3 CoW steps 728
10-4 Copy on-Demand steps 728
10-5 Modifying a copied grain on the source 729
10-6 Modifying a non-copied grain on the source 730
10-7 Modifying a non-copied grain on the target 731
10-8 Modifying an already copied grain on the target 731
10-9 Reading a grain on target 732
10-10 IBM Storage Virtualize software architecture 733
10-11 Incremental FlashCopy example 735
10-12 Consistency groups and mappings combinations 739
10-13 FlashCopy dependencies example 740
10-14 A reverse FlashCopy example for data restoration 744
10-15 Thin-provisioned target volume 748
10-16 FlashCopy snapshot preset example 754
10-17 Copy Services menu 755
10-18 Source and target volumes that are displayed in the FlashCopy window 756
10-19 FlashCopy Consistency Groups window 756
10-20 FlashCopy Mapping window 757
10-21 FlashCopy window 758
10-22 Creating a FlashCopy mapping with an existing target 759
10-23 Selecting source and target for a FlashCopy mapping 759
10-24 Viewing source and target at creation time 760
10-25 FlashCopy mapping preset selection 761
10-26 Select or not a consistency group for the FlashCopy mapping 762
10-27 Creating a FlashCopy mapping and creating targets 763
10-28 FlashCopy mapping preset selection 764
10-29 Select a consistency group for the FlashCopy mapping 765
10-30 Select the pool 766
10-31 Select the type of volumes for the created targets 766
10-32 Single-click snapshot creation and start 767
10-33 Selection single-click snapshot creation and start 768
10-34 Single-click clone creation and start 769
10-35 Selection single-click clone creation and start 769
10-36 Single-click backup creation and start 770
10-37 Selection single-click backup creation and start 771
10-38 Creating a FlashCopy Consistency group 771
10-39 Enter the name and ownership group of the new consistency group 772
10-40 Creating a FlashCopy mapping 772
10-41 Select source and target volumes for the FlashCopy mapping 773
10-42 FlashCopy mapping preset selection 774
10-43 FlashCopy mapping Consistency Group selection 775
10-44 Showing related volumes for a mapping, a consistency group or another volume 776
10-45 Showing related volumes list 776
10-46 Moving a FlashCopy mapping to a consistency group 777
10-47 Selecting the consistency group where to move the FlashCopy mapping 777
10-48 Removing FlashCopy mappings from a consistency group 778
10-49 Confirm the selection of mappings to be removed 779
10-50 Editing a FlashCopy mapping properties 779
10-51 Editing copy and cleaning rates of a FlashCopy mapping 780

10-52 Renaming FlashCopy mappings 780
10-53 Renaming the selected FlashCopy mappings 781
10-54 Renaming a consistency group 781
10-55 Renaming the selected consistency group 782
10-56 Deleting FlashCopy mappings 783
10-57 Confirming the selection of FlashCopy mappings to be deleted 783
10-58 Deleting a consistency group 784
10-59 Confirming the consistency group deletion 784
10-60 Starting FlashCopy mappings 785
10-61 FlashCopy mappings status and progress examples 785
10-62 Showing target volumes state and FlashCopy mappings status 786
10-63 Stopping FlashCopy mappings 787
10-64 Modifying resources allocation per I/O group 789
10-65 DNS settings 792
10-66 Enabling Cloud Tiering 794
10-67 Acknowledge that the system is not using encryption 795
10-68 Selecting your cloud service provider 795
10-69 Entering cloud service provider information 796
10-70 Cloud Connection summary 796
10-71 Enabled Transparent Cloud Tiering window 797
10-72 Cloud volumes menu 798
10-73 Cloud volumes window 798
10-74 Adding volumes to Cloud Tiering 799
10-75 Selecting if a full copy is made or if the system decides 799
10-76 Cloud Volumes list example 800
10-77 Available actions in the Cloud Volumes window 800
10-78 Deleting versions of a volume's snapshots 801
10-79 Selecting a volume to restore a snapshot from 801
10-80 Selecting a snapshot version to restore 802
10-81 Restoring a snapshot on an existing volume or on a new volume 802
10-82 Restoring a snapshot to a new volume 803
10-83 Restoring a snapshot summary 804
10-84 Multiple systems mirroring configuration example 809
10-85 Star topology 809
10-86 Triangle topology 810
10-87 Fully connected topology 810
10-88 Daisy-chain topology 810
10-89 Remote Copy consistency group 812
10-90 Write on volume in Metro Mirror relationship 815
10-91 Global Mirror write sequence 818
10-92 Colliding writes example 821
10-93 Global Mirror without change volumes 822
10-94 Global Mirror with Change Volumes 823
10-95 Metro Mirror or Global Mirror mapping state diagram 828
10-96 New admin model 842
10-97 Typical Ethernet network data flow 843
10-98 Optimized network data flow by using Bridgeworks SANSlide technology 844
10-99 Scaling of host I/O 846
10-100 Single link with only one Remote Copy portset configured in each system 851
10-101 One Remote Copy portset on each system and nodes with failover ports configured 852
10-102 Multinode systems single inter-site link with only one RC portset 853
10-103 Multinode systems single inter-site link with only one Remote Copy portset 855

10-104 Dual links with two Remote Copy portsets on each system configured 856
10-105 Multinode systems with dual inter-site links between the two systems. 857
10-106 Multinode systems with dual inter-site links between the two systems. 858
10-107 Two node systems with single inter-site link and Remote Copy portset configured. 859
10-108 Dual Links with two Remote Copy Port Groups with failover Portsets configured. 860
10-109 Deployment example 860
10-110 Deployment example 861
10-111 Remote Copy menu 863
10-112 Remote Copy panel 863
10-113 Select Remote Copy topology 864
10-114 Creating a Partnership details. 865
10-115 Fully configured FC partnership 866
10-116 Creating a Remote Copy Relationship in an existing consistency group 867
10-117 Creating Standalone Remote Copy relationships 867
10-118 Creating a Remote Copy relationship 868
10-119 Selecting the target system for the RC relationship 869
10-120 Selecting the master and auxiliary volumes 869
10-121 Add Change Volumes panel 870
10-122 Checking and adding the relationship. 871
10-123 Selecting if volumes are synchronized 871
10-124 Start copying Remote Copy relationship. 872
10-125 Creating a Remote Copy consistency group 872
10-126 Entering a name for the new consistency group. 873
10-127 Remote Copy Consistency Groups tab. 873
10-128 Renaming Remote Copy relationships 874
10-129 Renaming Remote Copy relationships 874
10-130 Renaming a Remote Copy consistency group 875
10-131 Entering new name for consistency group 875
10-132 Moving relationships to a consistency group 876
10-133 Selecting the consistency group to add the relationships to 876
10-134 Removing relationships from a consistency group 877
10-135 Confirm the removal of relationships from a consistency group 877
10-136 Starting Remote Copy relationships 878
10-137 Starting a Remote Copy consistency group 878
10-138 Switching Remote Copy relationship direction 879
10-139 Switching master-auxiliary direction of relationships changes the write access . . 879
10-140 Switched Remote Copy Relationship 880
10-141 Switching a consistency group direction 880
10-142 Switching direction of Consistency Groups changes the write access 881
10-143 Stopping a Remote Copy relationship 881
10-144 Grant access in read and write to the auxiliary volume. 882
10-145 Stopping a consistency group. 882
10-146 Grant access in read and write to the auxiliary volumes. 883
10-147 Deleting Remote Copy Relationships 883
10-148 Confirmation of relationships deletion. 884
10-149 Deleting a consistency group 884
10-150 Confirmation of a consistency group deletion 885
10-151 Modifying resources allocation 886
10-152 Current infrastructure 887
10-153 IP replication solution in Anadolu Sigorta 888
10-154 3-Site Replication overview. 891

10-155 Add replication relationship to 3-site consistency group 893
10-156 Process flow to remove a replication relationship from 3-site consistency group 894
10-157 Journaling mode 907
10-158 Cycling mode 907
10-159 Establish partnership 909
10-160 Create storage pool 910
10-161 Create provisioning policy 910
10-162 Assign provisioning policy on source system 911
10-163 Assign provisioning policy on target system 911
10-164 Links pools between systems 912
10-165 Create replication policy 912
10-166 Final setup 913
10-167 Create a new volume group on the production system and assigning a replication policy 914
10-168 Create a new volume group 914
10-169 Replication is configured automatically 915
10-170 Create volume while disconnected 916
10-171 Automatically applying replication policy 917
10-172 Verify software version 924
10-173 Creating a child pool 925
10-174 Child pool name and provisioning policy 925
10-175 Create a linked child pool 926
10-176 Create linked child pool details 926
10-177 Replication policy panel 927
10-178 Create replication policy 927
10-179 Select the primary and secondary systems and RPO 928
10-180 Verify that the replication policy is correctly created 928
10-181 Volume group panel 929
10-182 Create a volume group 929
10-183 Create an Empty Group 930
10-184 Access the volume group and selecting Policy tab 930
10-185 Assign replication policy 931
10-186 Verify that the replication policy is correctly assigned 931
10-187 Verify that the replication policy is correctly assigned on the target system 932
10-188 Volume panel 932
10-189 Create Volume panel 933
10-190 Define Volume Properties 933
10-191 Create Volumes 934
10-192 Volume creation progress bar 934
10-193 List volumes 935
10-194 Enable access from the secondary site 935
10-195 Enable access to the recovery copy 936
10-196 Restart replication from the secondary site 936
10-197 Restart replication from the primary site 937
10-198 Enable Restart replication 937
10-199 Initial copy incomplete status 938
10-200 Verify that the recovery point within the policy is 0 seconds behind the production copy 938
10-201 Update the existing partnership 941
10-202 Partnership properties 941
10-203 Verify that GM replication is converted to policy-based replication 942
10-204 Verify the current state of the consistency group 943

10-205 Stop Remote-Copy consistency group 944
10-206 State of the consistency group to "Idling" 944
10-207 Delete relationship 945
10-208 Confirm relationship deletion 945
10-209 Delete group 946
10-210 Monitoring volume groups. 947
10-211 Monitor Volume groups policies 947
10-212 Audit logs 949
11-1 LEDs on each node canister. 965
11-2 FlashSystem 7300 node canister 965
11-3 IBM FlashSystem 9500 rear view 966
11-4 Expansion canister status LEDs 970
11-5 Dense Drawer LEDs 971
11-6 Concept of SAS chaining 972
11-7 SAS cabling with numbered enclosures 973
11-8 Controller enclosure LED status indicator FlashSystem 7300 974
11-9 Expansion enclosure power supply unit 975
11-10 Removing a node by using the GUI 977
11-11 Service Assistant Tool post-node removal 978
11-12 Radio button for node actions 978
11-13 Exit service state. 979
11-14 Download Existing Package 982
11-15 Support Package Selection. 983
11-16 Filtering specific files for download 983
11-17 Settings menu 987
11-18 Update System window 987
11-19 Upload option for both the Upgrade Test Utility and update package 988
11-20 The update type selection 989
11-21 Update pause options 989
11-22 Issues that are detected by the Upgrade Test Utility 989
11-23 Description of the warning from the Upgrade Test Utility 990
11-24 Resuming the update 991
11-25 Update process paused for host path recovery 991
11-26 Node failover. 992
11-27 The update has finished 992
11-28 Upgrade Test Utility drive firmware warning 993
11-29 Internal Storage view versions 994
11-30 Select drives and update 994
11-31 Selecting files on Drive Update Checker 995
11-32 CLI command shows 995
11-33 All drives safe to update 996
11-34 CLI command shows for updating drives 996
11-35 Drive update started 997
11-36 Selecting Drive Upgrade running task view 997
11-37 Drive upgrade progress for a single drive upgrade. 998
11-38 All drives updated 998
11-39 Monitoring options. 1001
11-40 Messages in the event log 1003
11-41 Recommended Actions. 1003
11-42 Status alerts 1003
11-43 Grid options of the event log 1005
11-44 Event sense data and properties 1006
11-45 Configuring Call Home notifications 1007

11-46 Call Home Connect Cloud service welcome screen 1008
11-47 Transmission types 1009
11-48 Proxy server 1009
11-49 Location of the device 1010
11-50 Contact information 1010
11-51 Software entitlement 1011
11-52 Inventory Reporting and Configuration Reporting 1011
11-53 Configuring email servers and inventory reporting 1012
11-54 Call home settings summary 1012
11-55 Entering Edit mode 1013
11-56 Saving a modified configuration 1014
11-57 Disabling or enabling email notifications 1015
11-58 Support Assistance menu 1016
11-59 Enabling or disabling the support wizard 1017
11-60 Support wizard proxy setup 1017
11-61 Support wizard access choice 1018
11-62 Support status and session management 1018
11-63 SNMP configuration 1019
11-64 Add SNMP Server 1020
11-65 Syslog Servers menu 1021
11-66 Syslog configuration 1021
11-67 Audit Log from the Access menu 1023
11-68 Audit log 1023
11-69 How to change audit log column headings 1024
11-70 Support Package window 1025
11-71 Upload Support Package window 1026
11-72 Task detail window 1026
11-73 Support package upload to IBM 1027
11-74 Support package upload to IBM - success 1027
11-75 Downloaded Existing Package 1028
11-76 Filtering on snap to download 1029
11-77 ECuRep details 1030
11-78 ECuRep File upload 1031
11-79 Service Assistant Tool login 1032
11-80 Service Assistant Tool Login GUI 1033
11-81 Service Assistant Tool GUI 1033
11-82 SAT home window options 1034
11-83 SAT node details 1034
11-84 IBM Storage Insights System overview 1035
11-85 Component Health overview 1035
11-86 Ports in error 1036
11-87 Capacity area of the IBM Storage Insights system overview 1036
11-88 IBM Storage Insights Capacity view 1037
11-89 Pools list view 1037
11-90 System overview: Performance 1037
11-91 IBM Storage Insights: Performance view 1038
11-92 Filtered performance graph 1038
11-93 Performance List View 1039
11-94 Get Support 1039
11-95 Get Support window 1040
11-96 Create Ticket wizard 1040
11-97 Add a note or attachment window 1041
11-98 Selecting a Severity Level window 1041

11-99 Review the ticket window 1042
11-100 Ticket Creation confirmation window 1042
11-101 View Tickets 1043
11-102 Adding a log package to the ticket 1043
11-103 Confirming the log upload 1044
12-1 Accessing the System Certificates window 1048
12-2 Creating a self-signed system certificate 1049
12-3 Generating a certificate signing request 1051
12-4 Signed certificate installed 1052
12-5 Configure remote authentication window with Advanced Settings expanded 1056
12-6 Create User window 1059
12-7 Password policies - Password creation 1061
12-8 Password expiration and account lockout 1064
12-9 Allow locking of the superuser account 1065
12-10 Inactivity Logout settings window 1065
12-11 General flow of authentication 1066
12-12 Add DNS server 1067
12-13 Internal Proxy Server configuration 1067
12-14 Exporting the unique certificate 1068
12-15 ISV Certificates 1068
12-16 Add signer certificate 1069
12-17 MFA Configuration Details showing sample credentials 1070
12-18 Confirm Turn on multifactor authentication window 1070
12-19 MFA enabled system wide, navigate to user groups settings 1071
12-20 Create User Group window 1071
12-21 Create User window 1072
12-22 IBM FlashSystem GUI login screen 1073
12-23 Signing in by using user ID and password 1073
12-24 Choose a two-step verification method for authentication 1074
12-25 Touch Approval pending authentication challenge message 1074
12-26 Verification request in IBM Verify app on mobile phone 1075
12-27 Request verified in IBM Verify app 1076
12-28 Single Sign-on configuration 1077
12-29 Complete all fields as displayed 1078
12-30 Enable single sign-on 1079
12-31 Create User Group for SSO login 1079
12-32 Select User Directory 1080
12-33 Users and groups window 1080
12-34 ISV Add group window 1081
12-35 Sign In with SSO 1081
12-36 SSO provider login screen 1082
12-37 Choosing a second factor verification method 1082
12-38 Enter the TOTP 1082
12-39 SSO login successfully completed 1083
12-40 Creating the first ownership group 1085
12-41 Ownership groups management window 1085
12-42 Creating and assigning a user group 1086
12-43 Creating a user 1086
12-44 Unassigning user groups 1087
12-45 Creating a child pool 1087
12-46 Creating a child pool and assigning it to an ownership group 1087
12-47 Ownership group management window 1088
12-48 Enabling the ownership group attribute display 1088

12-49 Listing volumes for all ownership groups 1089
12-50 Ownership group administrator view. 1089
12-51 Renaming or removing an ownership group 1089
12-52 Adding a volume copy for migration 1090
12-53 Migrating to a child pool 1091
12-54 Assigning a child pool to an ownership group. 1091
12-55 Selecting a child pool to assign. 1091
12-56 Additional Resources to add window 1092
12-57 Resources of an ownership group 1092
12-58 Example of inconsistent volume ownership 1092
12-59 GUI TPI Configuration Page 1094
12-60 Warning - Permissions changed upon enabling TPI. 1094
12-61 TPI enabled. 1095
12-62 Request a time base role elevation. 1095
12-63 Duration of Role Elevation request 1096
12-64 Pending Role Elevation Request. 1096
12-65 Approve or Deny a pending Role Elevation Request 1097
12-66 Approved Role Elevation Request 1097
12-67 Encryption example 1101
12-68 Encryption activation during initial system setup 1105
12-69 Information storage system during initial system setup 1105
12-70 Selecting license activation method 1106
12-71 Successful encryption license activation during initial system setup 1106
12-72 Expanding Encryption Licenses section on the Licensed Functions window 1107
12-73 Select the node on which you want to enable the encryption. 1107
12-74 Successful encryption license activation on a running system 1108
12-75 Encryption license Activate License Automatically window 1108
12-76 Entering an authorization code 1109
12-77 Activating encryption. 1109
12-78 Successful encryption license activation. 1109
12-79 Authorization code failure 1110
12-80 Manual encryption license activation window 1111
12-81 Successful encryption license activation. 1111
12-82 License key failure 1112
12-83 Enable Encryption from the Suggested Tasks window. 1113
12-84 Enable Encryption from the Security window 1114
12-85 Enable Encryption wizard Welcome window. 1115
12-86 Selecting USB flash drives in the Enable Encryption wizard 1117
12-87 Writing the master access key to USB flash drives. 1117
12-88 Commit the encryption enablement 1118
12-89 Encryption enabled message that uses USB flash drives 1118
12-90 Encryption view showing by using USB flash drives as the enabled provider . . . 1119
12-91 Selecting Key server as the only provider in the Enable Encryption wizard 1121
12-92 Selecting SGKLM as the key server type 1121
12-93 Configuration of the primary SGKLM server 1122
12-94 Configuring multiple SGKLM servers 1123
12-95 Checking key server device group 1123
12-96 Uploading key servers or certificate authority SSL certificate. 1124
12-97 Downloading the IBM Storage Virtualize SSL certificate 1125
12-98 Finish enabling encryption by using SGKLM key servers. 1125
12-99 Encryption enabled with only SGKLM servers as encryption key providers. 1126
12-100 Selecting Key servers as the only provider in the Enable Encryption wizard . . . 1127
12-101 Selecting Gemalto SafeNet KeySecure as key server type 1128

12-102 Configuring multiple SafeNet KeySecure servers 1128
12-103 Key server credentials input (optional) 1129
12-104 Uploading SafeNet KeySecure key servers certificate 1129
12-105 Downloading the IBM Storage Virtualize SSL certificate 1130
12-106 Finish enabling encryption by using SafeNet KeySecure key servers 1130
12-107 Encryption enabled with 4 SafeNet KeySecure key servers 1131
12-108 Selecting Key servers and USB flash drives in the Enable Encryption wizard . . 1132
12-109 Selecting the key server type 1132
12-110 Prompt to insert USB flash drives 1133
12-111 Encryption enabled with both USB flash drives and key servers 1134
12-112 Enable key servers as a second provider 1135
12-113 Do not disable USB flash drive encryption key provider 1135
12-114 Configuration summary before committing 1136
12-115 Encryption enabled with two key providers available 1136
12-116 Example of encryption enabled with two key providers available 1137
12-117 Disable USB flash drive provider while changing to SGKLM provider 1138
12-118 IBM Storage Virtualize encryption configured with IBM SGKLM servers 1139
12-119 IBM SAN Volume Controller encryption configured with USB flash drives 1139
12-120 IBM SAN Volume Controller encryption configured with SafeNet KeySecure . . 1140
12-121 Create Pool window basic. 1141
12-122 Pool encryption state 1142
12-123 Mix and match encryption in a pool 1143
12-124 Create a child pool of an encrypted parent pool 1143
12-125 Child pool encryption state 1144
12-126 Array encryption state 1144
12-127 Drive encryption state 1145
12-128 MDisk encryption state 1145
12-129 Declaring MDisk as externally encrypted 1146
12-130 Overriding external encryption setting for external MDisk. 1147
12-131 MDisk self-encryption state 1147
12-132 Volume view customization. 1148
12-133 Volume encryption status depending on volume copies encryption 1148
12-134 Create an encrypted volume by selecting an encrypted pool 1149
12-135 Start rekey on SGKLM key server 1150
12-136 Start rekey on USB flash drives provider 1151
12-137 Writing new keys to USB flash drives 1151
12-138 Disabling encryption on a system with both providers 1152
12-139 Encryption disabled. 1153
13-1 REST API Explorer actions 1159
13-2 REST API Explorer authentication 1160
13-3 REST API Explorer /mkvdisk 1161
13-4 REST API Explorer /mkvdisk output 1162
A-1 PuTTY key generator 1197
A-2 Generating keys 1198
A-3 Confirming the security warning 1198
A-4 Opening the user section 1199
A-5 User properties 1200
A-6 Confirming the SSH key upload 1200
A-7 Key successfully imported 1201
A-8 PuTTY Configuration 1201
A-9 PuTTY Auto-login username 1202
A-10 SSH protocol Version 2 1202
A-11 SSH authentication 1203

A-12 Session information 1203
A-13 Connecting to a system 1204
A-14 Confirming the security alert 1204
A-15 PuTTY login 1205

Tables

1-1 IBM Storage Virtualize V8.6 supported product list 2
1-2 IBM SAN Volume Controller base models 47
1-3 Historical overview of IBM SAN Volume Controller models 48
1-4 IBM FlashSystems 7200 and 9200 product range 52
1-5 Software PIDs and SWMA feature codes 52
1-6 Billing calculations that are based on customer usage 57
1-7 IBM FlashSystem 9500 Utility Model UG8 billing feature codes 57
1-8 IBM FlashSystem 9500R Rack Solution combinations 60
1-9 Key to rack configuration 60
1-10 IBM FlashSystem 5200 host, drive capacity, and functions summary 73
1-11 Machine type and model comparison for the IBM FlashSystem 5000 76
1-12 IBM FlashSystem 5015 host, drive capacity, and functions summary 77
1-13 2.5-inch supported drives for the IBM FlashSystem 5000 family 77
1-14 3.5-inch supported drives for the IBM FlashSystem 5000 family 78
1-15 IBM FlashSystem 5035 host, drive capacity, and functions summary 78
1-16 IBM FlashSystem 5045 host, drive capacity, and functions summary 81
1-17 Volume types that are available in pools 86
1-18 FCM type capacities 101
1-19 NVMe drive size options 102
1-20 SCM drive options 103
1-21 SCU category definitions 116
2-1 Communication options 127
2-2 Communication options 127
2-3 HyperSwap attributes 145
2-4 Summary of supported drives, array types, and RAID levels 151
2-5 Summary of supported hardware platforms and RAID levels 151
2-6 Minimum overhead capacity requirements for DRPs 153
2-7 Limits of the cache data 156
2-8 Supported hosts and connections types for nondisruptive system migration 171
4-1 SCU ratio per storage type 322
4-2 Examples of allocation of bitmap memory 330
4-3 Supported network configurations for Call Home with cloud services 345
7-1 Configuration and restriction for Storage Virtualize product family 520
7-2 IBM Storage Virtualize NPIV ports 522
7-3 Configuration limitations 530
7-4 Emulex or Mellanox HBAs 531
8-1 New hardware-clustering options for IBM Storage Virtualize storage systems. 637
8-2 Example table for capturing external LUN information 646
9-1 Storage tier to Easy Tier mapping 672
9-2 Maximum physical capacity after data reduction 699
9-3 Minimum capacity in a single DRP 699
9-4 Using data reduction at two levels 710
10-1 Maximum Configuration updates 718
10-2 Copy rate values 734
10-3 FlashCopy mapping state summary 738
10-4 Sequence example of write IOs on a source with multiple targets 740
10-5 Summary table of the FlashCopy indirection layer algorithm t 742
10-6 Cleaning rate values 743

© Copyright IBM Corp. 2023.

10-7 Mapping events 746
10-8 FlashCopy and remote copy interaction 751
10-9 FlashCopy limitations in V8.6.0 752
10-10 Memory allocation for FlashCopy services 788
10-11 Intersystem heartbeat traffic in Mbps 807
10-12 Supported Global Mirror link latencies 819
10-13 Valid combination for a single volume 827
10-14 Metro Mirror configuration limits 827
10-15 Terminology for IP partnership 848
10-16 States of IP partnership 849
10-17 Memory allocation for Remote Copy services 885
10-18 Snapshot volumes by pool type 897
10-19 Comparison between volume group snapshot and SafeGuarded Copy snapshot 903
10-20 Comparison and contrast between Global Mirror and policy-based replication (PBR) 905
10-21 Policy-based replication attributes 908
10-22 Longest RTT supported by IBM FlashSystem storage-to-storage connectivity . . . 918
10-23 Required I/O group memory for policy-based replication 920
10-24 Default and maximum setup of the bitmap space 920
10-25 Supported targets and volume quantities and capacity limit 923
10-26 Volume group statistics 956
10-27 Node statistics 956
10-28 Volume statistics 960
11-1 Supported card configurations for IBM FlashSystem 7300 966
11-2 Supported card configurations for IBM FlashSystem 5200 967
11-3 Supported card configurations for IBM FlashSystem 9500 and SV3 967
11-4 Fibre Channel link LED statuses 968
11-5 Ethernet ports and their functions 968
11-6 Ethernet LED statuses 968
11-7 SAS LED statuses 969
11-8 Node canister LEDs 969
11-9 Battery LEDs 970
11-10 Expansion canister LEDs statuses 971
11-11 Expansion canister LEDs statuses 972
11-12 Files that are created by the backup process 981
11-13 The iogrp setup 999
11-14 Types of snaps 1027
13-1 REST API rate limits 1159
13-2 Options of the curl command 1172
13-3 Syntax comparison 1176
13-4 Variable parameters and their values for the example playbook 1187
C-1 Demonstration videos 1243

Examples

3-1 Reenabling the onboard Ethernet port 2 as the technician port 179
3-2 Listing node IPs currently not set 209
3-3 Running commands to change node IP 209
3-4 Changed node IP (output shortened for clarity) 209
3-5 Listing the I/O groups 212
3-6 Listing the candidate control enclosures 212
3-7 Adding a control enclosure 212
3-8 Adding an expansion enclosure 212
3-9 Listing I/O groups 215
3-10 Listing the candidate nodes 215
3-11 Adding a node as a spare 215
3-12 Adding a node to an I/O group 215
3-13 Single IO-group (two nodes) and one spare 215
3-14 Two IO-groups (four nodes) configured- no spare 216
3-15 Starting the IP quorum application on the Windows operating system 222
3-16 Viewing the local port mask 225
3-17 Setting a local port mask by running the chsystem command 225
4-1 Compact syslog message example 290
4-2 Full format syslog message example 290
5-1 The lsmdiskgrp output (some columns are not shown). 368
5-2 The mkmdiskgrp command. 370
5-3 Using chmdiskgrp to rename a storage pool 372
5-4 Changing the warning threshold level by using the CLI 373
5-5 Setting a storage pool throttle by using the mkthrottle command 375
5-6 Removing a pool throttle by running the rmthrottle command 375
5-7 Using lsmdisk (some columns are not shown) 376
5-8 The lsmdiskgrp output (partially shown) 377
5-9 The mkmdiskgrp command to create child pools 379
5-10 Running the chmdiskgrp command to rename a child pool 381
5-11 lsdrive output (some lines and columns are not shown) 387
5-12 Setting drive offline with CLI 389
5-13 Changing drive role with CLI 390
5-14 Changing slot LED to identification mode with CLI 391
5-15 Listing volumes dependent on drives with the CLI 393
5-16 Listing array recommendations by using the CLI 400
5-17 Creating DRAID with mkdistributedarray 400
5-18 Renaming array MDisk with charray 401
5-19 Replacing array member with CLI (some columns are not shown). 402
5-20 Expanding an array by using the CLI 403
5-21 Listing VDisks that depend on MDisk with CLI 405
5-22 lsarray output (truncated) 407
5-23 Listing controllers by using the CLI (some columns are not shown) 409
5-24 Changing a controller's name and site 410
5-25 The addmdisk command. 413
5-26 Changing the tier setting by using the CLI 414
5-27 Using chmdisk to modify encryption 414
5-28 Including a degraded MDisk by using the CLI 415
5-29 Listing VDisks that depend on an MDisk by using the CLI 416

© Copyright IBM Corp. 2023.

6-1 The lsvdisk command 488
6-2 The mkvdisk command 489
6-3 The lsvdisk command 490
6-4 Running the mkvdisk command 491
6-5 The mkvdisk (image mode) command 492
6-6 The lsvdisk command 493
6-7 The lsvdisk command 493
6-8 The addvdiskcopy command 495
6-9 Synchronization 495
6-10 The lsvdisk command 495
6-11 Volume name changes 498
6-12 A decompressed volume 498
6-13 Compressed copy 498
6-14 The lsvdisk command output 499
6-15 Splitting a volume 499
6-16 The lsvdisk command 499
6-17 The rmvdisk command 502
6-18 The rmvdisk -force command 502
6-19 The expandvdisksize command 503
6-20 The lsvdisk command 503
6-21 The mkvdiskhostmap command 505
6-22 The lshostvdiskmap -delim command 506
6-23 The mkvolumehostclustermap command 507
6-24 The lshostvdiskmap command 507
6-25 The lshostclustervolumemap command 507
6-26 The lsvdiskhostmap command 507
6-27 The rmvdiskhostmap command 508
6-28 The rmvolumehostclustermap command 508
6-29 The migratevdisk command 509
6-30 The lsmigrate command 509
6-31 The migratetoimage command 510
6-32 The shrinkvdisksize command 511
6-33 The lsmdiskmember command 511
6-34 The lsvdiskmember command 511
6-35 The lsvdisk -filtervalue command: Volumes in the pool 511
6-36 The lsvdisk command: Storage pool ID and name 512
6-37 Volume ID returned by the multipath -ll command 514
6-38 The lshostvdiskmap command 515
6-39 The lsvdiskmember command 515
6-40 The lsmdisk command 515
7-1 Listing the I/O groups in the system 524
7-2 Checking the NPIV mode by viewing the fctargetportmode field 525
7-3 Listing the virtual WWPNs 525
7-4 Changing the NPIV mode to enabled 525
7-5 Running the lstargetportfc command to get the primary host WWPNs (virtual WWPNs) 526
7-6 Enabling transitional mode for NPIV 526
7-7 Host attach WWPNs (virtual WWPNs) permitting host traffic 527
7-8 Established host zone 527
7-9 Transitional host zone (added host attach ports are in bold) 527
7-10 Host device pathing: Before and after 528
7-11 Enabling the NPIV 529
7-12 Final host zone 529

7-13 FC portset creation by way of CLI 540
7-14 Assigning FC port to FC portset 542
7-15 The use of chhost CLI command 548
7-16 Removing an FC port from an FC portset 550
7-17 Removing FC portset 553
7-18 I/O blocked over misconfigured port 554
7-19 I/O active over included FC ports 555
7-20 The lsiogrp command 569
7-21 Rescanning the SAN 597
7-22 Available WWPNs 597
7-23 Host creation 597
7-24 Creating an iSCSI host by running the mkhost command 598
7-25 Verifying the iSCSI host by running the lshost command 598
7-26 The mkhost command 598
7-27 The lshost command 599
7-28 The mkhost command 599
7-29 The lshost command 599
7-30 The mkhost command 600
7-31 The lshost command 600
7-32 Mapping a volume 601
7-33 Checking the mapped volume 601
7-34 Mapping the same volume to a second host 601
7-35 Ensuring that the same volume is mapped to multiple hosts 602
7-36 Unmapping a volume from a host 602
7-37 Renaming a host 602
7-38 Removing a host 602
7-39 Host details 602
7-40 Listing the newly available WWPN 603
7-41 Adding the newly discovered WWPN to the host definition 604
7-42 Adding a WWPN to the host definition by using the -force option 604
7-43 Host with the updated port count 604
7-44 Adding an iSCSI port to the defined host 604
7-45 The addhostport command 604
7-46 Running the lshost command to check the WWPNs 605
7-47 Running the rmhostport command to remove a WWPN 605
7-48 Removing the iSCSI port from the host 605
7-49 Removing the NQN port from the host 605
7-50 Creating a host cluster 606
7-51 Adding a host or hosts to a host cluster 606
7-52 Listing host cluster members by running the lshostclustermember command 606
7-53 Mapping a volume to a host cluster 606
7-54 Listing volumes that are mapped to a host 606
7-55 Removing a volume mapping 607
7-56 Removing a host cluster member 607
7-57 Removing a host cluster 607
7-58 Adding a host cluster to an ownership group 608
7-59 Removing a host cluster from an ownership group 608
7-60 RHEL release check 608
7-61 Discovering the hosts' WWPNs 609
7-62 Scanning and rebuilding the multipath 610
7-63 Creating a physical volume in LVM for further use 612
7-64 Installing iscsi-initiator-utils 613
7-65 Checking the initiator's IQN 613

7-66 iSCSI targets discovery 614
7-67 Logging in to the discovered targets/storage 615
7-68 Multipathing driver/device mapper output 615
7-69 Obtaining host WWPNs 617
7-70 Discovering the NVMe FC ports 617
7-71 Checking NVMe support for the lpfc driver 617
7-72 Checking the nvme-cli and nvme-connect availability 618
7-73 Obtaining the NQN 618
7-74 Verifying the remote/target ports and information about the FC-NVMe connection 618
7-75 Script for FC-NVMe discovery and connection 619
7-76 Discovery and connection script output 619
7-77 NVMe devices list that is visible from the host 621
7-78 Output of the multipath -ll command 621
7-79 Checking host RDMA cards 622
7-80 Installing the required RDMA packages on host 623
7-81 Finding RDMA NICs 623
7-82 Loading the required RDMA and NVMe drivers 623
7-83 Assigning IP addresses to RDMA NICs 623
7-84 Ensuring RDMA drivers are loaded 624
7-85 Obtaining the host nqn 624
7-86 Defining NVMe over RDMA host 624
7-87 Discovering the RDMA targets over NVMe from host 624
7-88 Connecting to target from host using nvme cli 625
7-89 List storage subsystem as seen by the host 625
7-90 Native nvme multipath off for RHEL 626
7-91 Checking multipath daemon status 626
7-92 Multipathing storage volume on host 626
7-93 Multipathed volume 627
7-94 checking host TCP cards 628
7-95 Installing the required packages on host 628
7-96 Finding NICs 628
7-97 Loading the required NVMe drivers 628
7-98 Assigning IP addresses to NICs 628
7-99 Ensuring that drivers are loaded 629
7-100 Obtaining the host nqn 629
7-101 Defining NVMe over TCP host 629
7-102 Discovering the NVMe targets from host 629
7-103 Connecting to target from host using nvme cli 630
7-104 List storage subsystem as seen by the host 630
7-105 Native nvme multipath off for RHEL 631
7-106 Checking multipath daemon status 631
7-107 Multipathing storage volume on host 631
7-108 Multipathed volume 631
8-1 Migration progress on the command line interface 654
9-1 Listing and changing tiers for MDisks (partially shown) 679
9-2 Changing the Easy Tier load 680
9-3 Listing and changing the Easy Tier status on pools 681
9-4 Checking and modifying the Easy Tier settings on a volume 683
9-5 The chsystem command 683
9-6 Running IBM STAT by using the Windows command prompt 688
9-7 Verifying the back-end UNMAP support status 693
9-8 Turning on host UNMAP support 694

9-9 Data Reduction Pool volume capacity reporting on the CLI 702
9-10 Data Reduction Pool capacity reporting on the CLI 703
9-11 The lsmdisk parameters for thin-provisioned MDisks 713
10-1 Listing the size of a volume in bytes and creating a volume of equal size 745
10-2 lsdnsserver command 792
10-3 Output from the lssystem command showing the system layer 808
10-4 New commands 842
10-5 Add stand-alone relationship to a 3-site consistency group 892
10-6 svctask with mkvolumegroup flag 895
10-7 svctask with addsnapshot flag 895
10-8 svctask with addsnapshot flag 898
10-9 lsvdisk with showhidden flag 898
10-10 svctask with rmsnapshot flag 900
10-11 Create thin clone 901
10-12 Create Thick Clone 902
10-13 Add snapshot for orphaned volume group 902
10-14 Create snapshot policy 904
10-15 List snapshot policy 904
10-16 Assign policy to a volume group 904
10-17 Information about the status of volume group using lsreplicationpolicy command 948
10-18 Details about the status of volume group using lsreplicationpolicy command 948
10-19 Information about the status of volume group using lsreplicationpolicy command 949
10-20 lssystemlimits command 950
10-21 lsvdisk command 951
10-22 lsvolumegroup command 951
10-23 lsmdiskgrp command 952
10-24 lssystem command 952
10-25 lsdumps command 955
10-26 Retrieve /dumps/iostats directory 955
10-27 Volume group statistics 956
10-28 Nodes statistics 957
10-29 Volume statistics 960
11-1 The lsnode output 976
11-2 The rmnodecanister command 976
11-3 The rmnode command 977
11-4 The lsnode output after removing a node canister 977
11-5 The addncontrolenclosure command 979
11-6 The svcinfo lsnode command 979
11-7 Saving the configuration by using the CLI 980
11-8 Listing the backup files by using the CLI 980
11-9 Saving the config backup files to your workstation 981
11-10 Example of using PSCP to download update files (output shortened for clarity) . . 985
11-11 The svc_snap command 1028
12-1 Generating a self-signed certificate 1050
12-2 TLS Certificate in PEM format 1051
12-3 Manually lock and unlock a user account 1062
12-4 Manually enable superuser account locking option 1062
12-5 Manually lock the superuser account 1063
12-6 MFA prompt at CLI login 1076
12-7 Creating Security Admin user fails with TPI enabled 1095
12-8 Eventlog entry 3375/009219 1097

12-9 Creating an unencrypted array by using CLI with IBM SAN Volume Controller . . . 1144
13-1 JSON notation for creating a thin provisioned mirrored VDisk 1158
13-2 Using ssh and grabbing selected information 1165
13-3 Using PuTTY PLink. 1165
13-4 Using paramiko within Python. 1166
13-5 Using Net::OpenSSH within Perl 1167
13-6 Basic use of pywbem 1168
13-7 Parsing EnumerateInstances() output for classes cluster, nodes, and storage pools. 1169
13-8 Querying only required data by using the ExecQuery() method 1170
13-9 Accessing performance metrics by using the PyWBEM module 1170
13-10 Creating a JSON Web Token (JWT). 1172
13-11 Get all M0Disks. 1173
13-12 Piping the output to python for getting better readable JSON output 1173
13-13 Using the REST API by using Python. 1174
13-14 Using the REST API by using Perl 1175
13-15 Using the REST API by using Windows PowerShell 1177
13-16 Using the REST API by using PowerShell Core 1178
13-17 Using the REST API by using Windows PowerShell - providing data as body. . . 1179
13-18 Using the REST API by using Windows PowerShell - passing boolean values . . 1179
13-19 Using the IBM Storage Control REST API by using Python 1180
13-20 Example displaying online help. 1186
13-21 YAML notation for obtaining an authentication token 1188
13-22 YAML notation for creating an empty host cluster 1188
13-23 YAML notation for creating a new FC host object. 1189
13-24 YAML notation to create a thin-provisioned volume 1189
13-25 YAML notation to map a volume to the hostcluster 1189
13-26 Complete playbook for specified use-case 1190
A-1 SSH keys generation with ssh-keygen 1205
A-2 SSH public key copy to an IBM Storage System 1206
A-3 Importing the SSH public key to a user 1206
A-4 Connecting to IBM Storage System with an SSH private key 1207

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

© Copyright IBM Corp. 2024.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at https://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®
Cognos®
Db2®
DS8000®
Easy Tier®
FlashCopy®
Guardium®
HyperSwap®
IBM®
IBM Cloud®
IBM FlashCore®
IBM FlashSystem®
IBM Research®
IBM Security®
IBM Spectrum®
Passport Advantage®
PIN®
PowerHA®
QRadar®
Redbooks®
Redbooks (logo) ®
Storwize®
Tivoli®
WebSphere®
XIV®

The following terms are trademarks of other companies:

Intel, Intel Xeon, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

ITIL is a Registered Trade Mark of AXELOS Limited.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Ansible, OpenShift, Red Hat, are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

VMware, VMware vSphere, and the VMware logo are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.

Other company, product, or service names may be trademarks or service marks of others.

Preface

IBM® Storage FlashSystem and IBM SAN Volume Controller solutions incorporate some of the top IBM technologies that are typically found only in enterprise-class storage systems, which raise the standard for storage efficiency in midrange disk systems. These cutting-edge storage systems extend the comprehensive storage portfolio from IBM and can help change the way organizations address the ongoing information explosion.

This IBM Redbooks® publication introduces the features and functions of an IBM Storage Virtualize V8.6 storage system through several examples. This book is aimed at pre-sales and post-sales technical support and marketing and storage administrators. It helps you understand the architecture, how to implement it, and how to take advantage of its industry-leading functions and features.

Demonstration videos: In addition to the written content, you can find the links for several demonstration videos throughout the book. Most of these videos are created by the authors of this IBM Redbooks publication as part of the project. We hope that these videos help reinforce the key concepts highlighted in this book. A full list of these videos can be found in Appendix C, “List of the demonstration videos” on page 1243. You can also visit https://www.redbooks.ibm.com/feature/storagevideos.

Note: Consider the following points:

- This book is focused on IBM Storage FlashSystem and IBM SAN Volume Controller systems that run IBM Storage Virtualize Version 8.6. Most configuration principles and best practices that are described in the book are still applicable for systems that cannot be upgraded to Version 8.6; however, some features might not be available.

In this book, we provide links to IBM Documentation and a description of the relevant section that provides more information. In most cases, our starting point is the IBM FlashSystem 9500 family web page, and the reader might need to select the product that applies to their environment.

© Copyright IBM Corp. 2024.

Authors

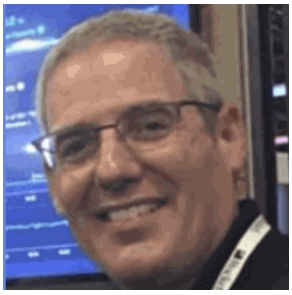
This book was produced by a team of specialists from around the world.



Tiago Moreira Candelaria Bastos is a storage area network and Storage Disk specialist at IBM Kyndryl Brazil. He has over 20 years in the IT arena and is an IBM Certified Master IT Specialist. Certified for IBM Storwize, he works on storage as service implementation projects. His areas of expertise include planning, configuring, and troubleshooting IBM DS8000, IBM FlashSystem, IBM SAN Volume Controller, and IBM XIV lifecycle management and copy services.



Nezih Boyacıoglu has 20 years of experience as a SAN Storage specialist and currently works for IBM Premier business partner Istanbul Pazarlama in Turkey. His IBM storage journey starts with Tivoli® Storage Manager (Spectrum Protect) and tape systems and his main focus for the last 10 years has been on the IBM Storage Virtualize family (IBM SAN Volume Controller, Storwize, and FlashSystem), and Storage Area Networks. He is an IBM Certified Specialist for Enterprise Storage Technical Support, Flash Technical Solutions, Virtualized Storage, and IBM Storage Virtualize software.



Andrew Greenfield is an IBM Global XIV® and Flash Solution Engineer who is based in Phoenix, Arizona. He holds numerous technical certifications from Cisco, Microsoft, and IBM. Andrew brings over 25 years of data center experience inside the Fortune 100 to the team. He graduated with magna cum laude, honors, from the University of Michigan, Ann Arbor. Andrew has also written for and contributed to several IBM Redbooks publications.



Byron M Grossnickle is a member of the IBM Advanced Technology Group and serves North America as a Subject Matter Expert on Storage Virtualize and IBM FlashSystem®. He helps technical sales teams engineer IBM storage and communicate its benefits. He also trains technical sales teams worldwide. Byron has worked in IBM Storage for 18 years. Prior to that, he engineered storage for the data centers of a large national telecommunications provider. Byron has also worked a number of years in IT in the healthcare industry.



Vasfi Gucer is a project leader with the IBM Redbooks Team. He has more than 20 years of experience in the areas of systems management, networking hardware, and software. He writes extensively and teaches IBM classes worldwide about IBM products. His focus has been primarily on storage and cloud computing for the last eight years. Vasfi is also an IBM Certified Senior IT Specialist, Project Management Professional (PMP), IT Infrastructure Library (ITIL) V2 Manager, and ITIL V3 Expert.



Jon Herd is an IBM Senior Executive Advocate working for the TLS EMEA Remote Technical Support and Client Care team based in IBM Germany. He covers the United Kingdom, Ireland and beyond, advising customers on a portfolio of IBM storage products, including FlashSystem products. He also works as a senior advisor to the TLS EMEA RTS/CC management on new products, strategy and new technologies that might affect the TLS business. Jon has been with IBM for more than 45 years, and has held various technical roles, including Europe, Middle East, and Africa (EMEA) level 2 support on mainframe servers and technical education development. He has written many IBM Redbooks on the FlashSystems products and is an IBM Redbooks Platinum level author. He holds IBM certifications in Product Services profession at a thought leader L3 level, and is a Technical Specialist at an experienced L1 level. He also is a certified Chartered Member of the British Computer Society (MBCS - CITP), a Certified Member of the Institution of Engineering and Technology (MIET), and a Certified Technical Specialist of the Open Group (TOG).



Mert Korcum is a brand technical specialist for storage with the infrastructure team in Turkey. He has been with IBM for more than 10 years and has held various technical roles in the IBM Support Team, focusing on power systems, SAN, storage software, open source, and Linux. Before his current role, he was a client availability leader, technical advocate, and client architect for all select dedicated accounts in Turkey. He is an open-minded lifelong learner who loves and learns all aspects of technology.



Sergey Kubin is a Senior Storage Support Engineer working in GBM Qatar. He holds an Electronics Engineer degree from Ural Federal University in Russia and has more than 15 years of experience in IT. In GBM, he provides support and guidance for customers using IBM and multi-vendor storage solutions. His expertise includes file and block storage, and storage area networks.



Youssef Largou is the founding director of PowerM, a platinum IBM Business Partner in Morocco. He has 22 years of experience in systems, HPC, middleware, and hybrid cloud, including IBM Power, IBM Storage, IBM Spectrum®, IBM WebSphere®, IBM Db2®, IBM Cognos®, IBM WebSphere Portal, IBM MQ, ESB, IBM Cloud® Pak, and Red Hat OpenShift. He has worked within numerous industries with many technologies. Youssef is an IBM Champion 2020, 2021, 2022 and 2023, an IBM Redbooks Platinum Author, and has designed many reference architectures. He has been recognized as an IBM Beacon Award Finalist in Storage, Software-Defined Storage, and LinuxONE five times. He holds an engineer degree in Computer Science from the Ecole Nationale Supérieure des Mines de Rabat and Excecutif MBA fromEMLyon.



Carsten Larsen is an IBM Certified Senior IT Specialist working for the Technical Services Support organization at IBM Denmark, where he delivers consultancy services to IBM clients within the storage arena. Carsten joined IBM in 2007 when he left HP, where he worked with storage arrays and UNIX for 10 years. While working for IBM, Carsten obtained several Brocade and NetApp certifications. Carsten is the author of several IBM Redbooks publications.



Guillaume Legmar has been involved with IBM storage solutions for more than 18 years. He is a part of the Montpellier Garage to develop demos about FlashSystem and cyber resiliency. He is also a member of the FlashSystem Beta team and virtualize beta team.



Hartmut Lonzer is a Storage Advisory Partner Technical Specialist for DACH. Before this position, he was OEM Alliance Manager for Lenovo in IBM Germany. He works at the IBM Germany headquarter in Ehningen. His main focus is on the IBM FlashSystem Family and the IBM SAN Volume Controller. His experience with the IBM SAN Volume Controller and IBM FlashSystem products goes back to the beginning of these products. Hartmut has been with IBM in various technical and sales roles now for 45 years.



Corne Lottering is a Storage Client Technical Specialist in the US, focusing on technical sales in Texas and Louisiana within the Public Sector industry. He has been with IBM for more than 20 years, and has experience in various storage technologies, including the IBM System Storage DS5000, IBM DS8000®, IBM® Storwize®, XIV, FlashSystems, IBM SAN switches, IBM Tape Systems, and Software Defined Storage software. Since joining IBM, he has fulfilled roles in support, implementation, and pre-sales support across various African and Middle Eastern countries. Corne is the author of several IBM Redbooks publications that are related to the midrange IBM System Storage DS Storage Manager range of products, and IBM FlashSystem products.



John Nycz is an Advanced Subject Matter Expert for IBM Storage Virtualize and IBM FlashSystem. He has more than 10 years of experience in the areas of systems management, networking hardware, and software. John has been with IBM for more than 20 years and has been a member of numerous development, project management, and support teams. For the last seven years, he has been a member of IBM's Storage Virtualization Support Team.



Denis Olshanskiy is a Storage Specialist with a demonstrated history of working in the IT and services industry. His areas of expertise include storage area networks (SANs), data centers, storage solutions, Arduino, and Linux. He has a master's degree in Mechatronics, Robotics, and Automation Engineering from Budapest University of Technology and Economics.



Uygar Sahin is a system engineer with 14 years of experience in the IT industry. He has a master's degree in artificial intelligence and anomaly detection from Kadir Has University. Uygar is currently working as a Senior Systems Specialist at Anadolu Sigorta. He has held system administrator roles in some of Turkey's largest financial companies. He also worked at IBM Turkey Lab Services. He was a speaker at IBM Technical University in 2020. He has extensive experience with IBM Storage Systems and IBM Storage Virtualization. Uygar is focused on UNIX/Linux systems, containerized systems, automation, open source, Power Systems, storage systems, and SAN.



Uwe Schreiber is a Solution Architect at SVA System Vertrieb Alexander GmbH. He has been working with Storage Virtualize and IBM SAN Volume Controller since 2002 (until 2011 as customer and since 2012 as a Business Partner employee). Uwe is an experienced professional providing technical pre-sales and post-sales solutions for IBM server and storage systems since 1995. He holds an engineering diploma in Computer Sciences from the University of Applied Science in Darmstadt, Germany.



Christian Schroeder provides support with passion to clients worldwide, covering IBM storage products of the Storage Virtualize and FlashSystem products family. He has been with IBM for more than 23 years, working in product support for various platforms as IBM System x servers & Bladecenter, SAN switches, IBM SAN Volume Controller, and IBM FlashSystem. Christian has co-authored IBM Redbooks publications that cover IBM System x servers and SAN Volume Controller V6.3.0 & V8.4.2.

Thanks to the following people for their contributions to this project:

Martin Keen, Jeffrey Bisti, Elias Luna
IBM USA

Pravin Mahajan
IBM India

Reza Fanaei Aghdam
Lenovo

Thanks to the authors of the previous edition of this book: *Implementation Guide for IBM Spectrum Virtualize Version 8.5*, SG24-8520, published on 02 June 2022:

Barry Whyte, Dharmesh Kamdar, Konrad Trojok, Mandy A Stevens, Paresh Chudasma, Reginald D'Souza, Ron Verbeek, Sabine Gronert, and Stephen Solewin

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an IBM Redbooks residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, IBM Redbooks
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on LinkedIn:

  https://www.linkedin.com/groups/2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/subscribe

- Stay current on recent Redbooks publications with RSS Feeds:

  https://www.redbooks.ibm.com/rss.html

1

Introduction and system overview

This chapter defines the concept of storage virtualization and provides an overview of its application in addressing the challenges of modern storage environment. It also contains an overview of each of the products that make up the IBM SAN Volume Controller and IBM Storage FlashSystem families.

This chapter includes the following topics:

- 1.1, “IBM Storage Virtualize” on page 2
- 1.2, “IBM SAN Volume Controller architectural overview” on page 8
- 1.3, “Changes and enhancements in version 8.6.0” on page 13
- 1.4, “IBM SAN Volume Controller family” on page 23
- 1.5, “IBM SAN Volume Controller models” on page 43
- 1.6, “IBM FlashSystem family” on page 48
- 1.7, “IBM FlashSystem 9500 overview” on page 53
- 1.8, “IBM FlashSystem 9500R Rack Solution overview” on page 59
- 1.9, “IBM FlashSystem 9000 Expansion Enclosure Models AFF and A9F” on page 62
- 1.10, “IBM FlashSystem 7300 overview” on page 66
- 1.11, “IBM FlashSystem 5200 overview” on page 71
- 1.12, “IBM FlashSystem 5000 family overview” on page 75
- 1.13, “Storage efficiency and data reduction features” on page 84
- Figure 1.14 on page 90
- 1.15, “Copy services” on page 96
- 1.16, “IBM FlashCore Module drives, NVMe SSDs, and SCM drives” on page 100
- 1.17, “Storage virtualization” on page 103
- 1.18, “Business continuity” on page 107
- 1.19, “Management and support tools” on page 113
- 1.20, “Licensing” on page 115

© Copyright IBM Corp. 2024. All rights reserved.

IBM Storage rebranding In January 2023, IBM announced that it renamed its Spectrum software-defined storage products to IBM Storage products. This change was made to simplify the product portfolio and make it easier for customers to find the products they need. For example, IBM Spectrum Virtualize was renamed to IBM Storage Virtualize. You can likely find documentation under both the Spectrum and Storage names for some time, as the transition to the new names occurs over a period of several months. However, all future documentation will be under the IBM Storage name.

1.1 IBM Storage Virtualize

IBM Storage Virtualize (previously known as IBM Spectrum Virtualize) is a key member of the IBM Storage portfolio. It is a highly flexible storage solution that enables rapid deployment of block storage including storage virtualization, for new and traditional workloads, on-premises, off-premises, or a combination of both.

Note: For more information, see IBM Storage FlashSystem and IBM SAN Volume Controller.

With the introduction of the IBM Storage family, the *software* that runs on IBM SAN Volume Controller and on IBM Storage FlashSystem (IBM FlashSystem) products is called IBM Storage Virtualize. The name of the underlying hardware platform has not changed.

IBM FlashSystem Storage Systems and IBM SAN Volume Controllers are built with award-winning IBM Storage Virtualize software that simplifies infrastructure and eliminates the differences in management, function, and even hybrid multicloud support.

IBM Storage Virtualize is an offering that is available for the IBM SAN Volume Controller and IBM FlashSystem family of storage solutions. It provides a way to manage and protect huge volumes of data from mobile and social applications, enable rapid and flexible cloud services deployments, and deliver the performance and scalability that is needed to gain insights from the latest analytics technologies.

Note: This version of the IBM Redbooks includes systems that can run IBM Storage Virtualize V8.6. Some products that are listed in the book are no longer sold by IBM but can still run the V8.6 software. Where this is applicable, it is mentioned in the text.

Table 1-1 shows the IBM Storage Virtualize V8.6 supported product list and whether the product is still currently sold or is designated as End-of-Marketing (EOM).

Table 1-1 IBM Storage Virtualize V8.6 supported product list

Product	Machine Type	Model	Comment
FS9500/R	4666, 4983	AH8, UH8	Current Product
FS7300	4657	924, U7D	Current Product
FS5200	4662	6H2, UH6	Current Product
FS5000 (FS5015, FS5035)	2072, 4680	2N2, 2N4, 3N2, 3N4	Current Product
FS5000 (FS5045)	4680	3P2, 3P4	Current Product

Product	Machine Type	Model	Comment
SVC	2145, 2147	SA2, SV3	Current Product
SVC	2145, 2147	SV2	EOM 01/2023
FS9200	9846, 9848, 4666	AG8, UG8	EOM 07/2022
FS7200	2076, 4664	824, U7C	EOM 07/2022
FS9100	9846, 9848	AF8, UF8	EOM 07/2022

Benefits of IBM Storage Virtualize

IBM Storage Virtualize delivers benefits that improve storage infrastructure in many ways, including the following examples:

- Reduces the cost of storing data by increasing the use and accelerating applications to speed business insights. To achieve this goal, the solution provides the following functions:
  - Uses data reduction technologies to increase the amount of data that you can store in the same space
  - Enables rapid deployment of cloud storage for disaster recovery (DR) along with the ability to store copies of local data
  - Moves data to the most suitable type of storage based on policies that you define by using IBM Storage Control to optimize storage
  - Improves storage migration performance so that you can do more with your data
- Protects data from theft or inappropriate disclosure while enabling a high availability (HA) strategy that includes protection for data and application mobility and DR. To achieve this goal, the solution provides the following functions:
  - Uses software-based encryption to improve data security
  - Provides fully duplexed copies of data and automatic switchover across data centers to improve data availability
  - Eliminates storage downtime with nondisruptive movement of data from one type of storage to another type.
- Simplifies data by providing a data strategy that is independent of your choice of infrastructure, which delivers tightly integrated functions and consistent management across heterogeneous storage. To achieve this goal, the solution provides the following functions:
  - Integrates with virtualization tools, such as VMware vCenter to improve agility with automated provisioning of storage and easy deployment of new storage technologies
  - Enables supported storage to be deployed with Kubernetes and Docker container environments, including Red Hat OpenShift
  - Consolidates storage, regardless of the hardware vendor for simplified management, consistent functions, and greater efficiency
  - Supports common capabilities across storage types, which provide flexibility in storage acquisition by allowing a mix of vendors in the storage infrastructure

Note: These benefits are a subset of the list of features and functions that are available with IBM Storage Virtualize software.

1.1.1 Storage virtualization terminology

Storage virtualization is a term that is used extensively throughout the storage industry. It can be applied to various technologies and underlying capabilities. In reality, most storage devices technically can claim to be virtualized in one form or another. Therefore, this chapter starts by defining the concept of storage virtualization as it is used in this book.

Storage virtualization is described in the following ways:

- Storage virtualization is a technology that allows different storage resources with different underlying capabilities to be managed and consumed by using a common set of interfaces.
- Storage virtualization is a logical representation of resources that is not constrained by physical limitations and hides part of the complexity. It also adds or integrates new functions with services, and can be nested or applied to multiple layers of a system.

The virtualization model consists of the following layers:

- Application. The user of the storage domain.
- Storage domain:
  - File, record, and namespace virtualization and file and record subsystem
  - Block virtualization
  - Block subsystem

Applications typically read and write data as vectors of bytes or records. However, storage presents data as vectors of blocks of a constant size (512 or in the newer devices, 4096 bytes per block).

The *file, record, and namespace virtualization* and *file and record subsystem* layers convert records or files that are required by applications to vectors of blocks, which are the language of the *block virtualization* layer. The block virtualization layer maps requests of the higher layers to physical storage blocks, which are provided by *storage devices* in the *block subsystem*.

Each of the layers in the storage domain abstracts the complexities of the lower layers and hides them behind an easy to use, standard interface that is presented to upper layers. The resultant decoupling of logical storage space representation and its characteristics that are visible to servers (storage consumers) from underlying complexities and intricacies of storage devices is a key concept of storage virtualization.

The focus of this publication is *block-level virtualization* at the *block virtualization layer*, which is implemented by IBM as IBM Storage Virtualize software that is running on IBM SAN Volume Controller and the IBM FlashSystem family. The IBM SAN Volume Controller is implemented as a clustered appliance in the storage network layer. The IBM FlashSystems are deployed as modular storage systems that can virtualize their internally and externally attached storage.

IBM Storage Virtualize uses the Small Computer System Interface (SCSI) protocol to communicate with its clients. It presents storage space as SCSI logical units (LUs), which are identified by SCSI logical unit numbers (LUNs).

Note: Although LUs and LUNs are different entities, the term LUN is often used to refer to a logical disk, which is also called an LU.

Most applications do not directly access storage; instead, they work with files or records by using a file system. However, the operating system of a host must convert these abstractions to the language of storage; that is, vectors of storage blocks that are identified by logical block addresses (LBAs) within an LU.

Inside IBM Storage Virtualize, each of the externally visible LUs is internally represented by a volume, which is an amount of storage that is taken out of a storage pool. Storage pools are made of managed disks (MDisks); that is, they are LUs that are presented to the storage system by external virtualized storage or arrays that consist of internal disks. LUs that are presented to IBM Storage Virtualize by external storage often correspond to RAID arrays that are configured on that storage.

The hierarchy of objects, from a file system block down to a physical block on a physical drive, is shown in Figure 1-1.

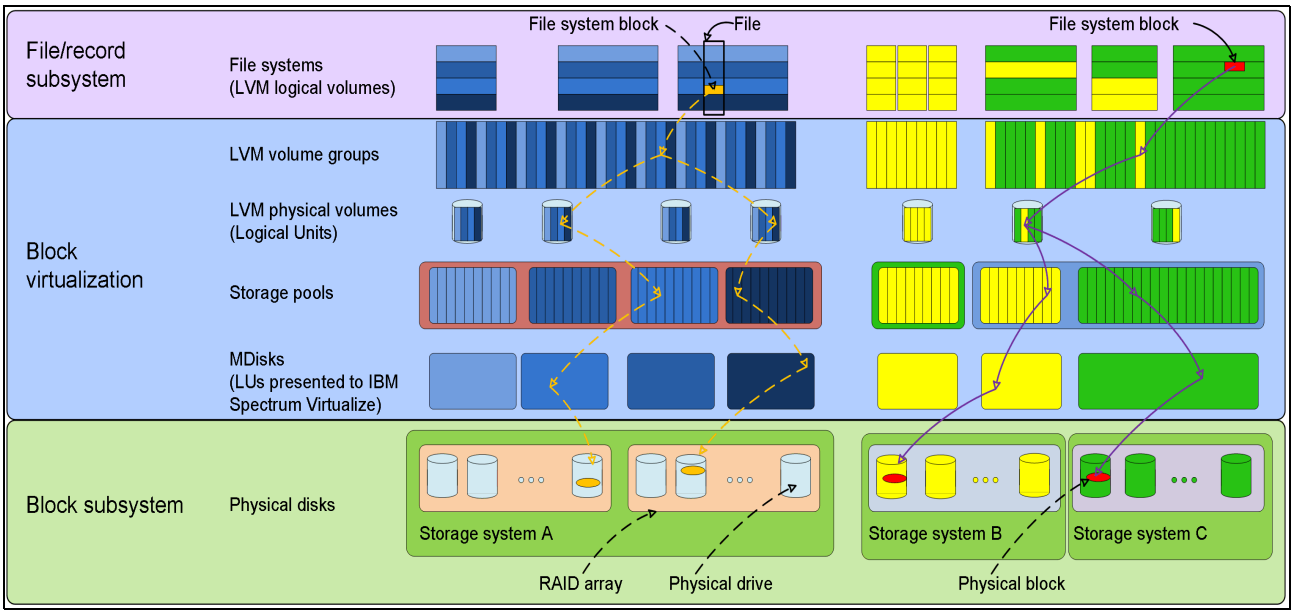


Figure 1-1 Block-level virtualization overview

With storage virtualization, you can manage the mapping between logical blocks within an LU that is presented to a host, and blocks on physical drives. This mapping can be as simple or as complicated as required by a use case. A logical block can be mapped to one physical block or for increased availability, multiple blocks that are physically stored on different physical storage systems, and in different geographical locations.

IBM Storage Virtualize uses the concept of an extent, which is a group of physical blocks, as a management construct to allow great flexibility in the usage of available storage.

Importantly, the mapping can be dynamic: With IBM Easy Tier®, IBM Storage Virtualize can automatically change underlying storage to which an extent is mapped to better match a host's performance requirements with the capabilities of the underlying storage systems.

IBM Storage Virtualize gives a storage administrator a wide range of options to modify volume characteristics: from volume resize to mirroring, creating a point-in-time (PiT) copy with IBM FlashCopy®, and migrating data across physical storage systems.

Importantly, all the functions that are presented to the storage users are independent from the characteristics of the physical devices that are used to store data. This decoupling of the storage feature set from the underlying hardware and ability to present a single, uniform

interface to storage users that masks underlying system complexity is a powerful argument for adopting storage virtualization with IBM Storage Virtualize.

Storage virtualization is implemented on many layers. Figure 1-1 on page 5 shows an example in which a file system block is mirrored by the host's operating system (left side of the figure) by using features of the logical volume manager (LVM) or the IBM Storage Virtualize system at the storage pool level (as shown on the right side of Figure 1-1 on page 5).

Although the result is similar in that the data block is written to two different arrays, the effort that is required for per-host configuration is disproportionately larger than for a centralized solution with organization-wide storage virtualization that is done on a dedicated system and managed from a single GUI.

IBM Storage Virtualize includes the following key features:

- Simplified storage management by providing a single management interface for multiple storage systems and a consistent user interface for provisioning heterogeneous storage.
- Online volume migration. IBM Storage Virtualize enables moving the data from one set of physical drives to another set in a way that is not apparent to the storage users and without over-straining the storage infrastructure. The migration can be done within a specific storage system from one set of disks to another set or across storage systems. Either way, the host that uses the storage is not aware of the operation, and no downtime for applications is needed.
- Enterprise-level Copy Services functions. Performing Copy Services functions within IBM Storage Virtualize removes dependencies on the capabilities and interoperability of the virtualized storage subsystems. Therefore, it enables the source and target copies to be on any two virtualized storage subsystems.
- Safeguarded Copy mechanism, which uses FlashCopy functions to take immutable copies, and aids in the recovery from ransomware or customer internal actions that might result in loss of data.
- Inline Data Corruption Detection is a feature that uses AI and ML services to detect data changes that can be indicative of threats or direct attacks on your data sets in near real-time.
- Improved storage space usage because of the pooling of resources across virtualized storage systems.
- Possible improved system performance because of volume striping across multiple virtualized arrays or controllers, and the benefits of cache of the IBM Storage Virtualize hardware.
- Improved data security by using data-at-rest encryption.
- Data replication, including replication to cloud storage by using advanced copy services for data migration and backup solutions.
- Data reduction techniques, such as deduplication and compression for space efficiency, which store more data on the storage that is available. IBM Storage Virtualize can enable significant savings that increase the effective capacity of storage systems up to 5x, and decrease the floor space, power, and cooling that are required by the storage system.

Note: IBM Real-time Compression (RtC) is available only for earlier generation systems. The newer IBM SAN Volume Controller systems, which include the SV3, SV2, and SA2 models, do *not* support RtC. However, they support software compression through Data Reduction Pools (DRP).

Summary

Storage virtualization is a fundamental technology that enables the realization of flexible and reliable storage solutions. It helps enterprises to better align IT architecture with business requirements, simplify their storage administration, and facilitate their IT departments efforts to meet business demands.

IBM Storage Virtualize running on IBM SAN Volume Controller and IBM FlashSystem family is a mature, 12th-generation virtualization solution that uses open standards and complies with the SNIA storage model. All products use in-band block virtualization engines that move the control logic (including advanced storage functions) from a multitude of individual storage devices to a centralized entity in the storage network.

IBM Storage Virtualize can improve the use of your storage resources, simplify storage management, and improve the availability of business applications.

Figure 1-2 shows the feature set that is provided by the IBM Storage Virtualize systems.

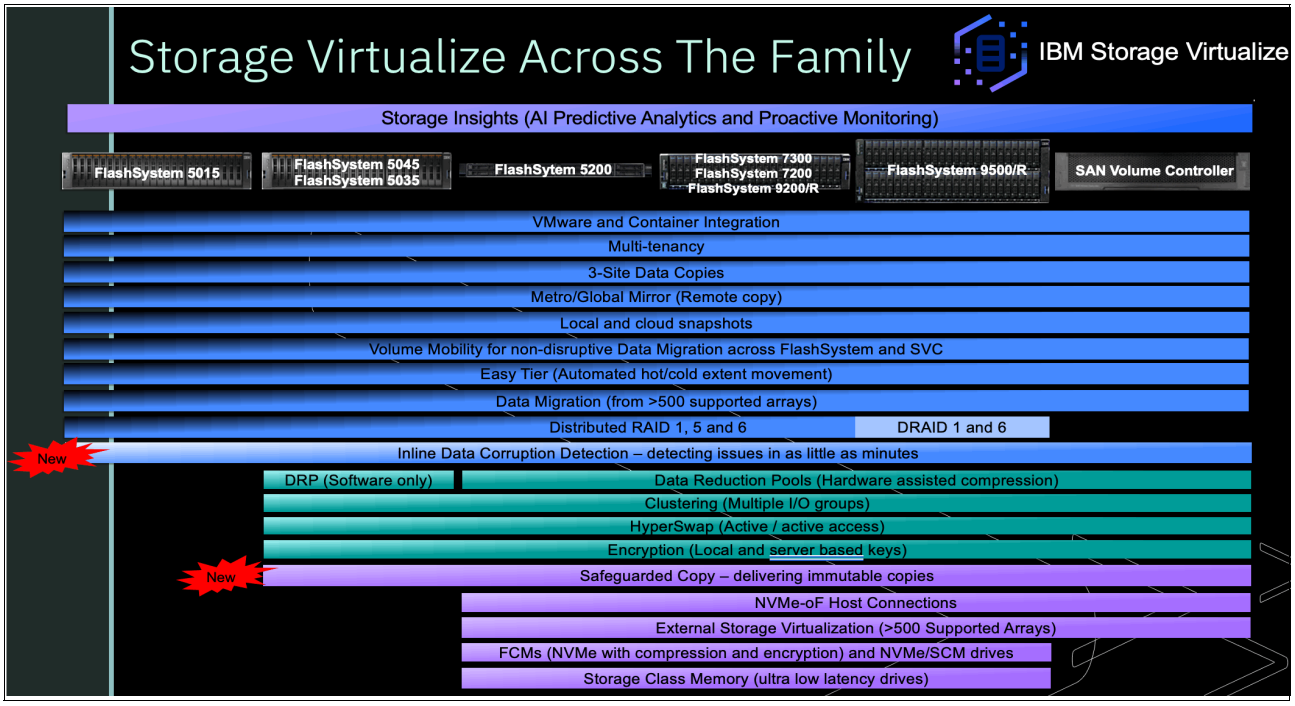


Figure 1-2 IBM Storage Virtualize features by product

1.2 IBM SAN Volume Controller architectural overview

IBM SAN Volume Controller is a SAN block aggregation virtualization appliance that is designed for attachment to various host computer systems.

The following major approaches are used today for the implementation of block-level aggregation and virtualization:

- Symmetric: In-band appliance

  Virtualization splits the storage that is presented by the storage systems into smaller chunks that are known as *extents*. These extents are then concatenated by using various policies to make virtual disks or *volumes*. With symmetric virtualization, host systems can be isolated from the physical storage. Advanced functions, such as data migration, can run without reconfiguring the host.

  With symmetric virtualization, the *virtualization engine* is the central configuration point for the SAN. The virtualization engine directly controls access to the storage and data that is written to the storage. As a result, locking functions that provide data integrity and advanced functions (such as cache and Copy Services) can be run in the virtualization engine. Therefore, the virtualization engine is a central point of control for device and advanced function management.

  Symmetric virtualization includes some disadvantages. The main disadvantage that is associated with symmetric virtualization is scalability. Scalability can cause poor performance because all input/output (I/O) must flow through the virtualization engine. To solve this problem, you can use an *n*-way cluster of virtualization engines that includes failover capability.

  You can scale the extra processor power, cache memory, and adapter bandwidth to achieve the level of performance that you want. More memory and processing power are needed to run advanced services, such as Copy Services and caching. IBM SAN Volume Controller uses symmetric virtualization. Single virtualization engines, which are known as *nodes*, are combined to create clusters. Each cluster can contain 2 - 8 nodes.

- Asymmetric: Out-of-band or controller-based

  With asymmetric virtualization, the virtualization engine is outside the data path and performs a metadata-style service. The metadata server contains all of the mapping and the locking tables, and the storage devices contain only data. In asymmetric virtual storage networks, the data flow is separated from the control flow.

  A separate network or SAN link is used for control purposes. Because the control flow is separated from the data flow, I/O operations can use the full bandwidth of the SAN. A separate network or SAN link is used for control purposes.

  Asymmetric virtualization can have the following disadvantages:

  – Data is at risk to increased security exposures, and the control network must be protected with a firewall.

  – Metadata can become complicated when files are distributed across several devices.

  – Each host that accesses the SAN must know how to access and interpret the metadata. Therefore, specific device drivers or agent software must be running on each of these hosts.

  – The metadata server cannot run advanced functions, such as caching or Copy Services, because it only “knows” about the metadata and not the data.

Figure 1-3 shows variations of the two virtualization approaches.

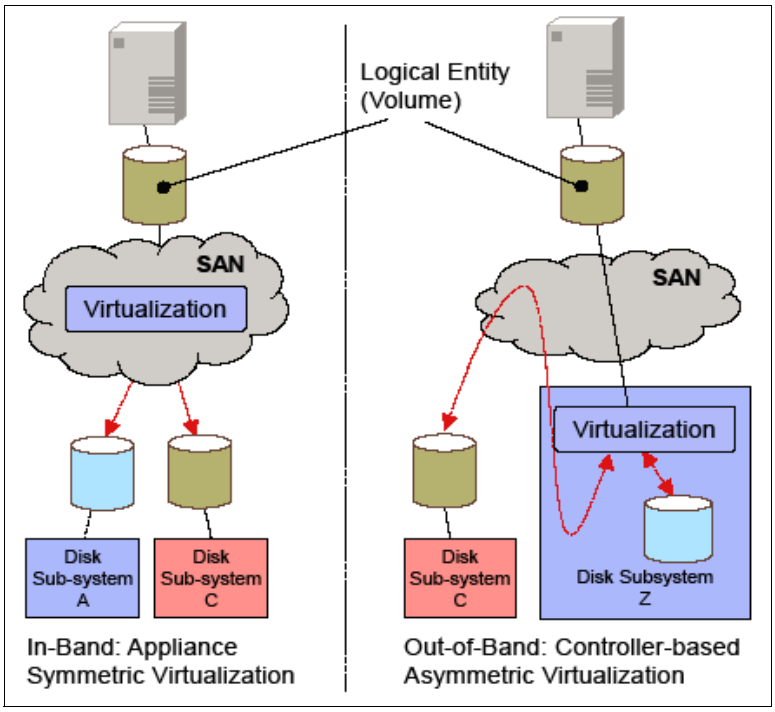


Figure 1-3 Overview of block-level virtualization architectures

Although these approaches provide essentially the same cornerstones of virtualization, interesting side-effects can occur.

1.2.1 Controller-based approach

The out-of-band, controller-based approach includes high functions, but it fails in terms of scalability or upgradeability. Because of the nature of its design, no true decoupling occurs by using this approach, which becomes an issue for the lifecycle of this solution, such as with a controller. Data migration issues and questions are challenging, such as how to reconnect the servers to the new controller, and how to reconnect them online without any effect on your applications.

By using this approach, you replace a controller and implicitly replace your entire virtualization solution. In addition to replacing the hardware, other actions (such as updating or repurchasing the licenses for the virtualization feature, and advanced copy functions) might be necessary.

1.2.2 SAN or fabric-based appliance solution

With an in-band, SAN or fabric-based appliance solution that is based on a scale-out cluster architecture, lifecycle management tasks, such as adding or replacing new disk subsystems or migrating data between them, are simple.

Servers and applications remain online, data migration occurs transparently on the virtualization platform, and licenses for virtualization and copy services require no update. No other costs are incurred when disk subsystems are replaced.

Only the fabric-based appliance solution provides an independent and scalable virtualization platform that can provide enterprise-class Copy Services that is open for future interfaces and protocols. By using the fabric-based appliance solution, you can choose the disk subsystems that best fit your requirements, and you are not locked into specific SAN hardware.

For these reasons, IBM chose the SAN-based appliance approach with inline block aggregation for the implementation of storage virtualization with IBM Storage Virtualize.

1.2.3 IBM SAN Volume Controller

IBM SAN Volume Controller includes the following key characteristics:

- It is highly scalable, which provides a growth path from two to eight (grow in a pair of nodes to allow for data protection and reliability).
- It is SAN interface-independent. It supports connections through Fibre Channel, NVM Express (NVMe) over Fibre Channel (FC-NVMe), NVM Express (NVMe) over RDMA, or a TCP network. Consider the following points:
  - FC-NVMe connections require 4-port 16 GB or 32 GB Fibre Channel adapters that must be purchased.
  - NVMe over RDMA connections support a 25 Gbps RoCE adapter or 100 Gbps RoCE adapter that must be purchased.
  - The IBM SAN Volume Controller 2145-SV3 and 2147-SV3 models support 100 Gbps adapter only.
- It is host-independent for fixed block-based Open Systems environments.
- It is external storage system independent, which provides a continuous and ongoing process to qualify more types of storage systems.
- Some nodes can use internal disks (flash drives) or externally direct-attached disks in expansion enclosures.

On the SAN storage that is provided by the disk subsystems, IBM SAN Volume Controller offers the following services:

- Creates a single pool of storage
- Provides LU virtualization
- Manages logical volumes
- Mirrors logical volumes

IBM SAN Volume Controller running IBM Storage Virtualize V8.6 also provides the following functions:

- Large scalable cache
- Copy Services
- IBM FlashCopy (PiT copy) function, including thin-provisioned FlashCopy to make multiple targets affordable
- IBM Transparent Cloud Tiering (TCT) function that enables the IBM SAN Volume Controller to interact with cloud service providers (CSPs)
- Metro Mirror (MM), which is a synchronous copy
- Global Mirror (GM), which is an asynchronous copy
- Policy-based replication
- Safeguarded Copy
- Inline Data Corruption Detection

- Data migration
- Storage space efficiency (thin-provisioning, compression, and deduplication)
- IBM Easy Tier to automatically migrate data between storage types of different performance that is based on disk workload
- Encryption of external attached storage
- Supports IBM HyperSwap®
- Supports VMware vSphere virtual volumes (vVols) and Microsoft Offloaded Data Transfer (ODX)
- Direct attachment of hosts
- Hot spare nodes with a standby function of single or multiple nodes
- Containerization connectivity with Container Storage Interface (CSI), which enables supported storage to be used as persistent storage in container environments
- Hybrid Multicloud function with IBM Storage Virtualize for Public Cloud

1.2.4 IBM SAN Volume Controller topology

External storage can be managed by IBM SAN Volume Controller in one or more pairs of hardware nodes. This configuration is referred to as a *clustered system*. These nodes are normally attached to the SAN fabric, with storage and host systems. The SAN fabric is zoned to enable the IBM SAN Volume Controller to communicate with external storage systems and hosts.

Within this software release, IBM SAN Volume Controller also supports iSCSI networks. This feature enables the hosts and storage systems to communicate with IBM SAN Volume Controller to build a storage virtualization solution.

It is important that hosts cannot see or operate on the same volume (LUN) that is mapped to the IBM SAN Volume Controller. Although a set of LUNs can be mapped to IBM SAN Volume Controller, and a separate set of LUNs can be mapped directly to one or more hosts, be sure that a separate set of LUNs is always used.

The zoning capabilities of the SAN switch must be used to create distinct zones to ensure that this rule is enforced. SAN fabrics can include standard FC, FC-NVMe, FCoE, iSCSI over Ethernet, or possible future types.

IBM Storage Virtualize 8.6.0 also supports the following NVMe protocols:

- NVMe/FC:
  - Requires version 8.2.1 and later
  - Requires 16/32 Gb FC adapters
  - Supports SLES/RH/ESX/Windows as initiators
  - Will support 64 Gb FC adapters in a future release
- NVMe/RoCE:
  - Requires version 8.5.0 or later
  - Supports RoCE (Mellanox CX-4/CX-6) adapters with 25 Gb and 100 Gb speeds on storage (target) side
  - Supports SLES/RH/ESX as host initiators OS, with RoCE 25 Gb, 40 Gb, and 100 Gb (Mellanox CX-4/CX-5/CX-6), and Broadcom adapters

- Requires RoCE supported Ethernet infrastructure.

- NVMe/TCP:
  - NVMe/NVMeOF protocol is designed to fully make use of the performance of all-flash storage.
  - The use of Ethernet storage connection technology in data centers is increasing.
  - NVMe/TCP is a ubiquitous transport and does not require special infrastructure, unlike NVMe/RoCE, which requires special infrastructure.
  - The increased speeds are comparable with NVMe/RoCE without the need for special networking hardware.

Figure 1-4 shows a conceptual diagram of a storage system that uses IBM SAN Volume Controller. It also shows several hosts that are connected to a SAN fabric or local area network (LAN).

In implementations that have HA requirements, which are most of the target clients for IBM SAN Volume Controller, the SAN fabric cloud represents a redundant SAN. A *redundant SAN* consists of a fault-tolerant arrangement of two or more SANs, which provide alternate paths for each SAN-attached device.

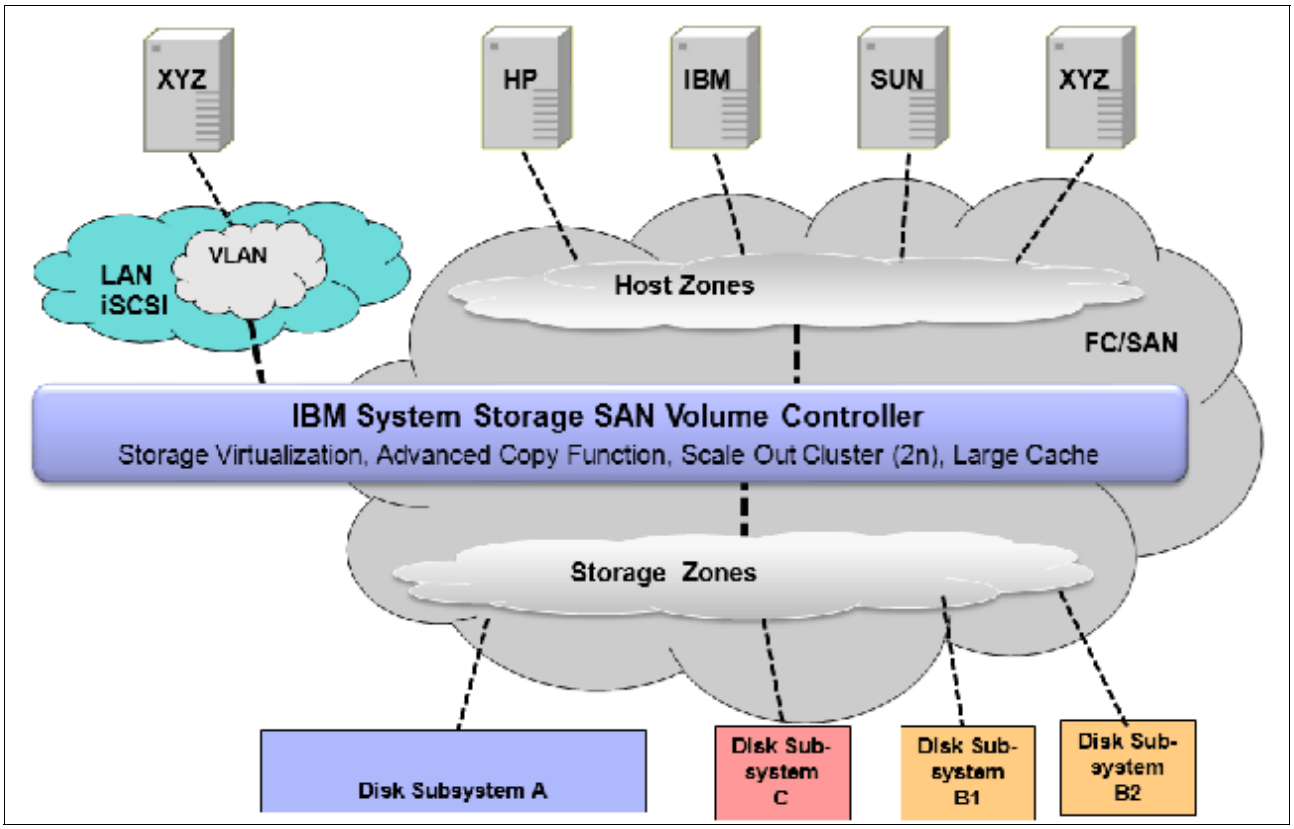


Figure 1-4 IBM SAN Volume Controller conceptual and topology overview

Both scenarios (the use of a single network and the use of two physically separate networks) are supported for iSCSI-based and LAN-based access networks to IBM SAN Volume Controller. Redundant paths to volumes can be provided in both scenarios.

For simplicity, Figure 1-4 on page 12 shows only one SAN fabric and two zones: host and storage. In a real environment, it is a best practice to use two redundant SAN fabrics. IBM SAN Volume Controller can be connected to up to four fabrics.

A clustered system of IBM SAN Volume Controller nodes that are connected to the same fabric presents *logical disks* or volumes to the hosts. These volumes are created from managed LUNs or MDisks that are presented by the storage systems.

The following distinct zones are shown in the fabric:

- A host zone, in which the hosts can see and address the IBM SAN Volume Controller nodes
- A storage zone, in which the IBM SAN Volume Controller nodes can see and address the MDisks or LUNs that are presented by the storage systems

As explained in 1.4, “IBM SAN Volume Controller family” on page 23, hosts are not permitted to operate on the RAID LUNs directly. All data transfer happens through the IBM SAN Volume Controller nodes. This flow is referred to as *symmetric virtualization*.

For iSCSI-based access, the use of two networks and separating iSCSI traffic within the networks by using a dedicated virtual local area network path for storage traffic prevents any IP interface, switch, or target port failure from compromising the iSCSI connectivity across servers and storage controllers.

1.3 Changes and enhancements in version 8.6.0

This section describes the latest changes and enhancements in IBM Storage Virtualize V8.6.0. The major functions that were included in V8.5.0 and V8.4.2 releases are also discussed.

1.3.1 IBM Storage Virtualize V8.6.0

IBM Storage Virtualize V8.6.0 provides more features and updates to the IBM Storage Virtualize family of products, which include IBM FlashSystems and the IBM SAN Volume Controller.

NVMe over TCP host connectivity

The Non-Volatile Memory express (NVMe) transport protocol provides enhanced performance on high-demand IBM Storage FlashSystem drives.

NVMe is a logical device interface specification for accessing non-volatile storage media. Host hardware and software use NVMe to fully use the levels of parallelism possible in modern solid-state drives (SSDs).

Storage Virtualize includes support for the following protocols:

- NVMe over Fibre Channel,
- NVMe over RDMA
- NVMe over TCP protocol.

For more information, see IBM System Storage Interoperation Center (SSIC)

Support Node CLI to shutdown individual Fibre Channel ports

This **chfcportfeature** command can enable or disable the Fibre Channel port features. The Fibre Channel port supports features such as fabric device management interface (FDMI) registration, discovery, and port state.

Support for iSCSI performance improvement

This procedure can provide a solution for Internet Small Computer Systems Interface (iSCSI) host performance problems when it is connected to a system and for its connectivity to the network switch.

Support for 1024 iSCSI hosts per I/O group

A host can be created with worldwide port names (WWPNs) or iSCSI names. The WWPN namespace and the iSCSI namespace within the system share the same internal system resources. Each system I/O group can have a maximum of 1024 iSCSI IQNs.

Support for migrating remote copy to policy-based replication

Policy-based replication uses volume groups and replication policies to automatically deploy and manage replication. Policy-based replication can simplify configuring, managing, and monitoring replication between two systems.

With policy-based replication, you can replicate data between systems with minimal management, higher throughput, and reduced latency compared to the remote-copy function.

A replication policy has the following properties:

- Replication policies cannot be changed after they are created. If changes are required, a new policy can be created and assigned to the associated volume group.
- Each system supports up to a maximum of 32 replication policies.

Support for SMTP authentication

Ability to configure Call Home with email notifications on any of the following email servers:

- Local email server

  Emails are sent, received, and archived on physical servers in an on-premises location.

- Cloud-based email server

  Emails are sent, received, and archived on cloud-based server providers such as Google Mail and Microsoft Office 365.

Support for TLS 1.3

Security of all key server communications is governed by TLS 1.2 and TLS 1.3 protocols. Encryption keys are distributed between nodes in the system by using TLS 1.2 and TLS1.3.

The system uses AES-256 encryption that uses OpenSSL library interfaces. To establish a connection between the key server and the system, the key server or services must support the configured TLS version.

Support for DNS check to be turned off

Use the **lsdnsserver** command to list information for any Domain Name System (DNS) servers in the system.

Support for FDMI information in Call Home

The Fabric Device Management Interface (FDMI) enables any storage endpoint to register itself to the Fibre Channel (FC) fabric and query the HBA and port details of the entire fabric.

Version 2 metadata volume for VMware virtual volumes

The system provides native support for VMware vSphere APIs for Storage Awareness (VASA) through a VASA Provider, also known as a Storage Provider. The VASA provider sends and receives information about storage that is used by VMware vSphere to the vCenter Server.

Through the VASA provider, the system also supports VMware vVols, which allows VMware vCenter to automate the creation, deletion and mapping of volumes.

Enhancement to IBM Storage Insights for threat detection

A key part of monitoring your system includes the detection of potential ransomware attacks. To ensure that you have the latest storage metadata for detecting those types of attacks, compression and cyber resiliency statistics for volumes are collected every 5 minutes.

With these statistics, IBM Storage Insights builds a historical model of a storage system and uses its built-in intelligence and formulas to identify when and where ransomware attacks might be occurring. For more information about statistics, see, Storage Insights Overview.

Support for non-superuser ability to manage the system

Each user of the management GUI must have a username and a password to sign on. Each user also has an associated role, such as monitor or security administrator. These roles are defined at the system level. For example, a user can be the administrator for one system, but the security administrator for another system.

This non-superuser level can monitor and manage certain functions on the system.

Downloading software patches through Call Home by using Restful API

This option enables Call Home with cloud services to connect to support and directly transfer the latest update available to the storage system.

To download a software update file through the Call Home by using the Restful API, the user enters the **`satask downloadsoftware`** command plus the specific update they require.

For further information about these features, see What's new in v8.6.0.

1.3.2 IBM Storage Virtualize V8.5.0

IBM Storage Virtualize V8.5.0 provides more features and updates to the IBM Storage Virtualize family of products of which IBM FlashSystems and the IBM SAN Volume Controller are part.

Version 8.5.0 includes major software and hardware changes and includes support for the following features:

- IBM FlashSystem 9500, 7300, and IBM SAN Volume Controller SV3 systems, which include the following features:
  - 100 Gbps Ethernet adapter
  - 48 NVMe drives per distributed RAID-6 array (IBM FlashSystem 9500 only)
  - Secure boot drives

- Multifactor authentication and single sign-on
- NVMe host attachment over RDMA
- Fibre Channel port sets
- I/O stats in microseconds
- Domain names for IP replication
- IBM Storage Virtualize 3-Site Orchestrator version 4.0
- Increased number of hosts per I/O group

Version 8.5.0 also includes the following features:

- Improved distributed RAID array recommendations
- Improved default time for updates
- Updates to OpenStack support summary

IBM FlashSystems 9500, 7300 and IBM SAN Volume Controller SV3

The following new hardware platforms were released with the IBM Storage Virtualize V8.5.0:

- IBM FlashSystems 9500

  The IBM FlashSystems 9500 is a 4U control enclosure and contains up to 48 NVMe-attached IBM FlashCore® Modules or other self-encrypted NVMe-attached SSDs.

  The NVMe-attached drives in the control enclosures provide significant performance improvements compared to SAS-attached flash drives. The system supports 2U and 5U all-Flash SAS attached expansion enclosure options.

- IBM FlashSystems 7300

  The FlashSystem 7300 is a 2U dual controller that contains up to 24 NVMe-attached IBM FlashCore Modules or other self-encrypted NVMe-attached SSDs or Storage Class Module drives.

  IBM FlashSystem 7300 system also supports 2U and 5U SAS-attached expansion enclosure options.

- IBM SAN Volume Controller SV3

  The IBM SAN Volume Controller SV3 system is a 2U single controller that combines software and hardware into a comprehensive, modular appliance that provides symmetric virtualization. The SV3 is run in pairs from an I/O group, which is the building block for any IBM SAN Volume Controller based virtualization setup.

  For more information about these new hardware platforms, see 1.5, “IBM SAN Volume Controller models” on page 43, and 1.6, “IBM FlashSystem family” on page 48.

New 100 Gbps Ethernet adapter

A maximum of six 100 Gbps Ethernet adapters can be installed in the IBM FlashSystems 9500 and 7300 and in a pair of IBM SAN Volume Controller SV3s.

In 100 Gbps Ethernet ports, iSCSI performance is equivalent to 25 Gbps iSCSI host attachment.

Support for 48 NVMe drives per distributed RAID-6 array

Enhanced support for 48 NVMe drives in the enclosure by using distributed RAID 6 technology is now available for the IBM FlashSystems 9500.

The following configurations are supported:

- Distributed RAID 6 arrays of NVMe drive expansion up to 48 member drives.
- Distributed RAID 6 arrays of FCM NVMe drive support expansion up to 48 member drives.
- Distributed RAID 6 arrays of extra large (38.4 TB) physical capacity FCM NVMe drives supports up to 24 member drives.

Support for secure boot drives

IBM Storage Virtualize V8.5.0 provides secure boot by pairing each boot drive with a Trusted Platform Module (TPM). TPM provides a secure cryptographic processor that verifies hardware and prevents unauthorized access to hardware and the operating system. On the system, TPM protects secure boot to ensure the code images that are installed are signed, trusted, and unchanged.

Multifactor authentication and single sign-on

Multifactor authentication requires users to provide multiple pieces of information when they log in to the system to prove their identity. Multifactor authentication uses any combination of two or more methods, called *factors*, to authenticate users to your resources and protect those resources from unauthorized access.

One of the key concepts of multifactor authentication is that each factor comes from a different category; that is, something the users knows, has, or is.

Single Sign-on (SSO) authentication requires users to register their credentials only once when the user signs on to the application for the first time. The user information is stored at the Identity Provider (IdP) that manages the user credentials and determines whether the user is required to authenticate again.

NVMe host attachment over RDMA

The new NVMe/RDMA function uses the RDMA capabilities of the host adapters, as does the existing iSER support. NVMe over RDMA connections support 25 Gbps RoCE adapters or 100 Gbps RoCE adapters.

Fibre Channel port sets

Port sets are groupings of logical addresses that are associated with the specific traffic types. The IBM Storage Virtualize V8.5.0 supports IP and Fibre Channel port sets for host attachment, and IP port sets for backend storage connectivity, and IP replication traffic.

I/O stats in microseconds

Real-time performance statistics provide short-term status information for the system. These statistics summarize the overall performance health of the system and can be used to monitor trends in bandwidth and CPU usage. IBM Storage Virtualize V8.5.0 gives you the ability to see the granularity down at microseconds intervals, which was not available on previous code levels.

Support for fully qualified domain names for IP replication

Metro Mirror and Global Mirror partnerships can be established over Ethernet links that use the IPv4 and IPv6 addresses that are associated with Ethernet ports. Before IBM Storage Virtualize V8.5.0, these addresses had to be physical dotted decimal IP types. IBM Storage Virtualize V8.5.0 allows the user to also specify domain names when IP partnerships are created. If you specify domain names, a DNS server must be configured on your system.

1.3.3 IBM Storage Virtualize V8.4.2

IBM Storage Virtualize V8.5.0 also provides all of the following previous functions and major enhancements of the V8.4.2 release:

- Code release schedule
- Safeguarded Copy
- Volume mobility
- Storage as a Service (STaaS)

Code release schedule

The new code release schedule includes the following features:

- Two types of releases: Long Term Releases (LTR) and Continuous Development Releases (CDRs)
- Four scheduled updates each year:
  - One LTR:
    - New features are formally announced.
    - Includes all the enhancements of the previous CDRs.
    - Adopts the major numbering scheme (8.4, 8.5, 8.6, and so on).
    - Includes all Program Temporary Fixes (PTFs) of problems that were found with existing features going forward.
  - Three CDRs:
    - No formal announcement.
    - New features are announced, but do not have PTFs.
    - PTFs are generally added to the next CDR (unless critical).
    - The fixes are incorporated into the next CDR.
    - A minor numbering scheme is used, such as 8.4.1 and 8.5.1.

Safeguarded Copy

Safeguarded Copy is a virtual air gap mechanism that uses FlashCopy functions to take immutable copies. This feature aids in the recovery from ransomware or internal “bad actors” who seek to destroy data.

Note: The Safeguarded Copy function is available with IBM Storage Virtualize software 8.5.0, but is not supported for the FlashSystem 5000, Storwize V5030E, Storwize V7000 Gen2, and Storwize V7000 Gen2+ models.

Safeguarded Copy is a feature or solution with which you can create point-in time copies of active production data that cannot be altered or deleted so that it is an immutable or protected copy. It requires a user with the correct privileged access to modify the Safeguarded Copy expiration settings. Lastly, Safeguarded Copy uses copy management software for testing and ease of recovery of copies.

Safeguarded Copy use cases

Safeguarded Copy includes the following use cases:

- Validation

  Regular analytics of the copy to provide early detection of a problem or reassurance that the copy is a good copy before further action is taken.

- Forensic

  Use a copy of the production system to investigate the problem and determine whether recovery action is necessary.
- Surgical

  Extract data from the copy and logically restore back to the production environment.
- Catastrophic

  Recover the entire environment back to the point in time of the copy because this option is the only available recovery option.
- Offline backup

  Performing an offline backup of data from a consistent point-in-time copy can be used to build a second line of defense, which provides a greater retention period and increased isolation and security.

Safeguarded Copy is the first step toward a cyber resiliency solution.

Focusing mainly on the feature set from a customer's perspective, consider the following three pillars, as shown in Figure 1-5:

- Separation of Duties
- Protected Copies
- Automation

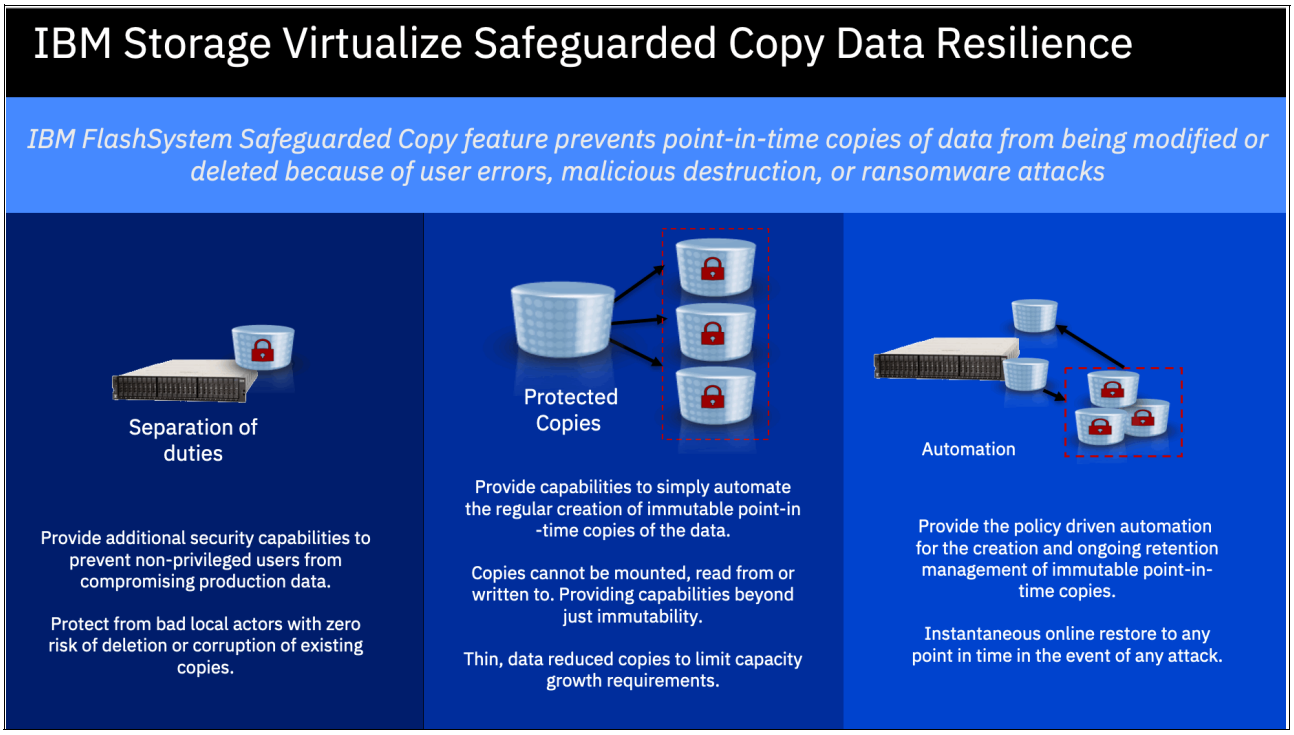


Figure 1-5 IBM Safeguarded Copy Data Resilience

Figure 1-5 shows the following IBM Storage Virtualize Safeguarded Copy Data Resilience examples:

- Separation of duties

  Traditional backup and restore capabilities are normally controlled by storage administrators and do not protect against intentional (for example, rogue employee) or non-intentional attacks.

Primary and backup are treated differently. Protecting current backups, or securing and hardening current backup, does not solve the problem.

- Protected copies of the data

  These backups must be immutable; for example, hidden, non-addressable, cannot be altered or deleted, only usable after recovery.

  Copies must deliver a higher level of security while meeting industry and business regulations.

- Automation:
  - Initially setting and managing of policies (number of copies, retention period)
  - Automating, managing, and restoring of those copies.

For more information, see the following resources:

- This IBM Documentation web page
- *IBM FlashSystem Safeguarded Copy Implementation Guide*, REDP-5654

Demonstration video: For more information, see the demonstration video IBM Storage for Data Resilience simply explained.

Volume mobility

Volume mobility is similar to nondisruptive volume moves between I/O groups except that you are migrating a volume between systems. For example, you move data from an IBM SAN Volume Controller to an FS9500. It allows a user to nondisruptively move data between systems that do not natively cluster. This feature is a major benefit for upgrading and changing Storage Virtualize systems.

By using volume migrations, you can also perform the following tasks:

- Nondisruptive volume migrations between independent systems or clusters
- Nondisruptive migrations between non clustering platforms
- Volume migration away from a cluster that is reaching maximum capacity limits

Volume mobility uses enhancements to SCSI (ALUA) path states. The migration is based on Remote Copy (Metro Mirror) functions.

Consider the following restrictions about volume mobility:

- No 3-site support.
- Not a DR or HA solution.
- No support for consistency groups, change volumes, or expanding volumes.
- Partnership requirements are equivalent to Metro Mirror.
- Performance considerations apply as with Metro Mirror.
- Reduced host interoperability support:
  - RHEL
  - SLES
  - ESXi
  - Solaris
  - HP-UX
  - SCSI only (No FC-NVMe)

- No SCSI persistent reservations or Offloaded Data Transfer (ODX).
- No support for IBM i, AIX®, Hyper-V, and Windows.
- Remote Mirror license is not required.

IBM Storage as a Service

The latest member in the IBM Storage Family is the newly announced IBM Storage as a Service (STaaS) offering.

Buyer behavior is shifting from technology-focused to service-level agreement (SLA)-driven, cloud-like simplicity. IT staff is being downsized and transitioning into generalist roles, rather than specializing in specific areas. Enterprise workloads need flexibility where applications are to be deployed. New and better infrastructures, such as 5G, allow this growth outside of the traditional data center.

The IBM STaaS offering is a pure OpEx solution and does not require initial capital.

Consumption-based solutions include the following important features:

- Flexible scale-up and scale-down model
- Cloud-like functions in most solutions
- All the deployment and managed support, optimization, and disposal services included
- Well-defined upgrade path
- Clear pricing terms
- Switch from capital expenditures (CapEx) funding to operating expenses (OpEx) funding
- Ease of billing and payment terms

With the IBM STaaS offering, the customer makes the following decisions:

- Which tier level is needed
- The amount of storage capacity is needed
- How long the customer wants to use this offering
- Connection type that is required
- Encryption option that is needed

Figure 1-6 on page 22 shows the STaaS tiers available for IBM FlashSystems.

	Extreme (Tier 1)	Premium (Tier 2)	Balanced (Tier 3)
Minimum capacity	25	50	100
Performance (IOPS/TB)*	4,500	2,250	600
Maximum read throughput (GB/s)**	45	45	35
Maximum write throughput (GB/s)**	12	12	10
Availability	99.9999% / optional 100% guarantee***	99.9999% / optional 100% guarantee***	99.9999% / optional 100% guarantee***
Connection type	Ethernet or FC	Ethernet or FC	Ethernet or FC
Encryption	Optional at no cost	Optional at no cost	Optional at no cost
Contract terms	1-5 years	1-5 years	1-5 years
Lifecycle management includes			
Installation	IBM	IBM	IBM
Predictive support	Storage Insights Pro	Storage Insights Pro	Storage Insights Pro
SW S&S	Yes	Yes	Yes
Remote firmware updates	Yes	Yes	Yes
IBM Technical Account Manager	Yes	Yes	Yes
IBM monitoring and capacity growth	Yes	Yes	Yes
HW service	24x7 same day onsite repair IBM performed	24x7 same day onsite repair IBM performed	24x7 same day onsite repair IBM performed

**IOPS/TB: This defines the minimum ratio of Input-Output Operations (IOPS) per physical used Terabyte (TB) with usage up to 90% of the usable capacity. IOPS/TB are based on a 70:30 Read / Write mixed workload with 50% cache hit and 16K I/O size using Fibre Channel. Remote replication may impact performance. Actual results will vary by workload, as such the IOPS/TB should be used for the purposes of selecting relative performance tiers.*

***Maximum read and write throughput based on 256KB I/O Size using Fibre Channel.*

**** 99.9999% objective. 100% guarantee requires HyperSwap configuration installed by IBM Lab Services. Terms and conditions apply.*

Figure 1-6 STaaS Tiers comparison

IBM Life-Cycle Management provides and includes the following features:

- Deployment, maintenance, and disposal of the hardware:
  - Offering is based on IBM FlashSystems technology.
  - All hardware is an IBM-owned asset that is installed on-premises at the customer data center.
- IBM monitoring of capacity growth and IBM installs more if needed (miscellaneous equipment specification [MES]).
- Periodic technology refresh with 90-day overlap for migration.

- IBM recycle process of old equipment.

STaaS pseudo machine type

The pseudo machine type for STaaS is 9601, with performance tiers and terms that are differentiated by models (base, medium, and high).

Feature codes to support these models are related to the setup type, annual capacity growth, and options for Ethernet, encryption, and decreasing capacity.

9601 models

The following 9601 models are available:

- Balanced performance: BT1, BT2, BT3, BT4, and BT5
- Premium performance: MT1, MT2, MT3, MT4, and MT5
- Extreme performance: HT1, HT2, HT3, HT4, and HT5

> **Note:** The numeric value in the model number is the duration of the STaaS contract in years.

The IBM STaaS offering also includes the new Storage Expert Care Premium Level of service, which features the resource of a dedicated Technical Account Manager (TAM), enhanced response to severity 1 and 2 issues, and predictive support by way of IBM Storage Insights.

For more information about the STaaS offering, see *IBM Storage as a Service Offering Guide*, REDP-5644.

1.4 IBM SAN Volume Controller family

This section explains the major underlying concepts of IBM SAN Volume Controller, provides an architectural overview and describes the terminology that is used in a virtualized storage environment. This section also introduces the software and hardware components and the other functions that are available with Version 8.6.

Figure 1-7 on page 24 shows the complete IBM SAN Volume Controller family that supports the IBM Storage Virtualize V8.6 software.

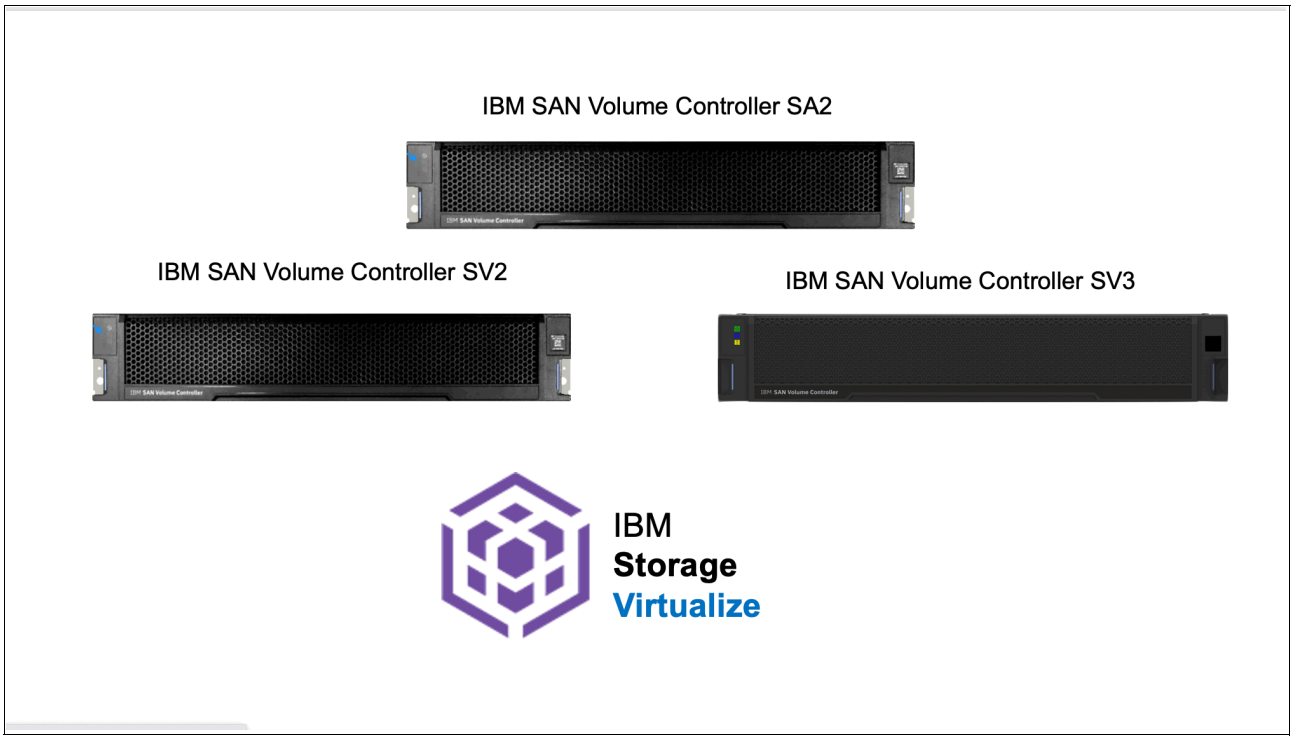


Figure 1-7 IBM SAN Volume Controller family

1.4.1 Components

IBM SAN Volume Controller provides block-level aggregation and volume management for attached disk storage. In simpler terms, the IBM SAN Volume Controller manages several back-end storage controllers or locally attached disks.

IBM SAN Volume Controller maps the physical storage within those controllers or storage systems into logical disk images or volumes that can be seen by application servers and workstations in the SAN. It logically sits between hosts and storage systems. It presents itself to hosts as the storage provider, also called the target, and presents itself to storage systems as one large host, also called the initiator.

The SAN is zoned such that the application servers cannot communicate with the back-end storage or controller. This configuration prevents any possible conflict between the IBM SAN Volume Controller and the application servers that are trying to manage the back-end storage.

The IBM SAN Volume Controller is based on the components that are described next.

1.4.2 Nodes

Each IBM SAN Volume Controller hardware unit is called a *node*. Each node is an individual server in an IBM SAN Volume Controller clustered system on which the Storage Virtualize software runs. The node provides the virtualization for a set of volumes, cache, and copy services functions.

The IBM SAN Volume Controller nodes are deployed in pairs (*io_group*), and one or multiple pairs constitute a *clustered system* or *system*. A system can consist of one pair and a maximum of four pairs.

One of the nodes within the system is known as the *configuration node*. The configuration node manages the configuration activity for the system. If this node fails, the system chooses a new node to become the configuration node.

Because the active nodes are installed in pairs, each node provides a failover function to its partner node if a node fails.

1.4.3 I/O groups

Each pair of IBM SAN Volume Controller nodes is also referred to as an *I/O group*. An IBM SAN Volume Controller clustered system can have 1 - 4 I/O groups.

A specific *volume* is always presented to a host server by a single I/O group of the system. The I/O group can be changed.

When a host server performs I/O to one of its volumes, all the I/Os for a specific volume are directed to one specific I/O group in the system. Under normal conditions, the I/Os for that specific volume are always processed by the same node within the I/O group. This node is referred to as the *preferred node* for this specific volume.

Both nodes of an I/O group act as the preferred node for their own specific subset of the total number of volumes that the I/O group presents to the host servers. However, both nodes also act as failover nodes for their respective partner node within the I/O group. Therefore, a node takes over the I/O workload from its partner node when required.

In an IBM SAN Volume Controller-based environment, the I/O handling for a volume can switch between the two nodes of the I/O group. Therefore, it is a best practice that servers are connected to two different fabrics through different FC host bus adapters (HBAs) to use multipath drivers to give redundancy.

The IBM SAN Volume Controller I/O groups are connected to the SAN so that all application servers that are accessing volumes from this I/O group can access this group. Up to 512 host server objects can be per I/O group. The host server objects can access volumes that are provided by this specific I/O group.

If required, host servers can be mapped to more than one I/O group within the IBM SAN Volume Controller system. Therefore, they can access volumes from separate I/O groups. You can move volumes between I/O groups to redistribute the load between the I/O groups. Modifying the I/O group that services the volume can be done concurrently with I/O operations if the host supports nondisruptive volume moves.

It also requires a rescan at the host level to ensure that the multipathing driver is notified that the allocation of the preferred node changed, and the ports (by which the volume is accessed) changed. This modification can be done in the situation where one pair of nodes becomes overused.

1.4.4 System

The system or clustered system consists of 1 - 4 I/O groups. Specific configuration limitations are then set for the entire system. For example, the maximum number of volumes that is supported per system is 10,000, or the maximum capacity of MDisks that is supported is ~28 PiB (32 PB) per system.

All configuration, monitoring, and service tasks are performed at the system level. Configuration settings are replicated to all nodes in the system. To facilitate these tasks, a management IP address is set for the system.

A process is provided to back up the system configuration data on to storage so that it can be restored if a disaster occurs. This method does not back up application data. Only the IBM SAN Volume Controller system configuration information is backed up.

For remote data mirroring, two or more systems must form a partnership before relationships between mirrored volumes are created.

For more information about the maximum configurations that apply to the system, I/O group, and nodes, see this IBM Support web page.

1.4.5 MDisks

The SAN Volume Controller system and its I/O groups view the storage that is presented to them by the back-end storage system as several disks or LUNs, which are known as *MDisks*. Because the IBM SAN Volume Controller does not attempt to provide recovery from physical disk failures within the back-end storage system, an MDisk must be provisioned from a RAID array.

These MDisks are placed into storage pools where they are divided into several extents. The application servers do not “see” the MDisks at all. Rather, they see logical disks, which are known as volumes. These disks are presented by the IBM SAN Volume Controller I/O groups through the SAN or LAN to the servers.

For more information, see V8.6.0.x Configuration Limits and Restrictions for IBM System Storage SAN Volume Controller.

When an MDisk is presented to the IBM SAN Volume Controller, it can be in one of the following states:

- Unmanaged MDisk

  An MDisk is reported as `unmanaged` when it is not a member of any storage pool. An unmanaged MDisk is not associated with any volumes and has no metadata that is stored on it.

  IBM SAN Volume Controller does not write to an MDisk that is in `unmanaged` mode except when it attempts to change the mode of the MDisk to one of the other modes. IBM SAN Volume Controller can see the resource, but the resource is not assigned to a storage pool.

- Managed MDisk

  Managed mode MDisks are always members of a storage pool, and they contribute extents to the storage pool. Volumes (if not operated in image mode) are created from these extents. MDisks that are operating in managed mode might have metadata extents that are allocated from them and can be used as *quorum disks*. This mode is the most common and normal mode for an MDisk.

- Image mode MDisk

  Image mode provides a direct block-for-block conversion from the MDisk to the volume by using virtualization. This mode is provided to satisfy the following major usage scenarios:

  – Image mode enables the virtualization of MDisks that contain data that was written directly and not through IBM SAN Volume Controller. Rather, it was created by a direct-connected host.

    This mode enables a client to insert an IBM SAN Volume Controller into the data path of a storage volume or LUN with minimal downtime. For more information about the data migration process, see Chapter 8, “Storage migration” on page 635.

    Image mode enables a volume that is managed by an IBM SAN Volume Controller to be used with the native copy services function that is provided by the underlying RAID controller. To avoid the loss of data integrity when the IBM SAN Volume Controller is used in this way, it is important that you disable the IBM SAN Volume Controller cache for the volume.

  – The IBM SAN Volume Controller can migrate to image mode, which enables the IBM SAN Volume Controller to export volumes and access them directly from a host without the IBM SAN Volume Controller in the path.

  Each MDisk that is presented from an external disk controller features an online path count that is the number of nodes that can access that MDisk. The *maximum count* is the maximum number of paths that is detected at any point by the system. The *current count* is what the system sees now. A current value that is less than the maximum can indicate that SAN fabric paths were lost.

Tier

It is likely that the MDisks (LUNs) that are presented to the IBM SAN Volume Controller system have different characteristics because of the disk or technology type on which they are placed. The following tier options are available:

- `tier0_flash`
- `tier1_flash`
- `tier_enterprise`
- `tier_nearline`
- `tier_scm`

The default value for a newly discovered unmanaged MDisk is `enterprise`. You can change this value by running the **`chmdisk`** command.

The tier of external MDisks is not detected automatically and is set to `enterprise`. If the external MDisk is made up of flash drives or nearline Serial Attached SCSI (SAS) drives and you want to use IBM Easy Tier, you must specify the tier when adding the MDisk to the storage pool. You can also run the **`chmdisk`** command to modify the tier attribute.

1.4.6 Cache

The primary benefit of storage cache is to improve I/O response time. Reads and writes to a magnetic disk drive experience seek time and latency time at the drive level, which can result in 1 ms - 10 ms of response time (for an enterprise-class disk).

Consider the following points:

- The IBM SAN Volume Controller Model SA2 and SV2 features 128 GB of memory with options for 768 GB of memory in a 2U 19-inch rack mount enclosure.

- The IBM SAN Volume Controller Model SV3 features 512 GB of memory with options for 1536 GB of memory in a 2U 19-inch rack mount enclosure.

The IBM SAN Volume Controller provides a flexible cache model. Cache is allocated in 4 kibibyte (KiB) segments. A *segment* holds part of one track. A *track* is the unit of locking and destaging granularity in the cache. The cache virtual track size is 32 KiB (eight segments).

A track might be only partially populated with valid pages. The IBM SAN Volume Controller combines writes up to a 256 KiB track size if the writes are in the same tracks before destaging. For example, if 4 KiB is written into a track, another 4 KiB is written to another location in the same track.

Therefore, the blocks that are written from the IBM SAN Volume Controller to the disk subsystem can be any size of 512 bytes - 256 KiB. The large cache and advanced cache management algorithms enable it to improve the performance of many types of underlying disk technologies.

The IBM SAN Volume Controller capability to manage in the background the destaging operations that are incurred by writes (in addition to still supporting full data integrity) assists with the IBM SAN Volume Controller capability in achieving good performance.

The cache is separated into two layers: upper cache and lower cache. Figure 1-8 shows the separation of the upper and lower cache.

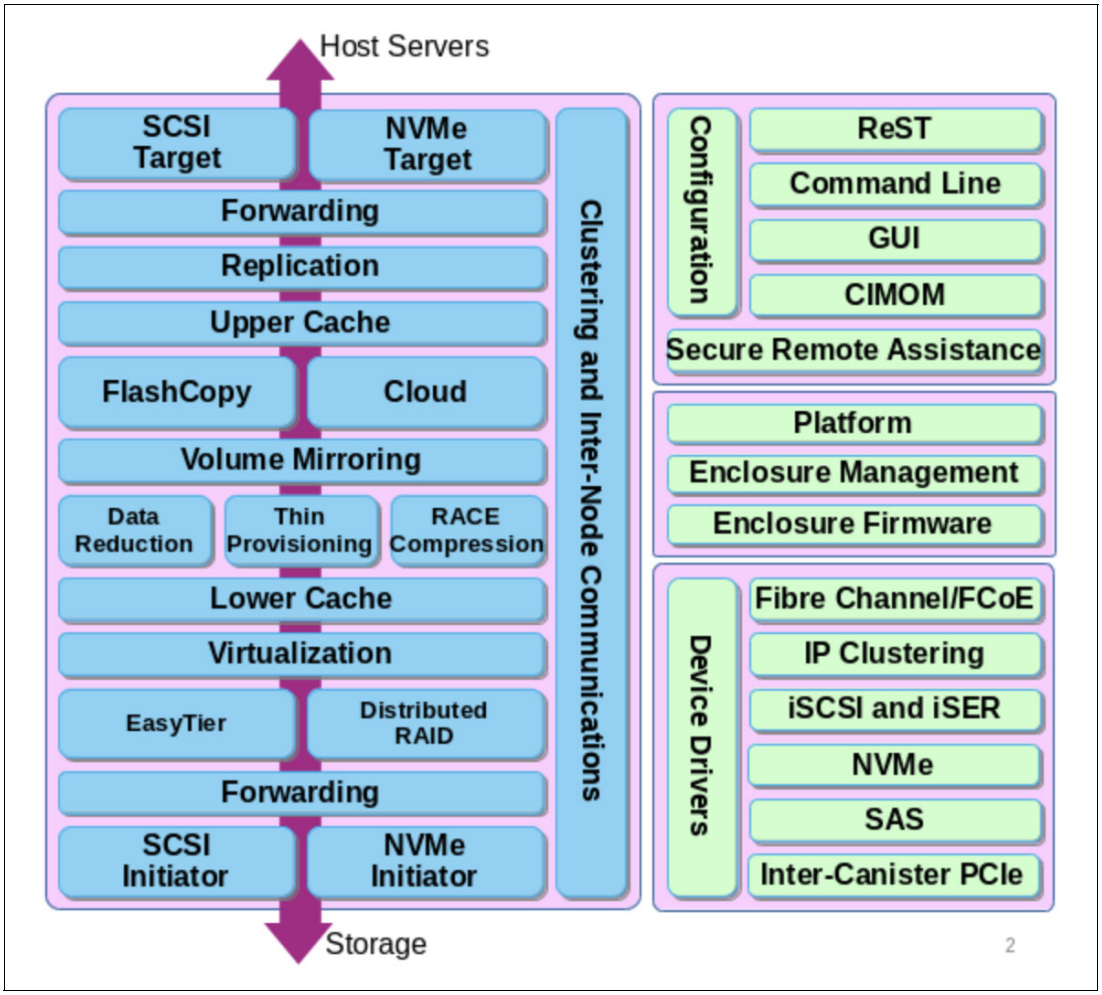


Figure 1-8 Separation of upper and lower cache

The upper cache delivers the following functions, which enable the IBM SAN Volume Controller to streamline data write performance:

- Fast write response times to the host by being as high up in the I/O stack as possible
- Partitioning

The lower cache delivers the following functions:

- Ensures that the write cache between two nodes is in sync
- Caches partitioning to ensure that a slow backend cannot use the entire cache
- Uses a destaging algorithm that adapts to the amount of data and the back-end performance
- Provides read caching and prefetching

Combined, the two levels of cache also deliver the following functions:

- Pins data when the LUN goes offline
- Provides the following:
  - Enhanced statistics for IBM Storage Control and IBM Storage Insights
  - Trace for debugging
- Reports medium errors
- Resynchronizes cache correctly and provides the atomic write function
- Ensures that other partitions continue operation when one partition becomes 100% full of pinned data
- Integrates with T3 recovery procedures
- Supports:
  - Fast-write (two-way and one-way), flush-through, and write-through
  - Two-way operation
  - None, read-only, and read/write as user-exposed caching policies
  - Flush-when-idle
  - Expanding cache as more memory becomes available to the platform
  - Credit throttling to avoid I/O skew and offer fairness/balanced I/O between the two nodes of the I/O group
- Enables switching the preferred node without needing to move volumes between I/O groups

Depending on the size, age, and technology level of the disk storage system, the total available cache in the IBM SAN Volume Controller nodes can be larger, smaller, or about the same as the cache that is associated with the disk storage.

Because hits to the cache can occur in the IBM SAN Volume Controller or the back-end storage system level of the overall system, the system as a whole can take advantage of the larger amount of cache wherever the cache is available.

In addition, regardless of their relative capacities, both levels of cache tend to play an important role in enabling sequentially organized data to flow smoothly through the system. The IBM SAN Volume Controller cannot increase the throughput potential of the underlying disks in all cases because this increase depends on the underlying storage technology and the degree to which the workload exhibits *hotspots* or sensitivity to cache size or cache algorithms.

However, the write cache is still assigned to a maximum of 12 GB and the compression cache to a maximum of 34 GB. The remaining installed cache is used as read cache (including allocation for features, such as IBM FlashCopy, GM, or MM). Data reduction pools share memory with the main I/O process.

1.4.7 Quorum disk

A *quorum disk* is an MDisk or a managed drive that contains a reserved area that is used exclusively for system management. A system automatically assigns quorum disk candidates. Quorum disks are used when a problem exists in the SAN fabric or when nodes are shut down, which leaves half of the nodes remaining in the system. This type of problem causes a loss of communication between the nodes that remain in the system and the nodes that do not remain.

The nodes are split into groups where the remaining nodes in each group can communicate with each other, but not with the other group of nodes that were formerly part of the system. In this situation, some nodes must stop operating and processing I/O requests from hosts to preserve data integrity while maintaining data access. If a group contains less than half the nodes that were active in the system, the nodes in that group stop operating and processing I/O requests from hosts.

It is possible for a system to split into two groups with each group containing half the original number of nodes in the system. A quorum disk determines which group of nodes stops operating and processing I/O requests. In this tiebreaker situation, the first group of nodes that accesses the quorum disk is marked as the owner of the quorum disk. As a result, all of the nodes that belong to the owner group continue to operate as the system and handle all I/O requests.

If the other group of nodes cannot access the quorum disk or discover that the quorum disk is owned by another group of nodes, it stops operating as the system and does not handle I/O requests. A system can have only one active quorum disk that is used for a tiebreaker situation. However, the system uses three quorum disks to record a backup of the system configuration data that is used if a disaster occurs. The system automatically selects one active quorum disk from these three disks.

The other quorum disk candidates provide redundancy if the active quorum disk fails before a system is partitioned. To avoid the possibility of losing all of the quorum disk candidates with a single failure, assign quorum disk candidates on multiple storage systems.

Quorum disk requirements: To be considered eligible as a quorum disk, a LUN must meet the following criteria:

- It is presented by a storage system that supports IBM SAN Volume Controller quorum disks.
- It is manually enabled as a quorum disk candidate by running the **`chcontroller -allowquorum yes`** command.
- It is in managed mode (no image mode).
- It includes sufficient free extents to hold the system state information and the stored configuration metadata.
- It is visible to all of the nodes in the system.

If possible, the IBM SAN Volume Controller places the quorum candidates on separate storage systems. However, after the quorum disk is selected, no attempt is made to ensure that the other quorum candidates are presented through separate storage systems.

Quorum disk placement verification and adjustment to separate storage systems (if possible) reduce the dependency from a single storage system, and can increase the quorum disk availability.

You can list the quorum disk candidates and the active quorum disk in a system by running the **`lsquorum`** command.

When the set of quorum disk candidates is chosen, it is fixed. However, a new quorum disk candidate can be chosen in one of the following conditions:

- When the administrator requests that a specific MDisk becomes a quorum disk by running the **`chquorum`** command.
- When an MDisk that is a quorum disk is deleted from a storage pool.
- When an MDisk that is a quorum disk changes to image mode.

An offline MDisk is not replaced as a quorum disk candidate.

For DR purposes, a system must be regarded as a single entity so that the system and the quorum disk can be collocated.

Special considerations are required for the placement of the active quorum disk for a stretched, split cluster or split I/O group configurations. For more information, see SAN Volume Controller.

Important: Running an IBM SAN Volume Controller system without a quorum disk can seriously affect your operation. A lack of available quorum disks for storing metadata prevents any migration operation.

Mirrored volumes can be taken offline if no quorum disk is available. This behavior occurs because the synchronization status for mirrored volumes is recorded on the quorum disk.

During the normal operation of the system, the nodes communicate with each other. If a node is idle for a few seconds, a heartbeat signal is sent to ensure connectivity with the system. If a node fails for any reason, the workload that is intended for the node is taken over by another node until the failed node is restarted and readmitted into the system (which happens automatically).

If the Licensed Internal Code on a node becomes corrupted, which results in a failure, the workload is transferred to another node. The code on the failed node is repaired, and the node is readmitted into the system (which is an automatic process).

IP quorum configuration

In a stretched configuration or IBM HyperSwap configuration, you must use a third, independent site to house quorum devices. To use a quorum disk as the quorum device, this third site must use FC or IP connectivity together with an external storage system. In a local environment, no extra hardware or networking, such as FC or SAS-attached storage, is required beyond what is normally always provisioned within a system.

To use an IP-based quorum application as the quorum device for the third site, no FC connectivity is used. Java applications are run on hosts at the third site. However, strict requirements on the IP network and some disadvantages with the use of IP quorum applications exist.

Unlike quorum disks, all IP quorum applications must be reconfigured and redeployed to hosts when certain aspects of the system configuration change. These aspects include adding or removing a node from the system, or when node service IP addresses are changed.

For stable quorum resolutions, an IP network must provide the following requirements:

- Connectivity from the hosts to the service IP addresses of all nodes. If IP quorum is configured incorrectly, the network must also deal with the possible security implications of exposing the service IP addresses because this connectivity can also be used to access the service GUI.
- Port 1260 is used by IP quorum applications to communicate from the hosts to all nodes.
- The maximum round-trip delay must not exceed 80 ms, which means 40 ms each direction.
- A minimum bandwidth of 2 MBps for node-to-quorum traffic.

Even with IP quorum applications at the third site, quorum disks at site one and site two are required because they are used to store metadata. To provide quorum resolution, run the **`mkquorumapp`** command or use the GUI in **Settings** → **Systems** → **IP Quorum** to generate a Java application that is then copied to and run on a host at a third site. A maximum of five applications can be deployed.

1.4.8 Storage pool

A *storage pool* is a collection of MDisks that provides the pool of storage from which volumes are provisioned. A single system can manage up to 1024 storage pools. The size of these pools can be changed (expanded or shrunk) at run time by adding or removing MDisks without taking the storage pool or the volumes offline.

At any point, an MDisk can be a member in one storage pool only, except for image mode volumes.

Figure 1-9 shows the relationships of the IBM SAN Volume Controller entities to each other.

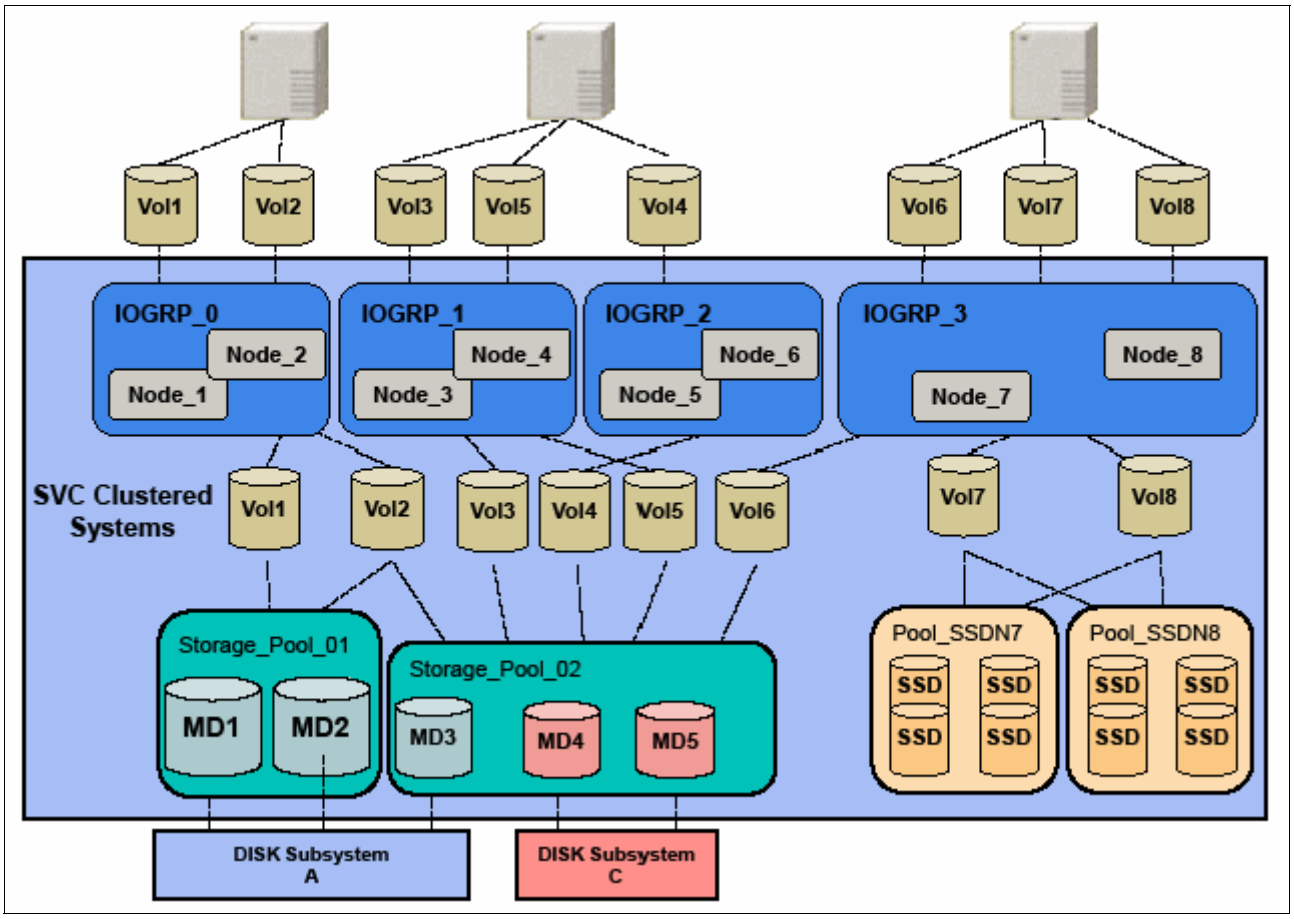


Figure 1-9 Overview of an IBM SAN Volume Controller clustered system with an I/O group

Each MDisk capacity in the storage pool is divided into several extents. The size of the extent is selected by the administrator when the storage pool is created and cannot be changed later. The size of the extent is 16 MiB - 8192 MiB.

It is a best practice to use the same extent size for all storage pools in a system. This approach is a prerequisite for supporting volume migration between two storage pools. If the storage pool extent sizes are not the same, you must use volume mirroring to copy volumes between pools.

The IBM SAN Volume Controller limits the number of extents in a system to 2^{22}=~4 million. Because the number of addressable extents is limited, the total capacity of an IBM SAN Volume Controller system depends on the extent size that is chosen by the IBM SAN Volume Controller administrator.

1.4.9 Volumes

Volumes are logical disks that are presented to the host or application servers by the IBM SAN Volume Controller.

The following types of volumes are available in terms of extents management:

- Striped

A striped volume is allocated one extent in turn from each MDisk in the storage pool. This process continues until the space that is required for the volume is satisfied.

It is also possible to supply a list of MDisks to use.

Figure 1-10 shows how a striped volume is allocated (assuming that 10 extents are required).

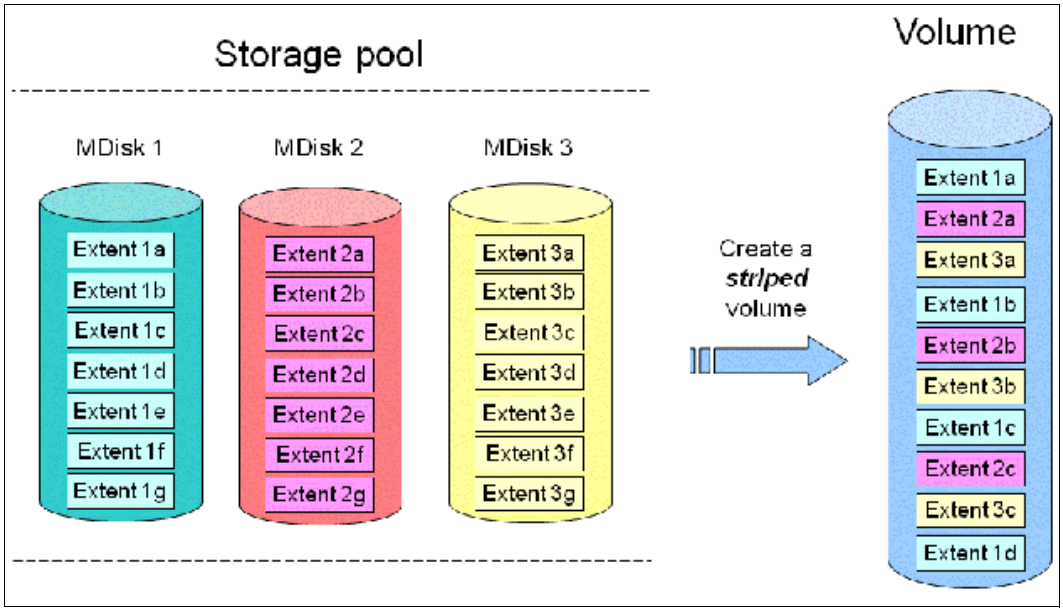


Figure 1-10 Striped volume

- Sequential

A sequential volume is where the extents are allocated sequentially from one MDisk to the next MDisk (see Figure 1-11).

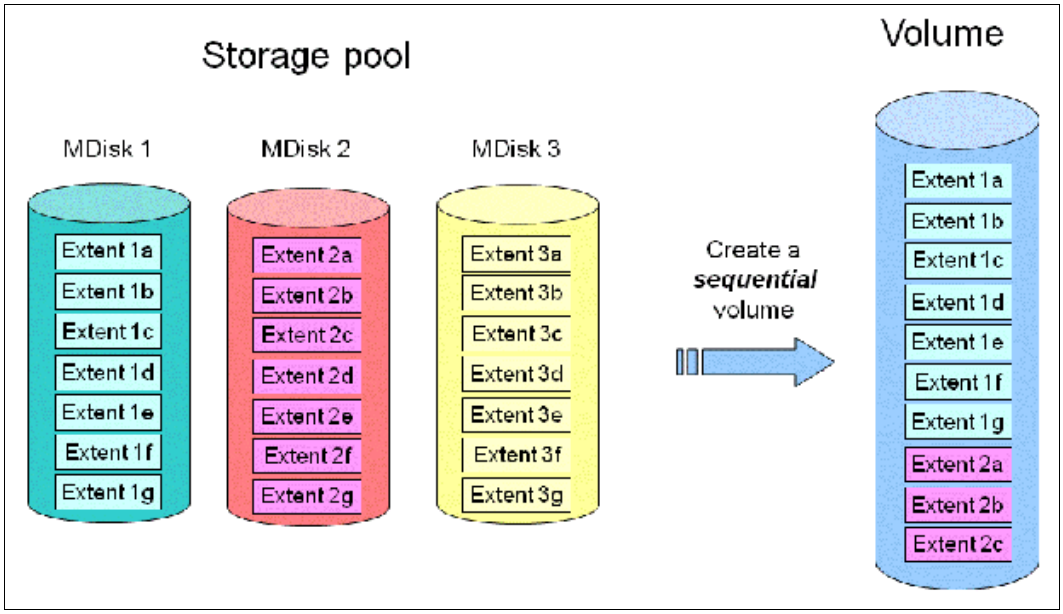


Figure 1-11 Sequential volume

- Image mode

  Image mode volumes (see Figure 1-12) are special volumes that have a direct relationship with one MDisk. The most common use case of image volumes is a data migration from your old, typically nonvirtualized, storage to the IBM SAN Volume Controller based virtualized infrastructure.

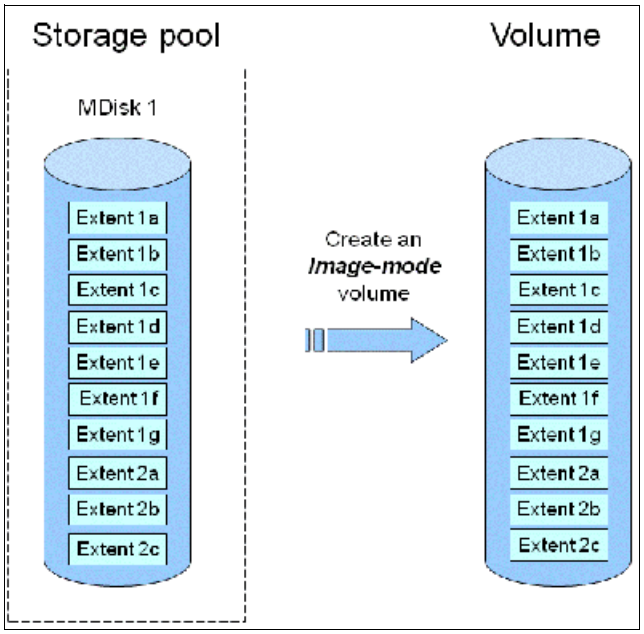


Figure 1-12 Image mode volume

When the image mode volume is created, a direct mapping is made between extents that are on the MDisk and the extents that are on the volume. The LBA x on the MDisk is the same as the LBA x on the volume, which ensures that the data on the MDisk is preserved as it is brought into the clustered system.

Because some virtualization functions are not available for image mode volumes, it is useful to migrate the volume into a new storage pool. After the migration completion, the MDisk becomes a managed MDisk.

If you add an MDisk that contains any historical data to a storage pool, all data on the MDisk is lost. Ensure that you create image mode volumes from MDisks that contain data before adding MDisks to the storage pools.

1.4.10 Easy Tier

Easy Tier is a performance function that automatically migrates extents of a volume from one MDisk storage tier to another MDisk storage tier.

Easy Tier monitors the host I/O activity and latency on the extents of all volumes with the Easy Tier function that is turned on in a multitier storage pool over a 24-hour period. Then, it creates an extent migration plan that is based on this activity, and then dynamically moves high-activity or hot extents to a higher disk tier within the storage pool. It also moves extents whose activity dropped off or cooled down from the high-tier MDisks back to a lower-tiered MDisk.

Easy Tier supports the new SCM drives with a new tier that is called `tier_scm`.

Turning on or off Easy Tier: The Easy Tier function can be turned on or off at the storage pool level and the volume level.

The automatic load-balancing function is enabled by default on each volume and cannot be turned off by using the GUI. This load-balancing feature is not considered an Easy Tier function, although it uses the same principles.

The management GUI supports monitoring Easy Tier data movement in graphical reports. The data in these reports helps you understand how Easy Tier manages data between the different tiers of storage, how tiers within pools are used, and the workloads among the different tiers. Charts for data movement, tier composition, and workload skew comparison can be downloaded as comma-separated value (CSV) files.

You can also offload the statistics file from the IBM SAN Volume Controller nodes and by using the IBM Storage Tier Advisor Tool (STAT) to create a summary report. The STAT can be downloaded for no initial cost from IBM Support.

For more information about Easy Tier, see Chapter 9, “Advanced features for storage efficiency” on page 669.

1.4.11 Hosts

A *host* is a logical object that represents a list of worldwide port names (WWPNs), NVMe qualified names (NQNs), or iSCSI or iSER names that identify the interfaces that the host system uses to communicate with the IBM SAN Volume Controller. Fibre Channel connections use WWPNs to identify host interfaces to the system. iSCSI or iSER names can be iSCSI qualified names (IQNs) or extended unique identifiers (EUIs). NQNs are used to identify hosts that use FC-NVMe connections.

Volumes can be mapped to a *host* to enable access to a set of volumes.

Node failover can be handled without having a multipath driver that is installed on the iSCSI server. An iSCSI-attached server can reconnect after a node failover to the original target IP address, which is now presented by the partner node. To protect the server against link failures in the network or HBA failures, a multipath driver *must* be used.

N_Port ID Virtualization (NPIV) is a method for virtualizing a physical Fibre Channel port that is used for host I/O. When NPIV is enabled, the partner node takes over the WWPN of the failing node. This takeover allows for rapid recovery of in-flight I/O when a node fails. In addition, path failures that occur because an offline node is masked from host multipathing.

Host cluster

A *host cluster* is a group of logical host objects that can be managed together. For example, you can create a volume mapping that is shared by every host in the host cluster. Host objects that represent hosts can be grouped in a host cluster and share access to volumes. New volumes can also be mapped to a host cluster, which simultaneously maps that volume to all hosts that are defined in the host cluster.

1.4.12 Array

An array is an ordered configuration, or group, of physical devices (drives) that is used to define logical volumes or devices. An array is a type of MDisk that is made up of disk drives (these drives are members of the array). A Redundant Array of Independent Disks (RAID) is a method of configuring member drives to create high availability (HA) and high-performance systems. The system supports *nondistributed* and *distributed* array configurations.

In *nondistributed* arrays, entire drives are defined as “hot-spare” drives. Hot-spare drives are idle and do not process I/O for the system until a drive failure occurs. When a member drive fails, the system automatically replaces the failed drive with a hot-spare drive. The system then resynchronizes the array to restore its redundancy.

However, all member drives within a *distributed* array have a rebuild area that is reserved for drive failures. All the drives in an array can process I/O data and provide faster rebuild times when a drive fails. The RAID level provides different degrees of redundancy and performance; it also determines the number of members in the array.

1.4.13 Encryption

The IBM SAN Volume Controller provides optional encryption of data at rest, which protects against the potential exposure of sensitive user data and user metadata that is stored on discarded, lost, or stolen storage devices. Encryption of system data and system metadata is not required; therefore, system data and metadata are not encrypted.

Planning for encryption involves purchasing a licensed function and then activating and enabling the function on the system.

To encrypt data that is stored on drives, the nodes that are capable of encryption must be licensed and configured to use encryption. When encryption is activated and enabled on the system, valid encryption keys must be present on the system when the system unlocks the drives or the user generates a new key.

Encryption keys can be managed by an external key management system when you use the following supported products:

- IBM Security® Guardium® Key Lifecycle Manager (formally IBM Security Key Lifecycle Manager)
- Thales CipherTrust Manager
- Gemalto KeySecure key servers

Encryption keys can also be stored on USB flash drives that are attached to a minimum of one of the nodes. For version 8.1 and later, IBM Storage Virtualize provides a combination of external and USB key repositories.

IBM Security Guardium Key Lifecycle Manager is an IBM solution that provides the infrastructure and processes to locally create, distribute, backup, and manage the lifecycle of encryption keys and certificates. Before activating and enabling encryption, you must determine the method of accessing key information during times when the system requires an encryption key to be present.

When Security Key Lifecycle Manager is used as a key manager for the IBM SAN Volume Controller encryption, you can encounter a deadlock situation if the key servers are running on encrypted storage that is provided by the IBM SAN Volume Controller. To avoid a deadlock situation, ensure that the IBM SAN Volume Controller can communicate with an encryption server to get the unlock key after a power-on or restart scenario. A maximum of 4 Security Key Lifecycle Manager servers are supported.

Although both Thales CipherTrust Manager and Gemalto KeySecure key servers support the same type of configurations, ensure that you complete the prerequisites on these key servers before you can enable encryption on the system.

Data encryption is protected by the Advanced Encryption Standard (AES) algorithm that uses a 256-bit symmetric encryption key in XTS mode, as defined in the Institute of Electrical and Electronics Engineers (IEEE) 1619-2007 standard as XTS-AES-256.[1] That data encryption key is protected by a 256-bit AES key wrap when it is stored in nonvolatile form.

Another data security enhancement, which is delivered with the Storage Virtualize version 8.4.2 and later, is the Safeguarded Copy function that can provide protected read-only air gap copies of volumes. This enhancement gives the customer effective data protection against cyberattacks.

For more information, see *IBM FlashSystem Safeguarded Copy Implementation Guide*, REDP-5654.

1.4.14 iSCSI and iSCSI Extensions over RDMA

iSCSI is an alternative means of attaching hosts and external storage controllers to the IBM SAN Volume Controller.

The iSCSI function is a software function that is provided by the IBM Storage Virtualize software, not hardware. In Version 7.7, IBM introduced software capabilities to enable the underlying virtualized storage to attach to the IBM SAN Volume Controller by using the iSCSI protocol.

The iSCSI protocol enables the transportation of SCSI commands and data over an IP network (TCP/IP), which is based on IP routers and Ethernet switches. iSCSI is a block-level protocol that encapsulates SCSI commands. Therefore, it uses an IP network rather than FC infrastructure.

The major functions of iSCSI include encapsulation and the reliable delivery of CDB transactions between initiators and targets through the IP network, especially over a potentially unreliable IP network.

Every iSCSI node in the network must have the following iSCSI components:

- An *iSCSI name* is a location-independent, permanent identifier for an iSCSI node. An iSCSI node has one iSCSI name, which stays constant for the life of the node. The terms *initiator name* and *target name* also refer to an iSCSI name.
- An *iSCSI address* specifies the iSCSI name of an iSCSI node and a location of that node. The address consists of a hostname or IP address, a TCP port number (for the target), and the iSCSI name of the node. An iSCSI node can have any number of addresses, which can change at any time, particularly if they are assigned by Dynamic Host Configuration Protocol (DHCP). An IBM SAN Volume Controller node represents an iSCSI node and provides statically allocated IP addresses.

[1] https://ieeexplore.ieee.org/document/4493450

IBM SAN Volume Controller models SV3, SV2, and SA2 support 25 Gbps Ethernet adapters that provide iSCSI and iSCSI Extensions over RDMA (iSER) connections. The IBM SAN Volume Controller model SV3 also supports 100 Gbps Ethernet adapters.

iSER is a network protocol that extends the iSCSI protocol to use RDMA. You can implement RDMA-based connections that use Ethernet networking structures and connections without upgrading hardware. As of this writing, the system supports RDMA-based connections with RDMA over Converged Ethernet (RoCE) or Internet-Wide Area RDMA Protocol (iWARP).

For host attachment, these 25 Gbps adapters support iSCSI and RDMA-based connections; however, for external storage systems, only iSCSI connections are supported through these adapters. When the 25 Gbps adapter is installed on nodes in the system, RDMA technology can be used for node-to-node communications.

Note: The 100 Gbps adapter on the IBM SAN Volume Controller model SV3 supports iSCSI. However, the performance is limited to 25 Gbps per port.

1.4.15 Data reduction pools

Data reduction pools (DRPs) represent an enhancement to the storage pool concept. The reason is that the virtualization layer is primarily a simple layer that runs the task of lookups between virtual and physical extents.

DRPs are a type of storage pool that implements various techniques, such as thin-provisioning, compression, and deduplication, to reduce the amount of physical capacity that is required to store data. Savings in storage capacity requirements translate into the reduction of the cost of storing the data.

By using DRPs, you can automatically de-allocate and reclaim the capacity of thin-provisioned volumes that contain deleted data and enable this reclaimed capacity to be reused by other volumes. Data reduction provides more capacity from compressed volumes because of the implementation of the new log-structured array.

Deduplication

Data deduplication is one of the methods of reducing storage needs by eliminating redundant copies of data. Data reduction is a way to decrease the storage disk infrastructure that is required, optimize the usage of storage disks, and improve data recovery infrastructure efficiency.

Existing data or new data is standardized into chunks that are examined for redundancy. If data duplicates are detected, pointers are shifted to reference a single copy of the chunk, and the duplicate data sets are then released.

To estimate the potential capacity savings that data reduction can provide on the system, use the Data Reduction Estimation Tool (DRET). This tool scans target workloads on all attached storage arrays, consolidates these results, and generates an estimate of potential data reduction savings for the entire system.

To download the DRE, see IBM Support.

1.4.16 IP replication

IP replication was introduced in Version 7.2. It enables data replication between IBM Storage Virtualize family members. IP replication uses the IP-based ports of the cluster nodes.

Similar to traditional FC-based mirroring, the IP replication function does not affect servers and applications. All remote mirroring modes, which include MM, GM, GMCV, and the new policy-based replication, are supported.

The configuration of the system can be done in the GUI. IBM Storage Virtualize family systems detect each other in the network and can be selected from the GUI.

IP replication includes Bridgeworks SANSlide network optimization technology, and it is available at no additional charge. Remote mirror is a chargeable option, but the price does not change with IP replication. Existing remote mirror users can access the function at no extra charge.

IP connections that are used for replication can have long latency, which can be caused by distance or by many “hops” between switches and other appliances in the network. Traditional replication solutions transmit data, wait for a response, and then transmit more data, which can result in network usage as low as 20% (based on IBM measurements). In addition, this scenario worsens the longer the latency.

Bridgeworks SANSlide technology, which is integrated with the IBM Storage Virtualize family, requires no separate appliances and incurs no extra cost and configuration steps. It uses artificial intelligence (AI) technology to transmit multiple data streams in parallel, adjusting automatically to changing network environments and workloads.

SANSlide improves network bandwidth usage up to 3x. Therefore, customers can deploy a less costly network infrastructure, or take advantage of faster data transfer to speed replication cycles, improve remote data currency, and enjoy faster recovery.

1.4.17 IBM Storage Virtualize copy services

IBM Storage Virtualize supports the following copy services functions:

- Remote copy, synchronous or asynchronous
- Policy-based replication
- FlashCopy (PiT copy) and Transparent Cloud Tiering (TCT)
- IBM HyperSwap

Copy services functions are implemented within a single IBM SAN Volume Controller, or between multiple members of the IBM Storage Virtualize family.

The copy services layer sits above and operates independently of the function or characteristics of the underlying disk subsystems that are used to provide storage resources to an IBM SAN Volume Controller.

1.4.18 Synchronous or asynchronous remote copy

The general application of remote copy seeks to maintain two copies of data. Often, the two copies are separated by distance, but not always. The remote copy can be maintained in synchronous or asynchronous modes. IBM Storage Virtualize, Metro Mirror, and Global Mirror are the IBM-branded terms for the functions that provide synchronous remote copy and asynchronous remote copy.

Synchronous remote copy ensures that updates are committed at the primary and secondary volumes before the application considers the updates complete. Therefore, the secondary volume is fully dated if it is needed in a failover. However, the application is fully exposed to the latency and bandwidth limitations of the communication link to the secondary volume. In a

truly remote situation, this extra latency can have a significant adverse effect on application performance.

Special configuration guidelines exist for SAN fabrics and IP networks that are used for data replication. Consider the distance and available bandwidth of the intersite links.

A function of Global Mirror for low bandwidth was introduced in IBM Storage Virtualize 6.3. It uses change volumes that are associated with the primary and secondary volumes. These points in time copies are used to record changes to the remote copy volume, the FlashCopy map that exists between the secondary volume and the change volume, and between the primary volume and the change volume. This function is called Global Mirror with change volumes (cycling mode).

Figure 1-13 shows an example of this function where you can see the relationship between volumes and change volumes.

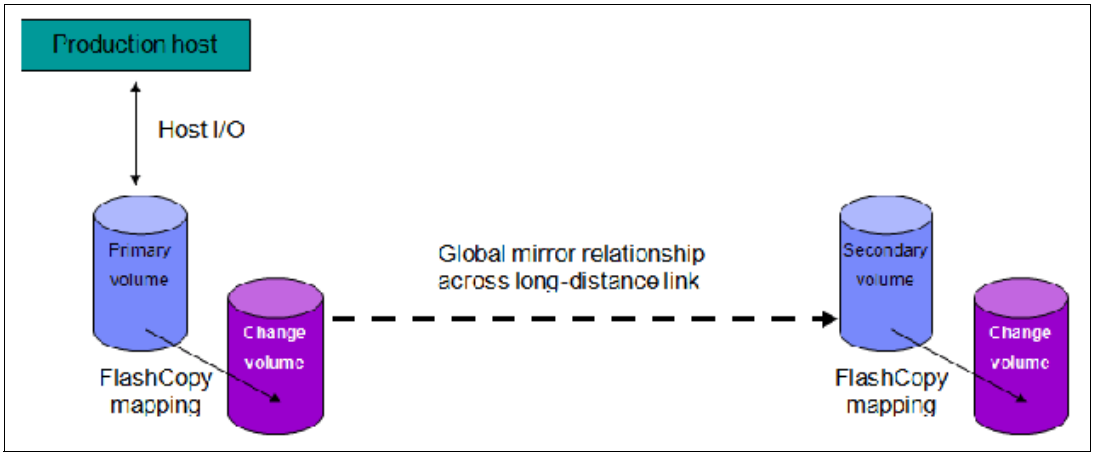


Figure 1-13 Global Mirror with change volumes

In asynchronous remote copy, the application acknowledges that the write is complete before the write is committed at the secondary volume. Therefore, on a failover, specific updates (data) might be missing at the secondary volume.

The application must have an external mechanism for recovering the missing updates, if possible. This mechanism can involve user intervention. Recovery on the secondary site involves starting the application on this recent backup, and then rolling forward or backward to the most recent commit point.

1.4.19 Policy-based replication

Policy-based replication (PBR) provides simplified configuration and management of asynchronous replication between two systems.

Policy-based replication uses volume groups to automatically deploy and manage replication. This feature can simplify configuring, managing, and monitoring replication between two systems. Policy-based replication can simplify asynchronous replication with the following key advantages:

- Uses volume groups instead of consistency groups. With volume groups, all volumes are replicated based on the assigned policy
- Simplifies administration by removing the need to manage relationships and change volumes

- Automatically manages provisioning on the remote system
- Supports easier visualization of replication during a site failover
- Automatically notifies you when the recovery point objective (RPO) is exceeded
- Easy-to-understand status and alerts on the overall health of replication

Policy-based replication is supported in version 8.5.2 or later on the following products:

- IBM SAN Volume Controller
- IBM Storage FlashSystem 9500
- IBM Storage FlashSystem 9200
- IBM Storage FlashSystem 9100
- IBM Storage FlashSystem 7300
- IBM Storage FlashSystem 7200
- IBM Storage FlashSystem 5200, requires a minimum of 128 GiB memory in each node canister
- IBM Storage Virtualize for Public Cloud

To learn more about concepts and objects that are related to PBR, see Policy-based replication in IBM Documentation.

Also, see *Policy-Based Replication with IBM Storage FlashSystem, IBM SAN Volume Controller and IBM Storage Virtualize,* IBM REDP5704.

1.4.20 FlashCopy and Transparent Cloud Tiering

FlashCopy and TCT are used to make a copy of a source volume on a target volume. After the copy operation starts, the original content of the target volume is lost, and the target volume has the contents of the source volume as they existed at a single point in time. Although the copy operation takes time, the resulting data at the target appears as though the copy was made instantaneously.

FlashCopy

FlashCopy is sometimes described as an instance of a time-zero (T0) copy or a point-in-time copy technology.

FlashCopy can be performed on multiple source and target volumes. FlashCopy enables management operations to be coordinated so that a common single PiT is chosen for copying target volumes from their respective source volumes.

When you use IBM Storage Virtualize, multiple target volumes can undergo FlashCopy from the same source volume. This capability can be used to create images from separate PiTs for the source volume, and to create multiple images from a source volume at a common PiT.

Reverse FlashCopy enables target volumes to become restore points for the source volume without breaking the FlashCopy relationship, and without waiting for the original copy operation to complete. IBM Storage Virtualize supports multiple targets and multiple rollback points.

Most customers want to integrate the FlashCopy feature for PiT copies and for quick recovery of their applications and databases. An IBM solution for this goal is provided by IBM Storage Protect and IBM Copy Data Management. For more information, see this IBM Storage web page.

Transparent Cloud Tiering

IBM Storage Virtualize Transparent Cloud Tiering (TCT) is an alternative for data protection, backup, and restores that interfaces to Cloud Services Providers (CSPs), such as IBM Cloud. The TCT function helps organizations to reduce costs that are related to power and cooling when offsite data protection is required to send sensitive data out of the main site.

TCT uses IBM FlashCopy techniques that provide full and incremental snapshots of several volumes. Snapshots are encrypted and compressed before being uploaded to the cloud. Reverse operations are also supported within that function. When a set of data is transferred out to the cloud, the volume snapshot is stored as object storage.

IBM Cloud Object Storage uses an innovative approach and a cost-effective solution to store a large amount of unstructured data. It also delivers mechanisms to provide security services, HA, and reliability.

You can use the management GUI for initial setup to define advanced security settings, and to view audit logs that record all backup and restore to cloud.

For more information, see IBM Cloud Object Storage.

1.4.21 IBM HyperSwap

The IBM HyperSwap function is an HA feature that provides dual-site access to a volume. When you configure a system with IBM HyperSwap topology, the system configuration is split between two sites for data recovery, migration, or HA use cases.

When an HyperSwap topology is configured, each node, external storage system, and host in the system configuration must be assigned to one of the sites in the topology. Both nodes of an I/O group must be at the same site. This site must be the same site as the external storage systems that provide the managed disks to that I/O group.

When managed disks are added to storage pools, their site attributes must match. This requirement ensures that each copy in a HyperSwap volume is fully independent and is at a distinct site.

When the system is configured between two sites, HyperSwap volumes have a copy at one site and a copy at another site. Data that is written to the volume is automatically sent to both copies. If one site is no longer available, the other site can provide access to the volume. If ownership groups are used to manage access to HyperSwap volumes, both volume copies and users who access them must be assigned to the same ownership group.

A 2-site HyperSwap configuration can be extended to a third site for DR that uses the IBM Storage Virtualize 3-Site Orchestrator. For more information, see *IBM Storage Virtualize 3-Site Replication*, SG24-8504.

1.5 IBM SAN Volume Controller models

The following IBM SAN Volume Controller models are supported by IBM Storage Virtualize V8.6:

- SV3
- SV2
- SA2

1.5.1 IBM SAN Volume Controller SV3

Figure 1-14 shows the front view of the IBM SAN Volume Controller SV3.



Figure 1-14 IBM SAN Volume Controller SV3 front view

Figure 1-15 shows the rear view of the IBM SAN Volume Controller SV3.



Figure 1-15 IBM SAN Volume Controller SV3 rear view

Figure 1-16 shows the internal hardware components of an IBM SAN Volume Controller SV3 node canister. To the left is the front of the canister where fan modules are located, followed by two Ice Lake CPUs and Dual Inline Memory Module (DIMM) slots. The battery backup units and the PCIe adapter cages are shown on the right side. Each of these adapters cages holds two PCIe adapters, except for cage number 2, which is dedicated to the compression card.

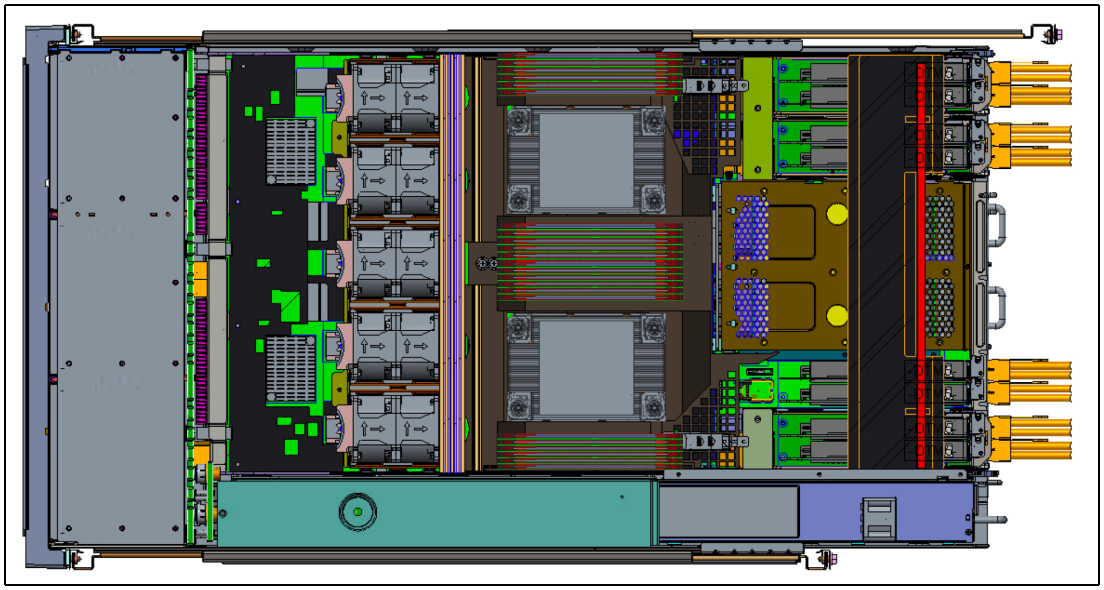


Figure 1-16 IBM SAN Volume Controller SV3 internal hardware components

Figure 1-17 shows the internal architecture of the IBM SAN Volume Controller SV3 model. You can see that the PCIe switch is still present, but has no outbound connections because these models do not support any internal drives. The PCIe switch is used for internal functions and monitoring purposes within the IBM SAN Volume Controller enclosure.

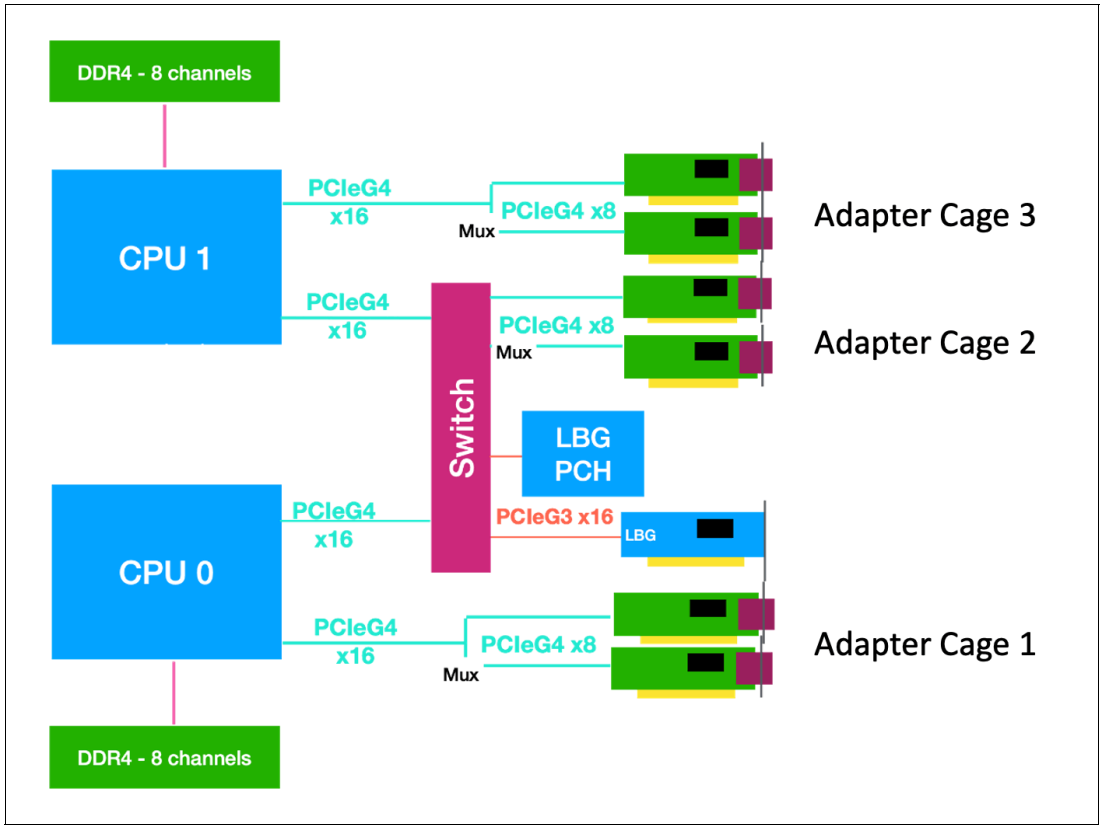


Figure 1-17 IBM SAN Volume Controller SV3 internal architecture

1.5.2 IBM SAN Volume Controller SV2 and SA2

Figure 1-18 shows the front view of the IBM SAN Volume Controller SV2 or SA2.



Figure 1-18 IBM SAN Volume Controller SV2 or SA2 front view

Figure 1-19 shows the rear view of the IBM SAN Volume Controller SV2 or SA2.

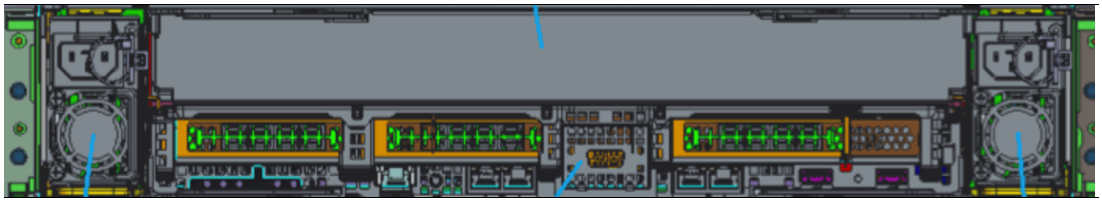


Figure 1-19 IBM SAN Volume Controller SV2 or SA2 rear view

Figure 1-20 shows the internal hardware components of an IBM SAN Volume Controller SV2 and SA2 node canister. To the left is the front of the canister where fan modules and battery backup are located, followed by two Cascade Lake CPUs and Dual Inline Memory Module (DIMM) slots, and PCIe risers for adapters on the right.

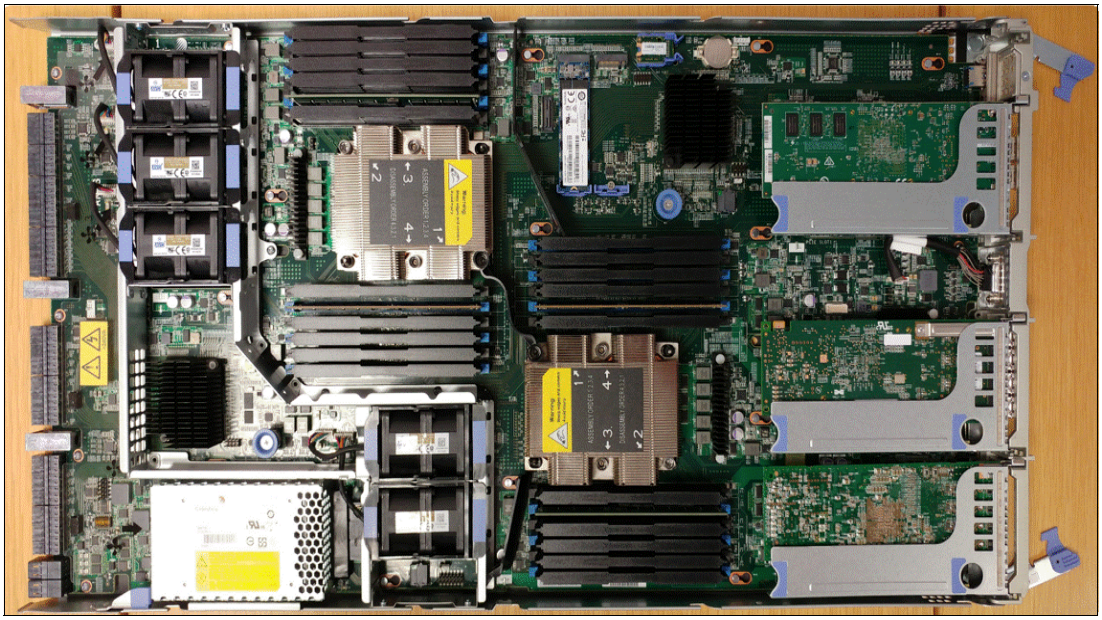


Figure 1-20 Internal hardware components

Figure 1-21 shows the internal architecture of the IBM SAN Volume Controller SV2 and SA2 models. You can see that the PCIe switch is still present, but has no outbound connections because these models do not support any internal drives. The PCIe switch is used for internal monitoring purposes within the IBM SAN Volume Controller enclosure.

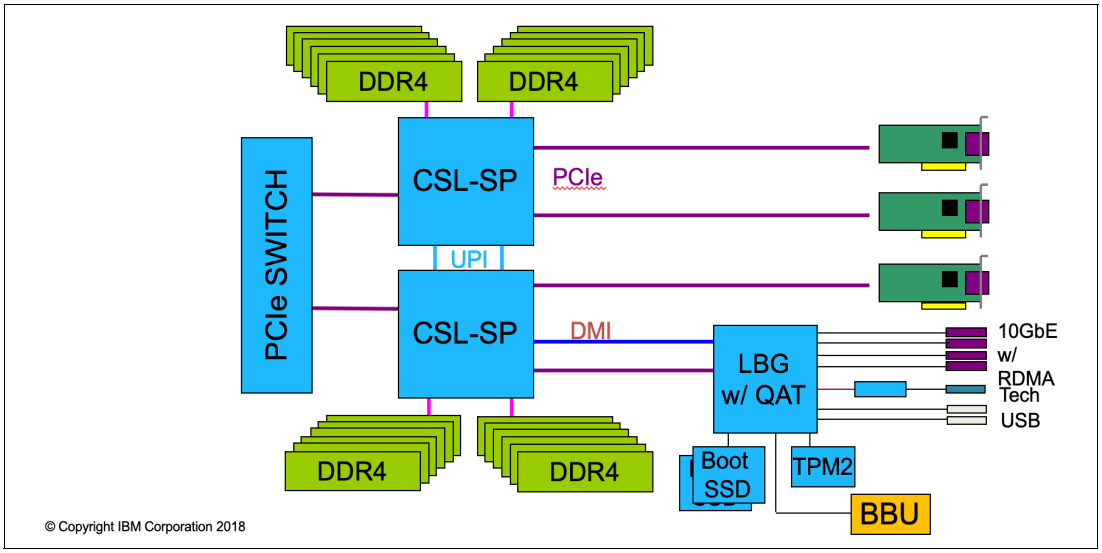


Figure 1-21 IBM SAN Volume Controller SV2 and SA2 internal architecture

Note: IBM SAN Volume Controller models SV3, SV2, and SA2 do not support any type of expansion enclosures.

1.5.3 IBM SAN Volume Controller model comparisons

All IBM SAN Volume Controller models are delivered in a 2U 19-inch rack-mounted enclosure. At the time of writing, three models of the IBM SAN Volume Controller listed in Table 1-2.are available.

More information: For the most up-to-date information about features, benefits, and specifications, see IBM SAN Volume Controller.

The information in this book is valid at the time of this writing and covers IBM Storage Virtualize V8.6. However, in future releases of the IBM SAN Volume Controller, you might see new features and improved specifications.

Table 1-2 IBM SAN Volume Controller base models

Feature	Models		
	2145/2147-SV2	**2145/2147-SA2**	**2145/2147-SV3**
Processor	Two Intel Cascade Lake 5218 Series, 16-cores, and 2.30 GHz (Gold)	Two Intel Cascade Lake 4208 Series, 8-cores, and 2.10 GHz (Silver)	Two Intel Ice Lake 4189 Series, 24-cores, and 2.4 GHz (Gold)
Base cache memory	128 GB	128 GB	512 GB
I/O ports and management	Four 10 Gb Ethernet ports for 10 Gb iSCSI connectivity and system management	Four 10 Gb Ethernet ports for 10 Gb iSCSI connectivity and system management	Two 1 Gb Ethernet ports for system management only (non-iSCSI ports)
Technician port	Single 1 Gb Ethernet	Single 1 Gb Ethernet	Single 1 Gb Ethernet
Maximum host interface adapters slots	3	3	6
USB ports	2	2	1
SAS chain	N/A	N/A	N/A
Max number of dense drawers per SAS chain	N/A	N/A	N/A
Integrated battery units	1	1	2
Power supplies and cooling units	2	2	2

The following optional features are available for IBM SAN Volume Controller SV2 and SA2:

- A 768 GB cache upgrade
- A 4-port 16 Gb FC/FC over NVMe adapter for 16 Gb FC connectivity
- A 4-port 32 Gb FC/FC over NVMe adapter for 32 Gb FC connectivity
- A 2-port 25 Gb iSCSI/iSER/RDMA over Converged Ethernet (RoCE)
- A 2-port 25 Gb iSCSI/iSER/internet Wide-area RDMA Protocol (iWARP)

The SV2 and SA2 systems have dual CPU sockets and three adapter slots along with four 10-GbE RJ45 ports on board.

Note: IBM SAN Volume Controller models SA2 and SV2 do not support FCoE.

The following optional features are available for IBM SAN Volume Controller SV3:

- A 1536 GB cache upgrade
- A 4-port 32 Gb FC/FC over NVMe adapter for 32 Gb FC connectivity
- A 2-port 25 Gb iSCSI/iSER/RDMA over Converged Ethernet (RoCE)
- A 2-port 25 Gb iSCSI/iSER/internet Wide-area RDMA Protocol (iWARP)
- A 2-port 100 Gb NVMe/iSCSI//RDMA over Converged Ethernet (RoCE)

Note: The 25 and 100 Gb adapters are NVMe capable; however, to support NVMe, a software dependency exists (at the time of this writing). Therefore, NVMe/NVMeoF is *not* supported on these cards.

All Ethernet cards can be used with the iSCSI protocol. The 25 Gb iWARP Ethernet cards can also be used for clustering. The 25 Gb and 100 Gb RoCE Ethernet cards can be used for NVMe RDMA.

The comparison of current and previous models of the IBM SAN Volume Controller is shown in Table 1-3. Expansion enclosures are not included in the list.

Table 1-3 Historical overview of IBM SAN Volume Controller models

Model	Cache (GB)	FC (Gbps)	iSCSI (Gbps)	Hardware base	Announced
2145-SV2	128 - 768	16 and 32	25, 50, and 100	Intel Xeon Cascade Lake	06 March 2020
2147-SV2	128 - 768	16 and 32	25, 50, and 100	Intel Xeon Cascade Lake	06 March 2020
2145-SA2	128 - 768	16 and 32	25, 50, and 100	Intel Xeon Cascade Lake	06 March 2020
2147-SA2	128 - 768	16 and 32	25, 50, and 100	Intel Xeon Cascade Lake	06 March 2020
2145-SV3	512 - 1536	32	25 and 100 by way of PCIe adapters only	Intel Xeon Ice Lake	08 March 2022
2147-SV3	512 - 1536	32	25 and 100 by way of PCIe adapters only	Intel Xeon Ice Lake	08 March 2022

Note: IBM SAN Volume Controller SV3, SV2, and SA2 do not support any type of SAS expansion enclosures.

1.6 IBM FlashSystem family

The IBM FlashSystem family, running IBM Storage Virtualize software, was simplified with innovations and enterprise-class features for deployments of all sizes, from entry to mid-range to high-end. A one-platform system allows for ease-of-use to manage seamlessly and securely data across your entire IT infrastructure.

IBM FlashSystem 5015, IBM FlashSystem 5035, IBM FlashSystem 5045 and IBM FlashSystem 5200 deliver entry enterprise solutions. IBM FlashSystem 7200 and 7300 provides a midrange enterprise solution. IBM FlashSystem 9200 and 9500 plus the rack-based IBM FlashSystem 9200R and 9500R provide four high-end enterprise solutions.

Although all the IBM FlashSystem family systems are running the same IBM Storage Virtualize software, the feature set that is available with each of the models is different.

Figure 1-22 shows the IBM FlashSystem family.

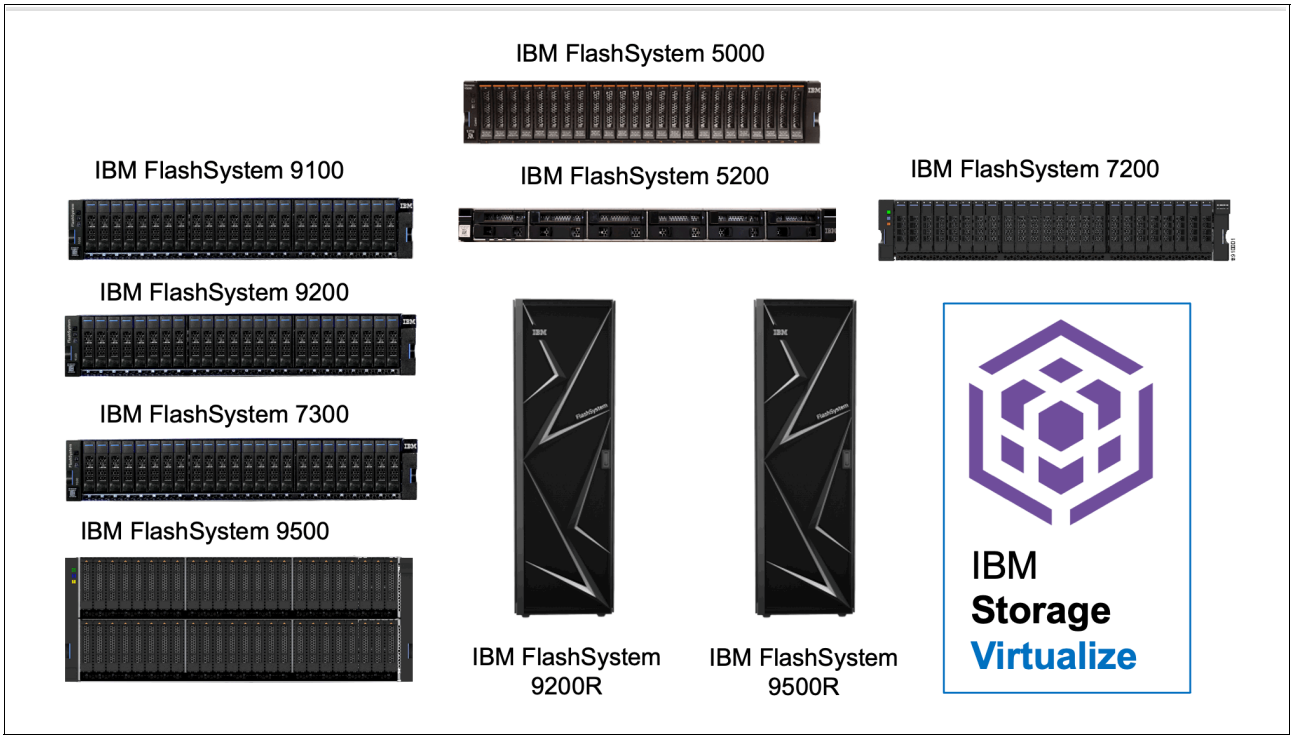


Figure 1-22 IBM FlashSystem family

For an overview of the IBM FlashSystem 5200, see the video IBM Storage Spectrum Virtualize 8.6 GUI, including Volume Group Snapshots (with Safeguarded Copy) and Policy Based Replication.

Note: The IBM FlashSystem 9100 Models AF7, AF8, UF7, and UF8 plus IBM FlashSystem 7200 Model 824 and U7C are no longer sold by IBM, but are included here for because they support IBM Storage Virtualize V8.6 software.

1.6.1 Storage Expert Care

IBM recently expanded the new service offering that is called *Storage Expert Care*, which allows flexible levels and duration of support contracts to supplement specific machine types and models in the IBM FlashSystems family.

IBM Storage Expert Care is designed to simplify and standardize the support approach on the IBM FlashSystem portfolio to keep the systems of a customer operating at peak performance.

The Storage Expert Care offering was originally released with the IBM FlashSystem 5200 and now also covers the IBM FlashSystems 7200, 7300, 9200/R, and 9500/R.

Customers can now choose their preferred level of support from up to three tiers (product-dependent), each priced as a simple percentage of the hardware sales price. This feature allows for easy, straightforward quoting from a single system.

These three tiers allow customers to select the best level of required service to support their environment, ranging from base-level service, through to premium-enhanced service. This Storage Expert Care offering is designed to improve product resiliency and reliability and reduce the operational costs that are associated with managing and maintaining increasingly complex and integrated IT environments.

The following tier selection and features are available:

- Basic:
  - Hardware maintenance with next business day onsite response
  - Software support and services
- Advanced:
  - Hardware maintenance with 24x7 same business day onsite response
  - Software support and services
  - Storage Insights predictive support
- Premium:
  - Hardware maintenance with 24x7 same business day onsite response
  - Software support and services
  - Machine setup services
  - Predictive support
  - Enhanced response time for defect support
  - Hardware remote code load
  - Access to a dedicated technical account manager
- Committed Maintenance - CMSL (optional):
  - Enables IBM Hardware Maintenance Services, which is committed maintenance to be included on top of IBM Storage Expert Care Advanced and Premium Tiers.

    Committed maintenance reduces the cost of downtime by providing a committed time frame to call back, arrive on site, or repair.
  - Reduces the loss of revenue, repair costs, and loss of consumer confidence and shareholder trust by making sure that your products are protected by committed maintenance.

Figure 1-23 shows a summary of the Storage Expert Care Tier Levels.

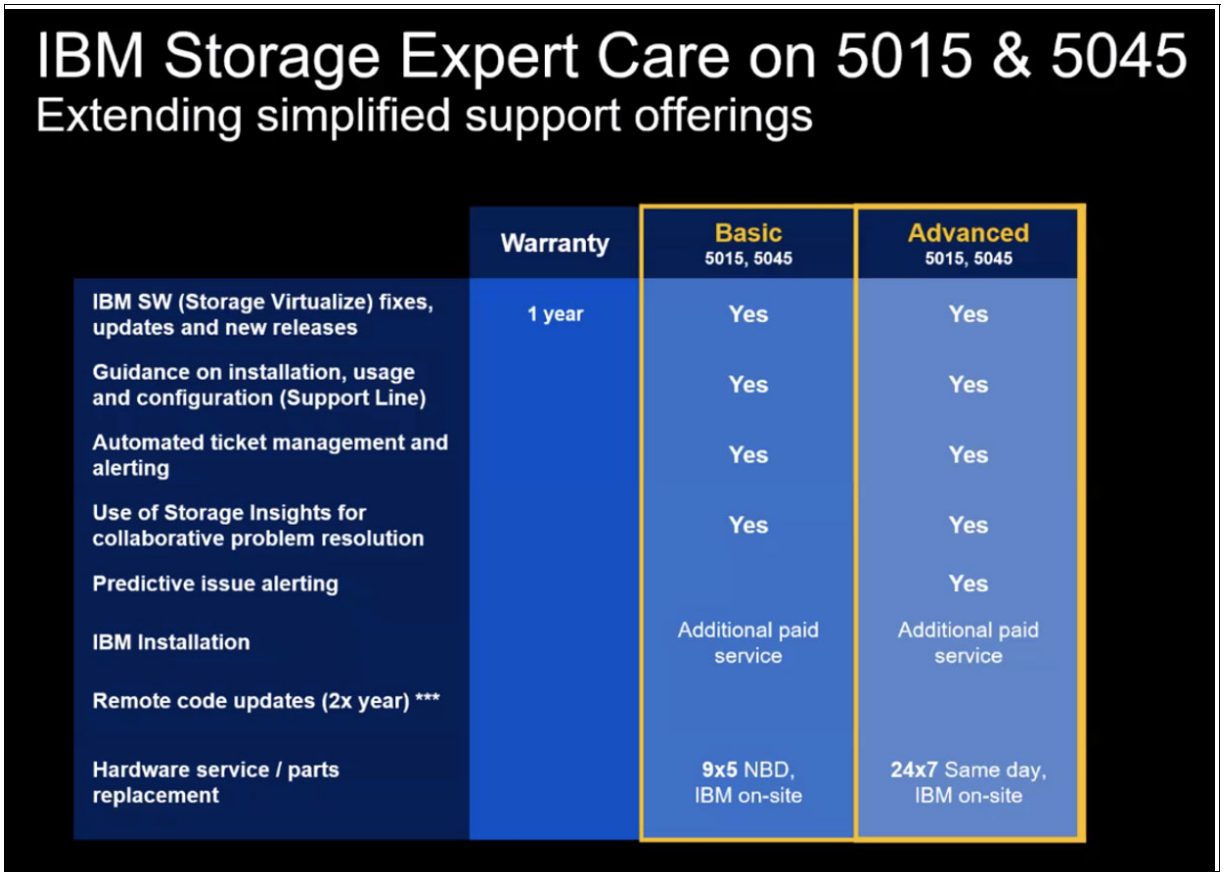


Figure 1-23 Storage Expert Care tier levels for IBM FlashSystem 5015 and IBM FlashSystem 5045

Note: Not all geographies and regions offer all the Storage Expert Care levels of support. If the Storage Expert Care is not announced in a specific country, the traditional warranty and maintenance options are still offered.

For more information about in which countries it is applicable, see the following announcement letters:

- FS5200 Announcement Letter
- FS7200 Announcement Letter
- FS9200 Announcement Letter
- FS7300 and FS9500 Announcement Letters

To support the new Storage Expert Care offering on the older IBM FlashSystems 7200 and 9200, new machine types and models were introduced for these products.

Table 1-4 lists the comparison of the old machine types with the traditional warranty and maintenance offering and the new Storage Expert Care offering.

Table 1-4 IBM FlashSystems 7200 and 9200 product range

Product	Previous Machine Type	Expert Care Machine Type	Model	Function
IBM FlashSystem 7200	2076	4664	824, U7C	Control enclosures
	2076	4664	12G, 24G, 92G	Expansion enclosures
IBM FlashSystem 9200	9846 / 9848	4666	AG8, UG8,	Control enclosures
	9846 / 9848	4666	AFF, A9F	Expansion enclosures

Table 1-5 lists the software PIDs and SWMA feature codes that must be added to the order, depending on the required level of cover.

Table 1-5 Software PIDs and SWMA feature codes

Product	Software type	Control enclosure	Expansion enclosure
FlashSystem 7200	New SW PIDs license	5639-F7K (controller)	5639-EC1 (expansion)
FlashSystem 9200	New SW PIDs license	5639-EA4 (controller)	5639-EB2 (expansion)
FlashSystem 7200	3-Month SWMA	5639-EA2 (controller)	5639-EC2 (expansion)
FlashSystem 9200	3-Month SWMA	5639-EA3 (controller)	5639-EB3 (expansion)

When selecting the level of Storage Expert Care, you must also select the duration of the contract, which can be 1 - 5 years. You can also opt for committed maintenance service levels (CMSL).

The contract and duration has its own machine types and models (in addition to the hardware machine type and model that are listed in Table 1-5):

- FS7200:
  - 4665-P01-05 for Premium
  - 4665-Pxx for Premium with CMSL
- FS9200:
  - 4673-P01-05 for Premium
  - 4673-Pxx for Premium with CMSL

For example, an FS9200 with Premium Expert care for three years is 4673-*PX3*, where:

- P: Premium Level service
- X: Reserved for committed services (CMSL) if added to the expert care contract
- 3: Denotes a three-year contract (if 0, no committed services were purchased)

For more information about IBM Storage Expert Care, see the following IBM Documentation web pages:

- IBM FlashSystem 5200 Storage Expert Care
- IBM FlashSystem 7200 Storage Expert Care
- IBM FlashSystem 9200 Storage Expert Care
- IBM FlashSystem 7300 Storage Expert Care
- IBM FlashSystem 9500 Storage Expert Care
- IBM FlashSystem 50x5 Storage Expert Care

1.7 IBM FlashSystem 9500 overview

This section describes the IBM FlashSystem 9500 architectural components, available models, and enclosure and software features.

1.7.1 IBM FlashSystem 9500 hardware components

IBM FlashSystem 9500 is an all-flash storage system that consists of a control enclosure that runs the IBM Storage Virtualize Software and manages your storage system, communicates with the hosts, and manages interfaces.

Figure 1-24 shows the IBM FlashSystem 9500 front and rear views.

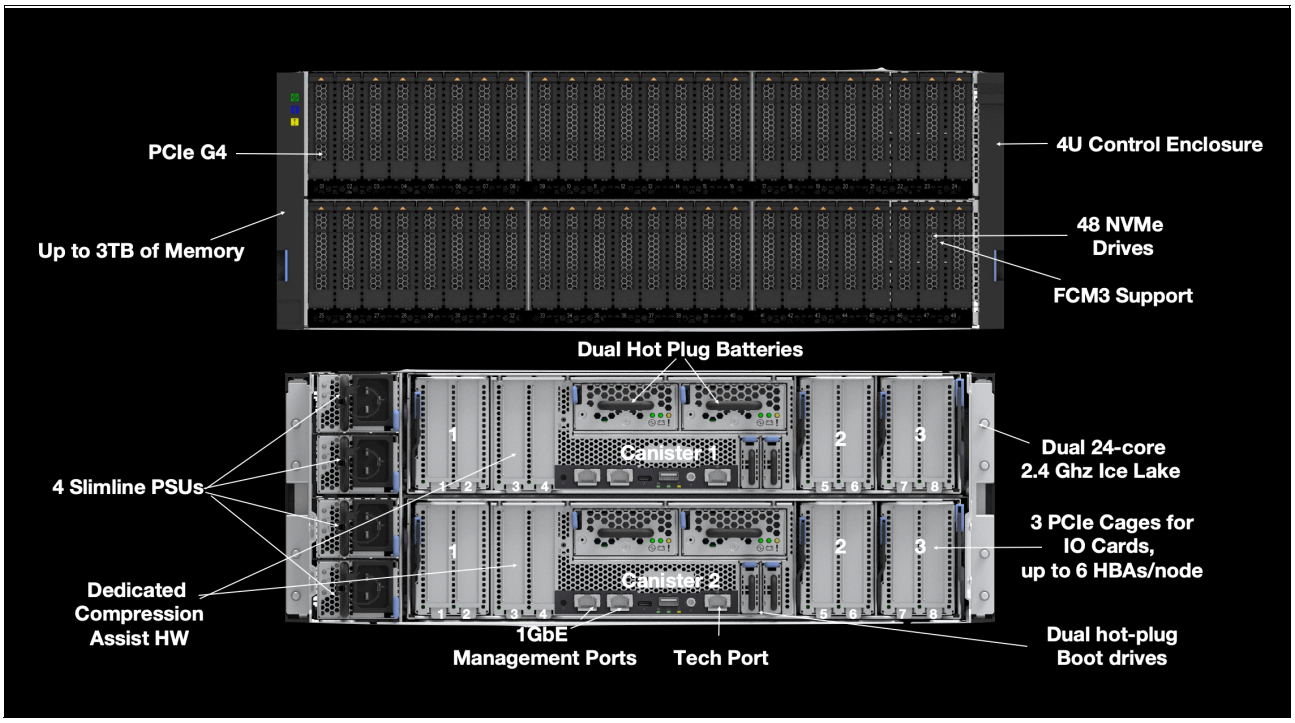


Figure 1-24 IBM FlashSystem 9500 front and rear views

IBM FlashSystem features the following 9500 components:

- IBM FlashSystem 9500 control enclosure:
  - Node canisters
  - Power supply units (PSUs): Uses C19 connectors not C13 type
  - Battery modules
  - Fan modules
  - Interface cards
  - Ice Lake CPUs and DIMM memory slots
  - USB ports
  - Ethernet ports by way of PCIe adapters
- Non-Volatile Memory Express (NVMe)-capable flash drives
- IBM FlashSystem 9000 expansion enclosures (serial-attached Small Computer System Interface [SCSI] [SAS]-attached)

As shown in Figure 1-24 on page 53, the IBM FlashSystem 9500 enclosure consists of redundant PSUs, node canisters, and fan modules to provide redundancy and HA.

Figure 1-25 shows the IBM FlashSystem 9500 internal architecture.

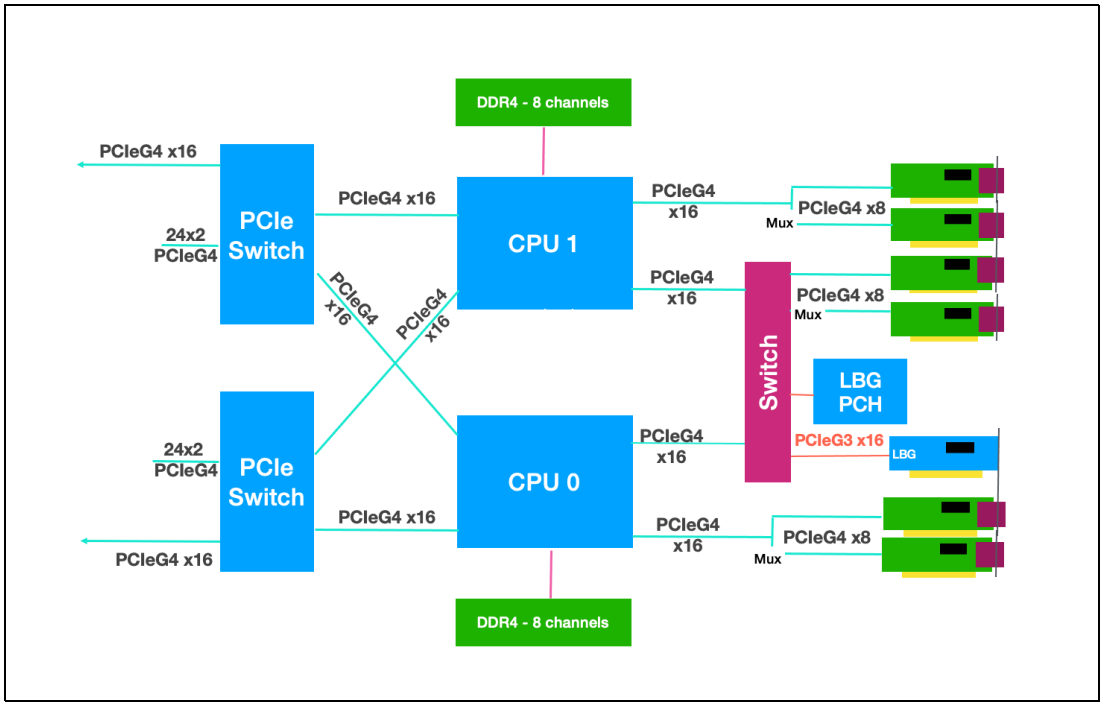


Figure 1-25 IBM FlashSystem 9500 internal architecture

Figure 1-26 shows the internal hardware components of a node canister. On the left is the front of the canister, where the NVMe drives and fan modules are installed, followed by two Ice Lake CPUs and memory DIMM slots, and Peripheral Component Interconnect Express (PCIe) cages for the adapters on the right. The dual battery backup units are in the center, between the PCIe adapter cages.

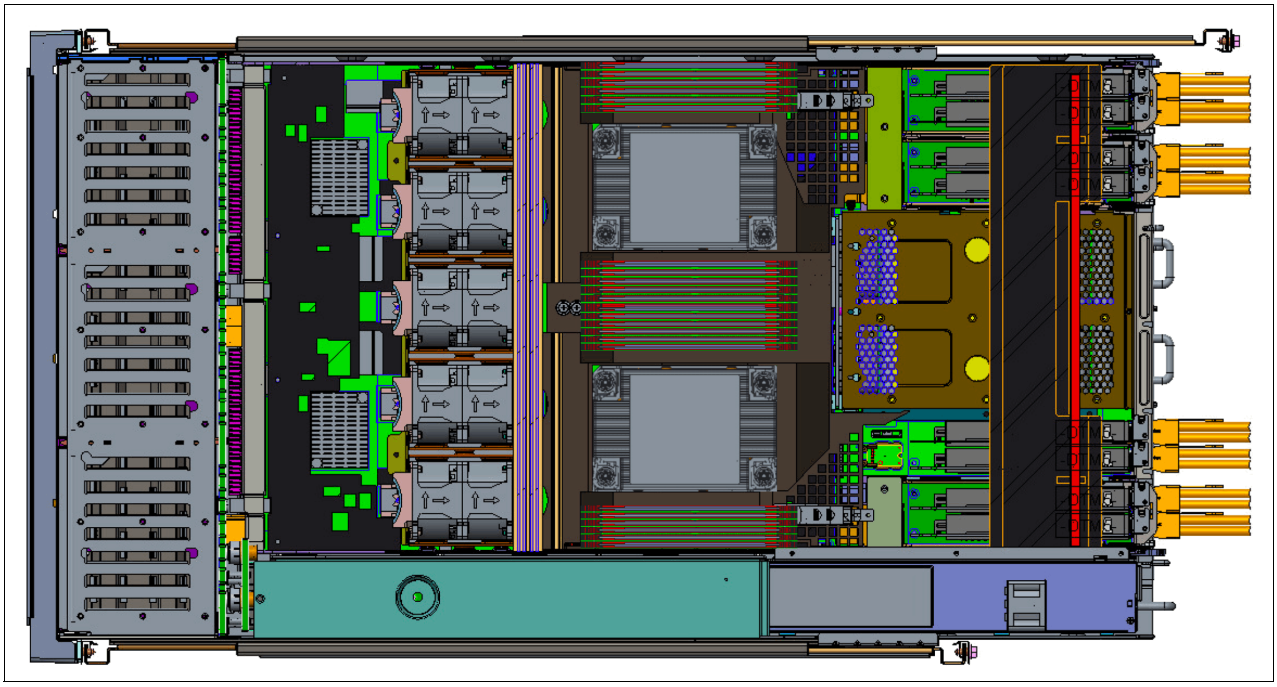


Figure 1-26 FS9500 internal hardware components

1.7.2 IBM FlashSystem 9500 control enclosure

The IBM FlashSystem 9500 system is a 4U model that can house a maximum of 48 NVMe-capable flash drives of various capacities and can be configured with a maximum of 3.0 TB of cache.

Note: Rules for the plugging of the NVMe drives in the control enclosure have been updated. See the “IBM FlashSystem 9500 NVMe drive options” on page 57.

An IBM FlashSystem 9500 clustered system can contain up to two IBM FlashSystem 9500 systems and up to 3,040 drives in expansion enclosures. The following clustering rules must be considered:

- IBM FlashSystem 9500 systems can be clustered only with another IBM FlashSystem 9500.
- IBM FlashSystem 9500 systems cannot be clustered with existing IBM FlashSystem 9200 or IBM FlashSystem 7200 or 7300 systems.

The IBM FlashSystem 9500 control enclosure node canisters are configured for active-active redundancy. The node canisters provide a web interface, Secure Shell (SSH) access, and Simple Network Management Protocol (SMNP) connectivity through external Ethernet interfaces. By using the web and SSH interfaces, administrators can monitor system performance and health metrics, configure storage, and collect support data, among other features.

IBM FlashSystem 9500 Control Enclosure Model AH8

IBM FlashSystem 9500 Control Enclosure machine type 4666 Model AG8 features the following components:

- Two node canisters, each with four 24-core 2.3 GHz Ice Lake CPUs with compression assist up to 100 gigabits per second (Gbps)
- Cache options from 512 GB (256 GB per canister) to 3.0 TB (1.5 TB per canister)
- Two 1 Gb Ethernet (GbE) onboard ports for IBM FlashSystem 9500 control enclosure management only
- Up to 12 (six per canister) I/O adapter features for:
  - Four-port 32 Gb FC-NVMe card
  - Two-port 25 GbE iSCSI/iSCSI Extensions for Remote Direct Memory Access (RDMA)
  - Two-port 25 GbE iSCSI/Hyperswap over iWARP9 (RPQ only) card
  - Two port 100 GbE iSCSI/NVMe RDMA card
  - 12 Gb SAS ports for expansion enclosure attachment
- 48 slots for 2.5-inch NVMe flash drives: Up to 12 SCM style drives
- 4U 19-inch rack mount enclosure with AC power supplies
- Four hot-swappable boot drives
- Four hot-swappable batteries and AC power supplies

Note: There is a machine type 4983 models AH8 and UHB, which is identical to the 4666, except it is sold with Licensed Internal Code (LIC) in line with the other products in the FlashSystems product portfolio. This ensures that all features are included in the product price except for the encryption.

IBM FlashSystem 9500 Utility Model UH8

IBM FlashSystem 9500 Utility Model UH8 provides a variable capacity storage offering. These models offer a fixed capacity with a base subscription of 35% of the total capacity.

IBM Storage Insights is responsible for monitoring the system and reporting the capacity that was used beyond the base 35%, which is then billed on the capacity-used basis. You can grow or shrink usage, and pay only for the configured capacity.

The IBM FlashSystem Utility Model is provided for customers who can benefit from a variable capacity system, where billing is based on actual provisioned space only. The hardware is leased through IBM Global Finance on a three-year lease, which entitles the customer to use approximately 30 - 40% of the total system capacity at no extra cost (depending on the individual customer contract). If storage needs increase beyond that initial capacity, usage is billed based on the average daily provisioned capacity per terabyte per month, on a quarterly basis.

Example: Total system capacity of 115 TB

A customer has an IBM FlashSystem 9500 Utility Model with 4.8 TB NVMe drives for a total system capacity of 115 TB. The base subscription for such a system is 40.25 TB. During the months where the average daily usage is less than 40.25 TB, no extra billing occurs.

The system monitors daily provisioned capacity and averages those daily usage rates over the month term. The result is the average daily usage for the month.

If a customer uses 45 TB, 42.5 TB, and 50 TB in three consecutive months, IBM Storage Insights calculates the overage as listed in Table 1-6, rounding to the nearest terabyte.

Table 1-6 Billing calculations that are based on customer usage

Average daily	Base	Overage	To be billed
45 TB	40.25 TB	4.75 TB	5 TB
42.5 TB	40.25 TB	2.25 TB	2 TB
50 TB	40.25 TB	9.75 TB	10 TB

The total capacity that is billed at the end of the quarter is 17 TB per month in this example.

Flash drive expansions can be ordered with the system in all supported configurations.

Table 1-7 lists the feature codes that are associated with the IBM FlashSystem 9500 Utility Model UH8 billing.

Table 1-7 IBM FlashSystem 9500 Utility Model UG8 billing feature codes

Feature code	Description
#AE00	Variable Usage 1 TB per month
#AE01	Variable Usage 10 TB per month
#AE02	Variable Usage 100 TB per month

These features are used to purchase the variable capacity that is used in the IBM FlashSystem 9500 Utility Models. The features (#AE00, #AE01, and #AE02) provide terabytes of capacity beyond the base subscription on the system. Usage is based on the average capacity that is used per month. The total of the prior three months' usage is totaled and the corresponding number of #AE00, #AE01, and #AE02 features that are ordered quarterly.

IBM FlashSystem 9500 NVMe drive options

The IBM FlashSystem 9500 control enclosure supports a maximum of 48 NVMe 2.5-inch drives, which can be the IBM FlashCore Module NVMe type drives or the industry-standard NVMe drives.

With partially populated control enclosures, some drive slot plugging rules must be followed to ensure the best possible operating conditions for the drives.

Figure 1-27 on page 58 shows the logical NMVe drive placement, starting from the center of the enclosure (slot 12) on the upper 24 slots. Any slots that do not have an NVMe drive present must have a blank filler installed.

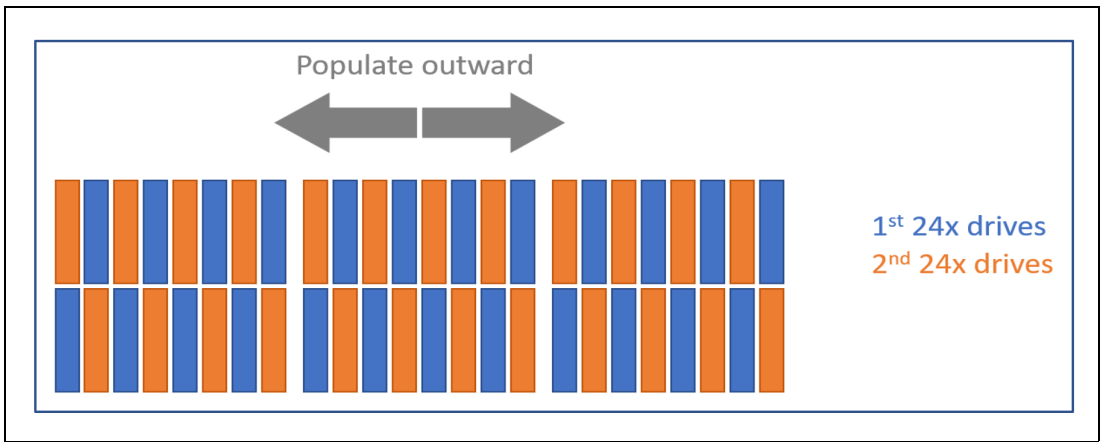


Figure 1-27 Logical NMVe drive placement

Figure 1-28 shows the actual drive population with numbering. This shows how the drives are populated from center out, and then distributing them from top and bottom, as the number of drives increase over time.

Note: The layout in Figure 1-28 has been split at slots 12 and 13 for better clarity on this page, but in reality slots 1 to 24 and slots 25 to 48 are contiguous.

Slot 1	Slot 2	Slot 3	Slot 4	Slot 5	Slot 6	Slot 7	Slot 8	Slot 9	Slot 10	Slot 11	Slot 12
36	11	34	9	32	7	30	5	28	3	26	1
Slot 25	Slot 26	Slot 27	Slot 28	Slot 29	Slot 30	Slot 31	Slot 32	Slot 33	Slot 34	Slot 35	Slot 36
24	47	22	45	20	43	18	41	16	39	14	37
Slot 13	Slot 14	Slot 15	Slot 16	Slot 17	Slot 18	Slot 19	Slot 20	Slot 21	Slot 22	Slot 23	Slot 24
25	2	27	4	29	6	31	8	33	10	35	12
Slot 37	Slot 38	Slot 39	Slot 40	Slot 41	Slot 42	Slot 43	Slot 44	Slot 45	Slot 46	Slot 47	Slot 48
13	38	15	40	17	42	19	44	21	46	23	48

Figure 1-28 NVMe drive population with numbering

1.7.3 IBM FlashSystem 9000 Expansion Enclosure Models AFF and A9F

IBM FlashSystem 9500 Model AH8 and IBM FlashSystem 9500 Utility Model UH8 support the expansion enclosures IBM FlashSystem 9000 Models AFF and A9F.

For more information, see 1.9, “IBM FlashSystem 9000 Expansion Enclosure Models AFF and A9F” on page 62.

Note: The IBM FlashSystem 9500 Model AH8 and IBM FlashSystem 9500 Model UH8 can support a maximum of one IBM FlashSystem 9000 Model A9F dense expansion or three IBM FlashSystem 9000 Model AFF enclosures per chain.

1.8 IBM FlashSystem 9500R Rack Solution overview

IBM FlashSystem 9500R is a pre-cabled, pre-configured rack solution that contains two IBM FlashSystem 9500 control enclosures. It uses IBM Storage Virtualize to linearly scale performance and capacity through clustering.

The IBM FlashSystem 9500R Rack Solution system features a dedicated FC network for clustering and optional expansion enclosures, which are delivered assembled in a rack. Available with two clustered IBM FlashSystem 9500 systems and up to four expansion enclosures, it can be ordered as an IBM FlashSystem 9502R, with the last number denoting the two AG8 controller enclosures in the rack.

The final configuration occurs on site following the delivery of the systems. More components can be added to the rack after delivery to meet the growing needs of the business.

Note: Other than the IBM FlashSystem 9500 control enclosures and its expansion enclosures, the extra components of this solution are not covered under Storage Expert Care. Instead, they have their own warranty, maintenance terms, and conditions.

Rack rules

The IBM FlashSystem 9500R Rack Solution product represents a limited set of possible configurations. Each IBM FlashSystem 9500R Rack Solution order must include the following components:

- Two 4666 Model AH8 control enclosures.
- Two IBM SAN24B-6 or two IBM SAN32C-6 FC switches.
- The following optional expansion enclosures are available by way of MES only, They cannot be ordered with the machine if ordered as a new build:
  - 0 - 4 4666 Model AFF expansion enclosures, with no more than one expansion enclosure per Model AG8 control enclosure and no mixing with the 9848/4666 Model A9F expansion enclosures.
  - 0 - 2 4666 Model A9F expansion enclosures, with no more than one expansion enclosure per Model AG8 control enclosure and no mixing with 9848/4666 Model AFF expansion enclosures.
- One 7965-S42 rack with the suitable power distribution units (PDUs) that are required to power components within the rack.
- All components in the rack much include feature codes #FSRS and #4651.
- For Model AH8 control enclosures, the first and largest capacity enclosure includes feature code #AL01 and #AL02 in capacity order. The 4666 / 4983 model AH8 control enclosure with #AL01 also must include #AL0R.

Following the initial order, each 4666 Model AH8 control enclosure can be upgraded through a MES.

More components can be ordered separately and added to the rack within the configuration limitations of the IBM FlashSystem 9500 system. Customers must ensure that the space, power, and cooling requirements are met. If assistance is needed with the installation of these additional components beyond the service that is provided by your IBM System Services Representative (IBM SSR), IBM Lab Services are available.

Note: There is a new machine type 4983 model AH8 being introduced which is physically identical to the 4666, except it will be sold with Licensed Internal Code (LIC) in line with the other products in the FlashSystems product line. This ensures that all features are included in the product price except for the encryption.

Table 1-8 lists the IBM FlashSystem 9500R Rack Solution combinations, the MTMs, and their associated feature codes.

Table 1-8 IBM FlashSystem 9500R Rack Solution combinations

Machine type and model (MTM)	Description	Quantity
7965-S42	IBM Enterprise Slim Rack	1
8960-F24	IBM SAN24B-6 FC switch (Brocade)	2[a]
8977-T23	IBM SAN32C-6 FC switch (Cisco)	2[b]
4666/4983-AH8	IBM FlashSystem 9500 control enclosure with 3-year Storage Expert Care	2
The following expansion enclosures are available by way of MES order only.		
4666-A9F	IBM FlashSystem 9000 5U large form factor (LFF) high-density expansion enclosures with 3-year Storage Expert Care	0 - 2[c]
4666-AFF	IBM FlashSystem 9000 2U small form factor (SFF) Expansion enclosure with 3-year Storage Expert Care	0 - 4[d]

a. For the FC switch, choose two of machine type (MT) 8977 or two of MT 8960.
b. For the FC switch, choose two of machine type (MT) 8977 or two of MT 8960.
c. For extra expansion enclosures, choose model AFF, model A9F, or none. You cannot use both.
d. For extra expansion enclosures, choose model AFF, model A9F, or none. You cannot use both.

1.8.1 IBM FlashSystem 9500R Rack Solution diagram

This section describes the rack diagram (see Figure 1-30 on page 61) that shows the IBM FlashSystem 9500R Rack Solution configuration.

Key to figures

The key to the symbols that are used in the figures in this section are listed in Table 1-9.

Table 1-9 Key to rack configuration

Label	Description
CTL*n*	► 9848/4666 AH8 control enclosure number *n* of 2 ► CTL1 and CTL2 are required
EXP*n*	9848/4666 expansion enclosure number *n* (optional)
FC SW*n*	► FC switch *n* of 2 ► These switches are both 8977-T32 or both 8960-F24
PDU A, PDU B	PDUs. Both have the same rack feature code: #ECJJ, #ECJL, #ECJN, or #ECJQ.

Figure 1-29 shows the legend that is used to denote the component placement and mandatory gaps for the figures that show the configurations.

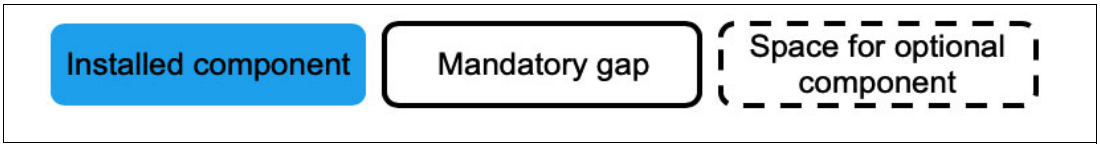


Figure 1-29 Legend to figures in this section

Figure 1-30 shows the standard IBM FlashSystem 9500R Rack Solution configuration in the rack.

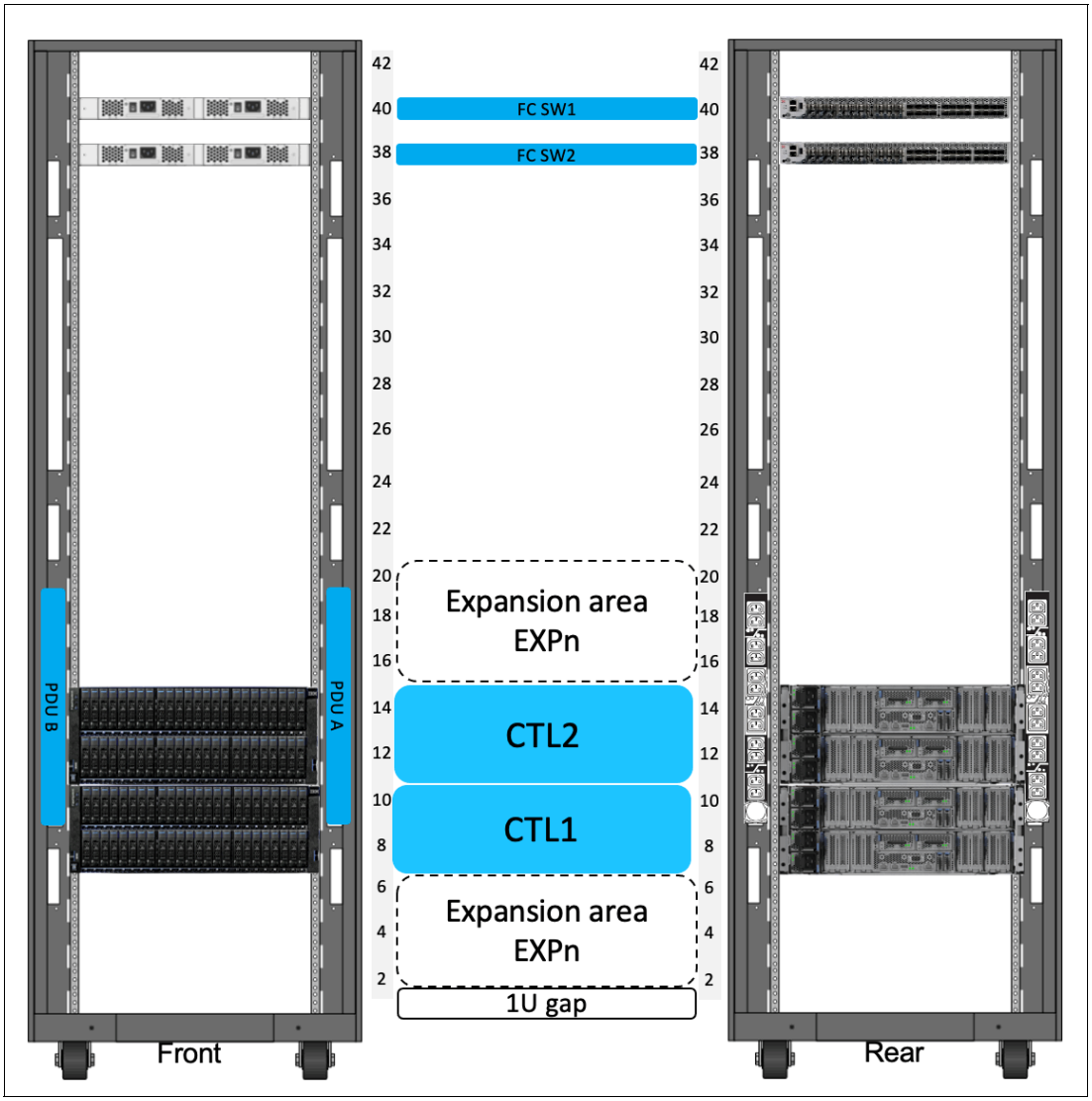


Figure 1-30 IBM FlashSystem 9500R Rack Solution configuration in the rack

Minimum configuration

Consider the following points:

- Control enclosures (CTL) 1 and 2 are mandatory.

- The product includes cables that are suitable for inter-system FC connectivity. You must order extra cables for host and Ethernet connectivity.
- The PDUs and power cabling that are needed depends on what expansion enclosures are ordered.

  Two PDUs with nine C19 outlets are required. This PDU also has three C13 outlets on the forward-facing side.
- FC SW1 and FC SW2 are a pair of IBM SAN32C-6 or IBM SAN24B-6 FC switches.
- You can allocate different amounts of storage (drives) to each CTL component.
- A gap of 1U is maintained below the expansion area to allow for power cabling routing.

Adding configuration with Model A9F and AFF expansion enclosures by way of MES

Consider the following points:

- The product includes cables that are suitable for inter-system FC connectivity. You must order extra cables for host and Ethernet connectivity.
- Any Model A9F or AFF expansion enclosures are installed in U2 - U6 and then in U16-20.
- The CTLs and EXPs are stacked and cabled to the PDU power, with the highest capacity at the bottom. You go upwards, with EXPn attached to CTLn in a bottom-up order by using an SAS adapter on CTLn and cables.
- For the 4666/AH8 CTL1 and CTL2, the following adapter cage rules apply:
  - Adapter cage 1 of each AH8 is dedicated to 32 Gb clustering usage.
  - Adapter cage 2 of each AH8 is used reserved for the compression adapter.
  - Adapter cage 3 is an SAS adapter if it is required, or one of your choice if it is not.
  - Adapter cage 4 is optional for FC 32 G or 25 Gb or 100 Gb Ethernet.

1.8.2 FC cabling and clustering

The IBM FlashSystem 9500R Rack Solution includes specialized internal cabling that is supplied by the manufacturing plant before the machine is shipped to the customer. Connecting the inter-system internal FC cables is done on site at installation time.

For more information, see Planning for clustered system connections within the rack.

1.9 IBM FlashSystem 9000 Expansion Enclosure Models AFF and A9F

IBM FlashSystem 9000 Expansion Enclosure Models AFF and A9F can be attached to an IBM FlashSystem 9200 or IBM FlashSystem 9500 control enclosure to increase the available capacity. They communicate with the control enclosure through a dual pair of 12 Gbps SAS connections. These expansion enclosures can house many flash (solid-state drive [SSD]) SAS type drives.

IBM FlashSystem 9000 Expansion Enclosure Model AFF

IBM FlashSystem 9000 Expansion Enclosure Model AFF holds up to 24 2.5-inch SAS flash drives in a 2U 19-inch rack mount enclosure.

An intermix of capacity drives can be used in any drive slot. The following attachment rules are applicable for each SAS chain:

- IBM FlashSystem 9200: A maximum of 10 IBM FlashSystem 9000 Expansion Enclosure Model AFF enclosures can be attached to the control enclosure to a total of 240 drives maximum.
- IBM FlashSystem 9500: A maximum of three IBM FlashSystem 9000 Expansion Enclosure Model AFF enclosures can be attached. This configuration provides extra capacity with a maximum of 72 drives.

Figure 1-31 shows the front view of the IBM FlashSystem 9000 Expansion Enclosure Model AFF.



Figure 1-31 IBM FlashSystem 9000 Expansion Enclosure Model AFF

IBM FlashSystem 9000 Expansion Enclosure Model A9F

The IBM FlashSystem 9000 Expansion Enclosure Model A9F holds a maximum of 92 3.5-inch SAS flash drives, or 2.5-inch SAS flash drives in carriers, in a 5U 19-inch rack mount enclosure.

An intermix of capacity drives is allowed in any drive slot and the following attachment rules are applicable for each SAS chain:

- IBM FlashSystem 9200: Up to four IBM FlashSystem 9000 Expansion Enclosure Model A9F enclosures can be attached to the control enclosure to a total of 368 drives maximum.
- IBM FlashSystem 9500: One IBM FlashSystem 9000 Expansion Enclosure Model A9F can be attached. This configuration provides extra capacity with a maximum of 92 drives.

Figure 1-32 shows the front view of the IBM FlashSystem 9000 Expansion Enclosure Model A9F.

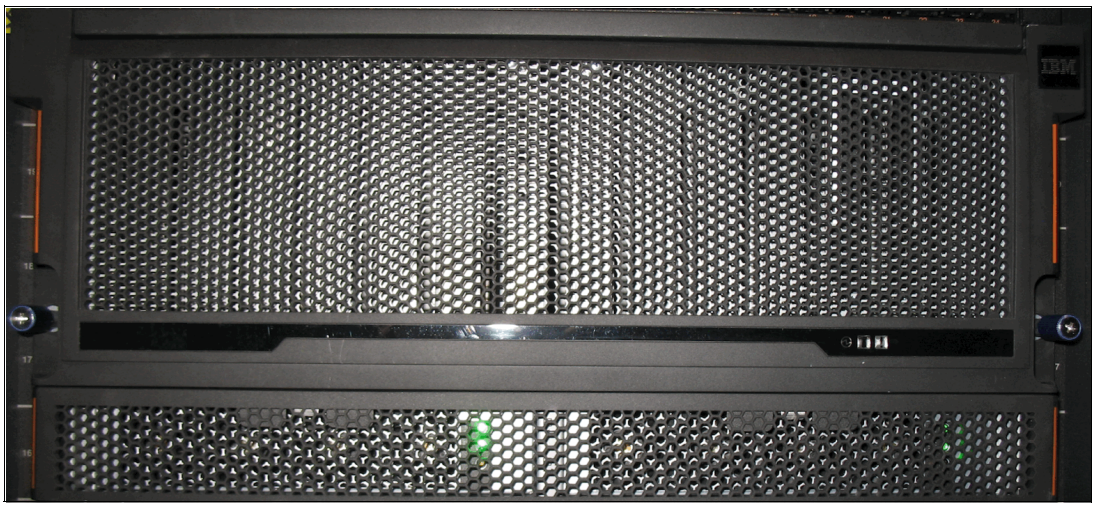


Figure 1-32 IBM FlashSystem 9000 Expansion Enclosure Model front view

SAS chain limitations

When attaching expansion enclosures to the control enclosure, you are not limited by the type of the enclosure. The only limitation for each of the two SAS chains is its chain weight. Each type of enclosure has its own chain weight:

- IBM FlashSystem 9000 Expansion Enclosure Model AFF has a chain weight of 1.
- IBM FlashSystem 9000 Expansion Enclosure Model A9F has a chain weight of 2.5.

For the IBM FlashSystem 9500, the maximum chain weight is 3.

Consider the following example. With the IBM FlashSystem 9500, you can have three IBM FlashSystem 9000 Expansion Enclosure Model AFF or one IBM FlashSystem 9000 Expansion Enclosure Model A9F expansions (3 x 1 or 1 x 2.5).

Figure 1-33 shows an IBM FlashSystem 9200 correctly cabled with SAS cables to two IBM FlashSystem 9000 Expansion Enclosure Model AFF enclosures and one IBM FlashSystem 9000 Expansion Enclosure Model A9F enclosure. The chain wait is a total of 4.5.

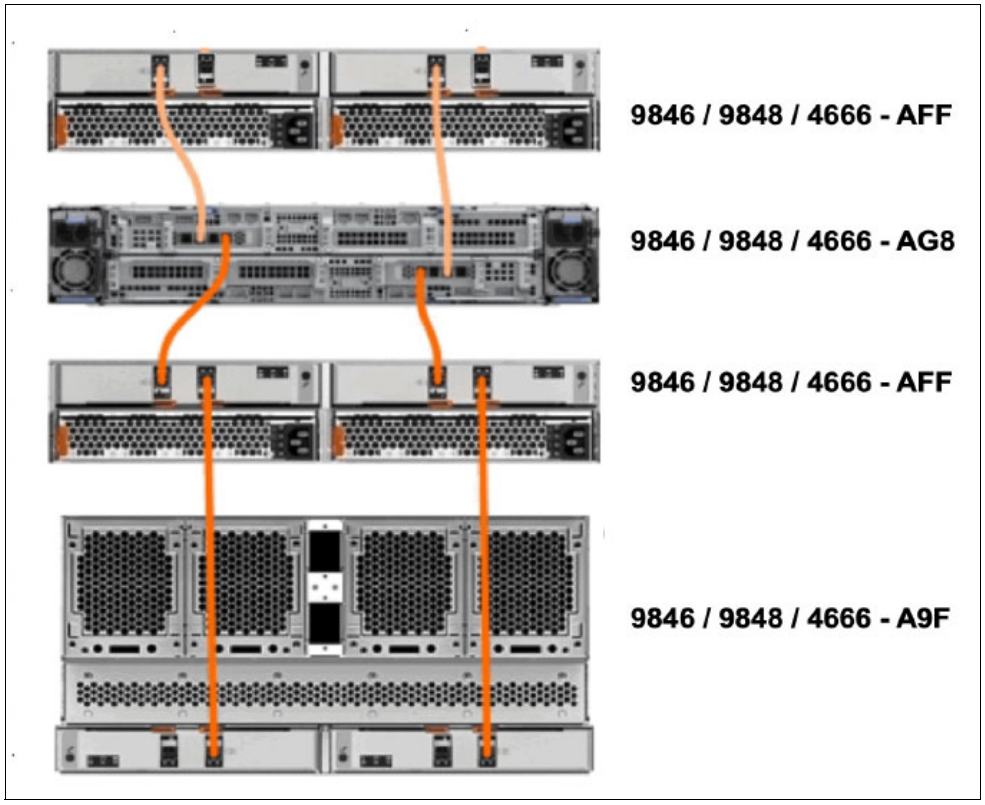


Figure 1-33 IBM FlashSystem 9200 system that is connected to an expansion enclosure

Figure 1-34 shows an IBM FlashSystem 9500 correctly cabled with SAS cables to three IBM FlashSystem 9000 Expansion Enclosure Model AFF enclosures with a chain weight of 3 and one IBM FlashSystem 9000 Expansion Enclosure Model A9F enclosure with a chain weight of 2.5.

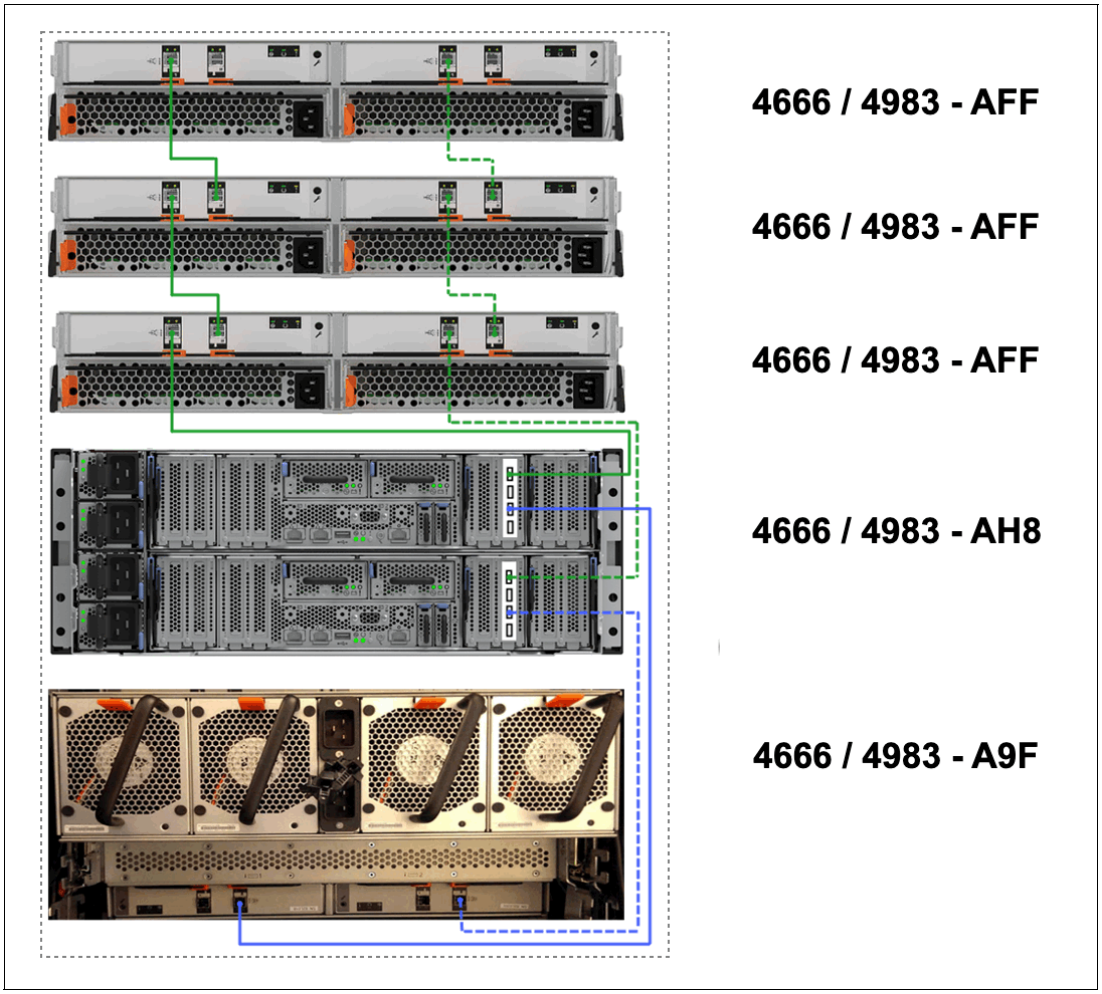


Figure 1-34 IBM FlashSystem 9500 system that is connected to an expansion enclosure

Note: The expansion enclosure rules are the same for the FlashSystem 9500 machine type 4983, which is functionally equivalent to the 4666, except it has Licensed Internal Code (LIC).

For more information, see V8.6.0 Configuration Limits and Restrictions for IBM FlashSystem 9500.

1.10 IBM FlashSystem 7300 overview

Each IBM FlashSystem 7300 system consists of a control enclosure and NVMe-attached flash drives. The control enclosure is the storage server that runs the IBM Storage Virtualize software that controls and provides features to store and manage data.

IBM FlashSystem 7300 has a machine type of 4657. This machine type includes the Storage Virtualize component as Licensed Machine Code (LMC). Therefore, it does not require the purchase of separate software maintenance (SWMA). IBM FlashSystem 7300 still requires licenses for the external virtualization of storage.

For more information, see section 1.20, “Licensing” on page 115.

Figure 1-35 shows the front and rear views of the IBM FlashSystem 7300 system.

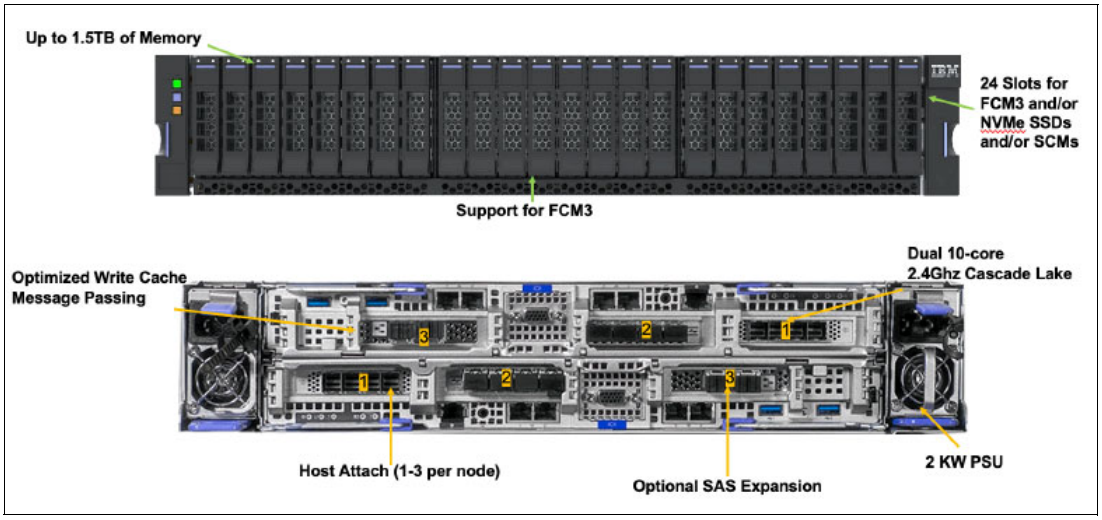


Figure 1-35 IBM FlashSystem 7300 front and rear views

The IBM FlashSystem 7300 features the following core components:

- IBM FlashSystem 7300 control enclosure:
  - PSUs
  - Node canisters
  - Battery modules
  - Fan modules
  - Interface cards
  - Cascade Lake CPUs and memory slots
- NVMe drives
- IBM FlashSystem 7000 SAS-attached expansion enclosures

3

Note: Consider the following points:

- The IBM FlashSystem 7300 (4657-924 and U7D) is available with Storage Expert Care as described in 1.6.1, “Storage Expert Care” on page 49.
- Machine type 4657 FlashSystem 7300 systems can be clustered only with other FlashSystem 7300 systems. Clustering with machine types 2076, 4664, 4666, 9846, or 9848 is not supported.

As shown in Figure 1-36, the IBM FlashSystem 7300 enclosure consists of redundant PSUs, node canisters, and fan modules to provide redundancy and HA.

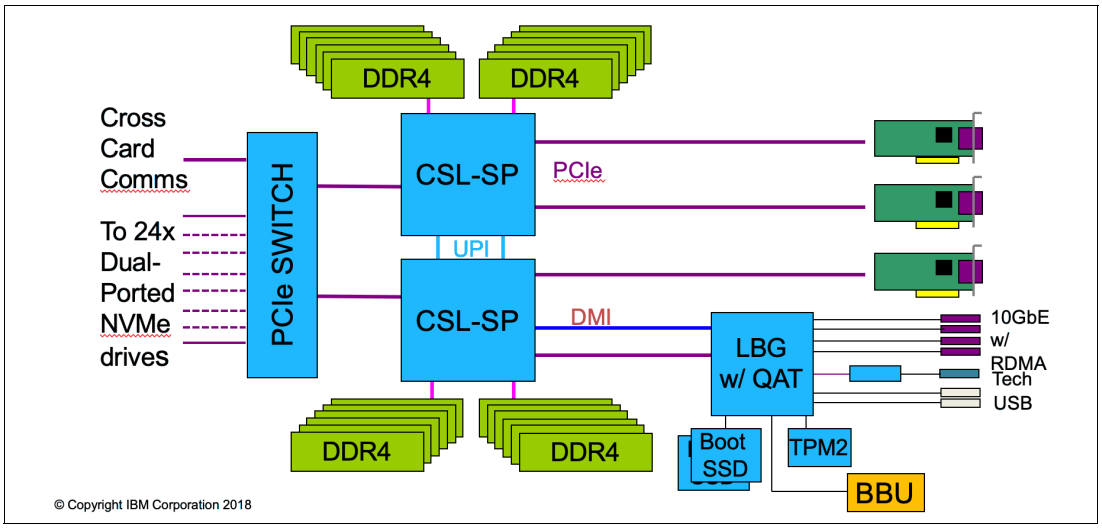


Figure 1-36 IBM FlashSystem 7300 internal architecture

Figure 1-37 shows the internal hardware components of a node canister. On the left side is the front of the canister where fan modules and battery backup are installed, followed by two Cascade Lake CPUs, Dual Inline Memory Module (DIMM) slots, and PCIe risers for adapters on the right side.

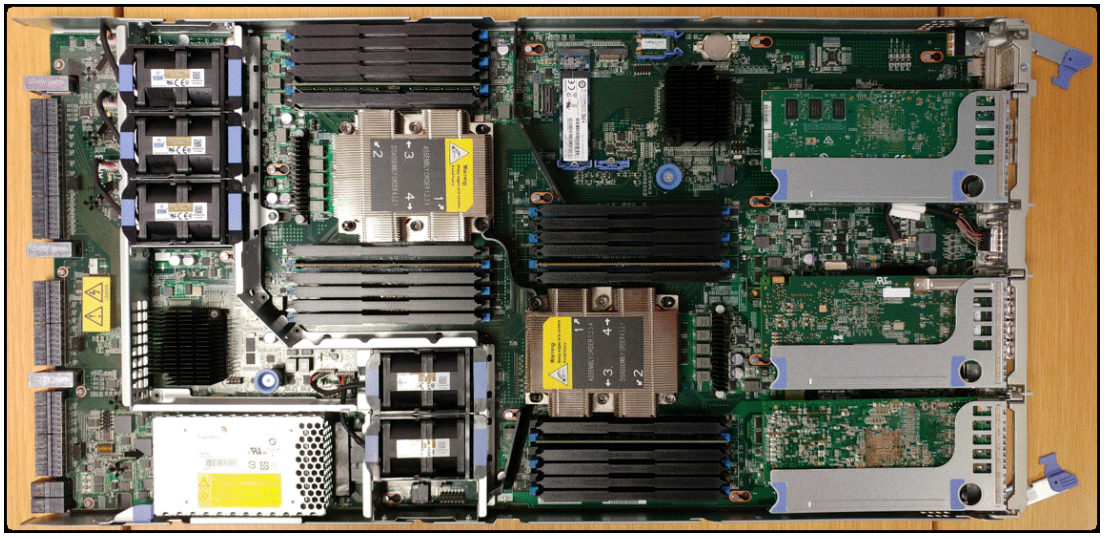


Figure 1-37 Internal hardware components

1.10.1 IBM FlashSystem 7300 control enclosures

IBM FlashSystem 7300 is a 2U model that can support up to 24 NVMe drives (FCM drives with hardware compression and encryption or industry-standard NVMe drives of various capacities or even SCM drives). IBM FlashSystem 7300 can be configured with up to 1.5 TB of cache.

For more information about the supported drive types, see 1.16, “IBM FlashCore Module drives, NVMe SSDs, and SCM drives” on page 100.

An IBM FlashSystem 7300 clustered system can contain up to four IBM FlashSystem 7300 systems and up to 3,040 drives. IBM FlashSystem 7300 systems can be added only into clustered systems that include other IBM FlashSystem 7300 systems.

IBM FlashSystem 7300 Model 924 system

The model 924 system offers the following features:

- Two node canisters, each with two 10-core processors and integrated hardware-assisted compression acceleration
- 256 GB cache (128 GB per canister) standard with options from 768 GB to 1536 GB (per system)
- Eight 10 Gb Ethernet ports standard for iSCSI connectivity (copper)
- 32 Gb FC connectivity options with SCSI and FC-NVMe support (fiber)
- 10/25 Gb and 100 Gb Ethernet ports for iSCSI and NVMe RDMA connectivity
- 12 Gb SAS ports for optional expansion enclosure attachment
- 24 slots for NVMe flash drives, including up to 12 storage-class memory drives
- 2U, 19-inch rack mount enclosure with AC power supplies
- Dual boot drives

IBM FlashSystem 7300 Model U7D system

The IBM 4657 Model U7D is the FlashSystem 7300 with a one-year warranty to be used in the Storage Utility Offering space. It is physically and functionally identical to the FlashSystem 7300 Model 924 except for variable capacity billing.

The variable capacity billing uses IBM Storage Insights to monitor the system use, which allows allocated storage use that is above a base subscription rate to be billed per TB, per month.

Allocated storage is identified as storage that is allocated to a specific host (and unusable to other hosts), whether data is written or not written. For thin-provisioning, the data that is written is considered used. For thick provisioning, total allocated volume space is considered used.

1.10.2 IBM FlashSystem 7000 Expansion Enclosures

The following types of expansion enclosures are available:

- IBM FlashSystem 7000 LFF Expansion Enclosure 4657 Model 12G
- IBM FlashSystem 7000 SFF Expansion Enclosure 4657 Model 24G
- IBM FlashSystem 7000 LFF Expansion Enclosure 4657 Model 92G

> **Note:** Attachment and intermixing of machine type 2076/4664 Expansion Enclosure Models 12G, 24G, and 92G with machine type 4657 FlashSystem 7300 controller and expansion models is *not* supported.

Expansion Enclosure 4657 Model 12G

The IBM FlashSystem 7000 LFF 12G Expansion Enclosure includes the following components:

- Two expansion canisters
- 12 Gb SAS ports for control enclosure and expansion enclosure attachment
- A total of 12 slots for 3.5-inch SAS drives

- 2U 19-inch rack-mounted enclosure with AC power supplies

Figure 1-38 shows the front view of a 4657 Model 12G.



Figure 1-38 IBM FlashSystem 7000 LFF Expansion Enclosure Model 12G

Expansion Enclosure 4657 Model 24G

IBM FlashSystem 7000 SFF Expansion Enclosure 4657 Model 24G includes the following components:

- Two expansion canisters
- 12 Gb SAS ports for control enclosure and expansion enclosure attachment
- A total of 24 slots for 2.5-inch SAS drives
- 2U 19-inch rack mount enclosure with AC power supplies

The SFF expansion enclosure is a 2U enclosure that includes the following components:

- A total of 24 2.5-inch drives (hard disk drives [HDDs] or SSDs).
- Two Storage Bridge Bay (SBB)-compliant Enclosure Services Manager (ESM) canisters.
- Two fan assemblies, which mount between the drive midplane and the node canisters. Each fan module is removable when the node canister is removed.
- Two power supplies.
- An RS232 port on the back panel (3.5 mm stereo jack), which is used for configuration during manufacturing.

Figure 1-39 shows the front of an SFF expansion enclosure.

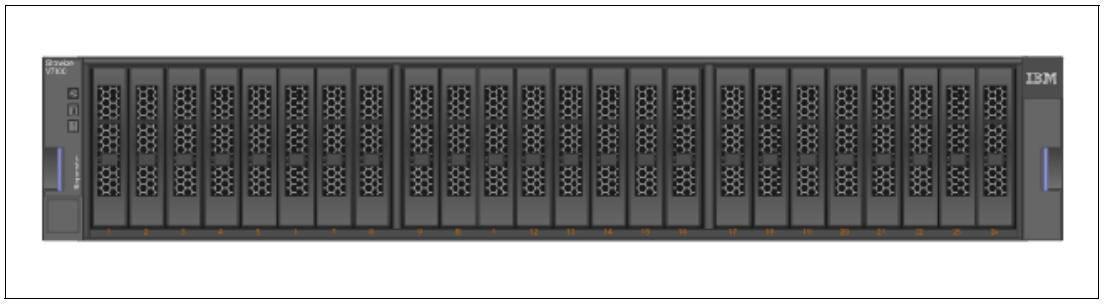


Figure 1-39 Front view of an IBM FlashSystem 7000 SFF expansion enclosure

Figure 1-40 shows the rear view of the expansion enclosure.

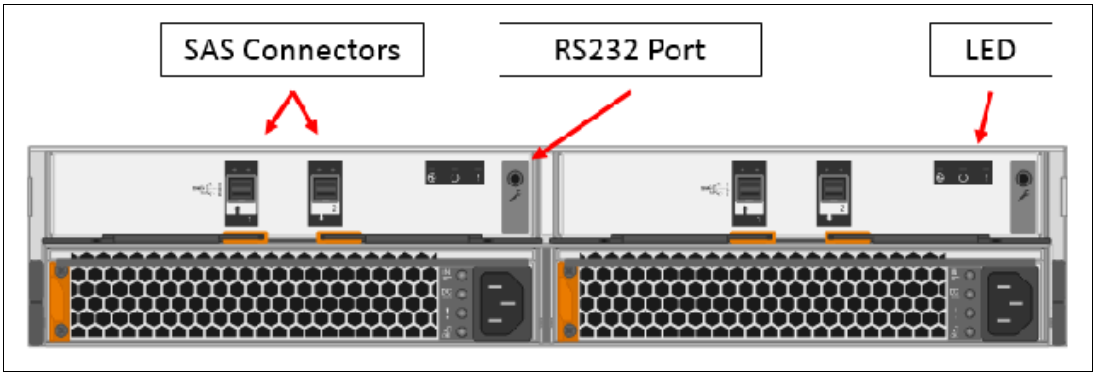


Figure 1-40 Rear of an IBM FlashSystem 7000 expansion enclosure

Dense Expansion Enclosure 92G

Dense expansion drawers, or dense drawers, are disk expansion enclosures that are 5U rack-mounted. Each chassis features two expansion canisters, two power supplies, two expander modules, and a total of four fan modules.

Each dense drawer can hold up to 92 drives that are positioned in four rows of 14 and another three rows of 12 mounted drives assemblies. Two SEMs are centrally located in the chassis: one SEM addresses 54 drive ports, and the other addresses 38 drive ports.

The drive slots are numbered 1 - 14, starting from the left rear slot and working from left to right, back to front.

Each canister in the dense drawer chassis features two SAS ports numbered 1 and 2. The use of SAS port1 is mandatory because the expansion enclosure must be attached to an IBM FlashSystem 7300 node or another expansion enclosure. SAS connector 2 is optional because it is used to attach to more expansion enclosures.

Each IBM FlashSystem 7300 system can support up to four dense drawers per SAS chain.

Figure 1-41 shows a dense expansion drawer.



Figure 1-41 IBM Dense Expansion Drawer

SAS chain limitations

When you attach expansion enclosures to the control enclosure, you are not limited by the type of the enclosure. The only limitation for each of the two SAS chains is its chain weight. Each type of enclosure has its own chain weight:

- Enclosures 12G and 24G have a chain weight of 1
- Enclosure 92G has a chain weight of 2.5

The maximum chain weight is 10.

For example, you can combine seven 24G and one 92G expansions (7x1 + 1x2.5 = 9.5 chain weight), or two 92G enclosures, one 12G, and four 24G (2x2.5 + 1x1 + 4x1 = 10 chain weight).

An example of chain weight 4.5 with one 24G enclosure, one 12G enclosure, and one 92G enclosure, all correctly cabled, is shown in Figure 1-42.

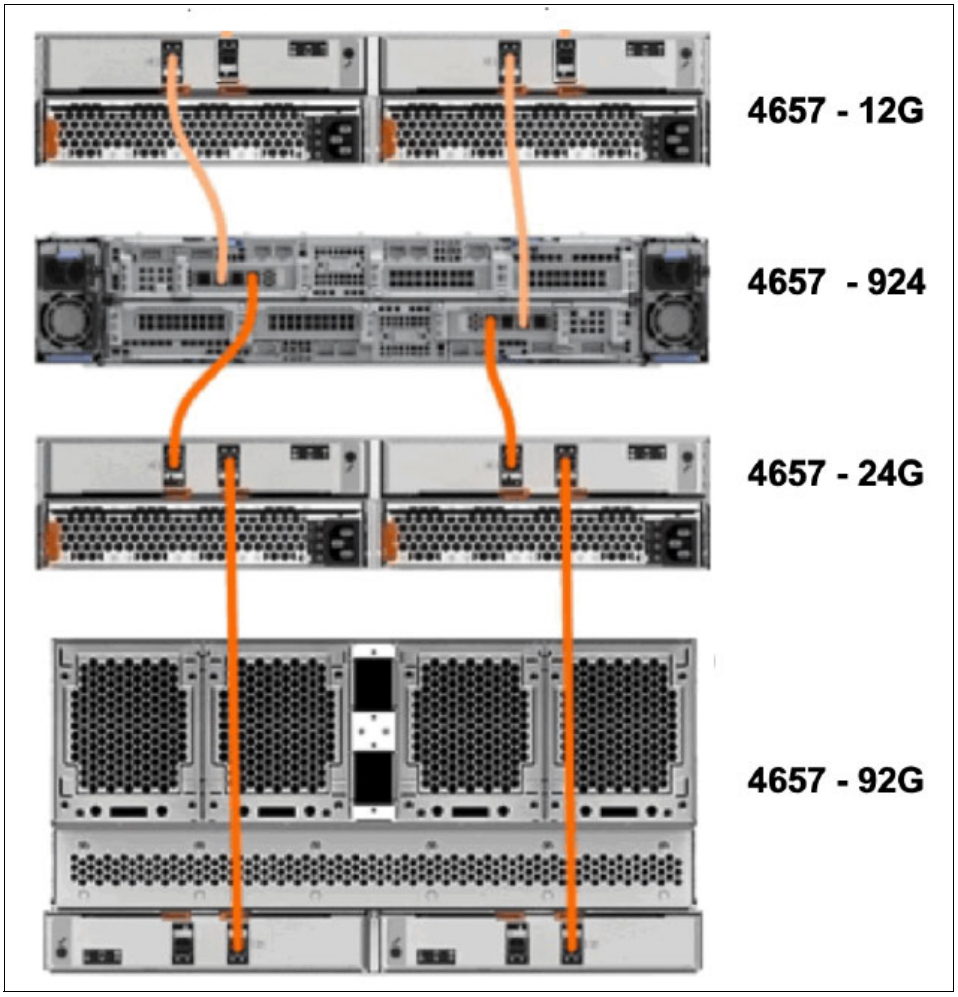


Figure 1-42 Connecting FS7300 SAS cables while complying with the maximum chain weight

1.11 IBM FlashSystem 5200 overview

With IBM FlashSystem 5200, you can be ready for a technology transformation without sacrificing performance, quality, or security and simplify your data management. This

powerful and compact solution is focused on affordability with a wide range of enterprise-grade features of IBM Storage Virtualize that can easily evolve and extend as businesses grows.

This system also has the flexibility and performance of flash and Non-Volatile Memory Express (NVMe) end to end, the innovation of IBM FlashCore technology, and Storage Class Memory (SCM) to help accelerate your business execution.

The innovative IBM FlashSystem family is based on a common storage software platform, IBM Storage Virtualize that provides powerful all-flash and hybrid-flash solutions that offer feature-rich, cost-effective, and enterprise-grade storage solutions.

Its industry-leading capabilities include a wide range of data services that can be extended to more than 500 heterogeneous storage systems, including the following examples:

- Automated data movement
- Synchronous and asynchronous copy services on-premises or to the public cloud
- HA configurations
- Storage automated tiering
- Data reduction technologies, including deduplication

IBM Storage Virtualize for Public Cloud works with IBM FlashSystem 5200 to deliver consistent data management between on-premises storage and public cloud. It is available on IBM Cloud, Amazon Web Services (AWS), and Microsoft Azure. You can move data and applications between on-premises and public cloud, implement new DevOps strategies, use public cloud for DR without the cost of a second data center, or improve cyber resiliency with air gap cloud snapshots.

IBM FlashSystem 5200 offers world-class customer support, product upgrades, and other programs. Consider the following examples:

- IBM Storage Expert Care service and support IBM Storage Expert Care service and support are simple. You can easily select the level of support and period that best fits your needs with predictable and up-front pricing that is a fixed percentage of the system cost.

  Note: For more information, see 1.6.1, “Storage Expert Care” on page 49.

- The IBM Data Reduction Guarantee helps reduce planning risks and lower storage costs with baseline levels of data compression effectiveness in IBM Storage Virtualize-based offerings.
- The IBM Controller Upgrade Program enables customers of designated all-flash IBM storage systems to reduce costs while maintaining leading-edge controller technology for essentially the cost of ongoing system maintenance.

The IBM FlashSystem 5200 control enclosure supports up to 12 2.5-inch NVMe-capable flash drives in a 1U high form factor.

Note: The IBM FlashSystem 5200 control enclosure supports IBM FCM3 drives if running IBM Storage Virtualize software V8.5 and later. These drives feature the same capacities as FCM2 drives, but have a higher internal compression ratio of up to 3:1. Therefore, it can effectively store more data, if the data is compressible.

One standard model of the IBM FlashSystem 5200 (4662-6H2) and one utility model (4662-UH6) are available.

Figure 1-43 shows the IBM FlashSystem 5200 control enclosure front view with 12 NVMe drives and a 3/4 ISO view.

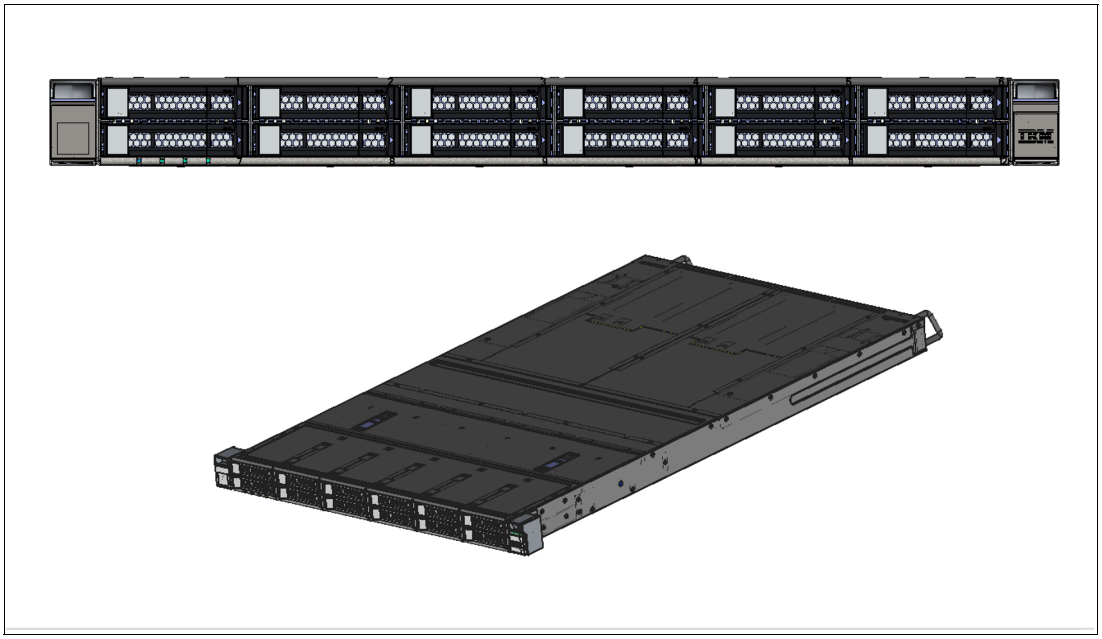


Figure 1-43 IBM FlashSystem 5200 control enclosure front and 3/4 ISO view

Table 1-10 lists the host connections, drive capacities, features, and standard options with IBM Storage Virtualize that are available on IBM FlashSystem 5200.

Table 1-10 IBM FlashSystem 5200 host, drive capacity, and functions summary

Feature or function	Description
Host interface	► 10 Gbps Ethernet (iSCSI) ► 25 Gbps Ethernet (iSCSI, iSER - iWARP, and RoCE) ► 16 Gbps Fibre Channel (FC and FC-NVMe) ► 32 Gbps Fibre Channel (FC and FC-NVMe)
Control enclosure supported drives (12 maximum)	► 2.5-inch NVMe self-compressing FCMs: 4.8 TB, 9.6 TB, 19.2 TB, and 38.4 TB ► NVMe flash drives: 800 GB, 1.92 TB, 3.84 TB, 7.68 TB, and 15.36 TB ► NVMe storage-class memory drives: 375 GB, 750 GB, 800 GB, and 1.6 TB
SAS Expansion Enclosures 760 per control enclosure 1,520 per clustered system Model 12G 2U 12 drives Model 24G 2U 24 drives Model 92G 5U 92 drives	► 2.5-inch flash drives supported: 800 GB, 1.6 TB, 1.92 TB, 3.84 TB, 7.68 TB, 15.36 TB, and 30.72 TB ► 2.5-inch disk drives supported: – 600 GB, 900 GB, 1.2 TB, 1.8 TB, and 2.4 TB 10k SAS disk – 2 TB 7.2 K nearline SAS disk ► 3.5-inch disk drives supported: 4 TB, 6 TB, 8 TB, 10 TB, 12 TB, 14 TB, 16 TB, and 18 TB 7.2 K nearline SAS disk
RAID levels.	Distributed RAID 5 and 6, TRAID 1 and 10

Feature or function	Description
Advanced features that are included with each system	► Virtualization of internal storage ► Data migration ► DRPs with thin provisioning ► UNMAP ► Compression and deduplication ► Metro Mirror (synchronous) and Global Mirror (asynchronous
More available advanced features	► Remote mirroring ► Policy-based replication ► IBM Easy Tier compression ► External virtualization ► Encryption ► FlashCopy ► IBM Storage Control ► IBM Storage Protect Snapshot

For more information, see V8.6.0 Configuration Limits and Restrictions for IBM FlashSystem 50x5 and 5200.

1.11.1 IBM FlashSystem 5200 expansion enclosures

The IBM FlashSystem 5200 Model 6H2 and UH6 attach to IBM FlashSystem 5200 Expansion Enclosure Models 4662-12G (2U 12-drive), 4662-24G (2U 24-drive), and 4662-92G (5U 92-drive), which support SAS flash drives and SAS HDD drives.

Note: IBM FlashSystem 5200 Models 6H2 and UH6 do *not* support attachment and intermixing of IBM Storwize V5100/V5000 Expansion Enclosure Models 12F, 24F, and 92F with IBM FlashSystem 5200 Expansion Enclosure Models 12G, 24G, and 92G.

The following 2.5-inch SFF flash drives are supported in the expansion enclosures:

- 400 and 800 GB
- 1.6, 1.92, 3.2, 3.84, 7.68, 15.36, and 30.72 TB

The following 3.5-inch LFF flash drives are supported in the expansion enclosures:

- 1.6, 1.92, 3.2, 3.84, 7.68, 15.36, and 30.72 TB
- 3.5-inch SAS disk drives (Model 12G):
  - 900 GB, 1.2 TB, 1.8 TB, and 2.4 TB 10,000 rpm
  - 4 TB, 6 TB, 8 TB, 10 TB, 12 TB, 14 TB, and 16 TB 7,200 rpm
- 3.5-inch SAS drives (Model 92G):
  - 1.6 TB, 1.92 TB, 3.2 TB, 3.84 TB, 7.68 TB, 15.36 TB, and 30.72 TB flash drives
  - 1.2 TB, 1.8 TB, and 2.4 TB 10,000 rpm
  - 6 TB, 8 TB, 10 TB, 12 TB, 14 TB, and 16 TB 7,200 rpm
- 2.5-inch SAS disk drives (Model 24G):
  - 900 GB, 1.2 TB, 1.8 TB, and 2.4 TB 10,000 rpm
  - 2 TB 7,200 rpm
- 2.5-inch SAS flash drives (Model 24G):
  - 400 and 800 GB
  - 1.6, 1.92, 3.2, 3.84, 7.68, 15.36, and 30.72 TB

For more information, see the video

IBM FlashSystem 5200 overview video: Check out this video "*IBM FlashSystem 5200 Overview*" at https://ibm.biz/Bdy6s2.

1.12 IBM FlashSystem 5000 family overview

IBM FlashSystem 5015, IBM FlashSystem 5035 and IBM FlashSystem 5045 are all-flash and hybrid flash solutions that provide enterprise-grade functions without compromising affordability or performance. They also include the rich features of IBM Storage Virtualize. IBM FlashSystem 5000 helps make modern technologies, such as artificial intelligence (AI), accessible to enterprises of all sizes.

IBM FlashSystem 5000 is a member of the IBM FlashSystem family of storage solutions. It delivers increased performance and new levels of storage efficiency with superior ease of use. This entry storage solution enables organizations to overcome their storage challenges.

The solution includes technology to complement and enhance virtual environments, which delivers a simpler, more scalable, and cost-efficient IT infrastructure. IBM FlashSystem 5000 features two node canisters in a compact, 2U 19-inch rack mount enclosure.

Figure 1-44 shows the IBM FlashSystem 5015, 5035 and 5045 SFF control enclosure front view.



Figure 1-44 IBM FlashSystem 5015, 5035 and 5045 SFF control enclosure front view

Figure 1-45 shows the IBM FlashSystem 5015, 5035 and 5045 LFF control enclosure front view.



Figure 1-45 IBM FlashSystem 5015, 5035 and 5045 LFF control enclosure front view

Table 1-11 lists the model comparison chart for the IBM FlashSystem 5000 family.

Table 1-11 Machine type and model comparison for the IBM FlashSystem 5000

MTM	Full name
2072-2N2	IBM FlashSystem 5015 LFF control enclosure
2072-2N4	IBM FlashSystem 5015 SFF control enclosure
2072-3N2	IBM FlashSystem 5035 LFF control enclosure
2072-3N4	IBM FlashSystem 5035 SFF control enclosure
2072-12G	IBM FlashSystem 5000 LFF expansion enclosure
2072-24G	IBM FlashSystem 5000 SFF expansion enclosure
2072-92G	IBM FlashSystem 5000 High-Density LFF expansion enclosure
4680-2P2	IBM FlashSystem 5015 LFF control enclosure (with Storage Expert Care)
4680-2P4	IBM FlashSystem 5015 SFF control enclosure (with Storage Expert Care)
4680-3P2	IBM FlashSystem 5045 LFF control enclosure (with Storage Expert Care)
4680-3P4	IBM FlashSystem 5045 SFF control enclosure (with Storage Expert Care)
4680-12H	IBM FlashSystem 5000 LFF expansion enclosure (with Storage Expert Care)
4680-24H	IBM FlashSystem 5000 SFF expansion enclosure (with Storage Expert Care)
4680-92H	IBM FlashSystem 5000 High-Density LFF expansion enclosure (with Storage Expert Care)

Note: IBM FlashSystems 5015 and 5035 (M/T2072) must use the 2072 XXG models of expansion enclosures. Similarly, the IBM FlashSystems 5105 and 504 (M/T 4680) must use the 4680 XXH models on expansion enclosures. The two types cannot be intermixed.

1.12.1 IBM FlashSystem 5015

IBM FlashSystem 5015 is an entry-level solution that is focused on affordability and ease of deployment and operation, with powerful scale-up features. It includes many IBM Storage Virtualize features and offers multiple flash and disk drive storage media and expansion options.

Table 1-12 lists the host connections, drive capacities, features, and standard options with IBM Storage Virtualize that are available on IBM FlashSystem 5015.

Table 1-12 IBM FlashSystem 5015 host, drive capacity, and functions summary

Feature / Function	Description
Host interface	► 1 Gb iSCSI (on the system board) ► 16 Gbps Fibre Channel ► 12 Gbps SAS ► 25 Gbps iSCSI (iWARP or RoCE) ► 10 Gbps iSCSI
Control enclosure and SAS expansion enclosures supported drives	► For SFF enclosures, see Table 1-13 on page 77 ► For LFF enclosures, see Table 1-14 on page 78
Cache per control enclosure/ clustered system	32 GB or 64 GB
RAID levels	DRAID 1, 5, and 6
Maximum expansion enclosure capacity	► Up to 10 standard expansion enclosures per controller ► Up to four high-density expansion enclosures per controller
Advanced functions that are included with each system	► Virtualization of internal storage ► DRPs with thin provisioning and UNMAP ► One-way data migration
More available advanced features	► Easy Tier ► FlashCopy ► Remote mirroring

Table 1-13 lists the 2.5-inch supported drives for IBM FlashSystem 5000 family.

Table 1-13 2.5-inch supported drives for the IBM FlashSystem 5000 family

2.5-inch (SFF)	Capacity					
Tier 1 flash	800 GB	1.9 TB	3.84 TB	7.68 TB	15.36 TB	30.72 TB
High-performance enterprise disk drives (10K rpm)	900 GB	1.2 TB	1.8 TB	2.4 TB		
High capacity nearline disk drives (7.2 K rpm)	2 TB					

Table 1-14 lists the 3.5-inch supported drives for IBM FlashSystem 5000 family.

Table 1-14 3.5-inch supported drives for the IBM FlashSystem 5000 family

3.5-inch (LFF)	Speed	Capacity							
High-performance, enterprise class disk drives	10,000 RPM	900 GB	1.2 TB	1.8 TB	2.4 TB				
High capacity archival class nearline disk drives	7,200 RPM	4 TB	6 TB	8 TB	10 TB	12 TB	14 TB	16 TB	18 TB

1.12.2 IBM FlashSystem 5035

IBM FlashSystem 5035 provides powerful functions, including powerful encryption capabilities and DRPs with compression, deduplication, thin provisioning, and the ability to cluster for scale-up and scale-out.

There following models of IBM FlashSystem 5035 are available:

- 2072-3N2 (2U 12 Drive Large Form Factor LFF)
- 2072-3N4 (2U 24 Drive Small Form Factor SFF)

The IBM FlashSystem 5000 expansion enclosures are available in the following form factors:

- 2U 12 Drive Large Form Factor LFF Model 12H,
- 2U 24 Drive Small Form Factor SFF Model 24H,
- 2U 92 Drive HD Form Factor LFF Model 92H

Available with the IBM FlashSystem 5035 model, DRPs help transform the economics of data storage. When applied to new or existing storage, they can increase usable capacity while maintaining consistent application performance. DPRs can help eliminate or drastically reduce costs for storage acquisition, rack space, power, and cooling, and can extend the useful life of storage assets.

DRPs feature the following capabilities:

- Block deduplication that works across all the storage in a DRP to minimize the number of identical blocks.
- New compression technology that ensures consistent 2:1 or better reduction performance across a wide range of application workload patterns.
- SCSI UNMAP support that de-allocates physical storage when operating systems delete logical storage constructs, such as files in a file system.

Table 1-15 lists the host connections, drive capacities, features, and standard options with IBM Storage Virtualize that are available on IBM FlashSystem 5035.

Table 1-15 IBM FlashSystem 5035 host, drive capacity, and functions summary

Feature / Function	Description
Host interface	► 10 Gb iSCSI (on the system board) ► 16 Gbps Fibre Channel ► 12 Gbps SAS ► 25 Gbps iSCSI (iWARP or RoCE) ► 10 Gbps iSCSI

Feature / Function	Description
Control enclosure and SAS expansion enclosures supported drives	► For SFF enclosures, see Table 1-13 on page 77 ► For LFF enclosures, see Table 1-14 on page 78
Cache per control enclosure/ clustered system	32 GB or 64 GB / 64 GB or 128 GB
RAID levels	DRAID 1, 5 (CLI only), and 6
Maximum expansion enclosure capacity	► Up to 20 standard expansion enclosures per controller ► Up to eight high-density expansion enclosures per controller
Advanced functions that are included with each system	► Virtualization of internal storage ► DRPs with thin provisioning ► UNMAP, compression, and deduplication ► One-way data migration ► Dual-system clustering
More available advanced features	► Easy Tier ► FlashCopy ► Remote mirroring ► Encryption

For more information, see V8.6.0.x configuration Limits and Restrictions for IBM FlashSystem 50x5 and 5200.

IBM FlashSystem 5035 hardware information

The IBM FlashSystem 5035 control enclosure features the following components:

- Two node canisters, each with a six-core processor
- 32 GB cache (16 GB per canister) with optional 64 GB cache (32 GB per canister)
- 10 Gb iSCSI (copper) connectivity standard with optional 16 Gb FC, 12 Gb SAS, 10 Gb iSCSI (optical), or 25 Gb iSCSI (optical)
- 12 Gb SAS port for expansion enclosure attachment
- 12 slots for 3.5-inch LFF SAS drives (Model 3N2) and 24 slots for 2.5-inch SFF SAS drives (Model 3N4)
- 2U, 19-inch rack mount enclosure with 100 - 240 V AC or -48 V DC power supplies

The IBM FlashSystem 5035 control enclosure models offer a high level of performance, scalability, and functions and include the following features:

- Support for 760 drives per system with the attachment of eight IBM FlashSystem 5000 High-Density LFF expansion enclosures and 1,520 drives with a two-way clustered configuration
- DRPs with deduplication, compression,[2] and thin provisioning for improved storage efficiency
- Encryption of data-at-rest that is stored within the IBM FlashSystem 5035 system

[2] Deduplication and compression require 64 GB of system cache.

Figure 1-46 lists the IBM FlashSystem 5035 SFF control enclosure with 24 drives.



Figure 1-46 Front view of an IBM FlashSystem 5035

Figure 1-47 shows the rear view of an IBM FlashSystem 5035 control enclosure.



Figure 1-47 Rear view of an IBM FlashSystem 5035

Figure 1-48 shows the available connectors and LEDs on a single IBM FlashSystem 5035 canister.

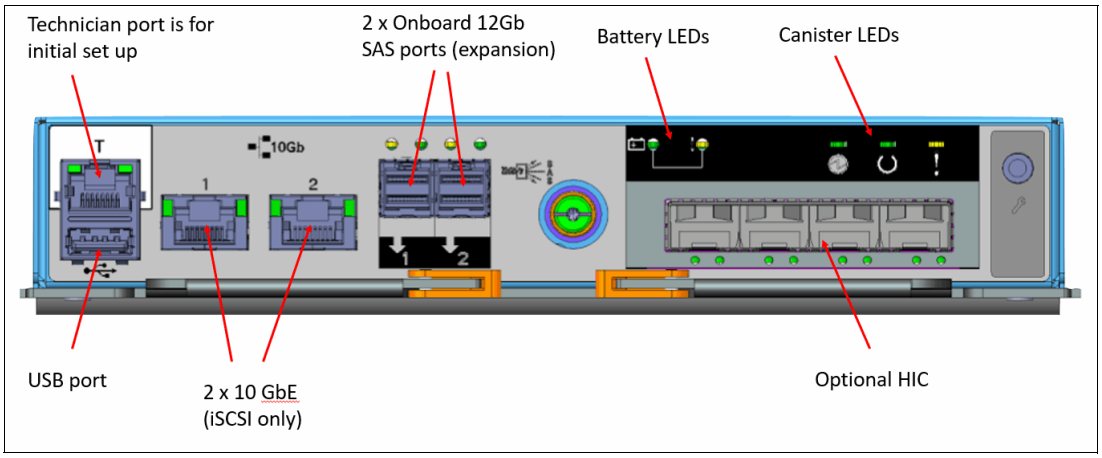


Figure 1-48 View of available connectors and LEDs on an IBM FlashSystem 5035 single canister

For more information, see V8.4.2.x Configuration Limits and Restrictions for IBM FlashSystem 5x00.

1.12.3 IBM FlashSystem 5045

IBM FlashSystem 5045 is a lightweight FlashSystem 5035 replacement, which uses the same base hardware as the FlashSystem 5035 but with software modifications to provide support for software features that are not available on FlashSystem 5035.

Additional supported software includes the following features:

- Safeguarded Copy
- FlashCopy 2 with internal scheduler

IBM FlashSystem 5045 also has a new machine type and 2 new models to support the IBM Storage Expert care service offerings:

- 4680-3P2 (2U 12 Drive Large Form Factor LFF)
- 4680-3P4 (2U 24 Drive Small Form Factor SFF)

The IBM FlashSystem 5000 expansion enclosures are available in the following form factors

- 2U 12 Drive Large Form Factor LFF Model 12H,
- 2U 24 Drive Small Form Factor SFF Model 24H,
- 2U 92 Drive HD Form Factor LFF Model 92H

IBM FlashSystem 5045 provides powerful functions, which include encryption capabilities and DRPs with compression, deduplication, thin provisioning, and the ability to cluster for scale-up and scale-out.

DRPs help transform the economics of data storage. When applied to new or existing storage, they can increase usable capacity and maintain consistent application performance. DRPs can help eliminate or drastically reduce costs for storage acquisition, rack space, power, and cooling, and can extend the useful life of storage assets.

DRPs feature the following capabilities:

- Block deduplication that works across all the storage in a DRP to minimize the number of identical blocks
- New compression technology that ensures consistent 2:1 or better reduction performance across a wide range of application workload patterns
- SCSI UNMAP support that de-allocates physical storage when operating systems delete logical storage constructs, such as files in a file system

Table 1-16 lists the host connections, drive capacities, features, and standard options with IBM Storage Virtualize that are available on IBM FlashSystem 5045.

Table 1-16 IBM FlashSystem 5045 host, drive capacity, and functions summary

Feature / Function	Description
Host interface	► 10 Gb iSCSI (on the system board) ► 16 Gbps Fibre Channel ► 12 Gbps SAS ► 25 Gbps iSCSI (iWARP or RoCE) ► 10 Gbps iSCSI
Control enclosure and SAS expansion enclosures supported drives	► For SFF enclosures, see Table 1-13 on page 77 ► For LFF enclosures, see Table 1-14 on page 78 Enterprise 15K rpm drives are not supported on IBM FlashSystem 5045
Cache per control enclosure/ clustered system	32 GB or 64 GB / 64 GB or 128 GB
RAID levels	DRAID 1, 5 (CLI only), and 6
Maximum expansion enclosure capacity	Maximum of 12 standard expansion enclosures per controller Maximum of 2 high-density expansion enclosures per controller

Feature / Function	Description
Advanced functions that are included with each system	► Virtualization of internal storage ► DRPs with thin provisioning ► UNMAP, compression, and deduplication ► One-way data migration ► Dual-system clustering ► Volume Mobility ► Easy Tier ► FlashCopy ► Remote mirroring
More available advanced features	Encryption

IBM FlashSystem 5045 does not cluster with any models other than another IBM FlashSystem 5045. As with the IBM FlashSystem 5035, a maximum of two I/O groups per cluster is allowed. IBM FlashSystem 5045 also has a reduction in Volume Group extents from 4 million to 1 million.

IBM FlashSystem 5045 uses a License Machine Code (LMC) model similar to that used by the FlashSystem 5200. This is a change from FlashSystem 5035 as all feature software (apart from encryption) is included with the base license. Encryption requires a hardware license key.

Expansion chain weight is reduced from 20U in FlashSystem 5035 to 12U for Flash System 5045.

This means that a single chain can contain up to 2x5U92 expansions or up to 6x2U12/24 expansions or a combination of 5U and 2U not exceeding 12U in total. This reduces the maximum number of expansion drive slots per chain, per I/O Group and per system.

- per chain: 2x92 + 1x24 + 24 = 232 drives
- per I/O Group: 2 chains per I/O Group so 4x92 + 2x24 + 24 = 440 drives
- per System: 2 I/O Groups so 2x440 = 880 drives

For more information, see V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 50x5 and 5200.

IBM FlashSystem 5045 hardware information

The IBM FlashSystem 5045 control enclosure features the following components:

- Two node canisters, each with a six-core processor
- 32 GB cache (16 GB per canister) with optional 64 GB cache (32 GB per canister)
- 10 Gb iSCSI (copper) connectivity standard with optional 16 Gb FC, 12 Gb SAS, 10 Gb iSCSI (optical), or 25 Gb iSCSI (optical)
- 12 Gb SAS port for expansion enclosure attachment
- 12 slots for 3.5-inch LFF SAS drives (Model 3N2) and 24 slots for 2.5-inch SFF SAS drives (Model 3N4)
- 2U, 19-inch rack mount enclosure with 100 - 240 V AC power supplies

The IBM FlashSystem 5045 control enclosure models offer a high level of performance, scalability, and functions and include the following features:

- Support for 760 drives per system with the attachment of eight IBM FlashSystem 5000 High-Density LFF expansion enclosures and 1,520 drives with a two-way clustered configuration
- DRPs with deduplication, compression,[3] and thin provisioning for improved storage efficiency
- Encryption of data-at-rest that is stored within the IBM FlashSystem 5045 system

Figure 1-46 on page 80 lists the IBM FlashSystem 5045 SFF control enclosure with 24 drives.



Figure 1-49 Front view of an IBM FlashSystem 5045

Figure 1-47 on page 80 lists the rear view of an IBM FlashSystem 5045 control enclosure.



Figure 1-50 Rear view of an IBM FlashSystem 5045

Figure 1-48 on page 80 lists the available connectors and LEDs on a single IBM FlashSystem 5045 canister.

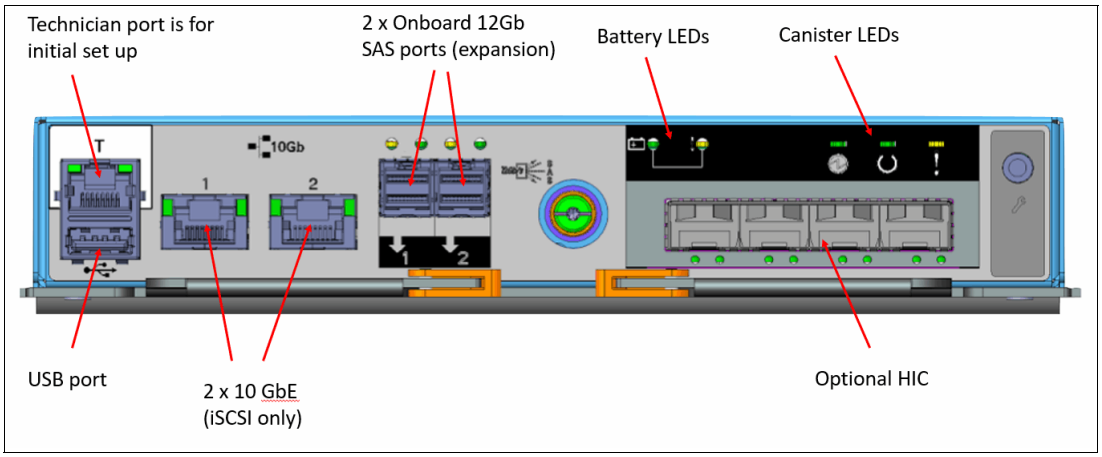


Figure 1-51 View of available connectors and LEDs on an IBM FlashSystem 5045 single canister

[3] Deduplication and compression require 64 GB of system cache.

For more information, see V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 50x5 and 5200.

1.12.4 IBM Storage Virtualize for Public Cloud

IBM Storage Virtualize for Public Cloud extends the IBM Storage Virtualize solution to a hybrid-cloud or cloud-based model, where servers, storage, and network infrastructure are delivered in a public cloud environment. It can be deployed on either IBM Cloud or Amazon Web Services (AWS) cloud infrastructures.

Demonstration videos: For more information on IBM Storage Virtualize for Public Cloud, see the following demonstration videos:

- Connecting IBM Storage to VMware vSphere
- https://ibm.biz/Bdy6sp
- "*IBM Spectrum Virtualize for Public Cloud V8.5 installation on AWS*" at https://ibm.biz/Bdy6sh.
- "*IBM Spectrum Virtualize for Public Cloud V8.5 installation on Azure*" at https://ibm.biz/Bdy6sJ.

1.13 Storage efficiency and data reduction features

IBM Storage Virtualize software that is running in the IBM FlashSystem storage systems or IBM SAN Volume Controllers offers several functions for storage optimization and efficiency.

1.13.1 IBM Easy Tier

Many applications exhibit a significant skew in the distribution of I/O workload. A small fraction of the storage is responsible for a disproportionately large fraction of the total I/O workload of an environment.

In a tiered storage pool, IBM Easy Tier acts to identify this skew and automatically place data in the suitable tier to take advantage of it. By moving the hottest data onto the fastest tier of storage, the workload on the remainder of the storage is reduced. By servicing most of the application workload from the fastest storage, Easy Tier acts to accelerate application performance.

Easy Tier is a performance optimization function that automatically migrates extents that belong to a volume among different storage tiers based on their I/O load. The movement of the extents is online and unnoticed from a host perspective.

As a result of extent movement, the volume no longer has all its data in one tier, but rather in two or three tiers. Each tier provides optimal performance for the extent, as shown in Figure 1-52.

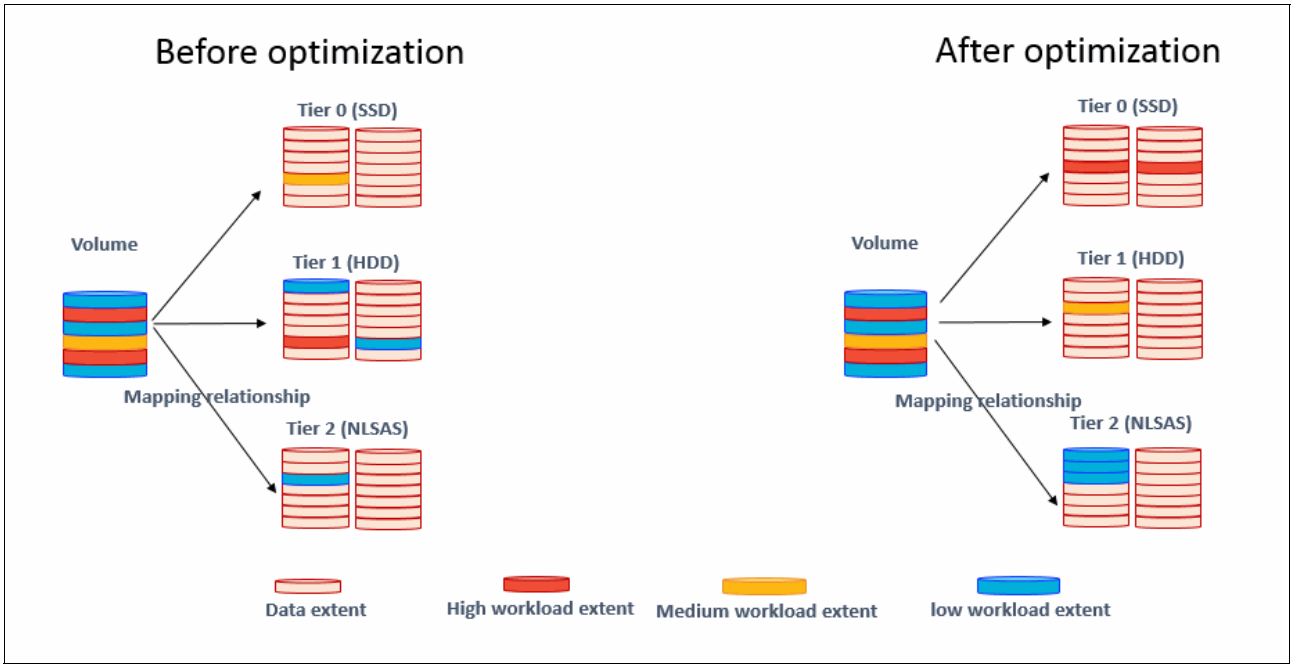


Figure 1-52 Easy Tier concept

Easy Tier monitors the I/O activity and latency of the extents on all Easy Tier enabled storage pools to create *heat maps*. Based on these maps, Easy Tier creates an extent migration plan and promotes (or moves) high activity or hot extents to a higher disk tier within the same storage pool. It also demotes extents that have a reduction in activity, or have cooled, by moving them from a higher tier MDisk back to a lower tier MDisk.

Storage pools that contain only one tier of storage can also benefit from Easy Tier if they include multiple disk arrays (or MDisks). Easy Tier has a balancing mode: It moves extents from busy disk arrays to less busy arrays of the same tier, which balances I/O load.

All MDisks (disk arrays) belong to one of the tiers. They are classified as SCM, Flash, Enterprise, or NL tier.

For more information about the EasyTier reports, see this IBM Documentation web page.

1.13.2 Data reduction and UNMAP

The UNMAP feature is a set of SCSI primitives that enables hosts to indicate to a SCSI target (storage system) that space that is allocated to a range of blocks on a target storage volume is no longer required. This command enables the storage controller to take measures and optimize the system so that the space can be reused for other purposes.

For example, the most common use case is a host application, such as VMware, which frees storage in a file system. Then, the storage controller can perform functions to optimize the space, such as reorganizing the data on the volume so that space is better used.

When a host allocates storage, the data is placed in a volume. To return the allocated space to the storage pools, the SCSI UNMAP feature is used. UNMAP enables host operating

systems to deprovision storage on the storage controller so that the resources can automatically be freed in the storage pools and used for other purposes.

A DRP increases infrastructure capacity use by using new efficiency functions and reducing storage costs. By using the end-to-end SCSI UNMAP function, a DRP can automatically de-allocate and reclaim the capacity of thin-provisioned volumes that contain deleted data so that this reclaimed capacity can be reused by other volumes.

At its core, a DRP uses a Log Structured Array (LSA) to allocate capacity. An LSA enables a tree-like directory to be used to define the physical placement of data blocks that are independent of size and logical location. Each logical block device has a range of logical block addresses (LBAs), starting from 0 and ending with the block address that fills the capacity.

When written, you can use an LSA to allocate data sequentially and provide a directory that provides a lookup to match an LBA with a physical address within the array. Therefore, the volume that you create from the pool to present to a host application consists of a directory that stores the allocation of blocks within the capacity of the pool.

In DRPs, the maintenance of the metadata results in *I/O amplification*. I/O amplification occurs when a single host-generated read or write I/O results in more than one back-end storage I/O request because of advanced functions. A read request from the host results in two I/O requests: a directory lookup and a data read. A write request from the host results in three I/O requests: a directory lookup, a directory update, and a data write. This aspect must be considered when sizing and planning your data-reducing solution.

Standard pools, which make up a classic solution that also is supported by the IBM FlashSystem and IBM SAN Volume Controller, do not use LSA. A standard pool works as a container that receives its capacity from MDisks (disk arrays), splits it into extents of the same fixed size, and allocates extents to volumes.

Standard pools do not cause I/O amplification and require less processing resource usage compared to DRPs. In exchange, DRPs provide more flexibility and storage efficiency.

Table 1-17 lists the volume capacity saving types that are available with standard pools and DRPs.

Table 1-17 Volume types that are available in pools

Saving type	Standard pool	DRP
Fully allocated	Yes	Yes
Thin-provisioned	Yes	Yes
Thin-provisioned compressed	No	Yes
Thin-provisioned deduplicated	No	Yes
Thin-provisioned compressed and deduplicated	No	Yes

Best practice: If you want to use deduplication, create thin-provisioned compressed and deduplicated volumes.

This book provides only an overview of DRP aspects. For more information, see *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430.

Fully allocated and thin-provisioned volumes

Volumes can be configured as thin-provisioned or fully allocated. Both versions can be configured in standard pools and DRP pools.

In IBM Storage Virtualized systems, each volume includes virtual capacity and real capacity parameters:

- *Virtual capacity* is the volume storage capacity that is available to a host. It is used by the host operating system to create a file system.
- *Real capacity* is the storage capacity that is allocated to a volume from a pool. It shows the amount of space that is used on a physical storage volume.

Fully allocated

Fully allocated volumes are created with the same amount of real capacity and virtual capacity. This type uses no storage efficiency features.

When a fully allocated volume is created on a DRP, it bypasses the LSA structure and works in the same manner as in a standard pool; Therefore, it has no effect on processing and provides no data reduction options at the pool level.

When fully allocated volumes are used on the IBM Storage Virtualized systems with FCM drives (whether a DRP or standard pool is used), capacity savings are achieved by compressing data with hardware compression that runs on the FCM drives. Hardware compression on FCM drives is always on and cannot be turned off. This configuration provides maximum performance with outstanding storage efficiency.

Thin-provisioned

A thin-provisioned volume presents a different capacity to mapped hosts than the capacity that the volume uses in the storage pool. Therefore, real and virtual capacities might not be equal. The virtual capacity of a thin-provisioned volume can be significantly larger than its real capacity. As more information is written by the host to the volume, more of the real capacity is used. The system identifies read operations to unwritten parts of the virtual capacity, and returns zeros to the server without the use of any real capacity.

In a shared storage environment, thin provisioning is a method for optimizing the use of available storage. Thin provisioning relies on the allocation of blocks of data on demand, versus the traditional method of allocating all of the blocks up front. This method eliminates almost all white space, which helps avoid the poor usage rates that occur in the traditional storage allocation method where large pools of storage capacity are allocated to individual servers but remain unused (not written to).

If a thin-provisioned volume is created in a DRP, the system monitors it for reclaimable capacity from host unmap operations. This capacity can be reclaimed and redistributed into the pool.

Space that is freed from the hosts is a process that is called UNMAP. A host can issue SCSI UNMAP commands when the user deletes files on a file system, which result in the freeing of all of the capacity that is allocated within that unmapping.

A thin-provisioned volume in a standard pool does not return unused capacity to the pool with SCSI UNMAP.

A thin-provisioned volume can be converted nondisruptively to a fully allocated volume, or vice versa, by using the volume mirroring function. For example, you can add a thin-provisioned copy to a fully allocated primary volume and then remove the fully allocated copy from the volume after they are synchronized.

Note: It is *not* recommended to use noncompressed thin-provisioned volumes on DRPs that contain FCM drives. The system GUI prevents the creation of such types of configurations.

1.13.3 Compression and deduplication

When DRPs are used on IBM Storage Virtualized systems, host data can be compressed or compressed and deduplicated before it is written to the disk drives.

The IBM FlashSystem and IBM SAN Volume Controller family DRP compression is based on the Lempel-Ziv lossless data compression algorithm that operates by using a real-time method. When a host sends a write request, the request is acknowledged by the write cache of the system and then staged to the DRP.

As part of its staging, the write request passes through the compression engine and is stored in a compressed format. Therefore, writes are acknowledged immediately after they are received by the write cache with compression occurring as part of the staging to internal or external physical storage. This process occurs transparently to host systems, which makes them unaware of the compression.

IBM Comprestimator tool

The IBM Comprestimator tool is available to check whether your data is compressible. It estimates the space savings that are achieved when compressed volumes are used. This utility provides a quick and easy view of showing the benefits of using compression. IBM Comprestimator can be run from the system GUI or command line interface (CLI), and it checks data that is already stored on the system. In DRPs, IBM Comprestimator is always on starting at code level 8.4, so you can display the compressibility of the data in the GUI and IBM Storage Insights at any time.

IBM Data Reduction Estimation tool

The IBM Data Reduction Estimation tool (DRET) is a host-based command-line utility for estimating the data reduction savings on block storage devices. To help with the profiling and analysis of user workloads that must be migrated to a new system, IBM provides the highly accurate DRET to support both deduplication and compression.

The tool scans target workloads on various earlier storage arrays (from IBM or another company), merges all scan results and then provides an integrated system-level data reduction estimate.

Both tools are available as stand-alone, host-based utilities that can analyze data on IBM or third-party storage devices. For more information, see Data Reduction Estimation Tool.

Deduplication can be configured with thin-provisioned and compressed volumes in DRPs for added capacity savings. The deduplication process identifies unique chunks of data, or byte patterns, and stores a signature of the chunk for reference when writing new data chunks.

If the new chunk's signature matches a signature, the new chunk is replaced with a small reference that points to the stored chunk. The matches are detected when the data is written. The same byte pattern might occur many times, which greatly reduce the amount of data that must be stored.

Compression and deduplication are not mutually exclusive: One, both, or none of the features can be enabled. If the volume is deduplicated and compressed, data is deduplicated first and

then compressed. Therefore, deduplication references are created on the compressed data that is stored on the physical domain.

1.13.4 Enhanced data security features

To protect data against the potential exposure of sensitive user data and user metadata that is stored on discarded, lost, or stolen storage devices, IBM FlashSystem storage systems and IBM SAN Volume Controller support encryption of data-at-rest.

Encryption is performed by the IBM FlashSystem or IBM SAN Volume Controller controllers for data that is stored:

- Within the entire system
- The IBM FlashSystem control enclosure
- All attached expansion enclosures
- As externally virtualized by the IBM FlashSystem or IBM SAN Volume Controller storage systems

Encryption is the process of encoding data so that only authorized parties can read it. Data encryption is protected by the Advanced Encryption Standard (AES) algorithm that uses a 256-bit symmetric encryption key in XTS mode, as defined in the IEEE 1619-2007 standard and NIST Special Publication 800-38E as XTS-AES-256.

Two types of encryption are available on devices that are running IBM Storage Virtualize: hardware and software. The method that is used for encryption is chosen automatically by the system based on the placement of the data:

- Hardware encryption: Data is encrypted by IBM FlashCore module (FCM) hardware and SAS hardware for expansion enclosures. It is used only for internal storage (drives).
- Software encryption: Data is encrypted by using the nodes' CPU (the encryption code uses the AES-NI CPU instruction set). It is used only for external storage that is virtualized by the IBM FlashSystem and IBM SAN Volume Controller managed storage systems.

Both methods of encryption use the same encryption algorithm, key management infrastructure, and license.

Note: Only data-at-rest is encrypted. Host-to-storage communication and data that is sent over links that are used for remote mirroring are *not* encrypted.

The IBM FlashSystem also supports self-encrypting drives, in which data encryption is completed in the drive.

Before encryption can be enabled, ensure that a license was purchased and activated.

The system supports two methods of configuring encryption:

- You can use a centralized external key server that simplifies creating and managing encryption keys on the system. This method of encryption key management is preferred for security and simplification of key management.

  A *key server* is a centralized system that generates, stores, and sends encryption keys to the system. Some key server providers support replicating keys among multiple key servers. If multiple key servers are supported, you can specify up to four key servers that connect to the system over a public network or separate private network.

- The system supports IBM Guardium Key Lifecycle Manager (formally IBM Security Key Lifecycle Manager) or Gemalto SafeNet KeySecure key servers to handle key management.
- The system also supports storing encryption keys on USB flash drives. USB flash drive-based encryption requires physical access to the systems and is effective in environments with a minimal number of systems. For organizations that require strict security policies regarding USB flash drives, the system supports disabling a canister's USB ports to prevent unauthorized transfer of system data to portable media devices. If you have such security requirements, use key servers to manage encryption keys.

Another data security enhancement that delivered with the IBM Storage Virtualize code is the new Safeguarded Copy feature that can provide protected read-only logical air gap copies of volumes. This enhancement gives the customer effective data protection against cyberattacks.

For more information, see "Safeguarded Copy" on page 18.

1.14 Application integration features

IBM FlashSystem storage systems include the following features, which enable tight integration with VMware:

- vCenter plug-in

  Enables monitoring and self-service provisioning of the system from within VMware vCenter.
- vStorage application programming interfaces (APIs) for Array Integration (VAAI) support:

  This function supports hardware-accelerated virtual machine (VM) copy/migration and hardware-accelerated VM initiation, and accelerates VMware Virtual Machine File System (VMFS).
- Microsoft Windows System Resource Manager (SRM) for VMware Site Recovery Manager

  Supports automated storage and host failover, failover testing, and failback.
- VMware vSphere virtual volume (vVol) integration for better usability

  The migration of space-efficient volumes between storage containers maintains the space efficiency of volumes. Cloning a VM achieves a full independent set of vVols, and resiliency is improved for VMs if volumes start running out of space.

VMware vSphere virtual volumes

The system supports vVols, which allow VMware vCenter to automate the management of system objects, such as volumes and pools. It is an integration and management framework that virtualizes the IBM FlashSystem storage systems that enables a more efficient operational model that is optimized for virtualized environments and centered on the application instead of the infrastructure.

The use of vVols simplifies operations through policy-driven automation that enables more agile storage consumption for VMs and dynamic adjustments in real time when they are needed. They also simplify the delivery of storage service levels to individual applications by providing finer control of hardware resources and native array-based data services that can be instantiated with VM granularity.

With vVols, VMware offers a paradigm in which an individual VM and its disks, rather than a logical unit number (LUN), becomes a unit of storage management for a storage system. It encapsulates VDisks and other VM files, and natively stores the files on the storage system.

By using a special set of APIs that are called vSphere APIs for Storage Awareness (VASA), the storage system becomes aware of the vVols and their associations with the relevant VMs. Through VASA, vSphere and the underlying storage system establish a two-way out-of-band communication to perform data services and offload certain VM operations to the storage system. For example, some operations, such as snapshots and clones, can be offloaded.

For more information about the actions that are required to implement this feature on the host side, see Understanding Virtual Volumes (vVols) in VMware vSphere 6.7/7.0/8.0 . Also, see *IBM Storage Virtualize and VMware: Integrations, Implementation and Best Practices*, SG24-8549.

IBM support for VASA is provided by IBM Storage Connect enabling communication between the VMware vSphere infrastructure and the IBM FlashSystem system. The IBM FlashSystem administrator can assign ownership of vVols to IBM Storage Connect by creating a user with the VASA Provider security role.

Although the system administrator can complete specific actions on volumes and pools that are owned by the VASA Provider security role, IBM Storage Connect retains management responsibility for vVols. For more information, see IBM Storage Connect documentation.

Note: At the time of this writing, vVols are *not* supported on DRPs. However, they are still valid with Version 8.6.0.

Demonstration videos: The following demonstration videos for VMware integration with IBM FlashSystem are also available:

- Connecting IBM Storage to VMware vSphere
- Managing datastores provisioned from IBM FlashSystem storage in the vSphere client
- Creating a datastore on IBM FlashSystem storage, directly from the vSphere client

1.14.1 Features for manageability

IBM FlashSystem storage systems and the IBM SAN Volume Controller offer the following manageability and serviceability features:

- Intuitive GUI
- IBM Call Home and Remote Support
- IBM Storage Insights
- IBM Storage Virtualize RESTful API

IBM Storage Virtualize family system GUI

The IBM storage systems that are running Storage Virtualize include a management GUI that runs on one of the node canisters in the control enclosure. This GUI helps you to monitor,

manage, and configure your system. You can access the GUI by opening any supported web browser and entering the management IP addresses.

The IBM storage systems use a GUI with the same design across all platforms for a consistent management experience. The GUI includes an improved overview dashboard that provides all of the information in an easy-to-understand format and enables visualization of effective capacity. By using the GUI, you can quickly deploy storage and manage it efficiently.

Figure 1-53 shows the Storage Virtualize GUI dashboard view. This default view is displayed after the user logs on to the system.

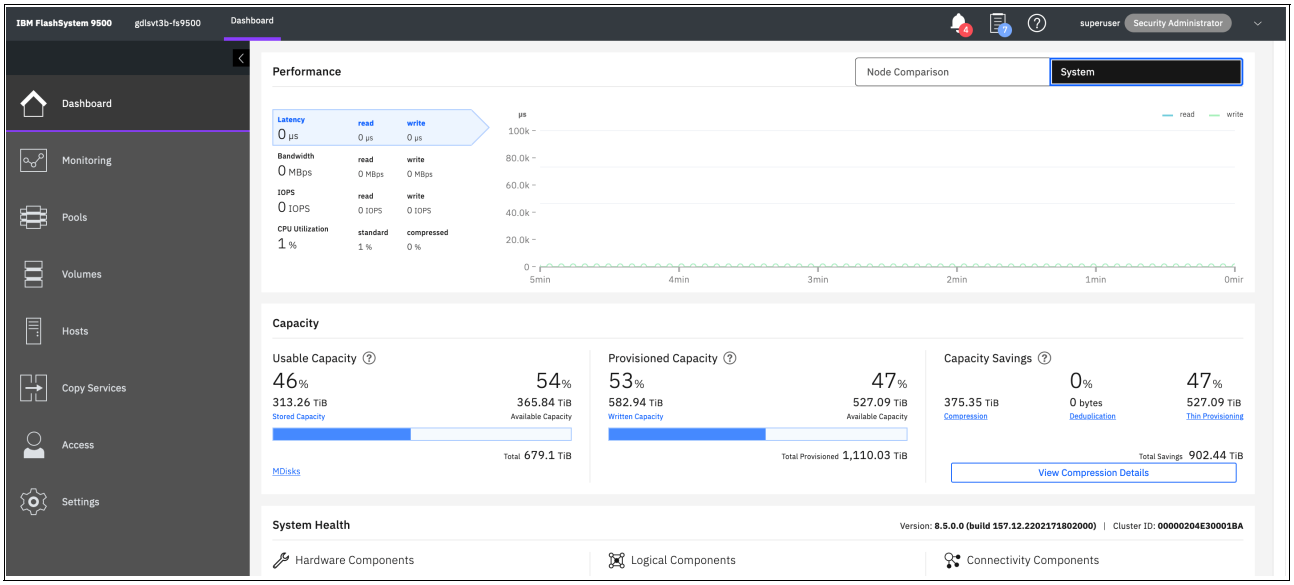


Figure 1-53 IBM Storage Virtualize GUI dashboard

The IBM FlashSystem storage systems and the IBM SAN Volume Controller also provide a CLI, which is useful for advanced configuration and scripting.

The systems support SNMP, email notifications that use Simple Mail Transfer Protocol (SMTP), and syslog redirection for complete enterprise management access.

IBM Call Home and Remote Support

IBM Call Home connects the system to IBM Service Personnel who can monitor and respond to system events to ensure that your system remains running.

The IBM Call Home function opens a service alert if a serious error occurs in the system, which automatically sends the details of the error and contact information to IBM Service Personnel.

If the system is eligible for support, a Cognitive Support Program (CSP) ticket is automatically created and assigned to the suitable IBM Support team. The information that is provided to IBM is an excerpt from the event log that contains the details of the error, and customer contact information from the system.

IBM Service Personnel contact the customer and arrange service on the system, which can greatly improve the speed of resolution by removing the need for the customer to detect the error and raise a support call themselves.

The system supports the following methods to transmit notifications to the support center:

- Call Home with cloud services

  Call Home with cloud services sends notifications directly to a centralized file repository that contains troubleshooting information that is gathered from customers. Support personnel can access this repository and automatically be assigned issues as problem reports.

  This method of transmitting notifications from the system to support removes the need for customers to create problem reports manually. Call Home with cloud services also eliminates email filters dropping notifications to and from support, which can delay resolution of problems on the system.

  This method sends notifications only to the predefined support center.

- Call Home with email notifications

  Call Home with email notification sends notifications through a local email server to support and local users or services that monitor activity on the system. With email notifications, you can send notifications to support and designate internal distribution of notifications, which alerts internal personnel about potential problems. Call Home with email notifications requires configuring at least one email server, and local users.

  However, external notifications to the support center can be dropped if filters on the email server are active. To eliminate this problem, Call Home with email notifications is not recommended as the only method to transmit notifications to the support center. Call Home with email notifications can be configured with cloud services.

IBM highly encourages all customers to take advantage of the Call Home feature so that you and IBM can collaborate for your success.

For more information, see IBM Storage Virtualize Products Call Home and Remote Support Overview.

IBM Call Home Connect Cloud

IBM Call Home Connect Cloud (CHCC) is IBM's web application that allows IBM hardware clients to view and monitor key status indicators about their Call Home-enabled IBM hardware assets. Offered at no additional cost to all IBM hardware clients, the website displays the following information:

- Critical cases and alerts
- Warranty and maintenance contract status
- Last contact status
- Current software and firmware levels and upgrade recommendations
- Asset details, such as:
  - Summary data about open and closed cases, with links to IBM Support to view them
  - Detailed alerts about various Call Home-related events
  - Links to product-specific support information

Clients can also use CHCC to perform the following tasks:

- Monitor Remote Code Load services on their assets in real time
- View the warranty and maintenance contract information for their assets
- Visualize and export Call Home data in several formats

Call Home Connect Cloud works offline, too, although real-time updates are not available while offline. For clients who want a more tailored mobile experience, IBM offers a mobile companion app, Call Home Connect Anywhere, for Android and iOS devices.

For more information about CHCC, see Introducing Call Home Connect Cloud.

IBM Storage Insights

IBM Storage Insights is an IBM Cloud software as a service (SaaS) offering that can help you monitor and optimize the storage resources in the system and across your data center. IBM Storage Insights monitors your storage environment and provides information about the statuses of multiple systems in a single dashboard.

You can view data from the perspectives of the servers, applications, and file systems. Two versions of IBM Storage Insights are available: IBM Storage Insights and IBM Storage Insights Pro.

When you order any IBM FlashSystem storage system or IBM SAN Volume Controller, IBM Storage Insights is available at no extra cost. With this version, you can monitor the basic health, status, and performance of various storage resources.

IBM Storage Insights Pro is a subscription-based product that provides a more comprehensive view of the performance, capacity, and health of your storage resources. In addition to the features that are offered by IBM Storage Insights, IBM Storage Insights Pro provides tools for intelligent capacity planning, storage reclamation, storage tiering, and performance troubleshooting services. Together, these features can help you reduce storage costs and optimize your data center.

Note: With some models of Storage Virtualize systems that offer the Premium Storage Expert Care level of support, Storage Insights Pro is included as part of the offering. For more information, see 1.6.1, “Storage Expert Care” on page 49.

IBM Storage Insights is a part of the monitoring and helps to ensure continued availability of the IBM FlashSystem storage or IBM SAN Volume Controller systems.

The tool provides a single dashboard that gives you a clear view of all your IBM block and file storage and some other storage vendors (the IBM Storage Insights Pro version is required to view other storage vendors’ storage). You can make better decisions by seeing trends in performance and capacity. With storage health information, you can focus on areas that need attention.

When IBM Support is needed, IBM Storage Insights simplifies uploading logs, speeds resolution with online configuration data, and provides an overview of open tickets, all in one place.

The following features are available with IBM Storage Insights:

- A unified view of IBM systems:
  - Provides a single view to see all your system’s characteristics
  - Displays all of your IBM storage inventory
  - Provides a live event feed so that you know in real time what is going on with your storage so that you can act quickly
- IBM Storage Insights collects telemetry data and Call Home data and provides real-time system reporting of capacity and performance.
- Overall storage monitoring by reviewing the following information:

  - The overall health of the system.
  - Monitoring of the configuration to see whether it meets best practices.
  - System resource management determines which system is overtaxed, and provides proactive recommendations to fix it.
- IBM Storage Insights provides advanced customer service with an event filter that you can use to accomplish the following tasks:
  - You and IBM Support can view, open, and close support tickets, and track trends.
  - You can use the autolog collection capability to collect the logs and send them to IBM before IBM Support investigates the problem. This capability can save time in resolving a case.

Figure 1-54 shows a view of the IBM Storage Insights dashboard.

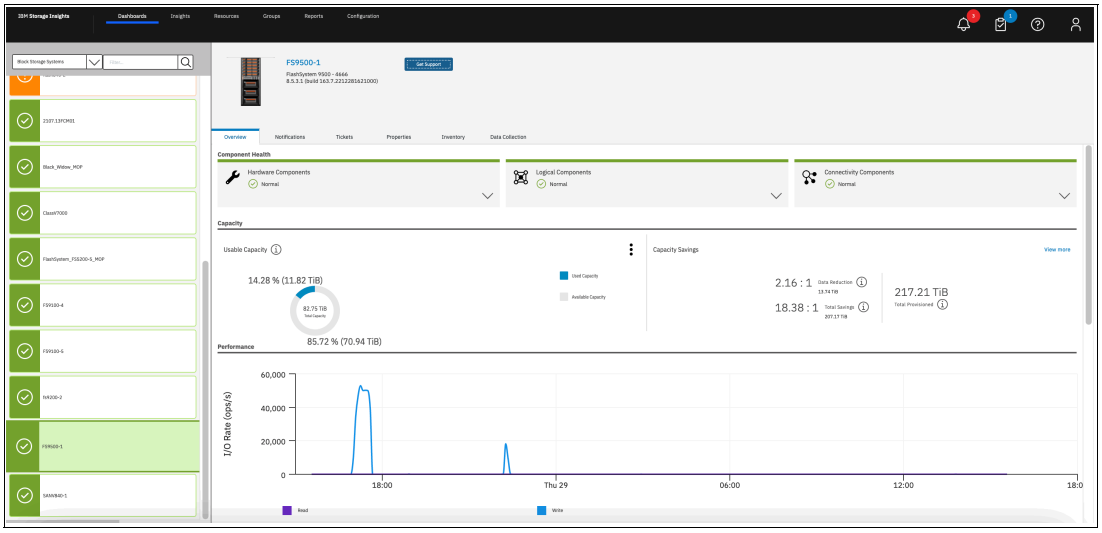


Figure 1-54 IBM Storage Insights dashboard

For IBM Storage Insights to operate, a lightweight data collector must be deployed in your data center to stream only system metadata to your IBM Cloud instance. The metadata flows in one direction: from your data center to IBM Cloud over HTTPS.

The application data that is stored on the storage systems cannot be accessed by the data collector. In IBM Cloud, your metadata is AES256-encrypted and protected by physical, organizational, access, and security controls.

IBM Storage Insights is ISO/IEC 27001 Information Security Management certified.

For more information about IBM Storage Insights, see the following websites:

- IBM Storage Insights Fact Sheet
- Functional demonstration environment (requires an IBMid)
- IBM Storage Insights security information
- IBM Storage Insights registration

IBM Storage Virtualize RESTful API

The IBM Storage Virtualize Representational State Transfer (REST) model API consists of command targets that are used to retrieve system information and to create, modify, and delete system resources. These command targets allow command parameters to pass through unedited to the IBM Storage Virtualize CLI, which handles parsing parameter

specifications for validity and error reporting. It also uses HTTPS to successfully communicate with the RESTful apiserver.

The RESTful apiserver does not consider transport security (such as Secure Sockets Layer (SSL)), but instead assumes that requests are started from a local, secured server. The HTTPS protocol provides privacy through data encryption. The RESTful API provides more security by requiring command authentication, which persists for 2 hours of activity or 30 minutes of inactivity, whichever occurs first.

Uniform Resource Locators (URLs) target different node objects on the system. The **HTTPS POST** method acts on command targets that are specified in the URL. To make changes or view information about different objects on the system, you must create and send a request to the system. You must provide specific elements for the RESTful apiserver to receive and transform the request into a command.

To interact with the system by using the RESTful API, make an HTTPS command request with a valid configuration node URL destination. Open TCP port 7443 and include the keyword **rest** and then use the following URL format for all requests:

`https://`*system_node_ip*`:7443/rest/`*command*

Where:

- *system_node_ip* is the system IP address, which is the address that is taken by the configuration node of the system.
- The port number is always 7443 for the IBM Storage Virtualize RESTful API.
- **rest** is a keyword.
- *command* is the target command object (such as **auth** or **lseventlog** with any parameters). The command specification uses the following format:

  command_name`,method="POST",headers={'`*parameter_name*`': '`*parameter_value*`', '`*parameter_name*`': '`*parameter_value*`',...}`

1.15 Copy services

IBM Storage Systems running IBM Storage Virtualize provide copy services functions that can be used to improve availability and support DR.

Volume mirroring

By using volume mirroring, a volume can have two physical copies in one IBM Storage System. Each volume copy can belong to a different pool and use a different set of capacity saving features.

When a host writes to a mirrored volume, the system writes the data to both copies. When a host reads a mirrored volume, the system picks one of the copies to read. If one of the mirrored volume copies is temporarily unavailable, the volume remains accessible to servers. The system remembers which areas of the volume are written, and resynchronizes these areas when both copies are available.

You can create a volume with one or two copies, and you can convert a non-mirrored volume into a mirrored volume by adding a copy. When a copy is added in this way, the system synchronizes the new copy so that the new copy is the same as the existing volume. Servers can access the volume during this synchronization process.

Volume mirroring can be used to migrate data to or from an IBM Storage Systems running IBM Storage Virtualize. For example, you can start with a non-mirrored image mode volume in the migration pool and then add a copy to that volume in the destination pool on internal storage. After the volume is synchronized, you can delete the original copy that is in the source pool. During the synchronization process, the volume remains available.

Volume mirroring also is used to convert fully allocated volumes to use data reduction technologies, such as thin-provisioning, compression, or deduplication, or to migrate volumes between storage pools.

FlashCopy

The FlashCopy or snapshot function creates a point-in-time (PiT) copy of data that is stored on a source volume to a target volume. FlashCopy is sometimes described as an instance of a time-zero (T0) copy. Although the copy operation takes some time to complete, the resulting data on the target volume is presented so that the copy appears to occur immediately, and all data is available immediately. Advanced functions of FlashCopy allow operations to occur on multiple source and target volumes.

Management operations are coordinated to provide a common, single PiT for copying target volumes from their respective source volumes to create a consistent copy of data that spans multiple volumes.

The function also supports multiple target volumes to be copied from each source volume, which can be used to create images from different PiTs for each source volume.

FlashCopy is used to create consistent backups of dynamic data and test applications, and to create copies for auditing purposes and for data mining. It can be used to capture the data at a specific time to create consistent backups of dynamic data. The resulting image of the data can be backed up; for example, to a tape device object storage or another disk/flash based storage technology. When the copied data is on tape, the data on the FlashCopy target disks becomes redundant and can be discarded.

Another possible FlashCopy application is creating test environments. FlashCopy can be used to test an application with real business data before the production version of the application is updated or replaced. With FlashCopy, a fully functional and space-efficient clone of a volume that contains real data can be created. It enables read and write access for the test environment while the real production environment data is safe and untouched. After testing is complete, the clone volume can be discarded or retained for future use.

FlashCopy can perform a restore from any FlashCopy mapping. Therefore, you can restore (or copy) from the target to the source of your regular FlashCopy relationships. When restoring data from FlashCopy, this method can be qualified as reversing the direction of the FlashCopy mappings. This approach can be used for various applications, such as recovering a production database application after an errant batch process that caused extensive damage.

Remote mirroring

You can use remote mirroring (also referred as Remote Copy [RC]) function to set up a relationship between two volumes, where updates made to one volume are mirrored on the other volume. The volumes can be on two different systems (intersystem) or on the same system (intrasystem).

For an RC relationship, one volume is designated as the primary and the other volume is designated as the secondary. Host applications write data to the primary volume, and updates to the primary volume are copied to the secondary volume. Normally, host applications do not run I/O operations to the secondary volume.

The following types of remote mirroring are available:

- Metro Mirror (MM)

  Provides a consistent copy of a source volume on a target volume. Data is written to the target volume synchronously after it is written to the source volume so that the copy is continuously updated.

  With synchronous copies, host applications write to the primary volume. The hosts do not receive a confirmation that the write operation is complete until the data is written to the secondary volume, which ensures that both volumes have identical data when the copy operation finishes. After the initial copy operation completes, the MM function always maintains a fully synchronized copy of the source data at the target site. The MM function supports copy operations between volumes that are separated by distances up to 300 km (186.4 miles).

  For DR purposes, MM provides the simplest way to maintain an identical copy on the primary and secondary volumes. However, as with all synchronous copies over remote distances, host application performance can be affected. This performance effect is related to the distance between primary and secondary volumes and depending on application requirements, its use might be limited based on the distance between sites.

- Global Mirror (GM)

  Provides a consistent copy of a source volume on a target volume. The data is written to the target volume asynchronously and the copy is continuously updated. When a host writes to the primary volume, a confirmation of I/O completion is received before the write operation completes for the copy on the secondary volume. Because of this situation, the copy might not contain the most recent updates when a DR operation is completed.

  If a failover operation is started, the application must recover and apply any updates that were not committed to the secondary volume. If I/O operations on the primary volume are paused for a short period, the secondary volume can become a match of the primary volume. This function is comparable to a continuous backup process in which the last few updates are always missing. When you use GM for DR, you must consider how you want to handle these missing updates.

  The secondary volume is typically less than 1 second behind the primary volume, which minimizes the amount of data that must be recovered if a failover occurs. However, a high-bandwidth link must be provisioned between the two sites.

- Global Mirror with Change Volumes (GMCV)

  Enables support for GM with a higher recovery point objective (RPO) by using change volumes. This function is for use in environments where the available bandwidth between the sites is smaller than the update rate of the replicated workload.

  With GMCV, or GM with cycling, change volumes must be configured for the primary and secondary volumes in each relationship. A copy is taken of the primary volume in the relationship to the change volume. The background copy process reads data from the stable and consistent change volume and copies the data to the secondary volume in the relationship.

  CoW technology is used to maintain the consistent image of the primary volume for the background copy process to read. The changes that occurred while the background copy process was active are also tracked. The change volume for the secondary volume can also be used to maintain a consistent image of the secondary volume while the background copy process is active.

  GMCV provides fewer requirements to inter-site link bandwidth than other RC types. It is mostly used when link parameters are insufficient to maintain the RC relationship without affecting host performance.

Intersystem replication is possible over an FC or IP link. The native IP replication feature enables replication between any family systems running IBM Storage Virtualize by using the built-in networking ports of the system nodes.

Note: Although all three types of RC are supported to work over an IP link, the recommended type is GMCV.

1.15.1 Policy-based replication

Policy-based replication uses volume groups and replication policies to automatically deploy and manage replication. Policy-based replication helps simplify configuring, managing, and monitoring replication between two systems.

Note: IBM FlashSystems 5015, 5035 and 5045 do not support policy-based replication.

With policy-based replication, you can replicate data between systems with minimal management, significantly higher throughput, and reduced latency compared to the remote-copy function. A replication policy has the following properties:

- A replication policy can be assigned to one or more volume groups.
- Replication policies cannot be changed after they are created. If changes are required, a new policy can be created and assigned to the associated volume group.

Each system supports up to a maximum of 32 replication policies:

- Replication policies

  Replication policies define the replication settings that are assigned to the volume groups. Replication policies replicate the volume groups and ensure that consistent data is available on the production and recovery system.

- Volume groups for policy-based replication

  Policy-based replication uses volume groups and replication policies to automatically deploy and manage replication. Policy-based replication significantly simplifies configuring, managing, and monitoring replication between two systems.

- Partnerships

  Two-site partnerships replicate volume data that is on one system to a remote system. Two-site partnerships are required for policy-based replication. Partnerships can be used for migration, 3-site replication, and disaster recovery situations.

For more information, see Getting Started with policy-based replication.

1.15.2 HyperSwap

The IBM HyperSwap function is an HA feature that provides dual-site, active-active access to a volume. It is available on systems that can support more than one I/O group.

With HyperSwap, a fully independent copy of the data is maintained at each site. When data is written by hosts at either site, both copies are synchronously updated before the write operation is completed. The HyperSwap function automatically optimizes to minimize data that is transmitted between two sites, and to minimize host read and write latency.

If the system or the storage at either site goes offline and an online and accessible up-to-date copy is left, the HyperSwap function can automatically fail over access to the online copy. The HyperSwap function also automatically resynchronizes the two copies when possible.

To construct HyperSwap volumes, active-active replication relationships are made between the copies at each site. These relationships automatically run and switch direction according to which copy or copies are online and up to date.

The relationships provide access to whichever copy is up to date through a single volume, which has a unique ID. This volume is visible as a single object across both sites (I/O groups), and is mounted to a host system.

A 2-site HyperSwap configuration can be extended to a third site for DR that uses the IBM Storage Virtualize 3-Site Orchestrator.

IBM Storage Virtualize 3-Site Orchestrator coordinates replication of data for disaster recovery and high availability scenarios between systems that are on three geographically dispersed sites. IBM Storage Virtualize 3-Site Orchestrator is a command line-based application that runs on a separate Linux host that configures and manages supported replication configurations on IBM Storage Virtualize products.

To extend a HyperSwap system or to create a HyperSwap configuration with IBM Storage Virtualize 3-Site Orchestrator requires more planning, installing, and configuring steps.

The HyperSwap function works with the standard multipathing drivers that are available on various host types, with no extra host support that is required to access the highly available volume. Where multipathing drivers support Asymmetric Logical Unit Access (ALUA), the storage system tells the multipathing driver which nodes are closest to it and must be used to minimize I/O latency. You tell the storage system which site a host is connected to, and it configures host pathing optimally.

1.16 IBM FlashCore Module drives, NVMe SSDs, and SCM drives

This section describes the three types of flash drives that can be installed in the control enclosures:

- FCM drives
- NVMe SSDs
- Storage-class memory (SCM) drives

Note: The SCM drives and XL FCM drives require IBM Storage Virtualize V8.3.1 or later to be installed on the IBM FlashSystem control enclosure.

The following IBM FlashSystem products can support all three versions of these drives:

- 9500
- 9500R Rack Solution
- 9200
- 9200R Rack Solution
- 7300
- 7200
- 5200
- 5100

They are not supported in any of the expansion enclosures.

Figure 1-55 shows an FCM (NVMe) with a capacity of 19.2 TB. The first generation of FCM drives were built by using 64-layer Triple Level Cell (TLC) flash memory and an Everspin MRAM cache into a U.2 form factor. FCM generation 2 and 3 are built using 96-layer Quad Level Cell (QLC) flash memory. These later generations also reformat some of the flash capacity to a pseudo-SLC mode (pSLC) to improve performance, reduce latency, and provide a dynamic read cache on the device.



Figure 1-55 IBM FlashCore Module (NVMe)

FCM drives are designed for high parallelism and optimized for 3D QLC and updated FPGAs. IBM also enhanced the FCM drives by adding read cache to reduce latency on highly compressed pages. Also added was four-plane programming to reduce the overall power during writes. FCM drives offer hardware-assisted compression up to 3:1 and are FIPS 140-2 complaint.

FCM drives carry IBM Variable Stripe RAID (VSR) at the FCM level and use DRAID to protect data at the system level. VSR and DRAID together optimize RAID rebuilds by offloading rebuilds to DRAID, and they offer protection against FCM failures.

Table 1-18 lists the capacities of the FCM type drives in terabytes usable (TBu).

Table 1-18 FCM type capacities

FCM module size	Physical size
Small	4.8 TBu
Medium	9.6 TBu
Large	19.2 TBu
XLarge	38.4 TBu

Industry-standard SSD NVMe drives

All the IBM FlashSystem models that are described in this book provide an option to use industry-standard SSD NVMe drives, which are sourced from Samsung and Toshiba and available in the several capacity variations, as listed in Table 1-19 on page 102.

Table 1-19 NVMe drive size options

Drive type	Physical size
NVMe Flash Drive	1.92 TBu
NVMe Flash Drive	3.84 TBu
NVMe Flash Drive	7.68 TBu
NVMe Flash Drive	15.36 TBu
NVMe Flash Drive	30.72 TBu

NVMe and adapter support

NVMe is a NUMA-optimized, high-performance, and highly scalable storage protocol that is designed to access nonvolatile storage media by using a host PCIe bus. NVMe uses low-latency and available parallelism, and reduces I/O impact.

NVMe supports multiple I/O queues with a maximum of 64 K queues, and each queue can support up to 64 K entries. Earlier generations of SAS and Serial Advanced Technology Attachment (SATA) support a single queue with only 254 and 32 entries and use many more CPU cycles to access data. NVMe handles a bigger workload for the same infrastructure footprint.

NVMe-oF is a technology specification that is designed to enable NVMe message-based commands to transfer data between a host computer and a target SSD or system. Data is transferred over a network, such as Ethernet, FC, or InfiniBand.

Storage-class memory

SCM drives use persistent memory technologies that improve endurance and reduce the latency of flash storage device technologies. All SCM drives use the NVMe architecture. IBM Research® is actively engaged in researching these new technologies.

For more information about nanoscale devices, see Thermal nanotechnology.

For a comprehensive overview of the flash drive technology, see the SNIA Educational Library web page.

These technologies fundamentally change the architecture of today's storage infrastructures. Figure 1-56 on page 103 shows the different types of storage technologies versus the latency for Intel drives.

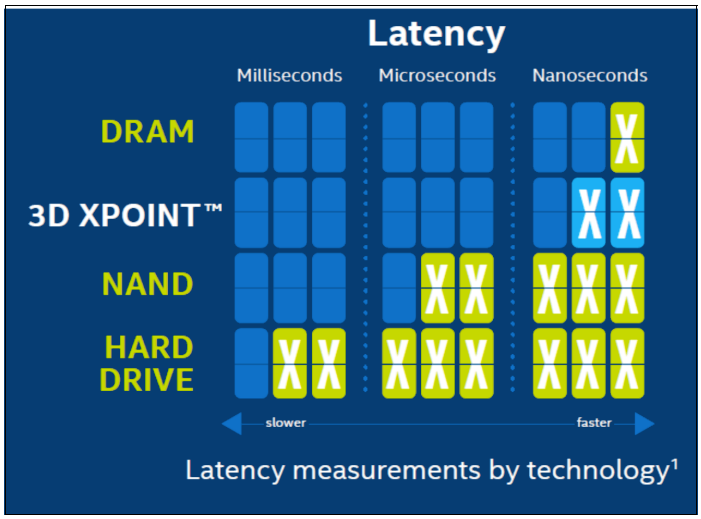


Figure 1-56 Storage technologies versus latency for Intel drives

IBM supports SCM class drives. Table 1-20 lists the SCM drive size options.

Table 1-20 SCM drive options

Drive type	Physical size (TBu)	Supported in
NVMe SCM Drive	.375 TBu	FS5200, FS7200
NVMe SCM Drive	.750 TBu	FS5200, FS7200
NVMe SCM Drive	.800 TBu	FS5200, FS7200
NVMe SCM Drive	1.6 TBu	FS5200, FS7200, FS7300, FS9500/R

The FS5200 and the FS7200 are not sold anymore, but they are still supported.

Easy Tier supports the SCM drives with a new tier that is called `tier_scm`.

Note: The SCM drive type supports only DRAID 6, DRAID 5, DRAID 1, and TRAID 1 or 10.

1.17 Storage virtualization

Storage virtualization is a term that is used extensively throughout the storage industry. It can be applied to various technologies and underlying capabilities. In reality, most storage devices technically can claim to be virtualized in one form or another. Therefore, this section starts by defining the concept of storage virtualization as it is used in this book.

We describe storage virtualization in the following ways:

- Storage virtualization is a technology that makes one set of resources resemble another set of resources (preferably with more wanted characteristics).
- Storage virtualization is a logical representation of resources that is not constrained by physical limitations and hides part of the complexity. It also adds or integrates new functions with services, and can be nested or applied to multiple layers of a system.

The virtualization model consists of the following layers:

- Application: The user of the storage domain.
- Storage domain:
  - File, record, and namespace virtualization, and file and record subsystem
  - Block virtualization
  - Block subsystem

Applications typically read and write data as vectors of bytes or records. However, storage presents data as vectors of blocks of a constant size (512 or in the newer devices, 4096 bytes per block).

The file, record, and namespace virtualization and file and record subsystem layers convert records or files that are required by applications to vectors of blocks, which are the language of the block virtualization layer. The block virtualization layer maps requests of the higher layers to physical storage blocks, which are provided by storage devices in the block subsystem.

Each of the layers in the storage domain abstracts the complexities of the lower layers and hides them behind a standard interface that is presented to upper layers. The resultant decoupling of logical storage space representation and its characteristics that are visible to servers (storage consumers) from underlying complexities and intricacies of storage devices is a key concept of storage virtualization.

The focus of this publication is block-level virtualization at the block virtualization layer, which is implemented by IBM as IBM Storage Virtualize software that is running on an IBM SAN Volume Controller and the IBM FlashSystem family. The IBM SAN Volume Controller is implemented as a clustered appliance in the storage network layer. The IBM FlashSystem storage systems are deployed as modular systems that can virtualize their internally and externally attached storage.

IBM Storage Virtualize uses the SCSI protocol to communicate with its clients and presents storage space as SCSI logical units (LUs), which are identified by SCSI LUNs.

Note: Although LUs and LUNs are different entities, the term *LUN* in practice is often used to refer to a logical disk, that is, an LU.

Although most applications do not directly access storage but work with files or records, the operating system of a host must convert these abstractions to the language of storage; that is, vectors of storage blocks that are identified by LBAs within an LU.

Inside IBM Storage Virtualize, each of the externally visible LUs is internally represented by a volume, which is an amount of storage that is taken out of a storage pool. Storage pools are made of MDisks; that is, they are LUs that are presented to the storage system by external virtualized storage or arrays that consist of internal disks. LUs that are presented to IBM Storage Virtualize by external storage usually correspond to RAID arrays that are configured on that storage.

With storage virtualization, you can manage the mapping between logical blocks within an LU that is presented to a host and blocks on physical drives. This mapping can be as simple or as complicated as required. A logical block can be mapped to one physical block, or for increased availability, multiple blocks that are physically stored on different physical storage systems, and in different geographical locations.

Importantly, the mapping can be dynamic: With Easy Tier, IBM Storage Virtualize can automatically change the underlying storage to which groups of blocks (extent) are mapped to

better match a host's performance requirements with the capabilities of the underlying storage systems.

IBM Storage Virtualize gives a storage administrator various options to modify volume characteristics, from volume resize to mirroring, creating a point-in-time (PiT) copy with FlashCopy, and migrating data across physical storage systems.

Importantly, all the functions that are presented to the storage users are independent from the characteristics of the physical devices that are used to store data. This decoupling of the storage feature set from the underlying hardware and ability to present a single, uniform interface to storage users that masks underlying system complexity is a powerful argument for adopting storage virtualization with IBM Storage Virtualize.

IBM Storage Virtualize includes the following key features:

- Simplified storage management by providing a single management interface for multiple storage systems, and a consistent user interface for provisioning heterogeneous storage.
- Online volume migration. IBM Storage Virtualize enables moving the data from one set of physical drives to another set in a way that is not apparent to the storage users and without over-straining the storage infrastructure. The migration can be done within a specific storage system (from one set of disks to another set) or across storage systems. Either way, the host that uses the storage is not aware of the operation, and no downtime for applications is needed.
- Enterprise-level copy services functions. Performing copy services functions within IBM Storage Virtualize removes dependencies on the capabilities and interoperability of the virtualized storage subsystems. Therefore, it enables the source and target copies to be on any two virtualized storage subsystems.
- Improved storage space usage because of the pooling of resources across virtualized storage systems.
- Opportunity to improve system performance as a result of volume striping across multiple virtualized arrays or controllers, and the benefits of cache that is provided by IBM Storage Virtualize hardware.
- Improved data security by using data-at-rest encryption.
- Data replication, including replication to cloud storage by using advanced copy services for data migration and backup solutions.
- Data reduction techniques for space efficiency and cost reduction. Today, open systems typically use less than 50% of the provisioned storage capacity. IBM Storage Virtualize can enable savings, increase the effective capacity of storage systems up to five times, and decrease the floor space, power, and cooling that are required by the storage system.

IBM FlashSystem and IBM SAN Volume Controller families are scalable solutions running on a HA platform that can use diverse back-end storage systems to provide all the benefits to various attached hosts.

External storage virtualization

You can use IBM Storage Systems running IBM Storage Virtualize to manage the capacity of other storage systems with external storage virtualization. When IBM Storage Virtualize virtualizes a storage system, its capacity is managed similarly to internal disk drives or flash modules. Capacity in external storage systems inherits all of the rich functions and ease of use.

You can use IBM Storage Systems running IBM Storage Virtualize to preserve your investments in storage, centralize management, and make storage migrations easier with

storage virtualization and Easy Tier. Virtualization helps insulate applications from changes that are made to the physical storage infrastructure.

To verify whether your storage can be virtualized by IBM FlashSystem or IBM SAN Volume Controller, see the IBM System Storage Interoperation Center (SSIC).

All the IBM Storage Systems that are running IBM Storage Virtualize can migrate data from external storage controllers, including migrating from any other IBM or third-party storage systems. IBM Storage Virtualize uses the functions that are provided by its external virtualization capability to perform the migration. This capability places external LUs under the control of an IBM FlashSystem or IBM SAN Volume Controller system. Then, hosts continue to access them through the IBM FlashSystem system or IBM SAN Volume Controller system, which acts as a proxy.

The migration process typically consists of the following steps:

1. Input/output (I/O) to the LUs that are on the external storage system must be stopped, and the mapping of the storage system must be changed so that the original LUs are presented directly to the IBM FlashSystem or IBM SAN Volume Controller Family machine and not to the hosts. IBM FlashSystem or IBM SAN Volume Controller discovers the external LUs and recognizes them as *unmanaged* external storage back-end devices (MDisks).
2. The unmanaged MDisks are imported to the IBM FlashSystem or IBM SAN Volume Controller image mode volumes and placed in a migration storage pool. This storage pool is now a logical container for the externally attached LUs. Each volume has a one-to-one mapping with an external LU. From a data perspective, the image mode volume represents the SAN-attached LUs exactly as they were before the import operation. The image mode volumes are on the same physical drives of the storage system, and the data remains unchanged.
3. Your hosts are configured for an IBM FlashSystem or IBM SAN Volume Controller attachment, and image-mode volumes are mapped to them. After the volumes are mapped, the hosts discover their volumes and are ready to continue working with them so that I/O can be resumed.
4. Image-mode volumes are migrated to the internal storage of IBM FlashSystem by using the volume mirroring feature. Mirrored copies are created online so that a host can still access and use the volumes during the mirror synchronization process.
5. After the mirror operations are complete, the image mode volumes are removed (deleted), and the external storage system can be disconnected and decommissioned or reused elsewhere.

The GUI of the IBM Storage Systems that are running IBM Storage Virtualize provides a storage migration wizard, which simplifies the migration task. The wizard features intuitive steps that guide users through the entire process.

Note: The IBM FlashSystem 5015, 5035 and 5045 systems do not support external virtualization for any purpose other than data migration.

Summary

Storage virtualization is a fundamental technology that enables the realization of flexible and reliable storage solutions. It helps enterprises to better align their IT architecture with business requirements, simplify their storage administration, and facilitate their IT departments efforts to meet business demands.

IBM Storage Virtualize that is running on the IBM FlashSystem family is a mature, 11th-generation virtualization solution that uses open standards and complies with the SNIA storage model. All the products are appliance-based storage, and use in-band block virtualization engines that move the control logic (including advanced storage functions) from many individual storage devices to a centralized entity in the storage network.

IBM Storage Virtualize can improve the use of your storage resources, simplify storage management, and improve the availability of business applications.

1.18 Business continuity

In today's online, highly connected, and fast-paced world, we expect that today's IT systems provide high availability (HA) and continuous operations, and that they can be quickly recovered if a disaster occurs. Yet, today's IT environment also features an ever-growing time to market pressure, with more projects to complete, more IT problems to solve, and a steep rise in time and resource limitations.

Thankfully, today's IT technology also features unprecedented levels of functions, features, and lowered cost. In many ways, it is easier than ever before to find IT technology that can address today's business concerns. This section describes some IBM FlashSystem storage solutions that can be applied to today's business continuity requirements.

1.18.1 Business continuity with the IBM SAN Volume Controller

A clustered system or system is a collection of servers that together provide a set of resources to a client. The key point is that the client has no knowledge of the underlying physical hardware of the system. The client is isolated and protected from changes to the physical hardware. This arrangement offers many benefits including, most significantly, HA.

Resources on the clustered system act as HA versions of unclustered resources. If a node (an individual computer) in the system is unavailable or too busy to respond to a request for a resource, the request is passed transparently to another node that can process the request. The clients are "unaware" of the exact locations of the resources that they use.

The IBM SAN Volume Controller is a collection of up to eight nodes, which are added in pairs that are known as *I/O groups*. These nodes are managed as a set (system), and they present a single point of control to the administrator for configuration and service activity.

The eight-node limit for an IBM SAN Volume Controller system is a limitation that is imposed by the Licensed Internal Code, and not a limit of the underlying architecture. Larger system configurations might be available in the future.

Although the IBM SAN Volume Controller code is based on a purpose-optimized Linux kernel, the clustered system feature is not based on Linux clustering code. The clustered system software within the IBM SAN Volume Controller (that is, the event manager cluster framework) is based on the outcome of the COMPASS research project. It is the key element that isolates the IBM SAN Volume Controller application from the underlying hardware nodes.

The clustered system software makes the code portable. It provides the means to keep the single instances of the IBM SAN Volume Controller code that are running on separate systems' nodes in sync. Therefore, restarting nodes during a code upgrade, adding nodes, removing nodes from a system, or failing nodes cannot affect IBM SAN Volume Controller availability.

All active nodes of a system must know that they are members of the system. This knowledge is especially important in situations where it is key to have a solid mechanism to decide which nodes form the active system, such as the split-brain scenario where single nodes lose contact with other nodes. A worst case scenario is a system that splits into two separate systems.

Within an IBM SAN Volume Controller system, the *voting set* and a quorum disk are responsible for the integrity of the system. If nodes are added to a system, they are added to the voting set. If nodes are removed, they are removed quickly from the voting set. Over time, the voting set and the nodes in the system can change so that the system migrates onto a separate set of nodes from the set on which it started.

The IBM SAN Volume Controller clustered system implements a dynamic quorum. Following a loss of nodes, if the system can continue to operate, it adjusts the quorum requirement so that further node failure can be tolerated.

The lowest Node Unique ID in a system becomes the boss node for the group of nodes. It determines (from the quorum rules) whether the nodes can operate as the system. This node also presents the maximum two-cluster IP addresses on one or both of its nodes' Ethernet ports to enable access for system management.

1.18.2 Business continuity with stretched clusters

Within standard implementations of the IBM SAN Volume Controller, all the I/O group nodes are physically installed in the same location. To supply the different HA needs that customers have, the *stretched system configuration* was introduced. In this configuration, each node (from the same I/O group) on the system is physically on a different site. When implemented with mirroring technologies, such as volume mirroring or copy services, these configurations can be used to maintain access to data on the system if power failures or site-wide outages occur.

Stretched clusters are considered HA solutions because both sites work as instances of the production environment (no standby location exists). Combined with application and infrastructure layers of redundancy, stretched clusters can provide enough protection for data that requires availability and resiliency.

In a stretched cluster configuration, nodes within an I/O group can be separated by a distance of up to 10 km (6.2 miles) by using specific configurations. You can use FC inter-switch links (ISLs) in paths between nodes of the same I/O group. In this case, nodes can be separated by a distance of up to 300 km (186.4 miles); however, potential performance impacts can result.

1.18.3 Business continuity with enhanced stretched cluster

Enhanced stretched cluster (ESC) further improves stretched cluster configurations with the *site awareness* concept for nodes, hosts, and external storage systems. It also provides a feature that enables you to manage effectively rolling disaster scenarios.

The site awareness concept enables more efficiency for host I/O traffic through the SAN, and an easier host path management.

The use of an IP-based quorum application as the quorum device for the third site does not require FC connectivity. Java applications run on hosts at the third site.

Note: Stretched cluster and ESC features are supported for IBM SAN Volume Controller only. They are not supported for the IBM FlashSystem family of products.

For more information and implementation guidelines about deploying a stretched cluster or ESC, see *IBM Spectrum Virtualize and SAN Volume Controller Enhanced Stretched Cluster with VMware*, SG24-8211.

1.18.4 Business continuity for IBM FlashSystem and IBM SAN Volume Controller

In this section, we discuss business continuity for IBM FlashSystem and IBM SAN Volume Controller.

1.18.5 Business continuity with HyperSwap

The HyperSwap HA feature in the IBM Storage Virtualize software enables business continuity during hardware, power, or connectivity failures, or disasters, such as fire or flooding. The HyperSwap feature is available on the IBM SAN Volume Controller and IBM FlashSystem products that are running IBM Storage Virtualize software.

The HyperSwap feature provides HA volumes that are accessible through two sites at up to 300 km (186.4 miles) apart. A fully independent copy of the data is maintained at each site.

When data is written by hosts at either site, both copies are synchronously updated before the write operation is completed. The HyperSwap feature automatically optimizes to minimize data that is transmitted between sites and to minimize host read and write latency.

HyperSwap includes the following key features:

- Works with IBM SAN Volume Controller and IBM FlashSystem products that are running IBM Storage Virtualize software.
- Uses intra-cluster synchronous RC (MM) capabilities along with change volume and access I/O group technologies.
- Makes a host's volumes accessible across two I/O groups in a clustered system by using the MM relationship in the background. They look like a single volume to the host.

- Works with the standard multipathing drivers that are available on various host types, with no other host support that is required to access the HA volume.

For more information about HyperSwap implementation use cases and guidelines, see the following publications:

- *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317
- *High Availability for Oracle Database with IBM PowerHA SystemMirror and IBM Spectrum Virtualize HyperSwap*, REDP-5459
- *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597

1.18.6 Business continuity with 3-site replication

A *3-site replication solution* was made available in limited deployments for Version 8.3.1, where data is replicated from the primary site to two alternative sites, and the remaining two sites are aware of the difference between themselves. This solution ensures that if a disaster occurs at any one of the sites, the remaining two sites can establish a `consistent_synchronized` RC relationship among themselves with minimal data transfer; that is, within the expected RPO.

IBM Storage Virtualize V8.4 and above expands the 3-site replication model to include HyperSwap, which improves data availability options in three-site implementations. Systems that are configured in a three-site topology have high DR capabilities, but a disaster might take the data offline until the system can be failed over to an alternative site.

HyperSwap allows active-active configurations to maintain data availability, which eliminates the need to failover if communications are disrupted. This solution provides a more robust environment, which allows up to 100% uptime for data, and recovery options that are inherent to DR solutions.

To better assist with 3-site replication solutions, IBM Storage Virtualize 3-Site Orchestrator coordinates replication of data for DR and HA scenarios between systems.

IBM Storage Virtualize 3-Site Orchestrator is a command-line based application that runs on a separate Linux host that configures and manages supported replication configurations on IBM Storage Virtualize products.

Figure 1-57 shows the two supported topologies for the three-site, replication-coordinated solutions.

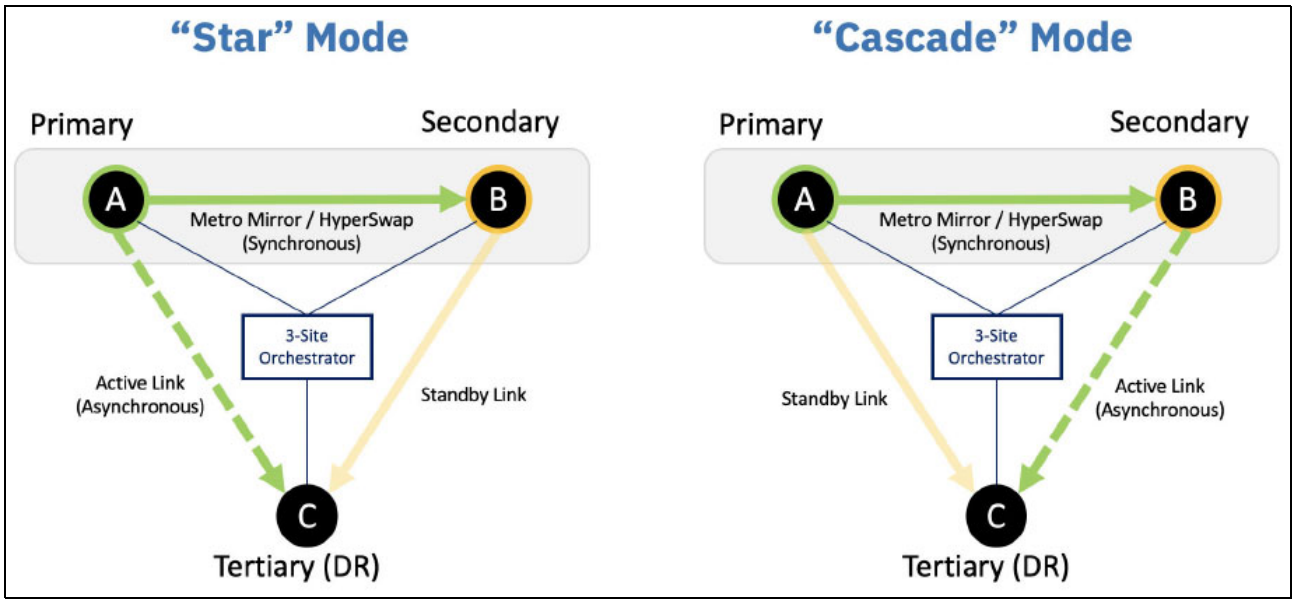


Figure 1-57 "Star" and "Cascade" modes in a three-site solution

For more information about this type of implementation, see *Spectrum Virtualize 3-Site Replication*, SG24-8474.

1.18.7 Automatic hot spare nodes (IBM SAN Volume Controller only)

In previous stages of IBM SAN Volume Controller development, the scripted *warm standby* procedure enables administrators to configure spare nodes in a cluster by using the concurrent hardware upgrade capability of transferring WWPNs between nodes. The system can automatically take on the spare node to replace a failed node in a cluster or to keep the entire system under maintenance tasks, such as software upgrades. These extra nodes are called *hot spare nodes*.

Up to four nodes can be added to a single cluster. When the hot-spare node is used to replace a node, the system attempts to find a spare node that matches the configuration of the replaced node perfectly.

However, if a perfect match does not exist, the system continues the configuration check until a matching criteria is found. The following criteria are used by the system to determine suitable hot-spare nodes:

- Requires an exact match:
  - Memory capacity
  - Fibre Channel port ID
  - Compression support
  - Site
- Recommended matching, but can be different:
  - Hardware type
  - CPU count
  - Number of Fibre Channel ports

If the criteria are not the same for both, the system uses lower criteria until the minimal configuration is found. For example, if the Fibre Channel ports do not match exactly but all the other required criteria match, the hot-spare node can still be used. The minimal configuration that the system can use as a hot-spare node includes identical memory, site, Fibre Channel port ID, and, if applicable, compression settings.

If the nodes on the system support and are licensed to use encryption, the hot-spare node must also support and be licensed to use encryption.

The hot spare node essentially becomes another node in the cluster, but is not doing anything under normal conditions. Only when it is needed does it use the N_Port ID Virtualization (NPIV) feature of the Storage Virtualize virtualized storage ports to take over the job of the failed node. It performs this takeover by moving the NPIV WWPNs from the failed node first to the surviving partner node in the I/O group and then over to the hot spare node.

Approximately 1 minute passes intentionally before a cluster swaps in a node to avoid any thrashing when a node fails. In addition, the system must be sure that the node definitely failed, and is not (for example) restarting. The cache flushes while only one node is in the I/O group, the full cache is returned when the spare swaps in.

This entire process is transparent to the applications; however, the host systems notice a momentary path that is lost for each transition. The persistence of the NPIV WWPNs lessens the multipathing effort on the host considerably during path recovery.

Note: A warm start of an active node (code assert or restart) does not cause the hot spare to swap in because the restarted node becomes available within 1 minute.

The other use case for hot spare nodes is during a software upgrade. Normally, the only impact during an upgrade is slightly degraded performance. While the node that is upgrading is down, the partner in the I/O group writes through cache and handles both nodes' workload. Therefore, to work around this issue, the cluster uses a spare in place of the node that is upgrading. The cache does not need to go into write-through mode and the period of degraded performance from running off a single node in the I/O group is significantly reduced.

After the upgraded node returns, it is swapped back so that you roll through the nodes as normal, but without any failover and failback at the multipathing layer. This process is handled by the NPIV ports; therefore, the upgrades must be seamless for administrators who are working in large enterprise IBM SAN Volume Controller deployments.

Note: After the cluster commits new code, it also automatically upgrades hot spares to match the cluster code level.

This feature is available to IBM SAN Volume Controller only. Although IBM FlashSystem systems can use NPIV and realize the general failover benefits, no hot spare canister or split I/O group option is available for the enclosure-based systems.

1.19 Management and support tools

The IBM Storage Virtualize system can be managed by using the included management software that runs on the controller hardware.

1.19.1 IBM Assist On-site and Remote Support Assistance

IBM Assist On-site and Remote Support Assistance are two tools that you can use to manage and support an IBM Storage Virtualize system.

IBM Assist On-site

With the IBM Assist On-site tool, a member of the IBM Support team can view your desktop and share control of your server to provide you with a solution. This tool is a remote desktop-sharing solution that is offered through the IBM website. With it, the IBM System Services Representative (IBM SSR) can remotely view your system to troubleshoot a problem.

You can maintain a chat session with the IBM SSR so that you can monitor this activity and understand how to fix the problem yourself or enable them to fix it for you.

For more information, see IBM remote assistance: Assist On-site.

When you access the website, you sign in and enter a code that the IBM SSR provides to you. This code is unique to each IBM Assist On-site session. A plug-in is downloaded to connect you and your IBM SSR to the remote service session. The IBM Assist On-site tool contains several layers of security to protect your applications and your computers. The plug-in is removed after the next restart.

You can also use security features to restrict access by the IBM SSR. Your IBM SSR can provide you with more information about the use of the tool.

Remote Support Assistance

The embedded part of the IBM Storage Virtualize V8.6 code is a software toolset that is called Remote Support Client. It establishes a network connection over a secured channel with Remote Support Server in the IBM network.

The Remote Support Server provides predictive analysis of the IBM Storage Controller running IBM Storage Virtualize software status and assists administrators with troubleshooting and fix activities. Remote Support Assistance is available at no extra charge, and no extra license is needed.

For more information, see Configuring support assistance.

1.19.2 Event notifications

IBM Storage Virtualize system can use SNMP traps, syslog messages, and a Call Home email or cloud-based notification, to notify you and the IBM Support Center when significant events are detected. Any combination of these notification methods can be used simultaneously.

Notifications are normally sent immediately after an event is raised. Each event that IBM Storage Controller detects is assigned a notification type of `Error`, `Warning`, or `Information`. You can configure the IBM Storage Controller to send each type of notification to specific recipients.

Simple Network Management Protocol traps

SNMP is a standard protocol for managing networks and exchanging messages. IBM Storage Virtualize can send SNMP messages that notify personnel about an event.

You can use an SNMP manager to view the SNMP messages that IBM Storage Virtualize sends. You can use the management GUI or the CLI to configure and modify your SNMP settings.

The MIB file for SNMP can be used to configure a network management program to receive SNMP messages that are sent by the IBM Storage Virtualize.

Syslog messages

The syslog protocol is a standard protocol for forwarding log messages from a sender to a receiver on an IP network. The IP network can be Internet Protocol Version 4 (IPv4) or Internet Protocol Version 6 (IPv6).

IBM SAN Volume Controller and IBM FlashSystems can send syslog messages that notify personnel about an event. The event messages can be sent in an expanded or concise format. You can use a syslog manager to view the syslog messages that IBM SAN Volume Controller or IBM FlashSystems sends.

IBM Storage Virtualize uses UDP to transmit the syslog message. You can use the management GUI or the CLI to configure and modify your syslog settings.

Call Home notification

Call Home notification improves the response time for issues on the system. Call Home notifications send diagnostic data to IBM Support personnel who can quickly determine solutions for these problems, which can disrupt operations on the system.

Call Home email

This feature transmits operational and error-related data to you and IBM through a Simple Mail Transfer Protocol (SMTP) server connection in the form of an event notification email. You can use the Call Home function if you have a maintenance contract with IBM or if IBM SAN Volume Controller or IBM FlashSystems are within the warranty period.

To send email, you must configure at least one SMTP server. You can specify as many as five other SMTP servers for backup purposes. The SMTP server must accept the relaying of email from the IBM SAN Volume Controller clustered system IP address. Then, you can use the management GUI or the CLI to configure the email settings, including contact information and email recipients. Set the reply address to a valid email address.

Send a test email to check that all connections and infrastructure are set up correctly. You can disable the Call Home function at any time by using the management GUI or CLI.

Cloud Call Home

Call Home with cloud services sends notifications directly to a centralized file repository that contains troubleshooting information that is gathered from customers. Support personnel can access this repository and be assigned issues automatically as problem reports. This method of transmitting notifications from the system to support removes the need for customers to create problem reports manually.

Call Home with cloud services also eliminates email filters dropping notifications to and from support, which can delay resolution of problems on the system. Call Home with cloud services uses Representational State Transfer (RESTful) APIs, which are a standard for transmitting data through web services.

For new system installations, Call Home with cloud services is configured as the default method to transmit notifications to support. When you update the system software, Call Home with cloud services is also set up automatically. Ensure that network settings are configured to allow connections to the support center. This method sends notifications to only the predefined support center.

To use Call Home with cloud services, ensure that all of the nodes on the system have internet access, and that a valid service IP is configured on each node on the system. In addition to these network requirements, you must configure suitable routing to the support center through a domain name service (DNS) or by updating your firewall configuration so it includes connections to the support center.

After a DNS server is configured, update your network firewall settings to allow outbound traffic to `esupport.ibm.com` on port 443.

If you are not using DNS but you have a firewall to protect your internal network from outside traffic, you must enable specific IP addresses and ports to establish a connection to the support center. For the list of IP addresses to allow through your firewall, see Preparing customer firewalls and proxies for the upcoming infrastructure changes - Call Home, Electronic Fix Distribution.

You can configure either of these methods or configure both for redundancy. DNS is the preferred method because it ensures that the system can still connect to the IBM Support center if the underlying IP addresses to the support center change.

IBM highly encourages all customers to take advantage of the Call Home feature so that you and IBM can collaborate for your success.

For more information, see IBM Storage Virtualize Products Call Home and Remote Support Overview.

1.20 Licensing

All IBM FlashSystem functional capabilities are provided through IBM Storage Virtualize software. Each platform is licensed as described in the following sections.

1.20.1 Licensing IBM SAN Volume Controller

The IBM SAN Volume Controller uses perpetual licenses. A perpetual license is the “traditional” model that is used to purchase software. You pay for your software license up-front and can use it indefinitely. For more information about the following storage capacity for external virtualization features, based on storage classification, see Table 1-21 on page 116.

- FlashCopy/TiB based
- Remote Mirroring/TiB based
- Encryption/Key license

1.20.2 Licensing IBM FlashSystem 9500/R, 9200/R, 7300, 7200, 5200 and 5045

The IBM FlashSystems 9500, 9500R, 9200, 9200R, 7300, 7200, 5200 and 5045 include all-inclusive licensing for all functions except encryption (which is a country-limited feature code) and external virtualization.

Note: All internal enclosures in the FS9xx0, 7300, 7200, and 5200 require a license. However, the software for the FS9500 (4983 model only), the FS5200, FS5045, and FS7300 is License Machine Code.

Any externally virtualized storage requires the External Virtualization license per storage capacity unit (SCU) that is based on the tier of storage that is available on the external storage system. In addition, if you use FlashCopy and Remote Mirroring on an external storage system, you must purchase a per-tebibyte license to use these functions.

The SCU is defined in terms of the category of the storage capacity, as listed in Table 1-21.

Table 1-21 SCU category definitions

License	Drive class	SCU ratio
SCM	SCM devices	SCU equates to 1.00 TiB usable of Category 1 storage.
Flash	All flash devices, other than SCM drives	SCU equates to 1.18 TiB usable of Category 2 storage.
Enterprise	10 K or 15 K RPM drives	SCU equates to 2 TiB usable of Category 3 storage.
NL	NL SATA drives	SCU equates to 4.00 TiB usable of Category 4 storage.

Encryption license (key-based)

Encryption is enabled on IBM FlashSystem systems by obtaining the Encryption Enablement feature. This feature enables encryption at the system level and externally virtualized storage subsystems.

The encryption feature uses a key-based license that is activated by using an authorization code. The authorization code is sent with the IBM FlashSystem Licensed Function Authorization documents that you receive after purchasing the license.

The Encryption USB Flash Drives (Four Pack) feature or an external key manager, such as the IBM Security Key Lifecycle Manager, are required for encryption keys management.

1.20.3 Licensing IBM FlashSystem 5035 and 5015

The base license that is provided with the system includes its basic functions. However, extra licenses can be purchased to expand the capabilities of the system. Administrators are responsible for purchasing extra licenses and configuring the systems within the license agreement, which includes configuring the settings of each licensed function.

IBM FlashSystem 5000 licenses (key-based)

The IBM FlashSystem 5015 and 5035 systems use key-based licensing in which an authorization code is used to activate licensed functions on the system. The authorization code is sent with the IBM FlashSystem 5000 Licensed Function Authorization documents that you receive after purchasing the license. These documents contain the authorization codes that are required to obtain keys (also known as *DFSA license keys*) for each licensed function that you purchased for your system. For each license that you purchase, a separate document with an authorization code is sent to you.

Each function is licensed to an IBM FlashSystem 5000 control enclosure. It covers the entire system (control enclosure and all attached expansion enclosures) if it consists of one I/O group. If the IBM FlashSystem 5030 and 5035 systems consists of two I/O groups, two keys are required.

The following functions require a license key before they can be activated on the system:

- Easy Tier

  Easy Tier automatically and dynamically moves frequently accessed data to flash (solid-state) drives in the system, which results in flash drive performance without manually creating and managing storage tier policies. Easy Tier makes it easy and economical to deploy flash drives in the environment. In this dynamically tiered environment, data movement is seamless to the host application, regardless of the storage tier in which the data is stored.

- Remote Mirroring

  The Remote Mirroring (also known as remote copy [RC]) function enables you to set up a relationship between two volumes so that updates that are made by an application to one volume are mirrored on the other volume.

  The license settings apply to only the system on which you are configuring license settings. For RC partnerships, a license is also required on any remote systems that are in the partnership.

- FlashCopy upgrade

  The FlashCopy upgrade extends the base FlashCopy function that is included with the product. The base version of FlashCopy limits the system to 64 target volumes. With the FlashCopy upgrade license activated on the system, this limit is removed. If you reach the limit that is imposed by the base function before activating the upgrade license, you cannot create more FlashCopy mappings.

To help evaluate the benefits of these new capabilities, Easy Tier and RC licensed functions can be enabled at no extra charge for a 90-day trial. Trials are started from the IBM FlashSystem management GUI and do not require any IBM intervention. When the trial expires, the function is automatically disabled unless a license key for that function is installed onto the machine.

If you use a trial license, the system warns you at regular intervals when the trial is about to expire. If you do not purchase and activate the license on the system before the trial license expires, all configurations that use the trial licenses are suspended.

Encryption license (key-based)

Encryption is enabled on IBM FlashSystem 5035 through the acquisition of the Encryption Enablement feature. This feature enables encryption on the entire IBM FlashSystem family system and externally virtualized storage subsystems.

Note: Encryption hardware feature is available on the IBM FlashSystem 5035 and 5045 only.

This encryption feature uses a key-based license and is activated with an authorization code. The authorization code is sent with the IBM FlashSystem 5000 Licensed Function Authorization documents that you receive after purchasing the license.

The Encryption USB flash drives (Four Pack) feature or IBM Security Key Lifecycle Manager are required for encryption keys management.

2

Installation and configuration planning

This chapter describes the steps that are required to plan the installation and configuration of IBM Storage Virtualize storage systems in your storage network. Not all features that are described in this chapter are available and supported on all IBM Storage Virtualize storage systems.

For more information about which product features are relevant to your IBM Storage Virtualize storage system, see 1.6, “IBM FlashSystem family” on page 48. For IBM SAN Volume Controller, see 1.4, “IBM SAN Volume Controller family” on page 23.

This chapter is *not* intended to provide in-depth information about the described topics. It provides only general guidelines. For an enhanced analysis, see *Performance and Best Practices Guide for IBM Storage FlashSystem and IBM SAN Volume Controller*, SG24-8543.

Note: Make sure that the planned configuration is reviewed by IBM or an IBM Business Partner before implementation. Such a review can increase the quality of the final solution and prevent configuration errors that might affect the solution delivery.

This chapter includes the following topics:

- 2.1, “General planning guidelines” on page 120
- 2.2, “Planning for availability” on page 121
- 2.3, “Physical installation planning” on page 122
- 2.4, “Planning for system management” on page 123
- 2.5, “Connectivity planning” on page 126
- 2.6, “Portsets” on page 127
- 2.7, “Fibre Channel SAN configuration planning” on page 130
- 2.8, “IP SAN configuration planning” on page 137
- 2.9, “Planning topology” on page 144
- 2.10, “Back-end storage configuration” on page 146
- 2.11, “Internal storage configuration” on page 148
- 2.12, “Storage pool configuration” on page 152
- 2.13, “Volume configuration” on page 157

© Copyright IBM Corp. 2024. All rights reserved.

- 2.14, “Host attachment planning” on page 162
- 2.15, “Planning copy services” on page 163
- 2.16, “Throttles” on page 168
- 2.17, “Data migration” on page 169
- 2.18, “Ansible automation for IBM Storage Virtualize systems” on page 171
- 2.19, “Container Storage integration” on page 171
- 2.20, “Safeguarded Copy” on page 172
- 2.21, “Performance monitoring with IBM Storage Insights” on page 173
- 2.22, “Configuration backup procedure” on page 176

2.1 General planning guidelines

To maximize the benefit from a system, installation planning must include several important steps. These steps ensure that the system provides the best possible performance, reliability, and ease of management for your application needs.

The general rule of planning is to define your goals and then plan a solution so that you can reach these goals.

Consider the following points when planning a system:

- Collect and document the following information about application servers (hosts) that you want to attach to the system and their data:
  - Amount of data in use for each host and growth plans.
  - Data profile: Compressibility and deduplicability.
  - Host traffic profile: Percentage of reads and writes, sequential and random access patterns, and data block size.
  - Host performance requirements: Input/output operations per second (IOPS) and bandwidth.
- Perform capacity and performance sizing of a system:
  - If any external back-end systems are going to be virtualized, assess their capacity and performance capabilities.
  - Calculate the number of drives or IBM FlashCore Module modules (FCM) that are needed to satisfy your capacity requirements by considering your data compression ratios and accounting for future growth.
  - Verify that the capacity assessment results satisfy your performance requirements.

  Note: Contact your IBM sales representative or IBM Business Partner to perform these calculations.

- Assess your recovery point objective (RPO)/recovery time objective (RTO) requirements and plan for high availability (HA) and Remote Copy (RC) functions. Decide whether you require a dual-site or three-site deployment, and whether you must implement RC and determine its type (synchronous or asynchronous). Review the extra configuration requirements that are imposed.
- Assess and determine your cyber resiliency and data encryption requirements:
  - Calculate the capacity requirements that are associated with Safeguarded Copy and the ability to recover from a cyberattack.

Note: Contact your IBM sales representative or IBM Business Partner to perform these calculations.

  - Ensure that sufficient licensing requirements are planned to allow for encryption of data or for implementation of Copy Service Manager or Copy Data Management software that is used along with Safeguarded Copy.

- Define the number of I/O groups (control enclosures) and expansion enclosures. The number of necessary enclosures depends on the solution type, overall performance, and capacity requirements.
- Plan for host attachment interfaces, protocols, and storage area network (SAN). Consider the number of ports, bandwidth requirements, and HA.
- Perform configuration planning by defining the number of internal storage arrays and external storage arrays that are to be virtualized. Define a number and the type of pools, the number of volumes, and the capacity of each volume.
- Define a naming convention for the system nodes, volumes, and other storage objects.
- Plan a management IP network and management users' authentication system.
- Plan for the physical location of the equipment in the rack.
- Verify that your planned environment is a supported configuration.

Note: Use IBM System Storage Interoperation Center (SSIC) to check compatibility.

- Verify that your planned environment does not exceed system configuration limits.

Note: For more information about your platform and code version, see the following suitable Configuration Limits and Restrictions link for your IBM Storage Virtualize system:

- V8.6.0.x Configuration Limits and Restrictions for IBM System Storage SAN Volume Controller
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 9500
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 7200 and 7300
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 50x5 and 5200

- Determine your storage solution service and support requirements to align with your IT environment Service Level Agreements.
- Review the planning aspects that are described in the following sections of this chapter.

2.2 Planning for availability

When planning the deployment of the IBM Storage Virtualize solution, avoid creating single points of failure (SPOFs). Plan your system availability according to the requirements of your solution. Depending on your availability needs, consider the following aspects:

- Single-site or multi-site configuration

  Multi-site configurations increase solution resiliency, and can be the basis of disaster recovery (DR) solutions. Systems can be configured as a multi-site solution with sites working in active-active mode.

Synchronous and asynchronous data replication is supported by multiple inter-site link options. Three-site replication deployments are also supported.

- Physical separation of system building blocks

  A dual-rack deployment might increase the availability of your system if your back-end storage, SAN, and local area network (LAN) infrastructure also do not use a single-rack placement scheme. You can further increase system availability by ensuring that enclosures are powered from different power circuits and in different fire protection zones.

- Quorum disk placement

  For an IBM FlashSystem deployment with multiple I/O groups, plan for a quorum device on an external back-end system or an IP quorum application. IP quorum applications must be deployed on hosts that do not depend on storage that is provisioned by a system. Multiple IP quorum application deployment is recommended.

  The SAN Volume Controller uses three managed disks (MDisks) (external storage Logical Units or LUs) or an IP quorum application as quorum devices for the clustered system. A best practice is to have each quorum device in a separate storage subsystem, if possible. Multiple IP quorum application deployment is also recommended.

- The use of spare nodes (IBM SAN Volume Controller only)

  You can purchase and configure a hot spare node to minimize the effect of hardware failures.

- Failure domain sizes

  A failure of a single managed disk (MDisk) within a storage pool can take the entire storage pool offline. To reduce the effect of an MDisk failure, consider reducing the number of back-end storage systems per storage pool and increasing the number of storage pools and reducing their size. However, this configuration limits the maximum performance of the pool (fewer back-end systems to share the load), increases storage management effort, can lead to less efficient storage capacity consumption, and might be subject to limitations by system configuration maximums.

- Consistency

  Strive to achieve consistent availability levels of all system building blocks. For example, if the solution relies on a single switch that is placed in the same rack as an IBM SAN Volume Controller node, investment in a dual-rack configuration for placement of the second node is not justified. Any incident that affects the rack that holds the critical switch brings down the whole system no matter where the second IBM SAN Volume Controller node is placed.

2.3 Physical installation planning

Consider several key factors when you plan the physical site of a system. The physical site must have the following characteristics:

- Sufficient rack space exists to install controller and disk enclosures.
- The site meets the power, cooling, and environmental requirements.

For more information about power and environmental requirements, see the IBM Documentation information that is relevant to your IBM Storage Virtualize platform. For example, to see the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Planning** → **Planning for hardware** → **Physical installation planning**, and then, select **Connections for control enclosures** and **SAS expansion enclosure requirements.**

Your system order includes a printed copy of the *Quick Installation Guide*, which also provides information about environmental and power requirements.

Create a cable connection table that follows your environment's documentation procedure to track the following connections that are required for the setup:

- Power
- Serial-attached Small Computer System Interface (SCSI) (SAS)
- Ethernet
- Fibre Channel (FC)

When planning for power, also plan for a separate independent power source for each of the two redundant power supplies of a system enclosure.

Distribute your expansion enclosures between control enclosures and SAS chains, as described in 11.1.3, "Enclosure SAS cabling" on page 972. For more information, see the IBM Documentation information that is relevant to your IBM Storage Virtualize platform.

For example, to see the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Installing** → **Connecting the components** → **Connecting expansion enclosures to the control enclosure**.

When planning SAN cabling, make sure that your physical topology adheres to zoning rules and recommendations.

The physical installation and initial setup of IBM FlashSystem 9500, and IBM SAN Volume Controller is performed by an IBM System Services Representative (IBM SSR).

IBM FlashSystem 7300, IBM FlashSystem 5200, IBM FlashSystem 5045, IBM FlashSystem 5035, and IBM FlashSystem 5015 are classified as Customer Setup Units (CSUs), and the physical installation and initial setup is the responsibility of the customer. IBM can be contracted to perform these services for a fee.

2.4 Planning for system management

Each node in the system node has a *technician port*. It is a dedicated 1 Gbps Ethernet port. The initialization of a system and its basic configuration is performed by using this port. After the initialization is complete, the technician port must remain disconnected from a network and used only to service the system.

On the IBM FlashSystem 5015 (as opposed to other platforms) the technician port is *not* dedicated. On those systems, after the initial configuration, it is converted to a regular Ethernet port that can be connected to the network and used for management tasks and to serve I/O to hosts with internet Small Computer Systems Interface (iSCSI).

For management, each system node requires at least one Ethernet connection. The cable must be connected to port 1. The following products require a 10 Gbps Ethernet connection, which does not negotiate speeds below 1 Gbps:

- IBM FlashSystem products, excluding IBM FlashSystem 9500
- IBM SAN Volume Controller 2145-SV2 and 2145-SA2 nodes

The following products require a 1 Gbps Ethernet connection on port 1:

- IBM SAN Volume Controller 2145-SV3 node
- IBM FlashSystem 9500

For increased availability, an optional management connection can be configured over Ethernet port 2.

Note: 10 Gbps management ports can also be used for iSCSI host I/O. However, 1 Gbps management ports can be used for management traffic only.

For configuration and management tasks, you must allocate an IP address to each node canister, which is referred to as the *service IP address*. IPv4 and IPv6 are supported.

In addition to a service IP address on each node, each system has a cluster management IP address. The cluster management IP address cannot be the same as any of the defined service IP addresses. The cluster management IP can automatically fail over between cluster nodes if maintenance actions or a node failure occur.

For example, a system that consists of two control enclosures requires a minimum of five unique IP addresses: one for each canister node and one for the system as a whole.

Ethernet ports 1 and 2 are not reserved only for management. They can also be used for iSCSI or IP replication traffic if they are configured to do so. However, management and service IP addresses cannot be used for host or back-end storage communication.

System management is performed by using an embedded GUI that is running on the nodes. The command line interface (CLI) is also available. To access the management GUI, enter the cluster management IP address in a supported web browser. To access the management CLI, enter the cluster management IP in a Secure Shell (SSH) client and use the default SSH protocol port (22/TCP).

By connecting to a service IP address with a browser or SSH client, you can access the Service Assistant Interface, which can be used for maintenance and service tasks.

When you plan your management network, note that the IP Quorum applications and Transparent Cloud Tiering (TCT) are communicating with a system through the management ports.

For more information about cloud backup requirements, see 10.5, “Transparent Cloud Tiering” on page 789.

2.4.1 User password creation options

IBM Storage Virtualize includes a password policy support feature with which system administrators set security requirements. These requirements are related to password creation and expiration, timeout for inactivity, and actions after failed logon attempts.

Password policy support allows administrators to set security rules that are based on your organization’s security guidelines and restrictions. The system supports the password and security-related rules that are described in the following subsections.

Password creation rules

Administrators can set and manage the following rules for all passwords that are created on the system:

- Specify password length requirements for all users.
- Require passwords to:
  - Use uppercase and lowercase characters

  - Contain special characters
- Prevent users from reusing recent passwords.
- Require users to change their password on the next login under any of the following conditions:
  - Their password expired.
  - An administrator created accounts with temporary passwords.
- Enable password history checking.
- Set the minimum required password age to prevent bypassing the password history restriction by rapidly changing passwords multiple times.

Note: A new policy does not apply retroactively to existing passwords.

Password expiration and account-locking rules

The administrator can create the following rules for password expiration:

- Set:
  - A password expiration limit
  - A password to expire immediately
  - Number of failed login attempts before the account is locked
  - A period for locked accounts
- Automatic log out for inactivity
- Locking superuser account access

Note: Systems that support a dedicated technician port can lock the superuser account. The superuser account is the default user that can complete installation, initial configuration, and other service-related actions on the system. If the superuser account is locked, service tasks cannot be completed.

For more information about implementing these features, see Chapter 4, “IBM Storage Virtualize GUI” on page 233.

2.4.2 Two person integrity

Use two-person integrity (TPI) to prohibit critical and risky tasks in the system from being run by a single security administrator and by requiring the involvement of two security administrators, when using the Safeguarded Copy feature.

TPI requires two security administrators to complete certain tasks. Protecting data is an important part of IBM Storage Virtualize, and TPI helps mitigate the chance of data loss, prevent inadvertent mistakes on operations, and enhance security.

Provide the following requirements to enable TPI:

- Two users with the security administrator role.
- The two users can be local, remote, or a combination of both.
- If you use remote users, a remote user group of security administrator role must be defined on the system and the remote authentication service must be enabled.

Requirements for disabling TPI:

- After TPI is enabled, a user with an approved TPI request can disable TPI.

When you enable TPI, the users that belong to user groups of security administrator role are assigned the restricted security administrator role instead. However, their user groups retain their security administrator role.

After TPI is enabled, a role elevation request and approval process is required to perform certain sensitive tasks:

- The restricted security administrator can issue a role elevation request on its own behalf to complete certain tasks in the system.
- Another restricted security administrator or a security administrator must approve the role elevation request.
- For example, this role elevation request and approval process is required to remove a Safeguarded snapshot.
- The restricted security administrators or security administrators can approve or deny role elevation requests, cancel role elevation requests, or revoke a role elevation request that was approved.

The following actions are available for a restricted security administrator that has an approved role elevation request:

- Create, change, or remove security administrator user groups
- Change the non-security administrator user group attribute on an existing local user to a security administrator user group
- Modify attributes on existing local users that are members of the security administrator user groups
- Change the role of existing non-security administrator user groups to the security administrator role
- Change the security administrator role of an existing user group to a non-security administrator role
- Remove and change Safeguarded backups and Safeguarded backup locations
- Delete Safeguarded snapshots
- Use a provisioning policy to define a set of rules that are applied when volumes are created within a storage pool or child pool
- Change the single sign-on credentials that are used for the system
- Remove the Safeguarded snapshot policy association from a volume group

2.5 Connectivity planning

An IBM Storage Virtualize system offers a wide range of connectivity options to back-end storage and hosts, such as FC technologies (“traditional” SCSI FC and Non-Volatile Memory Express over FC [FC-NVMe], which are also known as NVMe over Fabric [NVMe-oF]), IP network technologies (iSCSI, iSCSI Extensions for Remote Direct Memory Access [RDMA] [iSER] and NVMe over RDMA or TCP) and SAS technologies. The connection options and capabilities depend on the hardware configuration.

Table 2-1 lists the communication types that can be used for communicating between system nodes, hosts, and back-end storage systems on an IBM FlashSystem system. All types can be used concurrently.

Table 2-1 Communication options

Communication type	System to host	System to back-end storage	Node to node (intra-cluster)	System to system (replication)
SCSI FC	Yes	Yes	Yes	Yes
FC-NVMe	Yes	No	No	No
iSCSI	Yes	Yes	No	No[a]
iSER	Yes[b]	No	Yes	No[a]
NVMe-RDMA	Yes	No	No	No
NVMe-TCP	Yes	No	No	No
SAS	Yes[c]	Yes[d]		

a. Replication traffic can be sent over an IP network with native IP replication, which can be configured on an onboard 10 Gb Ethernet (GbE) ports and an optional 25 GbE ports.
b. ISER host attach is not supported on IBM FlashSystem 9500 and IBM FlashSystem 7300.
c. SAS host attachment is available on IBM FlashSystem 5015 and IBM FlashSystem 5035 only.
d. Back-end storage attachment is supported for data migration only.

Table 2-2 lists the same information for an IBM SAN Volume Controller system.

Table 2-2 Communication options

Communication type	System to host	System to back-end storage	Node to node (intra-cluster)	System to system (replication)
SCSI FC	Yes	Yes	Yes	Yes
NVMe over FC	Yes	No	No	No
iSCSI	Yes	Yes	No	No[a]
iSER	Yes[b]	No	Yes	No[a]
NVMe-RDMA	Yes	No	No	No
NMVe-TCP	Yes	No	No	No

a. Replication traffic can be sent over an IP network with native IP replication, which can be configured on an onboard 10-Gigabit Ethernet (GbE) ports and an optional 25 GbE ports.
b. ISER host attach is not supported on IBM SAN Volume Controller SV3.

2.6 Portsets

Portsets are groupings of logical addresses that are associated with the specific traffic types. The system supports portsets for host attachment, back-end storage connectivity, and IP replication traffic. The system supports a maximum of 72 portsets.

This section describes the different types of available portsets.

Ethernet portset

Each physical Ethernet port can have a maximum of 64 IP addresses with each address on a unique portset. However, for each port, the IP address can be shared between multiple unique portsets for different functions.

Each port can bind to only a single IP address per portset for specific Ethernet functions, such as host attachment (iSCSI), back-end storage connectivity (iSCSI only), and IP replication. For cloud environments, each Ethernet port supports two IP addresses and VLANs per port for multiple clients that share storage resources on the system.

A portset restricts a host to access only a specific set of IP addresses of a node. A host can access only those IP addresses that are configured on a portset and is mapped to that host.

A portset *object* is a system-wide object and might contain IP addresses from every I/O group. To access multiple nodes in a system, a portset must be configured with the IP addresses of the nodes that the host wants to access. A portset can be of Host Attach, Remote Copy, and Storage type. The default portset is the Host type. A portset of a specific type can be used only for that function. For example, a host attach type portset cannot be used for remote copy partnership.

Each portset is identified by a unique name. Portset 0, portset 1, portset 2, and portset 3 are the default portsets that are configured when the system is created or updated. A portset can be created and managed by using the CLI and GUI.

In general, portsets include the following characteristics:

- Portsets are system-wide objects where IP addresses from all nodes might be included in the portset for host, storage, and replication functions.
- Each IP address in a portset must be configured on a separate Ethernet port.
- The same ports can share IP addresses across different portsets that allow the same IP address to be used for host, storage, and remote-copy traffic. All shared IP addresses must use the same port and have the same VLAN, gateway, and prefix. When IP addresses are shared among multiple portsets, the system creates a logical copy of the IP address and its attributes, rather than a new IP address.
- Portsets that are owned by different ownership groups can share an IP address.
- The system supports a maximum of 64 portsets.
- A port can have 64 unique or shared IP addresses. All 64 IP addresses must be IPv4 or IPv6, or a mix of IPv4 and IPv6.
- Each port can be configured with only one unique routable IP address (gateway specified). The routable IP can be shared among multiple portsets.
- Portset 0 is a default portset that is automatically configured when the system is updated or created. Portset 0 is a host portset by default and cannot be deleted, even if it is empty. Portset 0 serves as the default portset for any IP addresses and host objects that are configured without a portset specified. Portset 0 allows administrators to continue with an original configuration that does not require multi-tenancy. After an update, all configured host objects are automatically mapped to portset 0.

A portset can be defined to an ownership group. When you define an ownership group for portsets, you can limit and restrict users to view and manage only specific portsets. Portset 0, Portset 3, and the replication portset are always globally owned and only global administrators can assign and modify IP addresses to the portsets.

In a typical configuration, a portset object is created and then, the IP address object and host object are configured. When an IP address or host is configured, a portset must be specified, or the default portset 0 is selected.

For more information about planning and the limitations of portsets, see the IBM Documentation information that is relevant to your IBM Storage Virtualize platform. For example, for the IBM FlashSystem 9500-related information, see IBM FlashSystem 9500 documentation and expand **Planning** → **Planning the configuration** → **Planning for multiple IP address for Ethernet connectivity**.

Fibre Channel portset

An FC portset can be defined to be a group of FC I/O ports and FC SCSI/NVME hosts. FC portsets can be used with Ethernet portsets. However, an FC portset does not support grouping FC I/O ports and Ethernet IP addresses. The FC portsets can be configured for host attachment only. A system can have a collective maximum of 72 FC or Ethernet portsets.

It is possible to create multiple FC portsets for host connectivity where different hosts can be associated with different FC portsets. Each FC I/O port can be added to multiple FC portsets. However, a host can be added to only one FC portset. Every portset can support up to four FC I/O ports.

Note: It is recommended to use portsets when the system has more than one FC port per canister.

FC portsets have the following characteristics:

- Many host objects can be mapped to a single FC portset.
- If you do not map a host object to a portset, the host object is automatically mapped to the default portset. The default portset for FC port sets is `portset64`.
- Host objects can access volumes through only the FC ports that are associated with the respective portset.
- The Port Type field value defines whether the portset is Ethernet or FC. If you are using the command line interface, specify Ethernet or FC in the **`port_type`** parameter of the **`mkportset`** command.
- The FC I/O ports are added to the portset by using FC I/O Port ID. It is applicable for all the nodes in the system.
- An FC port can be associated with more than one portset.

At the time of upgrade or during a new system installation, all FC I/O Ports in the system are added to the default FC portset. When the system detects a new FC I/O Port, the FC I/O ports in the system are added to the default FC portset (`portset64`). You can change an FC I/O port from the default portset to another portset.

This process includes the following steps:

1. Remove the FC I/O Port from the default portset.
2. Create a portset.
3. Add the FC I/O port to the new portset.

An FC portset can be defined to an ownership group. When you define an ownership group for portsets, you can limit and restrict users to view and manage only specific portsets. The *`Portset 64`*, *`Portset 3`*, and replication portset are always globally owned and only global administrators can assign and modify FC I/O port IDs to the portsets.

In a typical configuration, an FC portset object is created first, and then the FC port and host object are created and configured with an FC portset. Multiple FC portsets that are associated with different hosts can be used to separate ports for serving separate functional use cases.

For example, separate FC portsets can be created for connecting to SCSI hosts and NVMeF hosts. Similarly, separate FC portsets can be created to group ports that provide different network speeds. The system restricts a host from getting access to the FC port if the host is not configured with the portset. The `lsfabric` command lists the restricted login details.

2.7 Fibre Channel SAN configuration planning

Each node canister can be equipped with one, two, or three 4-port 16 Gbps adapters, and 1 - 6 32 Gbps FC adapters (the maximum number of adapters depends on the system hardware type) that are used for SCSI FC and FC-NVMe attachment.

Note: It is recommended that a minimum of two adapters of any type be installed per node canister to use the CPU cores for optimal bandwidth and performance. This configuration allows for the separation of ports for dedicated traffic use, but is not applicable with the FlashSystem 5200.

2.7.1 Physical topology

The switch configuration for a fabric must comply with the switch manufacturer's configuration rules, which can impose restrictions. For example, a switch manufacturer might limit the number of supported switches or ports in a SAN fabric. Operating outside of the switch manufacturer's rules is not supported.

In an environment where you have a fabric with mixed port speeds (8 Gb, 16 Gb, and 32 Gb), the best practice is to connect the system to the switch operating at the highest speed.

The connections between the system's enclosures (node-to-node traffic) and between a system and the virtualized back-end storage require the best available bandwidth. For optimal performance and reliability, ensure that paths between the system nodes and storage systems do not cross inter-switch links (ISLs). If you use ISLs on these paths, make sure that sufficient bandwidth is available. SAN monitoring is required to identify faulty ISLs.

No more than three ISL hops are permitted among nodes that are in the same system but in different I/O groups. If your configuration requires more than three ISL hops for nodes that are in the same system but in different I/O groups, contact your IBM Support Center.

Direct connection of the system FC ports to host systems or between nodes in the system without the use of an FC switch is supported. For more information, see the IBM Documentation information that is relevant to your IBM Storage Virtualize platform. For example, for IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Planning** → **Planning for connectivity** → **Planning for a direct-attached configuration**.

For more information about the planning and topology requirements for HyperSwap configurations, see the following publications:

- *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597

- *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317

For more information about planning and topology requirements for 3-site replication configurations, see *Spectrum Virtualize 3-Site Replication*, SG24-8504.

2.7.2 Zoning

A SAN fabric must have four distinct zone classes:

- Inter-node zones. For communication between nodes in the same system
- Storage zones. For communication between the system and back-end storage
- Host zones. For communication between the system and hosts
- Inter-system zones. For remote replication

Figure 2-1 shows the system zoning classes.

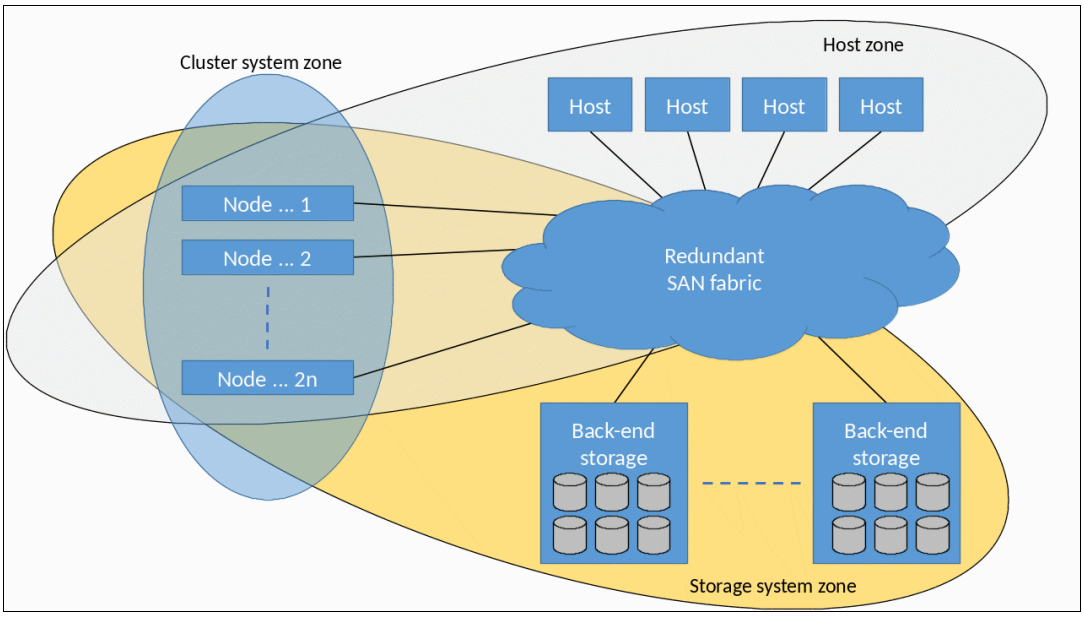


Figure 2-1 System zoning

The fundamental rules of system zoning are described in the following sections. However, you must review the latest zoning guidelines and requirements when designing zoning for the planned solution by reviewing the IBM Documentation information that is relevant to your IBM Storage Virtualize platform.

For example, for the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Configuring** → **Configuration details** → **SAN configuration and zoning rules summary**.

2.7.3 N_Port ID Virtualization

N_Port ID Virtualization (NPIV) is a method for virtualizing a physical FC port that is used for host I/O. By default, all new systems work in NPIV mode (the Target Port Mode attribute is set to Enabled).

NPIV mode creates a virtual worldwide port name (WWPN) for every physical system FC port. This WWPN is available for host connection only. During node maintenance, restart, or failure, the virtual WWPN from that node is transferred to the same port of the other node in the I/O group.

For more information about NPIV mode and how it works, see Chapter 7, “Hosts” on page 517.

Ensure that the FC switches give each physically connected system port the ability to create four more NPIV ports.

When performing zoning configuration, virtual WWPNs are used for host communication only; that is, “system to host” zones must include virtual WWPNs. Internode, intersystem, and back-end storage zones must use the WWPNs of physical ports. Ensure that equivalent ports (with the same port ID) are on the same fabric and in the same zone.

For more information about other host zoning requirements, see the IBM Documentation information that is relevant to your IBM Storage Virtualize platform.

For example, for the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Configuring** → **Configuration details** → **Zoning details** → **Zoning requirements for N_Port ID Virtualization**.

2.7.4 Inter-node zone

The purpose of intracluster or inter-node zones is to enable traffic between all node canisters within the clustered system. This traffic consists of heartbeats, cache synchronization, and other data that nodes must exchange to maintain a healthy cluster state.

For IBM FlashSystem products, traffic between nodes in one control enclosure is sent over a Peripheral Component Interconnect Express (PCIe) connection over an enclosure backplane. However, for redundancy, you must configure an inter-node SAN zone even if you have a single I/O group system. For a system with multiple I/O groups, all traffic between control enclosures must pass through a SAN.

For IBM SAN Volume Controller systems, a pair of nodes in an I/O group performs write-cache synchronization over the SAN. All delays in this SAN path directly affect system performance.

You can create up to two inter-node zones per fabric. Place a single port per node in each of them that is designated for intracluster traffic. Each node in the system must have at least two ports with paths to all other nodes in the system.

A system node cannot have more than 16 fabric paths to another node in the same system.

Mixed port speeds are not allowed for intracluster communication. All node ports within a clustered system must be running at the same speed.

2.7.5 Back-end storage zones

Create a separate zone for each back-end storage subsystem that is virtualized. Switch zones that contain back-end storage system ports must not have more than 40 ports. A configuration that exceeds 40 ports is *not* supported.

All nodes in a system must connect to the same set of back-end storage system ports on each device.

If the edge devices contain more stringent zoning requirements, follow the storage system rules to further restrict the system zoning rules.

Note: Cisco Smart Zoning and Brocade Peer Zoning are supported, with which you can insert target ports and multiple initiator ports in a single zone for easy of management but act the same as though each initiator and target are configured in isolated zones. The use of these zoning techniques is supported for host attachment and storage virtualization. As a best practice, use normal zones when configuring ports for clustering or for replication because these functions require the port to be an initiator and a target.

For more information about connecting back-end storage systems, see the IBM Documentation information that is relevant to your IBM Storage Virtualize platform. For example, for IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Configuring** → **Configuration details** → **External storage system configuration details (Fibre Channel)** and **Configuring** → **Configuring and servicing storage systems** → **External storage system configuration with Fibre Channel connections**.

2.7.6 Host zones

A host must be zoned to an I/O group to access volumes that are presented by this I/O group.

The preferred zoning policy is *single initiator zoning*. To implement it, create a separate zone for each host bus adapter (HBA) port, and place one port from each node in each I/O group that the host accesses in this zone. For deployments with more than 64 hosts that are defined in the system, this host zoning scheme must be used.

Note: Cisco Smart Zoning and Brocade Peer Zoning are supported, with which you can insert target ports and multiple initiator ports in a single zone for easy of management but act the same as though each initiator and target are configured in isolated zones. The use of these zoning techniques is supported for host attachment and storage virtualization. As a best practice, use normal zones when configuring ports for clustering or for replication because these functions require the port to be an initiator and a target.

For smaller installations, you can have up to 40 FC ports (including host HBA ports and the system's virtual WWPNs) in a host zone if the zone contains similar HBAs and operating systems. A valid zone can be 32 host ports plus eight system ports.

FC-NVMe applies the following limits to the host zone configuration:

- Zone up to four host ports to detect up to four ports on a node, and zone the same or more host ports to detect an extra four ports on the second node of the I/O group.
- Zone a total maximum of 16 hosts to detect a single I/O group.

Consider the following rules for zoning hosts over SCSI or FC-NVMe:

- For any volume, the number of paths through the SAN from the host to a system must not exceed eight. For most configurations, four paths to an I/O group are sufficient.

  In addition to zoning, you can use a port mask to control the number of host paths. For more information, see 3.3.7, "Configuring the local Fibre Channel port masking" on page 223.

- Balance the host load across the system's ports. For example, zone the first host with ports 1 and 3 of each node in the I/O group, zone the second host with ports 2 and 4, and so on. To obtain the best overall performance of the system, the load of each port must be equal. Assuming that a similar load is generated by each host, you can achieve this balance by zoning approximately the same number of host ports to each port.
- Spread the load across all system ports. Use all ports that are available on your machine.
- Balance the host load across HBA ports. If the host has more than one HBA port per fabric, zone each host port with a separate group of system ports.

All paths must be managed by the multipath driver on the host side. Make sure that the multipath driver on each server can handle the number of paths that is required to access all volumes that are mapped to the host.

2.7.7 Zoning considerations for Metro Mirror and Global Mirror

The SAN configurations that use inter-cluster Metro Mirror (MM) and Global Mirror (GM) relationships have the following extra switch zoning requirements:

- If two ISLs are connecting the sites, split the ports from each node between the ISLs; that is, one port from each node must be zoned across each ISL.
- Local-clustered system zoning continues to follow the standard requirement for all ports on all nodes in a clustered system to be zoned to one another.
- Review the latest requirements and recommendations in the IBM Documentation information that is relevant to your IBM Storage Virtualize platform. For example, for the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and select **Configuring** → **Configuration details** → **Zoning details** → **Zoning constraints for Metro Mirror and Global Mirror**.

When designing zoning for a geographically dispersed solution, consider the effect of the cross-site links on the performance of the local system.

The use of mixed port speeds for intercluster communication can lead to port congestion, which can negatively affect the performance and resiliency of the SAN. Therefore, it is not supported.

Note: If you limit the number of ports that are used for remote replication to two ports on each node, you can limit the effect of a severe and abrupt overload of the intercluster link on system operations.

If all node ports (N_Ports) are zoned for intercluster communication and the intercluster link becomes severely and abruptly overloaded, the local FC fabric can become congested so that no FC ports on the local system can perform local intracluster communication, which can result in cluster consistency disruption.

For more information about how to avoid such situations, see 2.7.8, "Port designation recommendations" on page 135.

For more information about zoning best practices, see *Performance and Best Practices Guide for IBM Storage FlashSystem and IBM SAN Volume Controller*, SG24-8543.

2.7.8 Port designation recommendations

If you have enough available FC ports on the system, designate different types of traffic to different ports. This configuration provides a level of protection against malfunctioning devices and workload spikes that might otherwise impact the system.

Note: It is not important which specific port is used for each kind of traffic if the general guidelines in this section are followed.

Intra-cluster communication must be protected because it is used for heartbeat and metadata exchange between all nodes of all I/O groups of the cluster.

In solutions with multiple I/O groups, upgrade nodes beyond the standard four FC port configuration. This upgrade provides an opportunity to dedicate ports to local node traffic, which separates them from other cluster traffic on the remaining ports.

Isolating remote replication traffic to dedicated ports is beneficial because it ensures that any problems that affect the cluster-to-cluster interconnect do not affect all ports on the local cluster.

It is recommended that two ports are used for remote copy.

For IBM Storage Virtualize systems with a total of 12 ports to isolate both node to node and system-to-system traffic, the example port designations that are shown in Figure 2-2 can be used. A similar pattern can be used in systems with more than 12 ports. however, the general guidelines must be met.

Card / Port	4 ports	8 ports	12 ports
Card 1 Port 1	Host/Storage/Inter-node	Host/Storage	Host/Storage
Card 1 Port 2	Host/Storage/Inter-node	Host/Storage	Host/Storage
Card 1 Port 3	Host/Storage/Replication*	Inter-node	Inter-node
Card 1 Port 4	Host/Storage/Replication*	Inter-node	Inter-node
Card 2 Port 1		Host/Storage	Host/Storage
Card 2 Port 2		Host/Storage	Host/Storage
Card 2 Port 3		Host/Storage/Replication*	Host/Storage/Replication*
Card 2 Port 4		Host/Storage/Replication*	Host/Storage/Replication*
Card 3 Port 1			Host/Storage
Card 3 Port 2			Host/Storage
Card 3 Port 3			Host/Storage
Card 3 Port 4			Host/Storage
localfcportmask	0011	00001100	000000001100
partnerfcportmask	1100	11000000	000011000000

* Use for host/storage when no replication is in place.

** Do not use the same port for replication and inter-node traffic.

*** For HyperSwap, dedicate ports for inter-node traffic.

Figure 2-2 Port masking configuration example for IBM FlashSystem

The example port designation that is shown in Figure 2-3 can be used for IBM SAN Volume Controller.

Card / Port	4 ports	8 ports	12 ports
Card 1 Port 1	Host/Storage/Inter-node	Host/Storage	Host/Storage
Card 1 Port 2	Host/Storage/Inter-node	Host/Storage	Host/Storage
Card 1 Port 3	Host/Storage/Replication*	Inter-node	Inter-node
Card 1 Port 4	Host/Storage/Replication*	Inter-node	Inter-node
Card 2 Port 1		Host/Storage	Host/Storage
Card 2 Port 2		Host/Storage	Host/Storage
Card 2 Port 3		Host/Storage/Replication*	Host/Storage/Replication*
Card 2 Port 4		Host/Storage/Replication*	Host/Storage/Replication*
Card 3 Port 1			Host/Storage
Card 3 Port 2			Host/Storage
Card 3 Port 3			Host/Storage
Card 3 Port 4			Host/Storage
localfcportmask	0011	00001100	000000001100
partnerfcportmask	1100	11000000	000011000000

* Use for host/storage when no replication is in place.

** Do not use the same port for replication and inter-node traffic.

*** For HyperSwap, dedicate ports for inter-node traffic.

Figure 2-3 Port masking configuration example for IBM SAN Volume Controller

When planning masking, consider the following examples:

- On an IBM FlashSystem system with a single control enclosure (I/O group) and without replication, no port dedication and masking are required. Inter-node traffic is sent over a backplane.
- In a HyperSwap system with two control enclosures or two I/O groups, dedicate ports for inter-node traffic and apply an FC mask.
- In a standard topology system with four I/O groups, the mask definition depends on the storage configuration, so more planning is required.

For more information about port masking, see Storage Virtualize Port Masking Explained.

To achieve traffic isolation, use a combination of SAN zoning and local and partner port masking. For more information about how to send port masks, see Chapter 3, “Initial configuration” on page 177.

Alternative port mappings that spread traffic across HBAs might allow adapters to come back online after a failure. However, they do not prevent a node from going offline temporarily to restart and attempt to isolate the failed adapter and then rejoin the cluster.

Also, the mean time between failures (MTBF) of the adapter is not significantly shorter than that of the non-redundant node components. The approach that is presented here accounts for all these considerations with the idea that increased complexity can lead to migration challenges in the future; therefore, a simpler approach is better.

The number of hosts must also be considered. Port masking is used to spread the host load across all ports, and limit the number of paths used per host. A host should have no more than two paths per node for best performance, and hosts should be evenly distributed across the available ports.

Note: For configurations with many hosts, it is important to make sure that sufficient adapters are installed while considering the number of logins that is allowed per port, which as of code level 8.6 is 512 logins per node.

2.8 IP SAN configuration planning

Starting with IBM Storage Virtualize code level 8.5.0, non-IBM provided cables and optics can be used. Therefore, customers who want to use their own cables and SFPs can do so if these components are equivalent in function and specification to the IBM-provided ones.

Depending on your IBM Storage Virtualize system, each IBM FlashSystem enclosure is equipped with two, four, or eight onboard 10 Gbps Ethernet network interface ports. IBM SAN Volume Controller 2145-SV2, and 2145-SA2 nodes are equipped with four on-board 10 Gbps Ethernet network interface ports. These ports can operate with link speeds of 1 Gbps and 10 Gbps. Any of these ports can be used for host I/O with the iSCSI protocol, external storage virtualization with iSCSI, and for native IP replication. Also, ports 1 and 2 can be used for managing the system. IBM SAN Volume Controller 2145-SV3 does not have any built-in ports that can be used for host I/O traffic.

Depending on your IBM Storage Virtualize system, each node can also be configured with up to six 2-port 25 Gbps RDMA-capable Ethernet adapters. These adapters can auto-negotiate link speeds of 1–25 Gbps. All their ports can be used for host attachment, external storage virtualization with iSCSI, node-to-node traffic, and IP replication. For host attachment, these 25 Gbps adapters support iSCSI, RDMA, and TCP-based connections. For external storage systems, only iSCSI connections are supported through these adapters. When the 25 Gbps adapter is installed on nodes in the system, RDMA technology can be used for node-to-node communications.

IBM SAN Volume Controller Model SV3, IBM FlashSystem 9500 and IBM FlashSystem 7300 can also support up to three 2-port 100 Gbps RDMA-capable Ethernet adapters per node. The 100 Gbps adapter supports iSCSI, NVMe over RDMA, and NVMe over TCP host attach protocols. For iSCSI, performance is limited to 25 Gbps per port.

Note: It is recommended that a minimum of two adapters of any type be installed per node canister to use the CPU cores for optimal bandwidth and performance. This configuration also allows for the separation of ports for dedicated traffic use, but it is not applicable to the FlashSystem 5200.

IBM Storage Virtualize systems support the 10 Gbps Finisar small form factor pluggable (SFP) (Finisar FTLX8574D3BCL) on the Mellanox and Chelsio 25 Gbps Ethernet adapters. The connections to the Ethernet adapter with 10 Gbps SFP parts are supported for host attachment, iSCSI back-end connectivity, and Ethernet clustering.

Note: At the time of this writing, only the 10 Gbps Finisar SFP is supported on the 25 GbE adapters. In all other instances, connecting a 10 Gbps switch to a 25 Gbps interface is supported through a SCORE request only. For more information, contact your IBM Sales representative or Business Partner.

You can set virtual local area network (VLAN) settings to separate network traffic for Ethernet transport. The system supports VLAN configurations for the system, host attachment, storage virtualization, and IP replication traffic. VLANs can be used with priority flow control (PFC) (IEEE 802.1Qbb).

All ports can be configured with an IPv4 address, an IPv6 address, or both. Each application of a port needs a separate IP. For example, port 1 of every node can be used for management, iSCSI, and IP replication, but three unique IP addresses are required.

If node Ethernet ports are connected to different isolated networks, a different subnet must be used for each network.

2.8.1 iSCSI and iSER protocols

The iSCSI protocol is a block-level access protocol that encapsulates SCSI commands into TCP/IP packets. Therefore, iSCSI uses an IP network rather than requiring the FC infrastructure.

The iSER is a network protocol that extends iSCSI to use RDMA. RDMA is provided by the Internet Wide Area RDMA Protocol (iWARP) or RDMA over Converged Ethernet (RoCE). It permits data to be transferred directly into and out of SCSI buffers, which provides faster connection and processing time than traditional iSCSI connections.

iSER requires 25 Gbps RDMA-capable Ethernet cards. RDMA links work only between RoCE ports or between iWARP ports: from a RoCE node canister port to a RoCE port on a host, or from an iWARP node canister port to an iWARP port on a host. Two types of 25 Gbps adapters are available for a system, and they cannot be interchanged without a similar RDMA type change on the host side.

iSCSI or iSER can be used for standard iSCSI communications for ports that do not use RDMA.

The 25 Gbps adapters include SFP28 fitted, which can be used to connect to switches that use OM3 optical cables.

For more information about the Ethernet switches and adapters that are supported by iSER adapters, see SSIC.

IBM Storage Virtualize systems support the 10 Gbps Finisar SFP (Finisar FTLX8574D3BCL) on the Mellanox and Chelsio 25 Gbps Ethernet adapters.

Note: At the time of this writing, only the 10 Gbps Finisar SFP is supported on the 25 Gbps Ethernet adapters. In all other instances, connecting a 10 Gbps switch to a 25 Gbps interface is supported only through a SCORE request. For more information, contact your IBM representative.

2.8.2 NVMe over RDMA (RoCE) and NVMe over TCP protocols

The Non-Volatile Memory express (NVMe) transport protocol provides enhanced performance on high-demand IBM Storage FlashSystem drives.

NVMe is a logical device interface specification for accessing non-volatile storage media. Host hardware and software use NVMe to fully use the levels of parallelism possible in modern solid-state drives (SSDs).

Depending on the host bus adapter (HBA) support within your storage systems, you can use NVMe over Fibre Channel, NVMe over RDMA (RoCE), or NVMe over TCP protocol. For more information about the adapters that are supported, see IBM System Storage Interoperation Center SSIC.

RDMA data transfer uses specialized network switches and requires less resources than FC-NVMe. RDMA allows higher throughput and better performance with lower latency. In addition, RDMA requires less expertise at the storage networking level than the Fibre Channel implementation, potentially reducing overall costs.

The advantage of using TCP data transfer is that unlike RDMA data transfer, TCP uses the existing Ethernet adapters on the host and network infrastructure.

Every physical Ethernet port supports four virtual ports: one for SCSI host connectivity, one for RDMA or TCP host connectivity, one for SCSI host failover, and one for RDMA or TCP host failover. Every NVMe virtual port supports the functions of NVMe discovery controllers and NVMe I/O controllers. Hosts create associations (NVMe logins) to the discovery controllers to discover volumes or to I/O controllers to complete I/O operations on NVMe volumes. Up to 128 discovery associations are allowed per port, not including N_Port ID virtualization (NPIV) failover. The number of ports per node depend on your storage system configuration.

When you use an Ethernet-based connection, you must consider the Ethernet protocol limitations:

- The behavior of a host that supports both Fibre Channel and Ethernet-based connections and accesses a single volume can be unpredictable and depends on the multi-pathing software.
- A maximum of four sessions can come from one iSCSI or NVMe initiator to an Ethernet-based target.

The following IP characteristics are for NVMe over RDMA and NVMe over TCP:

- IP addresses are used to discover storage volumes and to access the volumes by using I/O commands.
- Each node Ethernet port can be configured on the same subnet with the same gateway, or you can have each Ethernet port on separate subnets and use different gateways.
- If you are configuring a system to use node Ethernet ports 1 and 2 for NVMe over RDMA and NVMe over TCP I/O, ensure that the overall configuration also meets the system IP requirements that are listed previously.
- To ensure IP failover for NVMe over RDMA and NVMe over TCP operations, nodes in the same I/O group must be connected to the same physical segments on the same node ports. However, you can configure node Ethernet ports in different I/O groups to use different subnets and different gateways.

2.8.3 Priority flow control

Priority flow control (PFC) is an Ethernet protocol that you can use to select the priority of different types of traffic within the network. With PFC, administrators can reduce network congestion by slowing or pausing specific classes of traffic on ports, which provides better bandwidth for more important traffic. The system supports PFC on various supported Ethernet-based protocols on three types of traffic classes: system (inter-node), host attachment, and back-end storage traffic.

Note: PFC is not supported when you use NVMe over TCP connections.

You can configure a priority tag for each of these traffic classes. The priority tag can be any value 0–7. You can set identical or different priority tag values to all these traffic classes. You can also set bandwidth limits to ensure quality of service (QoS) for these traffic classes by using the Enhanced Transmission Selection (ETS) setting in the network.

To use PFC and ETS, ensure that the following tasks are completed:

- Configure a VLAN on the system to use PFC capabilities for the configured IP version.
- Ensure that the same VLAN settings are configured on all the entities, including all switches between the communicating end points.
- On the switch, enable Data Center Bridging Exchange (DCBx). DCBx enables switch and adapter ports to exchange parameters that describe traffic classes and PFC capabilities. For more information about these steps, see your switch documentation.
- For each supported traffic class, configure the same priority tag on the switch. For example, if you plan to have a priority tag setting of 3 for storage traffic, ensure that the priority is also set to 3 on the switch for that traffic type.
- If you plan to use the same port for different types of traffic, ensure that ETS settings are configured in the network.

NVMe over RDMA storage controllers support the use of Priority Flow Control (PFC) with ROCE v2 transport. Differentiated Services Code Point (DSCP) tagging is used for implementing the priority flow control. All RDMA ports are configured in DSCP trust mode and have a single PFC priority 3 enabled. The storage controller tags all outgoing frames with the DSCP value that was obtained from the initiator during connection establishment. To achieve end-to-end PFC configuration, initiators must use the Type of Service (TOS) value of 106 in RDMA connection requests, and you must configure Ethernet host ports and Ethernet switch ports to match this value (DSCP tag 26) to PFC class 3.

For more information, see the IBM Documentation information that is relevant to your IBM Storage Virtualize platform. For example, for the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Configuring** → **Configuring priority flow control**.

2.8.4 RDMA clustering

An IBM Storage Virtualize system can use 25 Gbps adapter cards for node-to-node traffic. A dual-site HyperSwap configuration can also use the cards for an inter-site link.

Note: The 100-Gbps Ethernet port does not support node-to-node communication.

A minimum of two dedicated RDMA-capable ports are required for node-to-node RDMA communications to ensure best performance and reliability. These ports must be configured for inter-node traffic only and cannot be used for host attachment, virtualization of Ethernet-attached external storage, or IP replication traffic.

Note: RDMA clustering is not supported on IBM FlashSystem 5015, IBM FlashSystem 5035, or IBM FlashSystem 5045.

The following limitations apply to a configuration of ports that are used for RDMA-clustering:

- Only IPv4 addresses are supported.
- Only the default value of 1500 is supported for the maximum transmission unit (MTU).

- Port masking is not supported on RDMA-capable Ethernet ports. Because of this limitation, do not exceed the maximum of four ports for node-to-node communications.
- For IBM SAN Volume Controller systems, hot-spare nodes are not supported on systems that use RDMA-capable Ethernet ports for node-to-node communications.
- Node-to-node communications that use RDMA-capable Ethernet ports are not supported in a network configuration that contains more than two hops in the fabric of switches.
- Some environments might not include a stretched layer 2 subnet. In such scenarios, a layer 3 network, such as in standard topologies or long-distance RDMA node-to-node HyperSwap configurations, is applicable.

  To support the layer 3 Ethernet network, the unicast discovery method can be used for RDMA node-to-node communication. This method relies on unicast-based fabric discovery rather than multicast discovery. To configure unicast discovery, see the man pages for the **`addnodediscoverysubnet`**, **`rmnodediscoverysubnet,`** or **`lsnodediscoverysubnet`** commands.

For more information, see the IBM Documentation information that is relevant to your IBM Storage Virtualize platform. For example, for the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Configuring** → **Configuration details** → **Configuration details for using RDMA-capable Ethernet ports for node-to-node communications**.

Note: Before you configure a system that uses RDMA-capable Ethernet ports for node-to-node communications in a standard or HyperSwap topology system, contact your IBM representative.

2.8.5 iSCSI back-end storage attachment

An IBM Storage Virtualize system supports the virtualization of external storage systems that are attached through iSCSI. Onboard 10 Gbps Ethernet ports or optional 25 Gbps Ethernet ports can be used. The 25 GbE network interface controllers (NICs) work in plain iSCSI mode without the use of any RDMA capabilities.

Consider the following points when planning for iSCSI virtualization:

- Unlike Fibre Channel connections, you must manually configure the connections between the source system and the target external storage systems.
- Discovery of the target IP addresses and session establishment is manually configured with either I/O group-wide (nodes from a specified I/O group discover and establish sessions) or system-wide (all nodes discover and establish sessions).
- After successful discovery, initiator sessions must be established from the source system to the target storage system manually through nodes in a specific I/O group or all nodes across the system.
- A one-to-one mapping of source ports to target ports is required. The same numbered ports on each source system or node connect to only one port on each target system or node. Multiple initiator ports on the same system node cannot connect to the same target port (many-to-one connectivity). Similarly, one initiator port cannot be connected to multiple target ports on the same target storage system (one-to-many connectivity). The CLI command fails if such an attempt is made.
- Zoning is not required.
- Direct attachment between the system and external storage systems is not supported. It requires Ethernet switches between the system and the external storage.

- To avoid a SPOF, a dual-switch configuration is recommended. For full redundancy, a minimum of two paths between each initiator node and target node must be configured with each path going through a separate switch.
- Extra paths can be configured to increase throughput if initiator and target nodes support more ports.

For more information about planning and implementing external storage virtualization with iSCSI, see *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

2.8.6 IP network host attachment

You can attach your IBM Storage Virtualize system to iSCSI, iSER, NVMe over RDMA, and NVMe over TCP hosts by using the Ethernet ports of the system.

The same ports can be used for iSCSI and iSER host attachment concurrently; however, a single host can establish an iSCSI or session, but not both.

Note: IBM SAN Volume Controller Model SV3, IBM FlashSystem FS9500, and IBM FlashSystem 7300 do not support ISER host attachment.

Hosts connect to the system through IP addresses, which are assigned to the Ethernet ports of the node. If the node fails, the address becomes unavailable and the host loses communication with the system through that node.

To allow hosts to maintain access to data, the node-port IP addresses for the failed node are transferred to the partner node in the I/O group. The partner node handles requests for its own node-port IP addresses and for node-port IP addresses on the failed node. This process is known as *node-port IP failover*.

In addition to node-port IP addresses, the iSCSI name and iSCSI alias for the failed node are transferred to the partner node. After the failed node recovers, the node-port IP address and the iSCSI name and alias are returned to the original node.

Note: The cluster name and node name form parts of the iSCSI name. Changing any of them requires reconfiguration all iSCSI hosts that communicate with the system.

With IBM Storage Virtualize v8.6.0, each system I/O group can have up to 1024 iSCSI IQNs.

iSER supports only one-way authentication through the Challenge Handshake Authentication Protocol (CHAP). iSCSI supports two types of CHAP authentication: one-way authentication (iSCSI target authenticating iSCSI initiators) and two-way (mutual) authentication (iSCSI target authenticating iSCSI initiators, and vice versa).

With the introduction of portsets, each physical Ethernet Port can have a maximum of 64 IP addresses with each IP on a unique portset. However, for each port, an IP address can be shared between multiple unique portsets for different functions.

Each port can bind to only a single IP address per portset for specific Ethernet functions, such as host attachment.

A host can access storage only by using IP addresses that are contained in the portset that is mapped to the host. Multiple hosts can be mapped to a single portset, but not vice versa. An IP address can belong to multiple portsets.

Consider the following points about host portsets:

- Portsets can have a maximum of four IP addresses per node.
- A single portset can contain IPv4, IPv6, or a mix of IPv4 and IPv6 addresses.
- For a host to log in to nodes on the system, the host must be mapped to a portset that contains at least one IP address from any of the nodes on that system.

When you configure NVMe over RDMA and NVMe over TCP host connections, attach the system to NVMe over RDMA or NVMe over TCP hosts by using the Ethernet ports on the systems.

NVMe over RDMA and NVMe over TCP connections route from hosts to the systems over the LAN. Each NVMe controller can request up to eight I/O queues for NVMe over RDMA or NVMe over TCP hosts.

For a full list of configuration and limitation rules for the IBM FlashSystem 9500 for example, see IBM FlashSystem 9500 Configuration Limits and Restrictions.

For general Ethernet port protocol limitations, see General Ethernet port configuration for host connections.

NVMe over RDMA and NVMe over TCP hosts support both the native multipath driver and a multipath daemon, if it is supported by the operating system.

For more information about iSCSI host attachment, see *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

Use SSIC to ensure that iSCSI initiators, host adapters, and Ethernet switches that are attached to the system are supported.

2.8.7 Native IP replication

Two systems can be linked over native IP links that are connected directly or by Ethernet switches to perform Remote Copy functions. Remote Copy over native IP provides a less expensive alternative to the use of FC configurations.

IP replication is supported on onboard 10 Gbps Ethernet ports and optional 25 Gbps Ethernet ports. However, when configured over 25 Gbps ports, it does not use RDMA capabilities. It also does not provide a performance improvement when compared to 10 Gbps ports.

As a best practice, use a different port for iSCSI host I/O and IP partnership traffic. Also, use a different VLAN ID for iSCSI host I/O and IP partnership traffic.

The portset feature can be used, which allows for multiple IP partnerships per IBM Storage Virtualize system. You can create IP partnerships at most between four systems.

A system can be part of three IP partnerships. IBM Storage Virtualize systems running versions earlier than 8.4.2.0 are still limited to one IP partnership. Partnerships on low-memory platform nodes share memory resources, which can lead to degraded performance.

Consider the following points about replication port sets:

- Replication portsets can have a maximum of 1 IP address per node.
- All IP addresses in replication port sets must be IPv4 or IPv6 addresses. You cannot mix IP protocol versions on replication port sets.

- Each IP partnership can be mapped to two port sets: one for each link between systems. For a partnership with a single link, a single portset can be defined in the `portset1` field on the Create Partnership page in the GUI.

  You can also use the **`-link1`** attribute in the **`mkippartnership`** command for partnerships with a single link to specify one of the portsets. For a partnership with dual links, a second portset must be defined in the `portset2` field in the GUI. Use the **`-link2`** attribute to specify the second portset for a dual-link configuration.
- Portsets replace the requirement for creating remote-copy groups for IP partnerships. Dedicated port sets can be created for remote copy traffic. The dedicated port sets provide a group of IP addresses for IP Partnerships.
- Each node can have one IP address that is assigned to a portset for remote-copy traffic. If the local system in the IP partnership contains four nodes, a portset can be created that defines four IP addresses (one per each node). Similarly, for the remote system with four nodes, a portset on that system can also have four IP addresses to handle remote-copy traffic exclusively.
- During updates of the software, any IP addresses that are assigned to remote copy groups with an IP partnership are automatically moved to a corresponding portset.

  For example, if remote-copy group 1 is defined on the system before the update, IP addresses from that remote-copy group are mapped to portset 1 after the update. Similarly, the IP address in remote-copy group 2 is mapped to portset 2.

Specific intersite link requirements must be met when you are planning to use IP partnership for RC. These requirements are described in the IBM Documentation that is relevant to your IBM Storage Virtualize platform.

For example, for the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and select **Configuring** → **IP partnership configuration** → **Intersite link planning**. Also, see Chapter 10, “Advanced Copy Services” on page 717.

2.8.8 Firewall planning

After your work to plan your IP network is complete, set up the suitable firewall rules for each data flow.

For a list of required and optional network flows for operating, see the IBM Documentation information relevant to your IBM Storage Virtualize platform. For example, for information that is related to the IBM FlashSystem 9500, see IBM FlashSystem 9500 documentation and expand **Planning** → **Planning for hardware** → **Physical installation planning** → **IP address allocation and usage**.

2.9 Planning topology

This section describes the planning topology.

2.9.1 High availability

IBM Storage Virtualize systems support two dual-site topologies: Standard topology, which includes synchronous or asynchronous replication, and HyperSwap. IBM SAN Volume Controller systems also support Enhanced Stretched Cluster (ESC). The key attributes of HyperSwap and ESC are listed in Table 2-3 on page 145.

Table 2-3 HyperSwap attributes

Item	HyperSwap	Enhanced stretched cluster
Minimum number of I/O groups that are required	2	1
Independent copies of data that are maintained	2 (four if volume mirroring to two pools in each site is configured	2
Cache that is retained if only one site is online	Yes	No
Stale consistent data is retained during resynchronization for DR	Yes	No
Ability to use MM, GM, or GM together with an HA solution	Yes	Yes
Maximum HA volume count	2000	7932
Licensing	Requires a Remote Mirroring license	Included in the base product

The HyperSwap topology uses extra system resources to support a full independent cache on each site, which enables full performance even if one site is lost.

The topologies differ in how the nodes are distributed across the sites. Consider the following points:

- For each I/O group in the system, the “stretched” topology has one node on one site and one node on the other site. The topology works with any number of I/O groups of 1 - 4.
- The “HyperSwap” topology places both nodes of an I/O group at the same site. Therefore, to get a volume resiliently stored at both sites, at least two I/O groups are required.

The ESC topology uses fewer system resources, which enable more HA volumes to be configured. However, during a disaster that makes one site unavailable, the IBM SAN Volume Controller system cache on the nodes of the surviving site are disabled. The HyperSwap topology uses extra system resources to support a full independent cache on each site, which enables full performance even if one site is lost.

For more information, see the following publications:

- *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597
- *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317
- *IBM SAN Volume Controller Stretched Cluster with PowerVM and PowerHA*, SG24-8142 *IBM Spectrum Virtualize and SAN Volume Controller Enhanced Stretched Cluster with VMware*, SG24-8211

2.9.2 Volume Mirroring

The information in Figure 2-4 on page 146 compares the various methods that you can use to mirror your volumes.

Description	Local volume mirroring	Metro Mirror	Global Mirror	HyperSwap®
Number of sites that host application data	1	2[1]	2[1]	2
Support multipathing failover between mirrored copies with no application impact	Yes	No	No	Yes
Maximum distance between copies	Within data center	Up to 300 km[2]	Up to 25000 km[2]	Up to 300 km[2]
Host completion that is delayed by secondary write operation	Yes	Yes	No	Yes
Bandwidth that is required on inter-site link	Not applicable	Maximum peak write bandwidth	Varies[3]	Maximum peak write bandwidth

[1] Multiple partnerships permit configurations with three or four sites.

[2] Subject to application and other constraints. See also the topic Long-distance links for Metro Mirror and Global Mirror partnerships.

[3] For more information, see the information about copy types in the topic Metro Mirror and Global Mirror relationships.

Figure 2-4 Volume Mirroring Comparison

2.9.3 Three-site replication

IBM Storage Virtualize systems support a 3-site replication topology, which includes 3-site replication with HyperSwap or 3-site replication with Metro Mirror (MM) configurations.

With the 3-site replication topology, data is replicated from the primary site or production site to two alternative sites. This feature ensures that if a disaster situation occurs at any one of the sites, the remaining two sites can establish a consistent replication operation with minimal data transfer. The Remote Copy (RC) relationships are synchronous or asynchronous, depending on which site failed.

The 3-site replication topology places three I/O groups at three different sites. It can ensure the availability of a minimum of two copies of data.

Note: Ensure that the planned configuration is reviewed by IBM or an IBM Business Partner before implementation. Such a review can increase the quality of the final solution and prevent configuration errors that might affect solution delivery.

For more information, see *Spectrum Virtualize 3-Site Replication*, SG24-8504.

2.10 Back-end storage configuration

External back-end storage systems, also known as controllers, provide their logical volumes (LUs). These LUs are detected by an IBM Storage Virtualize system as managed disks (MDisks) and can be used in storage pools to provision their capacity to system hosts.

Note: IBM FlashSystem 5015, IBM FlashSystem 5035, and IBM FlashSystem 5045 support external virtualization for migration purposes only.

The back-end storage subsystem configuration must be planned for all external storage systems that are attached. Apply the following general guidelines:

- Most of the supported FC-attached storage controllers must be connected through an FC SAN switch. However, a limited number of systems (including IBM FlashSystem 5000, IBM FlashSystem 7000, and IBM FlashSystem 9000 family) can be direct-attached by using FC.

- For migration purposes, any supported back-end storage systems can be attached by way of FC direct attachment.
- Connect all back-end storage ports to the SAN switch up to a maximum of 16 and zone them to all of the system to maximize bandwidth. The system is designed to handle many paths to the back-end storage.
- The cluster can be connected to a maximum of 1024 worldwide node names (WWNNs). The following general practices are used:
  - EMC DMX/SYMM, all HDS, and SUN/HP HDS clones use one worldwide node name (WWNN) per port. Each port appears as a separate controller to the system.
  - IBM, EMC CLARiiON, and HP use one WWNN per subsystem. Each port appears as a part of a subsystem with multiple ports, with up to a maximum of 16 ports (WWPNs) per WWNN.

  However, if you plan for a configuration that might be limited by the WWNN maximum, verify the WWNN versus WWPN policy with the back-end storage vendor.
- When defining a controller configuration, avoid hybrid configurations and automated tiering solutions. Create LUs for provisioning to the system from a homogeneous disk arrays or solid-state drive (SSD) arrays.
- Do not provision all available drives on the back-end storage capacity as a single LU. A best practice is to create one LU for eight hard disk drives (HDDs) or SSDs for the back-end system.
- If your back-end storage system is not supported by the round-robin path policy, ensure that the number of MDisks per storage pool is a multiple of the number of storage ports that are available. This approach ensures sufficient bandwidth for the storage controller, and an even balance across storage controller ports.
- An IBM Storage Virtualize system must exclusively access every LU that is provisioned to it from a back-end controller. Any specific LU cannot be presented to more than one system. The same back-end LU cannot be presented to a system and host.
- Data reduction (compression and deduplication) on the back-end controller is supported only with a limited set of IBM Storage systems.

In general, configure back-end controllers as though they are used as stand-alone systems. However, specific requirements or limitations might exist as to the features that are usable in the specific back-end storage system.

For more information about the requirements that are specific to your back-end controller, see the IBM Documentation that is relevant to your IBM Storage Virtualize platform. For example, for the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Configuring** → **Configuring and servicing storage systems**.

The system's large cache and advanced cache management algorithms also allow it to improve the performance of many types of underlying disk technologies. Because hits to the cache can occur in the upper level (the system) and the lower level (back-end controller) of the overall solution, the solution as a whole can use the larger amount of cache wherever it is. Therefore, the system's cache also provides more performance benefits for back-end storage systems with extensive cache banks.

However, the system cannot increase the throughput potential of the underlying disks in all cases. The performance benefits depend on the underlying back-end storage technology and the workload characteristics, including the degree to which the workload exhibits hotspots or sensitivity to cache size or cache algorithms.

2.11 Internal storage configuration

This section describes the internal storage configuration.

2.11.1 IBM FlashSystem Systems

For general-purpose storage pools with various I/O application workloads, follow the storage configuration wizard recommendations in the GUI. For specific applications with known I/O patterns, use the CLI to create arrays that suit your needs.

Distributed redundant array of independent disks (DRAID) configurations create large-scale internal MDisks. Different types of DRAID array level options are available.

The following restrictions and recommendations are applicable to DRAID arrays:

- DRAID 5 can contain as few as four drives initially, and expanded up to and contain a maximum of 128 drives.
- DRAID 6 can contain as few as six drives initially, and expanded up to and contain a maximum of 128 drives.
- DRAID 1 can contain 2 - 6 drives initially, and can be expanded up to 16 drives of the same capacity.
- The DRAID array recommendation for all types of internal storage except storage-class memory (SCM) is DRAID 6, which outperforms other available RAID levels in most applications while providing fault tolerance and high rebuild speeds.
- Up to 24 Storage Class Memory (SCM) drives are supported in IBM FlashSystem enclosures. DRAID 1 is recommended for best performance.

In specific IBM FlashSystem configurations (for example, small SCM or flash arrays), DRAID 1 is suggested to allow for high I/O performance because of all member drive participation in the I/O and the optimized I/O path for multi-core CPUs. It also adds fast rebuilt times on smaller arrays because of the distributed rebuild area.

DRAID 1 is the only DRAID that can be configured without a rebuild area, supports arrays with a minimum of two member drives, and is limited to 16 member drives (after expansion). Initially, start with six or fewer member drives. Based on the anticipated capacity (current and future), consider whether to start with a DRAID 1 array or plan for a DRAID 6 array (which can expand even more).

DRAID 1 is recommended as the default in the following scenarios:

- Two member drives array with no rebuild area
- 3–6 member drives with one rebuild area

Important: Consider the following points:

- DRAID 1 is not recommended with two member drives (and no rebuild area) for HDDs of any size.
- DRAID 1 is not recommended with two member drives (and no rebuild area) for SSDs (SAS, FCM, or NVMe) that are larger than 20 TB of physical capacity.
- DRAID 1 is not recommended with two member drives (and no rebuild area) for SCMs that are larger than 8 TB of physical capacity.
- DRAID 1 is not recommended with 3 - 6 member drives (and one rebuild area) for HDDs that are larger than 8 TB of physical capacity.
- DRAID 1 supports only a single rebuild area per 3 - 16 member drives.
- DRAID 1 is supported only for pools with extent size of 1024 MiB or greater.
- DRAID 1 arrays with 128 KB strip size are not supported.

Because of their mirrored nature, DRAID 1 arrays can use only half of the array's capacity for data. DRAID 6 can achieve better capacity utilization ratios.

Figure 2-5 on page 150 shows planning guidance for the recommended DRAID configuration that is based on the number of array member drives.

Number of drives	RAID Config	Usable Capacity	I/O Amplification for 70:30 (lower number is better)
★ 2	DRAID-1 2+No Spare	50%	1.3 - Poor redundancy
★ 3	DRAID-1 2+S	33%	1.3 - Performance optimized
★ 4	DRAID-1 3+S	37%	1.3 - Performance optimized
▲ 4	DRAID-5 2+P+S	50%	1.9 - Capacity optimized
★ 5	DRAID-1 4+S	40%	1.3 - Performance optimized
▲ 5	DRAID-5 3+P+S	60%	1.9 - Capacity optimized
★ 6	DRAID-1 5+S	41%	1.3 - Performance optimized
▲ 6	DRAID-6 3+P+Q+S	50%	2.5 - Best redundancy
7	DRAID-1 6+S	42%	1.3 - Performance optimized
★ 7	DRAID-6 4+P+Q+S	57%	2.5 - Best redundancy
8	DRAID-1 7+S	43%	1.3 - Performance optimized
★ 8	DRAID-6 5+P+Q+S	62%	2.5 - Best redundancy

★ #1 recommended configuration
▲ #2 recommended configuration

Figure 2-5 Distributed RAID planning guidance

For more information about internal storage configuration, see *Performance and Best Practices Guide for IBM Storage FlashSystem and IBM SAN Volume Controller*, SG24-8543.

Supported array types and RAID levels

IBM FlashSystem systems support FCM NVMe drives, industry standard NVMe drives, SCM drives with NVMe architecture, and SAS drives that are within expansion enclosures. The type and level of arrays vary depending on the type of drives in the I/O group.

Table 2-4 summarizes the supported levels. For storage arrays with fewer than seven drives, DRAID 1 is recommended because it offers enhanced resiliency over DRAID 6 arrays. DRAID 6 is recommended for storage arrays with seven or more drives because it can handle two concurrent drive failures.

Table 2-4 Summary of supported drives, array types, and RAID levels

Drive type	Non-DRAIDs	DRAIDs		
	RAID 0/1/10	DRAID 1	DRAID 5	DRAID 6
Industry standard NVMe drives or SAS drives (expansion enclosure)	x	x	x	x
FCM NVMe drives		x	x	x
SCM	x	x	x	x

Table 2-5 summarizes the supported RAID levels on each hardware platform. Any restrictions here must also be considered when planning the storage configuration.

Table 2-5 Summary of supported hardware platforms and RAID levels

Hardware Platform	DRAIDs		
	DRAID 1	DRAID 5	DRAID 6
IBM FlashSystem 9500	x		x
IBM FlashSystem 7300	x	x[a]	x
IBM FlashSystem 5200	x	x	x
IBM FlashSystem 5045	x	x	x
IBM FlashSystem 5035	x	x	x
IBM FlashSystem 5015	x	x	x

a. Available by way of SCORE request only.

For more information about any further restrictions in relation to RAID and DRAID array configurations, see the configuration limits and restrictions for your IBM Storage Virtualize system:

- V8.6.0.x Configuration Limits and Restrictions for IBM System Storage SAN Volume Controller
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 9500
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 7200 and 7300
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 50x5 and 5200

2.12 Storage pool configuration

The storage pool is at the center of the many-to-many relationship between the internal drive arrays or externally virtualized LUNs, which are represented as MDisks, and the volumes. It acts as a container of physical disk capacity from which chunks of MDisk space (known as *extents*) are allocated to form volumes that are presented to hosts.

The system supports two types of pools: standard pools and Data Reduction Pools (DRPs). The type of pool is configured when a pool is created and it cannot be changed later. The pool type determines the set of features that is available on the system.

Consider the following points:

- A feature that can be implemented only with standard pools is VMware vSphere integration with VMware vSphere virtual volumes (VVOLs).
- The following features can be implemented only with DRPs:
  - Automatic capacity reclamation with SCSI UNMAP (this feature returns capacity to the storage pool that is marked as no longer used by a host)
  - DRP compression (in-flight data compression)
  - DRP deduplication
  - FlashCopy with redirect-on-write (RoW)

Note: FlashCopy with RoW is usable only for volumes with supported deduplication without mirroring relationships and within the same pool and I/O group. Automatic mode selection (RoW/copy-on-write [CoW]) is based on these conditions.

In addition to providing data reduction options, DRP amplifies the I/O and CPU workload, which should be accounted for during performance sizing and planning.

Also, self-compressing drives (FCM drives) still perform compression independently of the pool type.

IBM Storage Virtualize provides a *Comprestimation Always On* feature, where the continuous comprestimation of all volumes is provided so that compressibility estimations are always available. This feature is on by default.

Another base storage pool parameter is the extent size. A storage pool extent size includes the following implications:

- The maximum volume, MDisks, and managed storage capacity depend on the extent size. The bigger the extent that is defined for the specific pool, the larger is the maximum size of this pool, the maximum MDisk size in the pool, and the maximum size of a volume that is created in the pool.
- The volume sizes must be a multiple of the extent size of the pool in which the volume is defined. Therefore, the smaller the extent size, the better control that you have over the volume size.

The system supports extent sizes of 16 - 8192 mebibytes (MiB). The extent size is a property of the storage pool, and it is set when the storage pool is created.

Note: The base pool parameters, pool type, and extent size are set during pool creation and cannot be changed later. If you need to change the extent size or pool type, all volumes must be migrated from a storage pool and then the pool must be deleted and re-created.

When you create a DRP, ensure that the usable capacity of the pool includes overhead capacity. *Overhead capacity* is an amount of usable capacity that contains the metadata for tracking unmap and reclaim operations within the pool. A general guideline is to ensure that the provisioned capacity with the DRP does not exceed 85% of the total usable capacity of the DRP. Table 2-6 lists the minimum DRP capacity that is required to create a volume within the pool.

Table 2-6 Minimum overhead capacity requirements for DRPs

Extent size	Overhead capacity requirements[a]
1 GB or smaller	1.1 TB
2 GB	2.1 TB
4 GB	4.2 TB
8 GB	8.5 TB

a. Standard-provisioned volumes are not included into the minimum overhead capacity values. When you are planning usable capacity for DRPs, determine the usable capacity that is needed for any standard-provisioned volumes first. Then, ensure that the minimum usable capacity values for the DRPs are included.

For more information about the relationship between a system's maximum configuration and extent size, see the configuration limits and restrictions for your IBM Storage Virtualize system:

- V8.6.0.x Configuration Limits and Restrictions for IBM System Storage SAN Volume Controller
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 9500
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 7200 and 7300
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 50x5 and 5200

When planning pools, the encryption is defined on a pool level and the encryption setting cannot be changed after a pool is created. If you create an unencrypted pool, it cannot be encrypted later. Your only option is to delete it and re-create it as encrypted.

When planning storage pool layout, consider the following aspects:

- With Storage Virtualize 8.5.0 and later, it is not possible to create a second DRAID array that consists of compressing drives, for example, FCMs, if the target storage pool contains one or more DRAID arrays of compressing drives.
- Consider the following points about pool reliability, availability, and serviceability (RAS):
  - The storage pool is a failure domain. If one array or external MDisk is unavailable, the pool and all volumes in it goes offline.
  - The number and size of storage pools affects system availability. The use of a larger number of smaller pools reduces the failure domain if one of the pools goes offline. However, increasing the number of storage pools affects the storage use efficiency, and the number is subject to the configuration maximum limit.
  - You cannot migrate volumes between storage pools with different types or extent sizes. However, you can use volume mirroring to create copies between storage pools.

- Consider the following points about pool performance:
  - Do not mix same-tier arrays or MDisks with different performance characteristics in one pool. For example, do not use DRAID 6 arrays of six tier 1 SSDs and DRAID 6 arrays of 24 tier 1 SSDs in the same pool. This technique is the only way to ensure consistent performance characteristics of volumes that are created from the pool.

    Arrays with different tiers in one pool can be used because their performance differences become beneficial when you use the Easy Tier function.
  - Create multiple storage pools if you must isolate specific workloads to separate storage.
  - Ensure that performance sizing was done for selected pool type and feature set.

2.12.1 Child pools

Instead of being created directly from MDisks, child pools are created from usable capacity that is assigned to a parent pool. As with parent pools, volumes can be created that specifically use the usable capacity that is assigned to the child pool. Child pools are similar to parent pools with similar properties and can be used for volume copy operation.

When a standard child pool is created, the usable capacity for a child pool is reserved from the usable capacity of the parent pool. The usable capacity for the child pool must be smaller than the usable capacity in the parent pool. After the child pool is created, the amount of usable capacity that is specified for the child pool is no longer reported as usable capacity of its parent pool.

When a data reduction child pool is created, the usable capacity for the child pool is the entire usable capacity of the data reduction parent pool without limit. After a data reduction child pool is created, the usable capacity of the child pool and the usable capacity of the parent pool are reported as the same.

Several administration tasks benefit from defining and working with a part of a pool. For example, the system supports VVOLs, which are used in VMware vCenter and vSphere APIs for Storage Awareness (VASA) applications. Before a child pool can be used for virtual volumes for these applications, the system must be enabled for virtual volumes.

Note: DRPs are still not supported for VVols.

Consider the following general guidelines when you create or work with a child pool:

- The management GUI displays child pools in relationship to the parent pool with which they are associated.
- Child pools can be created and changed by using the CLI or GUI.
- When child pools are used with standard pools, you can specify a warning threshold that alerts you when the used capacity of the child pool is reaching its upper limit. Use this threshold to ensure that access is not lost when the used capacity of the child pool is close to its usable capacity.
- On systems with encryption enabled, you can create the standard child pools to migrate existing volumes in a non-encrypted pool to encrypted child pools only when you virtualize external storage. When you create a standard child pool after encryption is enabled, an encryption key is created for the child pool, even when the parent pool is not encrypted. You can then use volume mirroring to migrate the volumes from the nonencrypted parent pool to the encrypted child pool.

Encrypted data reduction child pools can be created only if the parent pool is encrypted. The data reduction child pool inherits an encryption key from the parent pool.

- Ensure that any child pools that are associated with a parent pool have enough usable capacity for the volumes that are in the child pool before removing MDisks from a parent pool. The system automatically migrates all extents that are used by volumes to other MDisks in the parent pool to ensure that data is not lost.
- You cannot shrink the usable capacity of a child pool below its used capacity. The system also resets the warning level when the child pool is shrunk and issues a warning if the level is reached when the usable capacity is shrunk.
- The system supports migrating a copy of volumes between child pools within the same parent pool or migrating a copy of a volume between a child pool and its parent pool. Migrations between a source and target child pool with different parent pools are *not* supported.

  However, you can migrate a copy of the volume from the source child pool to its parent pool. The volume copy can then be migrated from the parent pool to the parent pool of the target child pool. Finally, the volume copy can be migrated from the target parent pool to the target child pool.

Child pools can be assigned to an ownership group. An *ownership group* defines a subset of users and objects within the system. You can create ownership groups to further restrict access to specific resources that are defined in the ownership group. Only users with Security Administrator roles can configure and manage ownership groups.

Ownership can be defined explicitly or it can be inherited from the user, user group, or other parent resources, depending on the type of resource. Ownership of child pools must be assigned explicitly, and they do not inherit ownership from other parent resources. New or existing volumes that are defined in the child pool inherit the ownership group that is assigned for the child pool.

For more information about ownership groups, see 12.5, “Ownership groups principles of operations” on page 1083.

Child pools can be created for use with the Safeguarded Copy feature. The Safeguarded Copy feature supports the ability to create cyber-resilient point-in-time copies of volumes that cannot be changed or deleted through user errors, malicious actions, or ransomware attacks. With the Safeguarded Copy feature, child pools provide a Safeguarded backup location for a group of volumes that are associated with the parent pool.

The Safeguarded backup location can contain many snapshots of volume data, each created at a wanted interval and with the required retention period to satisfy your recovery point objective. After the Safeguarded backup location is created, you must create a volume group and assign a Safeguarded backup policy to the volume group.

Consider the following points when you create or work with a child pool for Safeguarded Copy:

- In the management GUI, child pools that are used as Safeguarded backup locations are marked with a shield icon.
- You can create and manage child pools in the management GUI. However, users with privileges of Administrator or less cannot change or delete any pools that contain Safeguarded backups.

2.12.2 Storage pool and cache relationship

The system uses cache partitioning to limit the potential negative effects that a poorly performing storage controller can have on the clustered system. The cache partition allocation size is based on the number of configured storage pools. This design protects against an individual overloaded back-end storage system from filling the system write cache and degrading the performance of the other storage pools.

Table 2-7 lists the limits of the write-cache data that can be used by a single storage pool.

Table 2-7 Limits of the cache data

Number of storage pools	Upper limit
1	100%
2	66%
3	40%
4	30%
5 or more	25%

No single partition can occupy more than its upper limit of write cache capacity. When the maximum cache size is allocated to the pool, the system starts to limit incoming write I/Os for volumes that are created from the storage pool. The host writes are limited to the destage rate on a one-out-one-in basis.

Only writes that target the affected storage pool are limited. The read I/O requests for the throttled pool continue to be serviced normally. However, because the system is offloading cache data at the maximum rate that the back-end storage can sustain, read response times are expected to be affected.

All I/O that is destined for other, non-throttled, storage pools continues as normal.

2.12.3 Provisioning policies

A *provisioning policy* is an entity that defines a set of rules for allocating capacity from a pool. Any volumes that are added to the pool adopt the capacity-saving methods that are defined in the provisioning policy. Use the CLI to create and manage provisioning policies.

When a storage pool is assigned with a provisioning policy, all volumes that are created in that pool by using the **mkvolume** command are automatically created with the capacity savings that are defined by the policy. Provisioning policies allow a simpler and consistent provisioning process that can be used with external automation software.

Provisioning policies support the following capacity-saving options:

- None

  Use the **-capacitysaving none** option to indicate that volumes use only the built-in capacity-saving functions of the drives in the pool, if available. For example, specific drive types, such as FlashCore Module drives, are self-compressing. Capacity savings are not applied if this option is used with drives that do not support any built-in capacity savings.

- Thin-provisioning

  Use the **-capacitysaving thin** option to indicate that all volumes in a pool use thin-provisioning to save capacity. Thin provisioning allocates capacity only when data is written to storage.

- Compression

  Use the **-capacitysaving compressed** option to indicate that all volumes in a pool use compression to save capacity. Compression reduces the capacity that data requires by removing unnecessary content from the data.

- Deduplication

  Compression and thin-provisioning capacity saving can be used with deduplication to increase capacity savings for DRPs. Standard pools do not support deduplication. Use the **-deduplicated** option when you create a provisioning policy to add deduplication to your capacity saving for a DRP.

A provisioning policy can be renamed after the policy is created. However, the capacity savings option of a provisioning policy cannot be changed. To change the capacity savings to be used for new volumes, create and assign a new provisioning policy to the storage pool, then delete the old policy if it is no longer required. Also, the capacity savings for a volume can be changed after a volume is created. A provisioning policy for a storage pool is only applied when new volumes are created.

A provisioning policy can be associated with multiple pools. However, it is not possible to add multiple policies to one pool.

2.13 Volume configuration

When planning a volume, consider the required performance, availability, and capacity. Every volume is assigned to an I/O group that defines which pair of system nodes services I/O requests to the volume.

Note: No fixed relationship exists between I/O groups and storage pools.

When a host sends I/O to a volume, it can access the volume with either of the nodes in the I/O group, but each volume has a *preferred node*. Many of the multipathing driver implementations that the system supports use this information to direct I/O to the preferred node. The other node in the I/O group is used only if the preferred node is not accessible.

During volume creation, the system selects the node in the I/O group that has the fewest volumes to be the preferred node. After the preferred node is chosen, it can be changed manually, if required.

Strive to distribute volumes evenly across available I/O groups and nodes within the system.

For more information about volume types, see Chapter 6, “Volumes” on page 419.

2.13.1 Planning for image mode volumes

Use image mode volumes to present to hosts data that is written to the back-end storage before it was virtualized. An image mode volume directly corresponds to the MDisk from which it is created.

Image mode volumes are a useful tool in storage migration and during system implementation to a working environment.

2.13.2 Planning for standard-provisioned volumes

A standard-provisioned volume, also called a *fully allocated volume*, presents the same capacity to mapped hosts that the volume uses in the storage pool. No data reduction is performed on a pool level. However, if a standard-provisioned volume is provisioned from a pool with data reducing storage, such as self-compressing drives (FCM drives), the data is still compressed on a drive level.

Standard-provisioned volumes provide the best performance because they do not cause I/O amplification. They also require less CPU time compared to other volume types.

2.13.3 Planning for thin-provisioned volumes

A thin-provisioned volume presents a different capacity to mapped hosts than the capacity that the volume uses in the storage pool. Space is not allocated on a thin-provisioned volume if an incoming host write operation contains all zeros.

By using the thin-provisioned volume feature that is called *zero detect*, you can reclaim unused allocated disk space (zeros) when you convert a fully allocated volume to a thin-provisioned volume by using volume mirroring.

When thin-provisioned volumes are used, the system must maintain extra metadata that describes the contents of thin-provisioned volumes. As a result, the I/O rates that are obtained from thin-provisioned volumes can be lower than the rates that are obtained from standard-provisioned volumes that are allocated on the same MDisks.

IBM Storage Virtualize systems support thin-provisioned volumes in standard pools and in DRPs.

DRPs enhance capacity efficiency for thin-provisioned volumes by monitoring the host's capacity usage. When the host indicates that the capacity is no longer needed, the capacity is released and can be reclaimed by the DRP to be redistributed automatically. Standard pools cannot reclaim capacity.

Note: Avoid the use of thin-provisioned volumes on a data-reducing backend (such as self-compressing drives) when DRPs are implemented.

2.13.4 Planning for compressed volumes

With compressed volumes, data is compressed as it is written to disk, which saves more space. When data is read to hosts, the data is decompressed.

Compression is available through data reduction support as part of the system. If you want volumes to use compression as part of data reduction support, which is compressed, volumes must belong to DRPs.

If you use compressed volumes over a pool with self-compressing drives, the drive still attempts compression because it cannot be disabled on the drive level. However, performance is not affected because of the algorithms that FCM drives use to manage compression.

Before implementing compressed volumes, analyze the data to discover your average compression ratio and ensure that performance sizing was done for compression.

IBM Storage Virtualize provides a Comprestimation Always On feature, which ensures the continuous comprestimation of all VDisks is provided and always available. This feature is enabled by default.

Special considerations must be taken when implementing compression on IBM FlashSystem 5035 and IBM FlashSystem 5045, which does not have compression accelerator hardware and uses the node canister's CPU for compression and decompression. Therefore, strict performance planning and sizing is required.

Note: If you use compressed volumes over FCM drives, the compression ratio on a drive level must be assumed to be 1:1 to avoid array overprovisioning and running out of space.

2.13.5 Planning for deduplicated volumes

Deduplication can be configured for volumes that use different capacity-saving methods, such as thin provisioning. Deduplicated volumes must be created in DRPs for added capacity savings. Deduplication is a type of data reduction that eliminates duplicate copies of data. Deduplication of user data occurs within a DRP and only between volumes or volume copies that are marked as deduplicated.

With deduplication, the system identifies unique chunks of data that are called *signatures* to determine whether new data is written to the storage. Deduplication is a hash-based solution, which means chunks of data are compared to their signatures rather than to the data itself.

If the signature of the new data matches a signature that is stored on the system, the new data is replaced with a reference. The reference points to the stored data instead of writing the data to storage. This process saves the capacity of the back-end storage by not writing new data to storage. It might also improve the performance of read operations to data that has a signature.

The same data pattern can occur many times, and deduplication decreases the amount of data that must be stored on the system. A part of every hash-based deduplication solution is a repository that supports looking up matches for incoming data.

The system contains a database that maps the signature of the data to the volume and its virtual address. If an incoming write operation does not have a signature that is stored in the database, a duplicate is not detected and the incoming data is stored on back-end storage.

To maximize the space that is available for the database, the system distributes this repository between all nodes in the I/O groups that contain deduplicated volumes. Each node carries a distinct portion of the records that are stored in the database.

If nodes are removed or added to the system, the database is redistributed between the nodes to ensure full use of the available memory. Only specific models with specific hardware support deduplication. Verify whether your model and hardware components can use these functions.

When you create a volume, you can specify to include deduplication with other supported capacity savings methods. Deduplicated volumes must be created in DRPs. If you have volumes in standard pools, you can migrate them to DRPs to add deduplication to increase capacity savings for the volume.

Before implementing deduplication, analyze the data to estimate the deduplication savings and ensure that system performance sizing was done for deduplication.

You can use the Data Reduction Estimation Tool (DRET) to estimate how much capacity you might save if a standard volume that a host can access was a deduplicated volume. The tool scans target workloads on all attached storage arrays, consolidates these results, and generates an estimate of potential data reduction savings for the entire system. For more information, see IBM Data Reduction Estimator Tool (DRET) for SVC, Storwize and FlashSystem products and IBM FlashSystem Comprestimator.

Note: The DRET also provides some analysis of potential compression savings for volumes. However, use the management GUI or the CLI to run the integrated Comprestimator Utility to gather data for potential compression savings for volumes in DRPs.

2.13.6 Planning for Volume groups

A volume group is a container for managing a set of related volumes as a single object. The volume group provides consistency across all volumes in the group.

Volume groups can be used with the following functions:

Safeguarded Copy function

One implementation of volume groups is to group volumes to be configured as Safeguarded. Safeguarded copy function is a cyber-resiliency feature that creates immutable copies of data that cannot be changed or manipulated. A Safeguarded volume group describes a set of source volumes that can span different pools and are backed up collectively with the Safeguarded Copy function. Safeguarded snapshots are supported on the system through an internal scheduler that is defined in the snapshot policy or can be configured with an external snapshot scheduling application such as IBM Copy Services Manager.

Policy-based replication

You can use volume groups for policy-based replication. Policy-based replication is configured on all volumes in a volume group by assigning a replication policy to that volume group. The system automatically replicates the data and configuration for volumes in the group based on the values and settings in the replication policy. As part of policy-based replication, a recovery volume group is created automatically on the recovery system. Recovery volume groups cannot be created, changed, or deleted. A single replication policy can be assigned to multiple volume groups to simplify replication management. When additional volumes are added to the group, replication is automatically configured for these new volumes. Policy-based replication supports configuration changes while the partnership is disconnected. After the partnership is reconnected, the system automatically reconfigures the recovery system.

Snapshot function

Snapshots are the read-only point-in-time copies of a volume group that cannot be directly accessible from the hosts. To access the snapshot contents, you can create a clone or thin clone of a volume group snapshot. You can use the command line interface or management GUI to configure volume groups to use snapshot policies for multiple volumes for consistent management. Safeguarded snapshots with internal scheduler can be created by using the snapshot function. The snapshot function satisfies some of the use cases that are addressed by FlashCopy. Snapshots are a volume-centric approach that simplifies the overall management of point-in-time copies on the system. With snapshots, administrators can save

a mutually consistent image of the volumes in the volume group or save a snapshot of an individual volume. The main use case of snapshot is corruption protection. It protects the user data from deliberate or accidental data corruption from the host systems. Snapshot function can be used in parallel with the existing FlashCopy function. However, a snapshot is not compatible with FlashCopy function. When the volume has a snapshot, it cannot be part of any new FlashCopy mappings.

Snapshot policies

A snapshot policy is a set of rules that controls the creation, retention, and expiration of snapshots.

Administrators define snapshot policies to set up automatic backups for groups of storage volumes. They can choose how often backups are taken and for how long they are kept, which ensures data protection aligns with security and recovery goals. A snapshot policy has the following properties:

- It can be assigned to one or more volume groups.
- Only one snapshot policy can be scheduled to one volume group.
- The system supports a maximum number of 32 snapshot policies.

The system supports an internal scheduler to manage and create snapshot policies on the system. The management GUI supports selecting either a user-defined policy or a predefined policy and the user-defined policies can be created by using the management GUI or by using the **`mksnapshotpolicy`** command. Predefined policies contain specific retention and frequency values for common use-cases. Both predefined and user-defined policies are managed on the IBM Storage Virtualize system. The following predefined policies are supported:

- **Predefinedsspolicy0**

  Select this policy for the most frequent copies and retention. For this policy, snapshots are created daily and retained for a week. Use this policy for volume data that requires the highest recovery point objective (RPO). For example, volume data that is frequently updated and critical to your business can benefit from frequent copies and retention. Customer accounts, orders, or proprietary information are examples of data that can need more frequent backups. For more information, refer to your organization's business continuity plan.

- **Predefinedsspolicy1**

  Select this policy for less frequent copies and medium retention. For this policy, snapshots are created weekly and retained for a month. Use this policy for application data that is updated frequently and requires a high RPO, but might not contain business-critical data.

- **Predefinedsspolicy2**

  Select this policy for less frequent copies and longer retention. For this policy, snapshots are created monthly and retained for a year. Use this policy for older data that is not updated frequently but still requires retention, such as past customer accounts or employee records.

The volumes in a volume group are supposed to be mutually consistent. When a group of thin-clone volumes or clone volumes are populated, the snapshot function ensures that the images are mutually consistent. When volumes are added or removed from a group, the host applications ensure the volume groups are mutually consistent.

2.14 Host attachment planning

The system supports the attachment of a various host hardware types that are running different operating systems with FC SAN or IP SAN. For more information about instructions that are specific to your host setup, see the IBM Documentation information that is relevant to your IBM Storage Virtualize platform. For example, for the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Configuring** → **Host attachment**.

2.14.1 Queue depth

Typically, hosts issue subsequent I/O requests to storage systems without waiting for the completion of previous ones. The number of outstanding requests is called *queue depth*.

Sending multiple I/O requests in parallel (asynchronous I/O) provides significant performance benefits compared to sending them one-by-one (synchronous I/O). However, if the number of queued requests exceeds the maximum that is supported by the storage controller, you experience performance degradation.

For more information about how to calculate the correct host queue depth for your environment, see the IBM Documentation information that is relevant to your IBM Storage Virtualize and host platform, and connection protocol. For example, for the IBM FlashSystem 9500 related information, see IBM FlashSystem 9500 documentation and expand **Configuring** → **Host attachment**.

2.14.2 SAN boot support

The system supports SAN boot or startup for selected configurations of hosts running AIX, Linux, Microsoft Windows, and other operating systems. For more information about whether your configuration is supported for SAN boot, see the SSIC.

2.14.3 Planning for large deployments

Depending on your IBM Storage Virtualize system, each I/O group can have up to 2048 host objects defined. This limit is the same whether hosts are attached by using FC, iSCSI, or a combination of both. To allow more than 2048 hosts to access the storage, you must divide them into groups of 2048 hosts or less and map each group to a single I/O group only. With this approach, depending on your IBM Storage Virtualize system, you can configure up to 8192 host objects on a system with four I/O groups (eight nodes).

For best performance, split each host group into two sets. For each set, configure the preferred access node for volumes that are presented to the host set to one of the I/O group nodes. This approach helps to evenly distribute load between the I/O group nodes.

Note: A volume can be mapped to only a host that is associated with the I/O group to which the volume belongs.

For more information about a system's maximum configuration, see the suitable Configuration Limits and Restrictions link for your IBM Storage Virtualize system:

- V8.6.0.x Configuration Limits and Restrictions for IBM System Storage SAN Volume Controller
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 9500

- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 7200 and 7300
- V8.6.0.x Configuration Limits and Restrictions for IBM FlashSystem 50x5 and 5200

2.14.4 Planning for SCSI UNMAP

UNMAP is a set of SCSI primitives that hosts use to indicate to a SCSI target that space that is allocated to a range of blocks on a target storage volume is no longer required. With this command, the storage controller takes measures and optimizes the system so that the space can be reused for other purposes.

IBM Storage Virtualize systems support end-to-end UNMAP compatibility, which means that a command that is issued by a host is processed and sent to the back-end storage device or drive.

UNMAP processing can be controlled by using the following settings:

- The first setting advertises UNMAP support to hosts.
- The second setting controls whether IBM Storage Virtualize sends UNMAP commands to back-end storage (drives and external controllers).

Thorough planning is required if you want to use host UNMAP support. Enabling it allows you to fully benefit from capacity reclamation features in DRPs; however, host UNMAP requests might overload the IBM Storage Virtualize backend if it includes spinning disks, especially NL-SAS drives, which cause serious performance problems.

Back-end UNMAP is enabled by default on all IBM Storage Virtualize platforms. It is a best practice to keep it turned on for most use cases. Host UNMAP is off by default.

2.15 Planning copy services

IBM Storage Virtualize systems offer a set of copy services, such as IBM FlashCopy (snapshots) and Remote Copy (RC), in synchronous and asynchronous modes. For more information about copy services, see Chapter 10, “Advanced Copy Services” on page 717.

2.15.1 FlashCopy guidelines

With the FlashCopy function, you can perform a point-in-time (PiT) copy of one or more volumes. The FlashCopy function creates a PiT or time-zero (T0) copy of data that is stored on a source volume to a target volume by using a copy-on-write (CoW) and copy-on-demand mechanism.

While the FlashCopy operation is performed, the source volume is stopped briefly to initialize the FlashCopy bitmap and then, I/O can resume. Although several FlashCopy options require the data to be copied from the source to the target in the background (which can take time to complete), the resulting data on the target volume is presented so that the copy appears to complete immediately.

The FlashCopy function operates at the block level below the host operating system and cache; therefore, those levels must be flushed by the operating system for a FlashCopy copy to be consistent.

When you use the FlashCopy function, observe the following guidelines:

- The FlashCopy source *and* target volumes should use the same preferred node.

- If possible, keep the FlashCopy source and target volumes on separate storage pools.

A FlashCopy with redirect-on-write (RoW) mechanism is available with DRPs. FlashCopy with RoW uses the DRP internal deduplication-referencing capabilities to reduce overhead by creating references instead of copying the data. It provides for better performance and reduces back-end I/O amplification for FlashCopies and snapshots.

Note: FlashCopy with RoW is usable only for volumes with supported deduplication without mirroring relationships and within the same pool and I/O group. Automatic mode selection (RoW/CoW) is based on these conditions.

You can expand the source and target volumes in a FlashCopy mapping at any time. However, for incremental FlashCopy mappings, expand the target volume before the source volume. If you are expanding volumes in a non-incremental FlashCopy mapping, the source and target volumes can be expanded in any order.

Before preparing or starting a new FlashCopy mapping, ensure that the source and target volumes contain the same volume capacity. If you are restarting a FlashCopy mapping, verify the volume capacity of the mapping.

If the source and target do not have the same volume capacity, Mismatched Capacity displays in the management GUI.

Before restarting the FlashCopy mapping, expand the source or target volumes so they have the same volume capacity. You can change volume capacity for source and target volumes in FlashCopy mappings in the management GUI and CLI.

For more information about planning for the FlashCopy function, see *Performance and Best Practices Guide for IBM Storage FlashSystem and IBM SAN Volume Controller*, SG24-8543.

2.15.2 Planning for Metro Mirror and Global Mirror

MM is a copy service that provides a continuous, synchronous mirror of one volume to a second volume. A secondary volume can be on the same system or on another system. The systems can be up to 300 km (186.4 miles) apart.

Because the mirror is updated synchronously, no data is lost if the primary system becomes unavailable. MM is typically used for DR purposes, where it is important to avoid any data loss.

GM is a copy service that is similar to MM, but copies data asynchronously. You do not have to wait for the write to the secondary system to complete. For long distances, performance is improved compared to MM. However, you might lose data if a failure occurs.

GM uses one of two methods to replicate data. A noncycling mode, and a multi-cycling mode, which is also called GM with change volumes. Multicycling GM is designed to replicate data while adjusting for bandwidth constraints. It is suitable for environments where it is acceptable to lose a few minutes of data if a failure occurs.

For environments with higher bandwidth, noncycling GM can be used so that less than a second of data is lost if a failure occurs. GM also works well when sites are more than 300 kilometers (186.4 miles) apart.

When copy services are used, all components in the SAN must sustain the workload that is generated by application hosts and the data replication workload. Otherwise, the system can

automatically stop copy services relationships to protect your application hosts from increased response times.

While planning RC services, consider the following aspects:

- Copy services topology

  One or more clusters can participate in a copy services relationship. One typical and simple use case is DR, where one site is active and another performs only a DR function. In such a case, the solution topology is simple, with one cluster per site and uniform replication direction for all volumes. However, multiple other topologies are possible that you can use to design a solution that optimally fits your set of requirements.

- GM versus MM

  Decide which type of copy services that you are going to use. This decision should be requirement-driven. With MM, you prevent any data loss during a system failure, but it has more stringent requirements, especially regarding intercluster link bandwidth and latency, and remote site storage performance. Also, MM incurs a performance penalty because writes are not confirmed to the host until a data reception confirmation is received from the remote site.

  With GM, you can relax constraints on the system requirements at the cost of the use of asynchronous replication, which enables the remote site to lag behind the local site. The choice of the replication type has major effects on all other aspects of the copy services planning.

  The use of GM and MM between the same two clustered systems is supported. Also, the RC type can be changed from one to another one.

  For native IP replication, use the RC mode of Multicycling GM (or Global Mirror with Change Volumes (GMCV).

- Intercluster link

  The local and remote clusters can be connected by an FC or IP network. Each of the technologies has its own requirements concerning supported distance, link speeds, bandwidth, and vulnerability to frame or packet loss.

  When planning the intercluster link, consider the peak performance that is required. This consideration is especially important for MM configurations.

  The bandwidth between sites must be sized to meet the peak workload requirements. When planning the inter-site link, consider the initial sync and any future resync workloads. It might be worthwhile to secure more link bandwidth for the initial data synchronization.

  If the link between the sites is configured with redundancy so that they can tolerate single failures, you must size the link so that the bandwidth and latency requirements are met even during single failure conditions.

  When planning the inter-site link, note whether it is dedicated to the inter-cluster traffic or is going to be used to carry any other data. Sharing the link with other traffic might affect the link's ability to provide the required bandwidth for data replication.

- Volumes and consistency groups

  Determine whether volumes can be replicated independently. Some applications use multiple volumes and require that the order of writes to these volumes is preserved in the remote site. Notable examples of such applications are databases.

  If an application requires that the write order is preserved for the set of volumes that it uses, create a consistency group for these volumes.

2.15.3 Policy-based replication

Policy-based replication uses volume groups and replication policies to automatically deploy and manage replication. Policy-based replication can simplify configuring, managing, and monitoring replication between two systems.

With policy-based replication, you can replicate data between systems with minimal management, significantly higher throughput and reduced latency compared to the remote-copy function.

A replication policy has the following properties:

- A replication policy can be assigned to one or more volume groups.
- Replication policies cannot be changed after they are created. If changes are required, a new policy can be created and assigned to the associated volume group.
- Each system supports up to a maximum of 32 replication policies.

Replication policies

Replication policies define the replication settings that are assigned to the volume groups. Replication policies replicate the volume groups and ensure that consistent data is available on the production and recovery system.

Replication policy is assigned to the sites asynchronously, and it ensures that any change in the production system automatically reflects in the recovery system. Replication policy can be assigned, configured, and managed from a single system. It asynchronously copies the data based on configured rules. Replication policies help to safeguard your business by maintaining continuous data copies. In the event of an outage, you can quickly recover data and keep your network operational.

To configure replication policies, define a recovery point objective (RPO), locations, and topology based on your requirement. RPO is the maximum data loss that is acceptable during failure. Locations are the production and recovery systems. Topology is the type of replication that can be selected, currently 2-site asynchronous replication is supported.

A replication policy defines three key attributes:

- A set of locations. It defines the I/O groups on the partnered systems that contain a replicated copy of the volume group.
- A topology. It represents the organization of the systems and the type of replication that is performed between each location, that is, how the data is replicated between the locations.
- A name. It uniquely identifies the replication policy on both systems.

Before you can implement policy-based replication, determine your current recovery point objectives, applications requirements, and verify your networking configuration.

The following configuration limits and restrictions apply to policy-based replication:

- The name of a volume group cannot be changed while a replication policy is assigned.
- The name of a volume cannot be changed while the volume is in a volume group with a replication policy assigned.
- Ownership groups are not supported with policy-based replication.
- Policy-based replication is not supported on HyperSwap topology systems.
- Policy-based replication cannot be used with the following types of volumes:
  - Image mode

  - HyperSwap
  - Configured to use Transparent Cloud Tiering (TCT)
  - VMware vSphere virtual volumes (vVols)
- The following actions cannot be performed on a volume while the volume is in a volume group with a replication policy assigned:
  - Resize (expand or shrink)
  - Migrate to image mode, or add an image mode copy
  - Move to a different I/O group

Consider the following characteristics for host applications that use policy-based replication:

- Policy-based replication can be used with all host operating systems that are supported by IBM Storage Virtualize systems. The full interoperability list can be found at IBM System Storage Interoperation Center. For more information, see SSIC.
- For the best possible performance, configure the multipath drivers on the operating system to use SCSI ALUA or NVMe ANA.

With policy-based replication, you can replicate a thin-clone or the target volumes of a FlashCopy mapping to the recovery system for application consistency. You cannot replicate Safeguarded copies.

Policy-based replication is supported on systems that are connected over Fibre channel (FC) or IP connections.

Connectivity requirements for policy-based partnerships are similar to remote-copy based partnerships, except IP connectivity is required between management IP addresses of partnered systems to manage replication. The management traffic for policy-based replication uses authentication certificates to prevent unauthorized access and ensure secure communications between the partnered systems either in Fibre channel or in IP partnerships. Therefore, ensure that valid authentication certificates are installed on both the systems.

Also, replication that uses IP partnerships can encrypt data in flight (EDiF) by enabling secured IP partnership.

Fibre Channel port allocation for policy-based replication uses the same requirements of the remote-copy function.

Planning migration from MM or GM to policy-based replication

Before you migrate from Metro Mirror or Global Mirror to policy-based replication, you must meet certain requirements.

Ensure that you meet the following requirements before configuring a volume, which is in a Metro Mirror or Global Mirror relationship, to use policy-based replication:

- The relationship can be in a consistency group if it does not need to be moved between I/O groups by using the **`movevdisk`** command first.
- The relationship type must be Metro Mirror or Global Mirror.
- The volume being moved must be the primary volume within the Metro Mirror or Global Mirror relationship.
- A change volume must not be associated with the primary volume.
- The remote-copy relationship cannot have its direction switched or be made a secondary after a stop with access enabled when a migration is in progress.
- The primary volume cannot be made a policy-based replication recovery volume.
- Remote-copy cannot be configured on a policy-based replication volume.

- To replicate from the remote-copy secondary volumes, the remote-copy replication configuration must be removed first.
- Only policy-based replication is supported for volumes in Metro Mirror or Global Mirror relationships. To ensure compatibility, the volume's I/O group must match the production I/O group defined in the replication policy. If not, use the **movevdisk** command to relocate the volume to the correct group.

Moving a volume between I/O groups by using the **movevdisk** command has the following requirements before it can proceed:

- The relationship state must be *consistent_synchronized*.
- The relationship cannot be in a consistency group.
- The relationship type must be Metro Mirror or Global Mirror.
- The relationship must not have a change volume that is associated with the primary volume.
- The volume being moved must be the primary volume in the Metro Mirror or Global Mirror relationship.

For more information about the implementation of the Policy-Based Replication function, see *Policy-Based Replication with IBM Storage FlashSystem, IBM SAN Volume Controller and IBM Storage Virtualize*, REDP-5704.

2.16 Throttles

A throttle is a mechanism to control the amount of resources that are used when the system is processing I/Os on supported objects. The system supports throttles on hosts, host clusters, volumes, copy offload operations, and storage pools. If a throttle limit is defined, the system processes the I/O for that object, or delays the processing of the I/O to free resources for more critical I/O operations.

When you configure throttles on the system, consider the following guidelines:

- The throttle limit is a per node limit. For example, if a throttle limit is set for a volume at 100 IOPS, each node on the system that has access to the volume allows 100 IOPS for that volume. Any I/O operation that exceeds the throttle limit is queued at the receiving nodes.
- The multipath policies on the host determine how many nodes receive I/O operations and the effective throttle limit.
- If more than one throttle applies to an I/O operation, the lowest and most stringent throttle is used.

Throttles can be defined for storage pools to control I/O operations on back-end storage systems. Storage pool throttles can be used to avoid overwhelming the back-end storage and can be used with virtual volumes.

You can set throttles for both parent and child pools. Throttles work in a hierarchy when they are applied to the parent *and* child pool. Throttles that are defined in the parent pool apply to volumes in parent and child pools. However, throttles that are defined on the child pool apply only to the volumes in that child pool.

For mirrored volumes that are in different pools, only the throttling that applies to the primary copy is considered, regardless to which copy the I/O is directed. In this case, any throttles that are defined on the secondary copy of the volume are ignored.

Only throttles on the primary copy are used, whether the pool is a parent or child pool. When throttles are defined in parent and child pools, the more restrictive throttle applies to the volume in the child pool, but the I/O is counted against both pools.

Throttles for volumes, hosts, host clusters, copy offload, and storage pools can be configured in the management GUI and by using the CLI.

2.17 Data migration

Data migration is an important part of an implementation; therefore, you must prepare a detailed data migration plan. You might need to migrate your data for one of the following reasons:

- Redistribute a workload within a clustered system across back-end storage subsystems.
- Move a workload:
  - On to newly installed storage
  - Off old or failing storage ahead of decommissioning it
  - To rebalance a changed load pattern
- Migrate data from:
  - An older disk subsystem
  - One disk subsystem to another subsystem

Because multiple data migration methods are available, choose the method that best fits your environment, operating system platform, type of data, and the application's service-level agreement (SLA).

Data migration methods can be divided into three classes based on:

- The host operating system; for example, by using the system's logical volume manager (LVM)
- Specialized data migration software
- The system data migration features

For more information about system data migration tools, see Chapter 8, "Storage migration" on page 635 and Chapter 10, "Advanced Copy Services" on page 717.

With data migration, apply the following guidelines:

- Choose the data migration method that best fits your operating system platform, type of data, and SLA.
- Choose where you want to place your data after migration in terms of the storage tier, pools, and back-end storage.
- Check whether enough free space is available in the target storage pool.
- To minimize downtime during the migration, plan all of the required changes, including zoning, host definition, and volume mappings.
- Prepare a detailed operation plan so that you do not overlook anything at data migration time. Have the plan peer-reviewed and formally accepted by a suitable technical design authority within your organization (especially for a large or critical data migration).
- Perform and verify a backup before you start any data migration.

- You might want to use the system as a data mover to migrate data from a nonvirtualized storage subsystem to another nonvirtualized storage subsystem. In this case, you might have to add checks that relate to the specific storage subsystem that you want to migrate.

  Be careful when you use slower disk subsystems for the secondary volumes for high-performance primary volumes because the system's cache might not buffer all of the writes. Flushing cache writes to slower back-end storage might affect the performance of your hosts.
- Consider storage performance. The migration workload might be much higher than expected during normal operations of the system. If application data is on the system to which you are migrating, the application performance might suffer if the system is overloaded. Consider the use of host or volume level throttles when performing migration on a production environment.

Non-Disruptive System Migration can be used to migrate data from one IBM Storage Virtualize system to another nondisruptively. You can create this specific remote-copy relationship that copies data from source volumes on a system that you are decommissioning to auxiliary volumes that are on another system. The nondisruptive system migration is a remote-copy relationship type that is dedicated to volume migration between systems.

Select Non-Disruptive System Migration to migrate volume data between systems by creating a remote-copy relationship. Although Non-Disruptive System Migration does not require a Remote Mirroring license, specific remote-copy functions are restricted, such as creating consistency groups. Before you can migrate data with Non-Disruptive System Migration, a partnership between both systems must be created.

Remote-copy relationships that are used for migration feature the following restrictions:

- Stop-with-access is prohibited. The volumes cannot act independently if they use the same UUID.
- Migration relationships cannot be added to a consistency group.
- You cannot change the relationship to another type of remote-copy relationship.
- A migration relationship cannot be converted into a 3-site relationship.
- Associating change volumes to a migration relationship is prohibited.
- Volumes in migration relationships cannot be resized until the migration is completed or canceled.
- ODX must be disabled on both systems while migrating volumes. Migration relationships cannot be created while ODX is enabled.
- ODX must be disabled on both systems while migrating volumes.
- Migrating volumes that are mapped to NVMe-attached hosts is not supported.
- Migrating SAN-boot volumes is not supported.

Prerequisites

Before you can use nondisruptive system migration function, ensure that the following prerequisites are met:

- Both systems are running 8.4.2 or later.
- An FC or IP partnership is created between the two systems that you want to migrate volumes between. The maximum supported round-trip time (RTT) between the two systems is 3 milliseconds. The partnership must have sufficient bandwidth to support the write throughput for all the volume you are migrating.

- Any hosts that are mapped to volumes that you are migrating are correctly zoned to both systems. Hosts must appear in an online state on both systems. The system supports the hosts and connection types that are listed in Table 2-8 for nondisruptive system migration.

Table 2-8 Supported hosts and connections types for nondisruptive system migration

Supported hosts	Connection types
VMware ESXi 6.x	iSCSI and FC
VMware ESXi 7.x	iSCSI, iSER, and FC
RHEL 7.x, RHEL 8.x	iSCSI and FC
SLES 12.x, SLES 15.x	iSCSI and FC
Solaris 10, Solaris 11	FC
HP-UX 11iV3	FC
AIX 7.3 TL1 or later	iSCSI and Fibre Channel

2.18 Ansible automation for IBM Storage Virtualize systems

IBM Storage Virtualize systems for hybrid multicloud include integration with Red Hat Ansible Automation Platform with which you can create an Ansible Playbook. This playbook automates the tasks that are repeated across your organization in a consistent way. This feature improves outcomes and reduces risk.

With IBM Storage Virtualize systems and Red Hat Ansible Automation Platform, you can easily automate tasks, such as configuration management, provisioning, workflow orchestration, application deployment, and life-cycle management.

For more information, see 13.3, “Automation with Red Hat Ansible” on page 1180.

2.19 Container Storage integration

IBM Storage Virtualize systems implement the IBM block storage Container Storage Interface (CSI) driver that is used by Kubernetes persistent volumes (PVs) to dynamically provision for block storage that is used with stateful containers.

The IBM block storage CSI driver is based on an Open Source IBM project (CSI driver), which is included as a part of IBM storage orchestration for containers.

By using CSI drivers for IBM storage systems, Kubernetes PVs can be dynamically provisioned for block storage to be used with stateful containers, such as database applications (IBM Db2, MongoDB, PostgreSQL, and so on) running in Red Hat OpenShift Container Platform or Kubernetes clusters. Storage provisioning can be fully automatized with support of cluster orchestration systems to automatically deploy, scale, and manage containerized applications.

The CSI driver requires that Port 22 is opened on the worker nodes operating system firewall. Also, be sure that multipathing is installed and running.

For more information, see IBM block storage CSI driver documentation.

For more information about the implementation of the IBM block storage CSI drivers with IBM Storage Virtualize systems, see 7.12, “Container Storage Interface drivers” on page 632.

2.20 Safeguarded Copy

The Safeguarded Copy function isolates backup copies from production data. Therefore, if a cyberattack or ransomware attack occurs, you can quickly recover and restore data from Safeguarded copies.

The Safeguarded Copy function is supported on the following products:

- IBM FlashSystem 9500
- IBM FlashSystem 7300
- IBM FlashSystem 5200
- IBM FlashSystem 5045
- IBM SAN Volume Controller with FlashCopy license

The systems support Safeguarded snapshots, which use the snapshot function to create point-in-time copies of volume groups that are immutable and can be scheduled with an internal scheduler. The system also supports IBM Spectrum Copy Data Management and IBM Copy Services Manager as external scheduling applications.

IBM Copy Services Manager coordinates and automates Safeguarded Copy function across multiple systems.

IBM Storage Copy Data Management provides management of application-consistent Safeguarded copies that use FlashCopy.

Separation of duties

The Safeguarded Copy function employs separation of duties, which provides more security capabilities to prevent nonprivileged users from compromising production data. Operations related to Safeguarded backups are restricted to only a subset of users with specific roles on the system.

Administrator

Users with the Administrator role can provision and configure Safeguarded copies and related objects, such as volume groups. However, these users cannot remove or change existing Safeguarded snapshots. For auditing, it is recommended that you create a new Administrator user to configure the Safeguarded snapshots or Safeguarded Copy function. Users with this role are limited in how they can manage and interact with Safeguarded Copy operations.

Security Administrator

Users with the Security administrator role can manage users and security on the entire system and can remove and change Safeguarded backups and Safeguarded backup locations.

Superuser

Users with superuser privileges can configure all objects and complete maintenance tasks on the system. These users can remove and change both Safeguarded backups and Safeguarded policies. For more security, this account can be disabled on the system. However, it can be reenabled for remote support assistance or maintenance tasks.

Restricted Security Administrator and TPI

When TPI is enabled on the IBM Storage Virtualize system, the security administrator role is changed to restricted security administrator. TPI requires two security administrators to work together to complete critical or risky tasks. For example, a restricted security administrator with an elevated role can remove Safeguarded snapshots. See 2.4.2, “Two person integrity” on page 125.

System requirements

Ensure that the following system requirements are met:

- All IBM Storage Virtualize systems are running the 8.4.2 or later release.
- For all new and existing systems, ensure capacity planning is completed to accommodate the Safeguarded Copy requirement.

> **Note:** Contact your IBM sales representative or IBM Business Partner to perform these capacity calculations.

For more information about the implementation of the Safeguarded Copy function, see *Implementation Guide for SpecV/FlashSystem Safeguarded Copy*, REDP-5654.

2.21 Performance monitoring with IBM Storage Insights

IBM Storage Insights is integral to monitoring and ensuring the continued availability of the system.

IBM Storage Insights provides a unified view of your IBM storage systems. By using it, you can see all of your IBM storage inventory as a live event feed so that you know what is going on with your storage.

In addition, IBM Storage Insights provides advanced customer service with an event filter that provides the following functions:

- The ability for you and IBM Support to view support tickets and open and close them, and to track trends.
- With the auto log collection capability, you can collect the logs and send them to IBM before IBM Support examines the problem. This feature can reduce the time to solve the case by as much as 50%.

There are two versions of the IBM Storage Insights service: IBM Storage Insights and IBM Storage Insights Pro. When you order your IBM Storage Virtualize system, IBM Storage Insights is available at no cost. With this version, you can monitor the basic health, status, and performance of various storage resources.

IBM Storage Insights Pro is a subscription-based product that provides a more comprehensive view of the performance, capacity, and health of your storage resources. In addition to the features offered by IBM Storage Insights, IBM Storage Insights Pro provides tools for intelligent capacity planning, storage reclamation, storage tiering, and performance troubleshooting services. Together, these features can help you reduce storage costs and optimize your data center.

Figure 2-6 on page 174 shows the architecture of the IBM Storage Insights application, the supported products, and the three main teams who can benefit from the use of the tool.

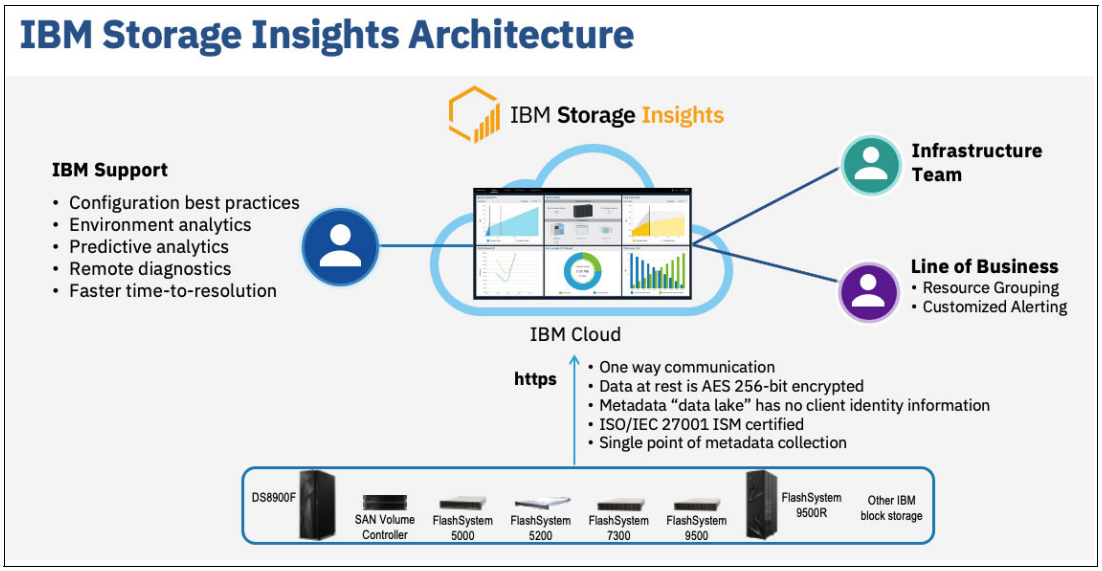


Figure 2-6 IBM Storage Insights architecture

IBM Storage Insights provides a lightweight data collector that is deployed on a server or virtual machine that runs AIX, Linux, or 64-bit Windows systems. On the server or virtual machine, you must provide at least 1 GB of RAM and 3 GB of disk space.

The data collector sends the asset, capacity, configuration, and performance metadata that is collected for your storage systems for analysis. The metadata is used to analyze storage usage, performance, and to upload logs when you open or update support tickets. Asset, configuration, and performance information for fabrics, switches, and ESXi hosts is also collected and analyzed.

With IBM Storage Virtualize v8.6.0, a new feature called In-line Data Corruption Detection (IDCD) is available, which assists with the detection of potential ransomware attacks. To ensure that you have the latest storage metadata for detecting those types of attacks, compression and cyber resiliency statistics for volumes are collected every 5 minutes. With these statistics, IBM Storage Insights builds a historical model of a storage system and uses its built-in intelligence and formulas to identify when and where ransomware attacks might be occurring.

The metadata flows in one direction, that is, from your data center to IBM Cloud over HTTPS. In the IBM Cloud your metadata is protected by physical, organizational, access, and security controls. IBM Storage Insights is ISO/IEC 27001/27017/27018/27701 Information Security Management certified.

To monitor storage systems, you must provide a username and password to log in to the storage systems. The role or user group that is assigned to the username must have the suitable privileges to monitor the data that is collected. For IBM Storage Virtualize version 8.3.1.2 or later and Storage Insights or IBM Spectrum Control 5.3.7 or later, data collection can be done by using the Monitor role, which is the least privileged role.

Figure 2-7 shows the data flow from systems to the IBM Storage Insights cloud.

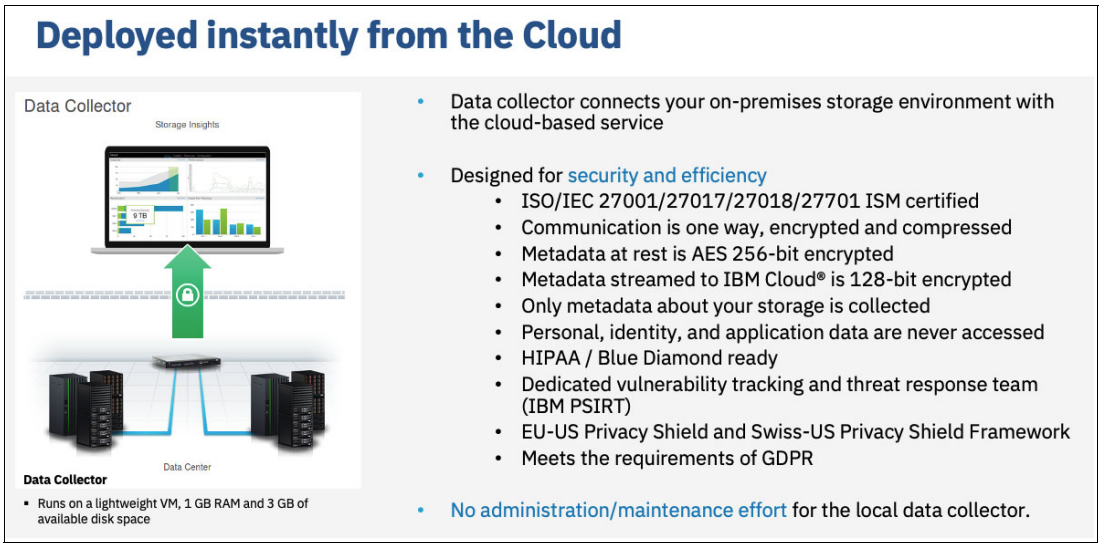


Figure 2-7 Data flow from the storage systems to the IBM Storage Insights cloud

Metadata about the configuration and operations of storage resources is collected and includes the following information:

- Inventory and configuration metadata such as name, model, firmware, type, and more
- Inventory and configuration metadata for internal components such as volumes, pools, disks, ports, and more
- Capacity metrics such as capacity, usable capacity, used capacity, compression ratios, and more
- Performance metrics such as read and write data rates, I/O rates, response times, and more
- Diagnostic data, system failure logs, maintenance levels, and more support-related information

The application data that is stored on the storage systems cannot be accessed by the data collector.

Access to the metadata that is collected is restricted to the following users:

- The customer who owns the dashboard
- The administrators who are authorized to access the dashboard, such as the customer's operations team
- The IBM Cloud team that is responsible for the daily operation and maintenance of IBM Cloud instances
- IBM Support for investigating and closing service tickets

In addition, you can now use IBM Call Home with cloud services to create a direct connection between your IBM Storage Virtualize storage systems and IBM Storage Insights with no data collector required.

When you enable Call Home with cloud services on an IBM Storage Virtualize storage system, you have the option to integrate that storage system with IBM Storage Insights. With this integration, you can collect the same metadata as a data collector and send it securely to IBM Cloud for analysis and reporting in IBM Storage Insights.

If the IBM Storage Virtualize system was not integrated with IBM Storage Insights during the initial system setup, an inline notification to add the storage system to IBM Storage Insights is displayed on the Call Home page and the Dashboard. On the Dashboard, an inline notification is displayed whenever the system is upgraded to a later version from the current version. If you close the notification, the system snoozes the notification and reminds you after 30 days.

The ability to use Call Home with cloud services for monitoring is available for IBM SAN Volume Controller and the IBM FlashSystem family that meet the following requirements:

- IBM Storage Virtualize 8.5.3 and later
- At least 128 GB of system memory

Note: The ability to integrate IBM Storage Virtualize storage systems with IBM Storage Insights by using Call Home with cloud services is available for storage systems that are not being monitored. Storage systems that are currently monitored by IBM Storage Insights with data collectors are not yet eligible.

For more information about IBM Storage Insights and to sign up and register for the free service, see the following resources:

- Fact Sheet
- Demonstration
- Security Guide
- Registration

For more information, see 11.12, “IBM Storage Insights monitoring” on page 1035.

2.22 Configuration backup procedure

Save the configuration before and after any major configuration changes to the system. Saving the configuration is a crucial part of management, and various methods can be applied to back up your system configuration.

A best practice is to implement an automatic configuration backup by using the configuration backup command. Make sure that you save the configuration to a host system that does not depend on the storage that is provisioned from a system whose configuration is backed up.

Only the data that describes the system configuration is backed up. Back up your application data by using the suitable backup methods.

For more information, see 11.4, “Configuration backup” on page 979.

3

Initial configuration

This chapter describes the initial configuration of the following systems:

- IBM FlashSystem 9500
- IBM FlashSystem 7300
- IBM FlashSystem 5200
- IBM FlashSystem 5045
- IBM FlashSystem 5015
- IBM SAN Volume Controller

It also provides step-by-step instructions about how to perform the initial setup process and defines the base settings of the system, which are done during the implementation phase before volumes are created and provisioned.

This chapter includes the following topics:

- 3.1, “Prerequisites” on page 178
- 3.2, “System initialization” on page 179
- 3.3, “System setup” on page 186

© Copyright IBM Corp. 2024. All rights reserved.

3.1 Prerequisites

Before initializing and setting up the system, ensure that the following prerequisites are met:

- The physical components fulfill all the requirements and are correctly installed, including:
  - The control enclosures or IBM SAN Volume Controller nodes are physically installed in the racks.
  - The Ethernet and Fibre Channel (FC) cables are connected.
  - The expansion enclosures (if available) are physically installed and attached to the control enclosures that use them.
  - The control enclosures or IBM SAN Volume Controller nodes and optional expansion enclosures are powered on.

    Note: IBM SAN Volume Controller nodes need enough time to charge the batteries. How long it takes to recharge depends on how long it was waiting idle in stock and not in production. You cannot start the nodes without a fully charged battery.

- The web browser that is used for managing the system is supported by the management GUI. For the list of supported browsers, see Web browser requirements to access the management GUI.
- The required information for remote management of the system, including:
  - The IPv4 (or IPv6) addresses that are assigned for the system's management interfaces:
    - The unique cluster IP address, which is the address that is used for the management of the system.
    - Unique service IP addresses, which are used to access node service interfaces. You need one address for each IBM SAN Volume Controller node or IBM FlashSystem node (two per control enclosure).
    - The IP subnet mask for each subnet that is used.
    - The IP gateway for each subnet that is used.
  - The licenses that might be required to use specific functions. Whether these licenses are required depends on the hardware that is used. For more information, see 1.20, "Licensing" on page 115.
  - Information that is used by a system when performing Call Home functions, such as:
    - The company name and system installation address.
    - The name, email address, and phone number of the storage administrator whom IBM can contact if necessary.
  - The following information is optional:
    - The Network Time Protocol (NTP) server IP address.
    - The Simple Mail Transfer Protocol (SMTP) server IP address, which is necessary only if you want to enable Call Home or want to be notified about system events through email.
    - The IP addresses for Remote Support Proxy Servers, which are required only if you want to use them with the Remote Support Assistance feature.

Note: IBM FlashSystem 9500, and IBM SAN Volume Controller are installed by an IBM System Services Representative (IBM SSR). Provide all the necessary information to the IBM SSR by completing the following planning worksheets:

- IBM FlashSystems
- IBM SAN Volume Controller

After the IBM SSR completes their portion of the setup, see 3.3, “System setup” on page 186 to continue the setup process.

3.2 System initialization

This section provides step-by-step instructions about how to create the system cluster.

Demonstration videos

You can view the following demonstration videos:

- IBM Storage Virtualize V8.6 Initial setup: SSR configuration tasks
- IBM Storage Virtualize V8.6 Initial setup: Customer configuration tasks
- IBM Storage Virtualize V8.6 Initial setup: Setting up a cluster from the service IP

To start the initialization procedure, connect a workstation to the technician port. The *technician port* is a dedicated 1 Gb Ethernet (GbE) port at the rear of each of the nodes canisters. On all platforms except IBM FlashSystem 5015, it can be used only to initialize or service the system. It cannot be connected to an Ethernet switch because it supports only a direct connection, and it remains disconnected after the initial setup is done.

On IBM FlashSystem 5015, the technician port is enabled initially. However, the port is switched to internet Small Computer Systems Interface (iSCSI) host attachment mode after the setup wizard is complete.

To re-enable the onboard Ethernet port 2 on a system to be used as the technician port, run the command that is shown in Example 3-1.

Example 3-1 Reenabling the onboard Ethernet port 2 as the technician port

```
IBM_IBM FlashSystem 5015:superuser>satask chserviceip -techport enable -force
```

The location of the technician port on an IBM FlashSystem 9500 is shown in Figure 3-1.

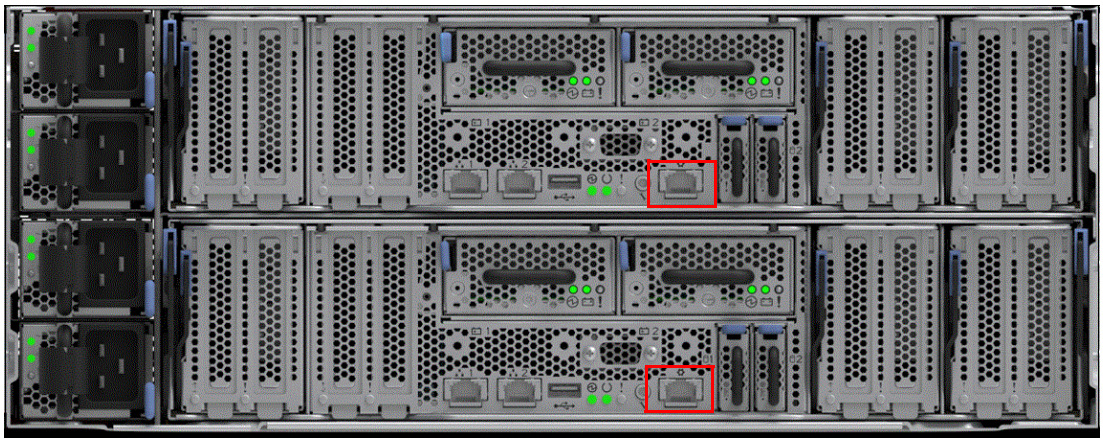


Figure 3-1 Technician port FlashSystem 9500

The location of the technician port of an IBM FlashSystem 7300 is shown in Figure 3-2.

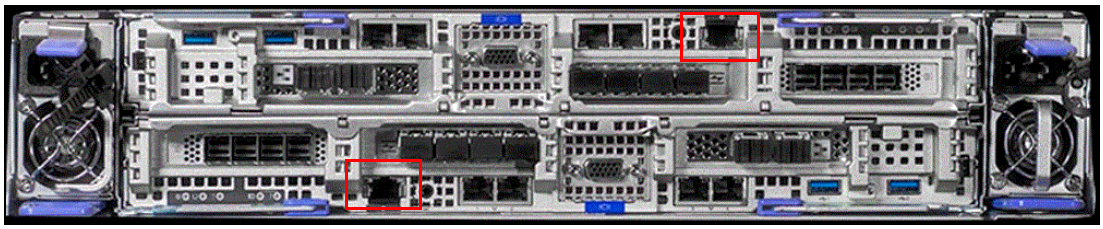


Figure 3-2 Technician port FlashSystem 7300

The location of the technician port of an IBM FlashSystem 5200 is shown in Figure 3-3

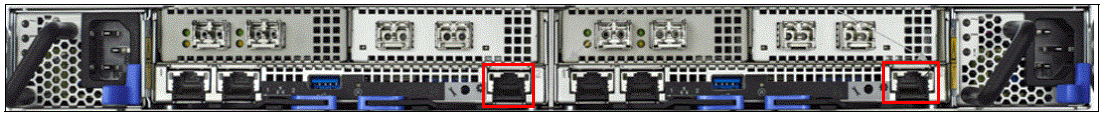


Figure 3-3 Technician port FlashSystem 5200

The location of the technician port of an IBM FlashSystem 5045 is shown in Figure 3-4.



Figure 3-4 Technician port FlashSystem 5045

The location of the technician port of an IBM FlashSystem 5015 is shown in Figure 3-5.

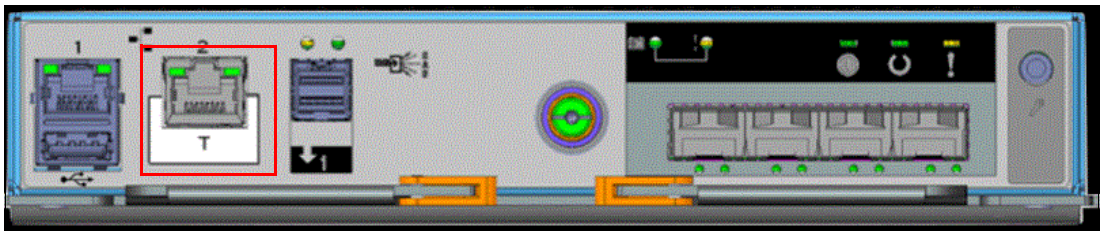


Figure 3-5 Technician port FlashSystem 5015

The location of a technician port on the IBM SAN Volume Controller 2145-SV3 is shown in Figure 3-6.

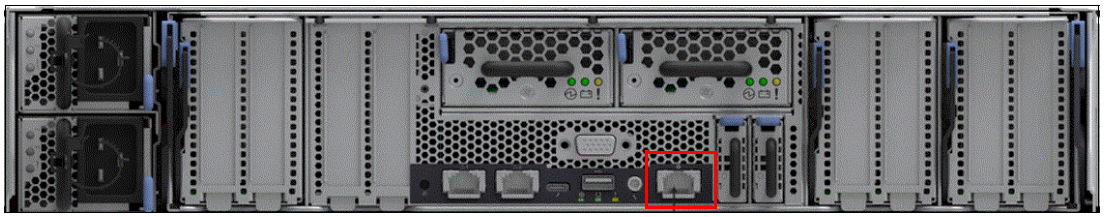


Figure 3-6 Technician port IBM SAN Volume Controller 2145-SV3

The location of a technician port on the IBM SAN Volume Controller 2145-SV2 is shown in Figure 3-7.



Figure 3-7 Technician port IBM SAN Volume Controller 2145-SV2

The technician port runs an IPv4 DHCP server, and it can assign an address to any device that is connected to this port. Ensure that your workstation Ethernet adapter is configured to use a DHCP client if you want the IP to be assigned automatically.

If you prefer not to use DHCP, you can set a static IP on the Ethernet port from the `192.168.0.x/24` subnet; for example, `192.168.0.2` with the netmask `255.255.255.0`.

The default IP address of a technician port on a node canister is `192.168.0.1`. Do *not* use this IP address for your workstation.

Note: Ensure that the technician port is not connected to the organization's network. No Ethernet switches or hubs are supported on this port.

3.2.1 System initialization process

Before a system is initialized, each node canister of a new system remains in the candidate state and cannot process I/O. During initialization, the nodes in one control enclosure are joined in a cluster, which is later configured to process data (for an IBM SAN Volume Controller system, the cluster consists at that moment of one node only).

If your systems feature more than one control enclosure or more IBM SAN Volume Controller nodes, all of the other enclosures or nodes except the first enclosure or node must not be initialized. The remaining control enclosures or IBM SAN Volume Controller nodes are added to the cluster by using a cluster management interface, GUI or CLI, after the first enclosure or node is set up.

You must specify IPv4 or an IPv6 system management addresses, which are assigned to Ethernet port 1 on each node and used to access the management GUI and CLI. After the system is initialized, you can specify other IP addresses.

Note: Do not perform the system initialization procedure on more than one node canister of one control enclosure. After initialization completes, use the management GUI or CLI to add control enclosures to the system.

To initialize a new system, complete the following steps:

1. Connect your workstation to a technician port of any canister of the control enclosure or the IBM SAN Volume Controller system. Ensure that you obtained a valid IPv4 address with DHCP.
2. Open a supported web browser and browse to `http://install`. The browser is automatically redirected to the System Initialization wizard. You can also use the IP address `http://192.168.0.1` if you are not automatically redirected.

   Note: During the system initialization, you might be prompted to accept untrusted certificates because the system certificates are self-signed. If you are directly connected to the service interface, the identity of the certificate issuer is confirmed. Therefore, you can safely accept the certificates.

   If the system is not in a state that allows initialization, the system does not start the System Initialization wizard, and you are redirected to the Service Assistant interface. Use the displayed error codes to troubleshoot the problem.

3. The window that is used to log in to Service Assistant opens (Figure 3-8). This window is the first step of initializing the system. Enter the default superuser password (`passw0rd`) and click **Log in**.

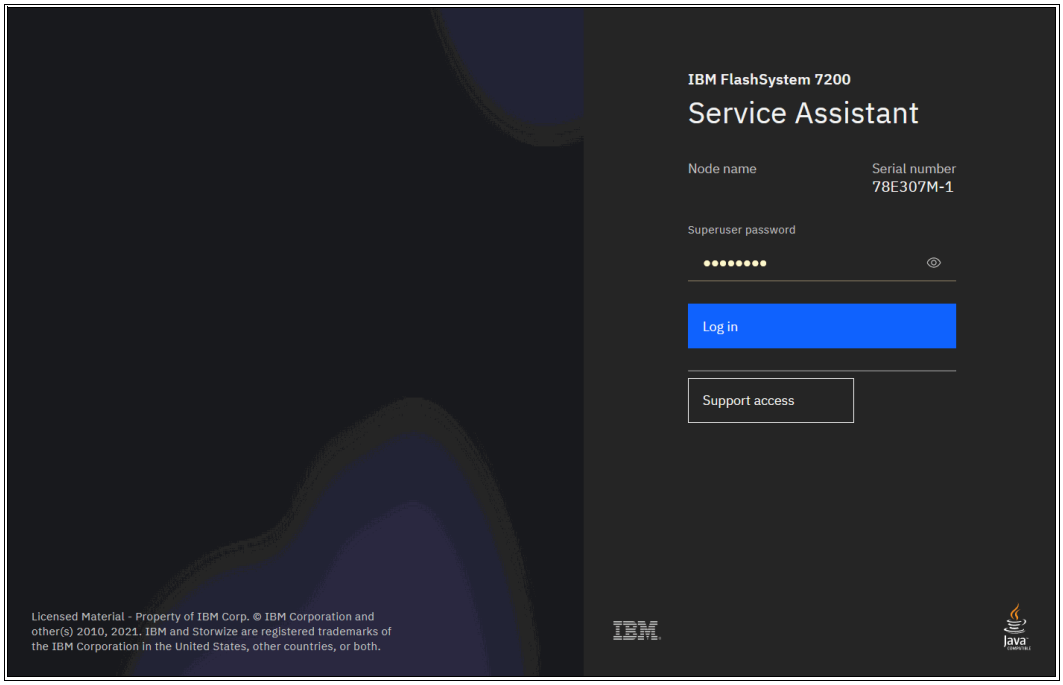


Figure 3-8 Logging in to Service Assistant by way of the technician port

4. The System Initialization wizard shows the detected canisters, as shown in Figure 3-9. Click **Proceed** to continue (this window is not shown for IBM SAN Volume Controller nodes).

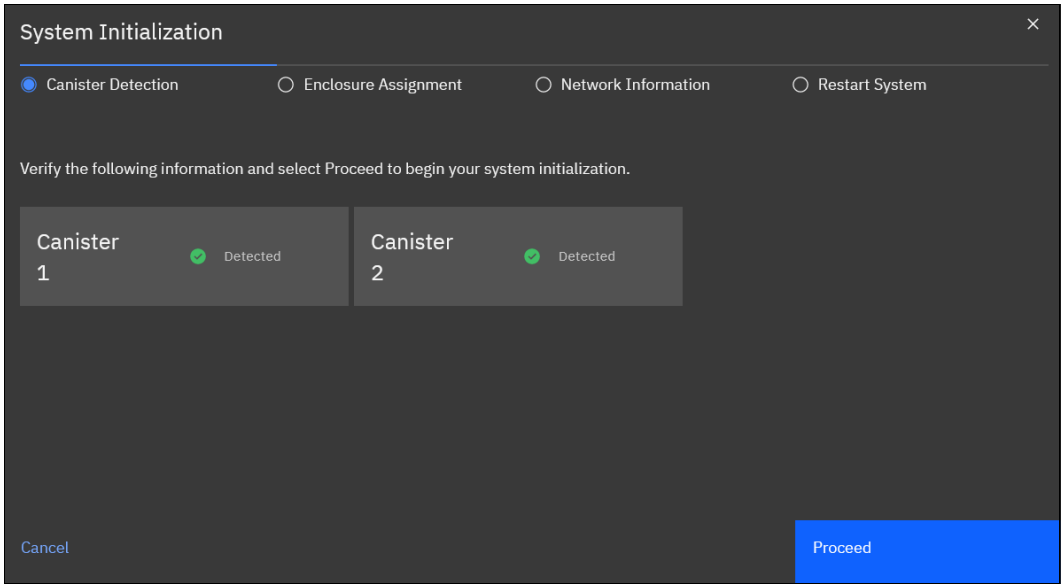


Figure 3-9 System Initialization: Canister detection

5. The System Initialization wizard shows the enclosure assignment. Select **As the first enclosure in a new system**, as shown in Figure 3-10. Click **Next**.

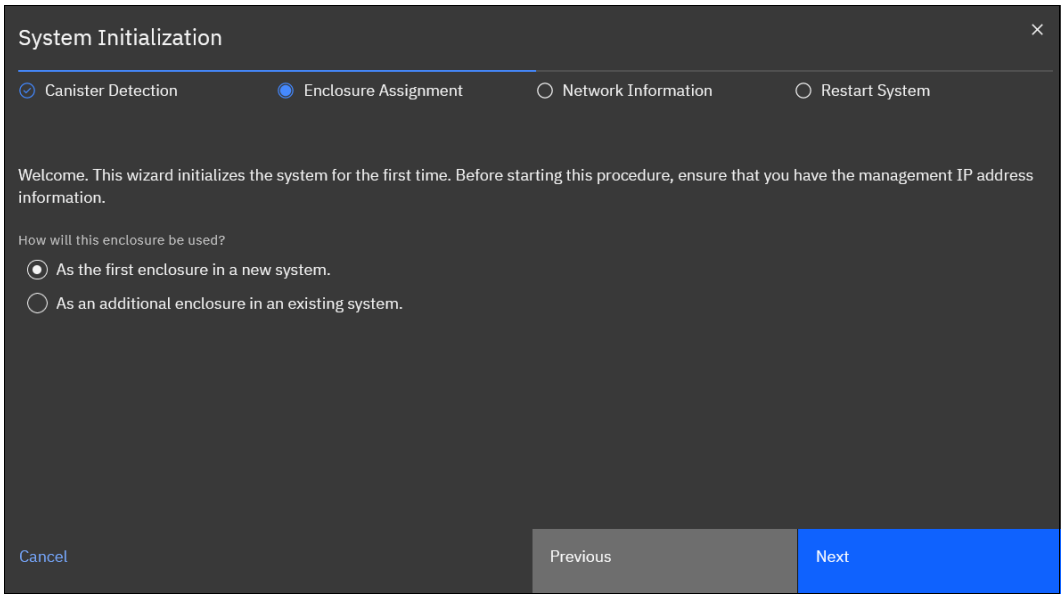


Figure 3-10 System Initialization: Initialize the first enclosure

If you select **As an additional enclosure in an existing system**, you are prompted to disconnect from the technician port and use the GUI of the system to which the new nodes are to be added.

The window looks a bit different for IBM SAN Volume Controller. Figure 3-11 shows that you are prompted for nodes and not for enclosures. You are also prompted to disconnect from the technician port and use the GUI of the system to which the new nodes are to be added if you select **As an additional node in an existing system**.

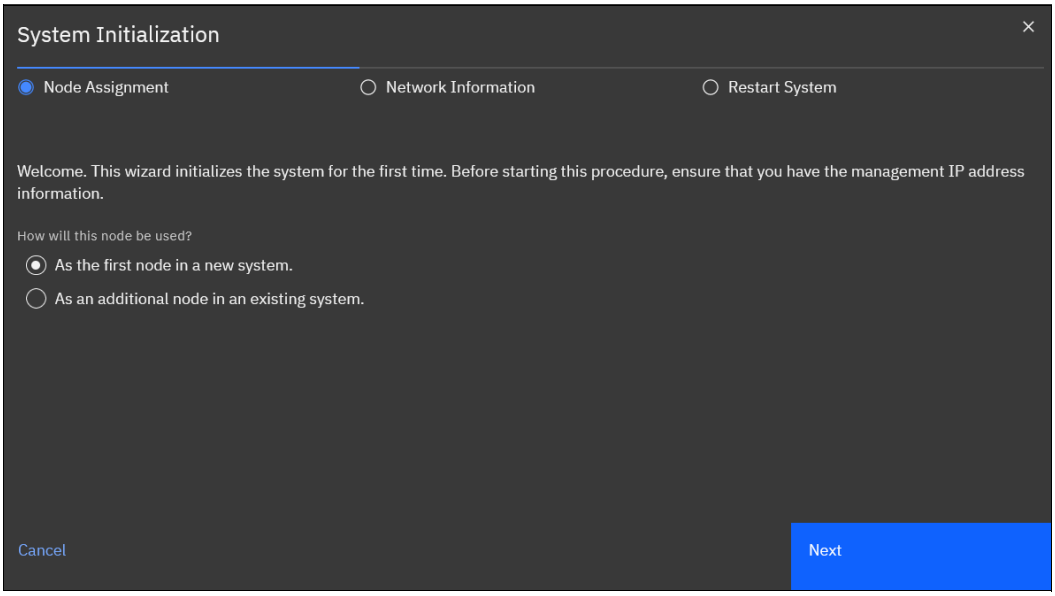


Figure 3-11 System Initialization: Initialize the first IBM SAN Volume Controller node

6. Enter the management IP address information for the new system as shown in Figure 3-12. Set the IP address, network mask, and gateway. Click **Next**.

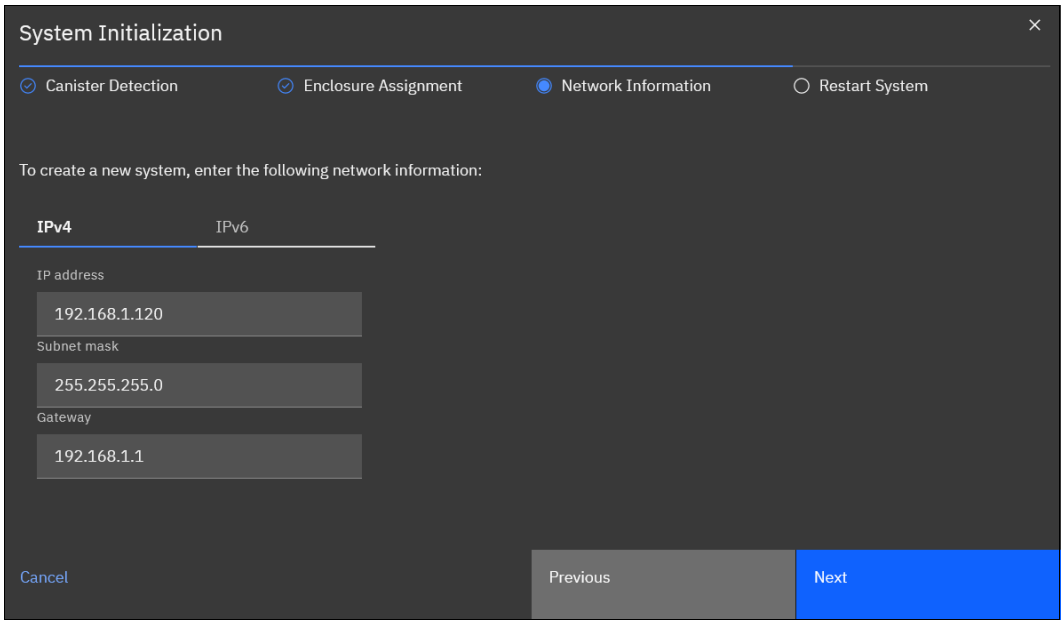


Figure 3-12 System Initialization: Enter Management IP

7. A window that includes a restart timer opens (Figure 3-13). When the timeout is reached, the window is updated to reflect success or failure. Failure occurs if the system is disconnected from the network, which prevents the browser from updating with the IBM FlashSystem web server.

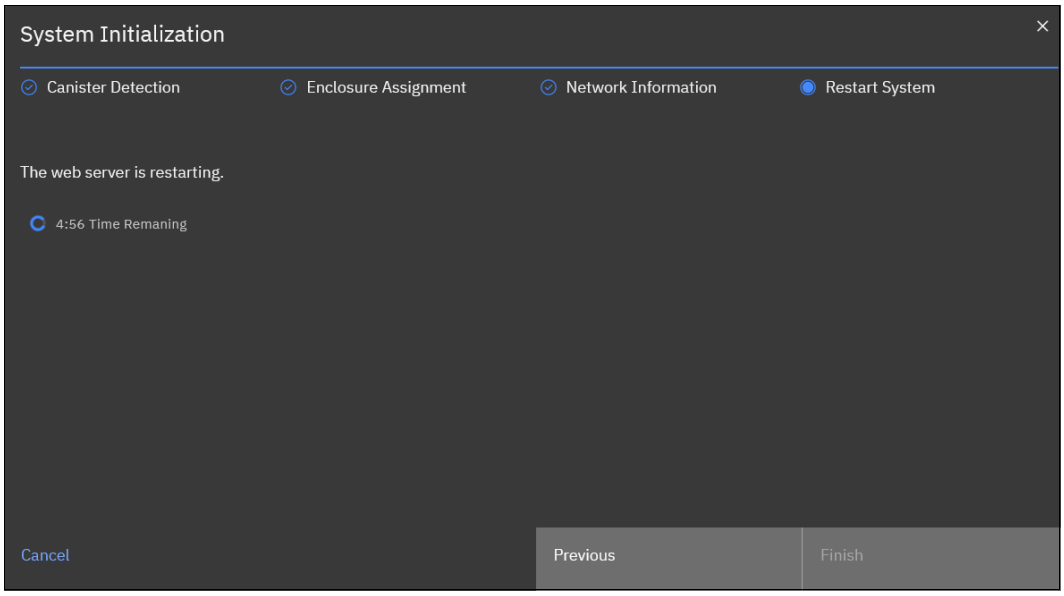


Figure 3-13 System Initialization: Web-server restart timer counting down from 5 minutes

8. The System Initialization completed wizard is shown in Figure 3-14. Click **Finish**.

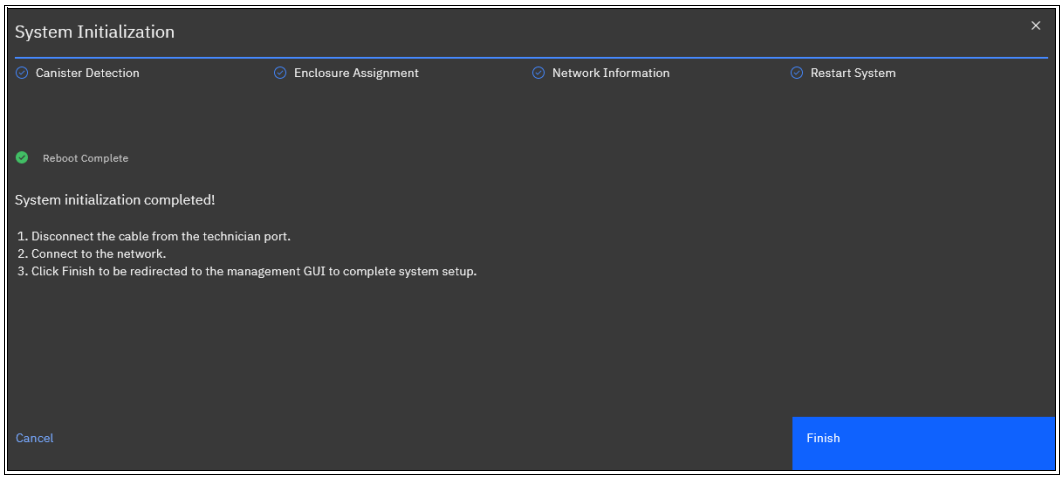


Figure 3-14 System Initialization completed

Follow the instructions, and direct your browser to the management IP address to access the system GUI after you click **Finish**.

If you decide, System Setup is also available directly from the technician port. The System Setup wizard is available through both the management IP address and the technician port.

3.3 System setup

This section provides instructions about how to define the basic settings of the system by using the System Setup wizard.

3.3.1 System Setup wizard

Complete the System Setup wizard to define the basic settings of the system. After the initialization is complete, you are redirected to a management GUI from your workstation, or you browse to the management IP address of a freshly initialized system from another workstation.

Note: Experienced users can disable the System Setup wizard and complete the configuration manually. However, this method is *not* recommended for most use cases. Consider the following points:

- To disable the System Setup wizard on a new system, run the following command:

  ```
  chsystem -easysetup no
  ```

- During the setup wizard, you are prompted to change the default *superuser* password. If the wizard is bypassed, the system blocks the configuration functions until it is changed. All attempts at configuration return the following error:

  ```
  CMMVC9473E The command failed because the superuser password must be changed
  before the system can be configured
  ```

- All configuration settings that are done by using the System Setup wizard can be changed later by using the system GUI or CLI.

The first time that you connect to the management GUI, you can be prompted to accept untrusted certificates because the system certificates are self-signed. If your company policy requests certificates that are signed by a trusted certificate authority (CA), you can install them after you complete the System Setup.

For more information about how to perform this task, see 12.1.1, “Support for non-superuser ability to manage the system” on page 1046.

To finish the System Setup wizard, complete the following steps:

1. Log in to system GUI. Until the wizard is complete, you can use only the *superuser* account, as shown in Figure 3-15. Click **Sign in**.

> **Note:** The default password for the *superuser* account is `passw0rd` (with the number zero, not the uppercase O). The default password must be changed by using the System Setup wizard or after the first CLI login. The new password cannot be set to the default password.

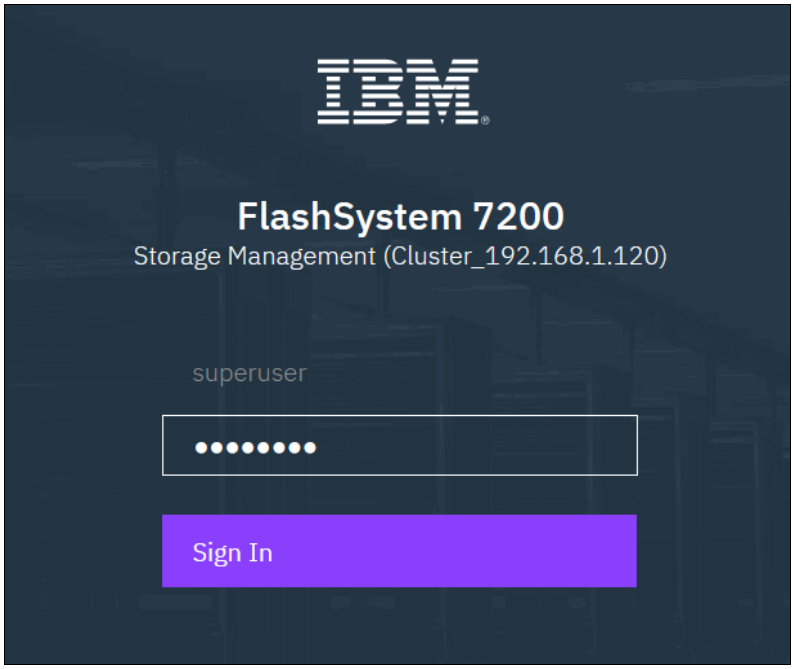


Figure 3-15 Logging in for the first time

2. The Welcome window is shown (see Figure 3-16). Click **Next**.

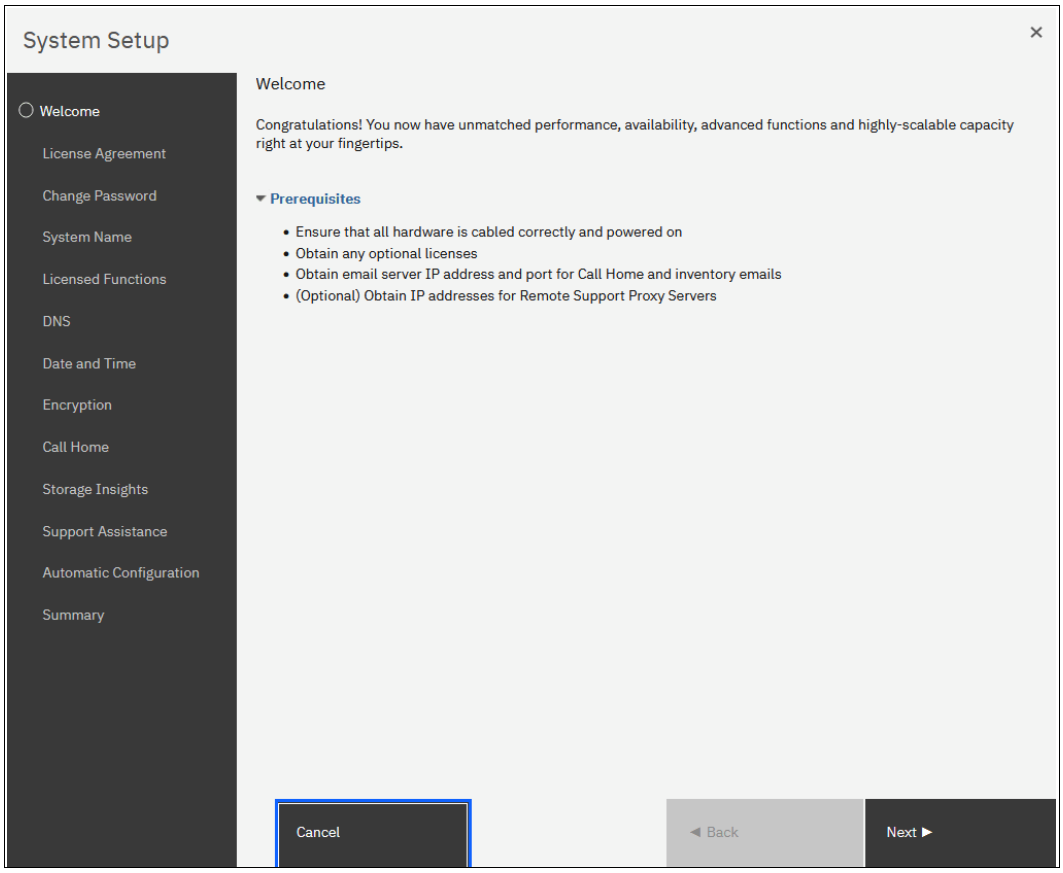


Figure 3-16 Welcome page in System Setup

3. Carefully read the license agreement. Select **I agree with the terms in the license agreement** if you want to continue the setup, as shown in Figure 3-17. Click **Next**.

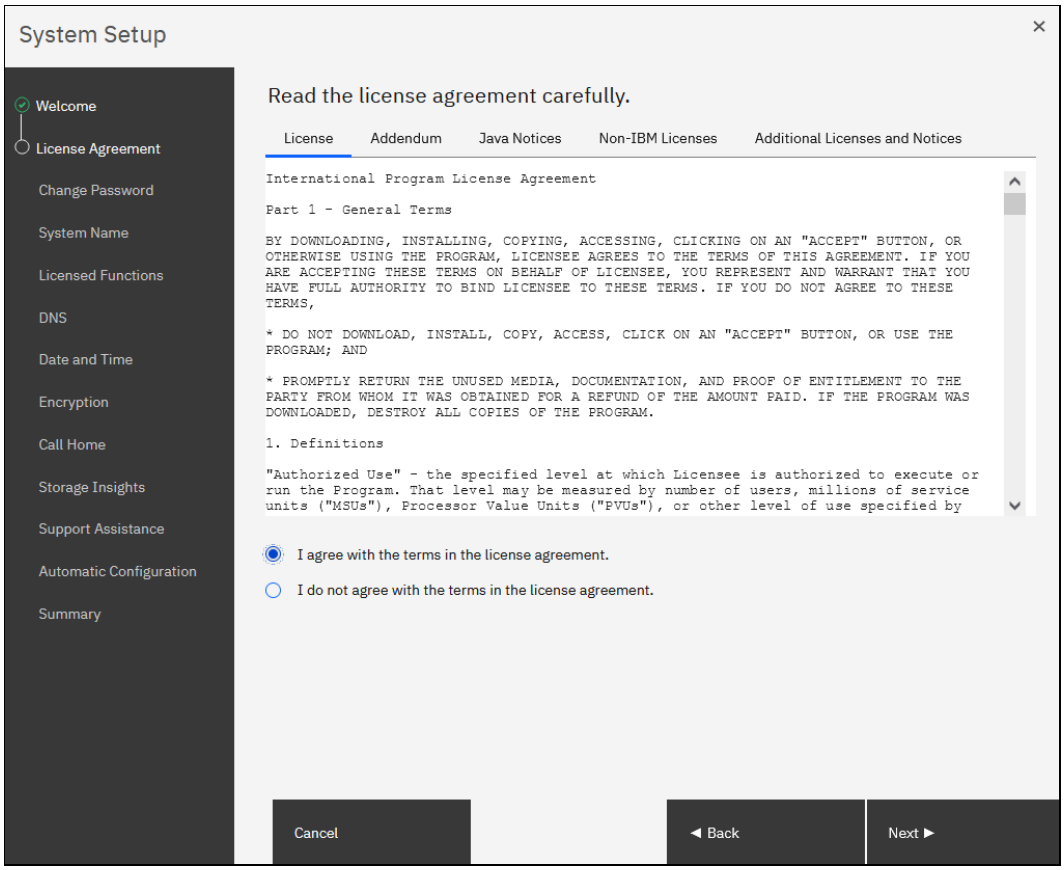


Figure 3-17 License agreement

4. Enter a new password for *superuser*, as shown in Figure 3-18 on page 190. A valid password is 8 - 64 characters and cannot begin or end with a space. Also, the password cannot be set to match the default password.

   For more information, see 12.3, “Configuring users and password policy” on page 1052. Click **Apply and Next**.

> **Note:** All configuration changes that are made by using the System Setup wizard are applied immediately, including the password change. The user sees the system running commands during the System Setup wizard.

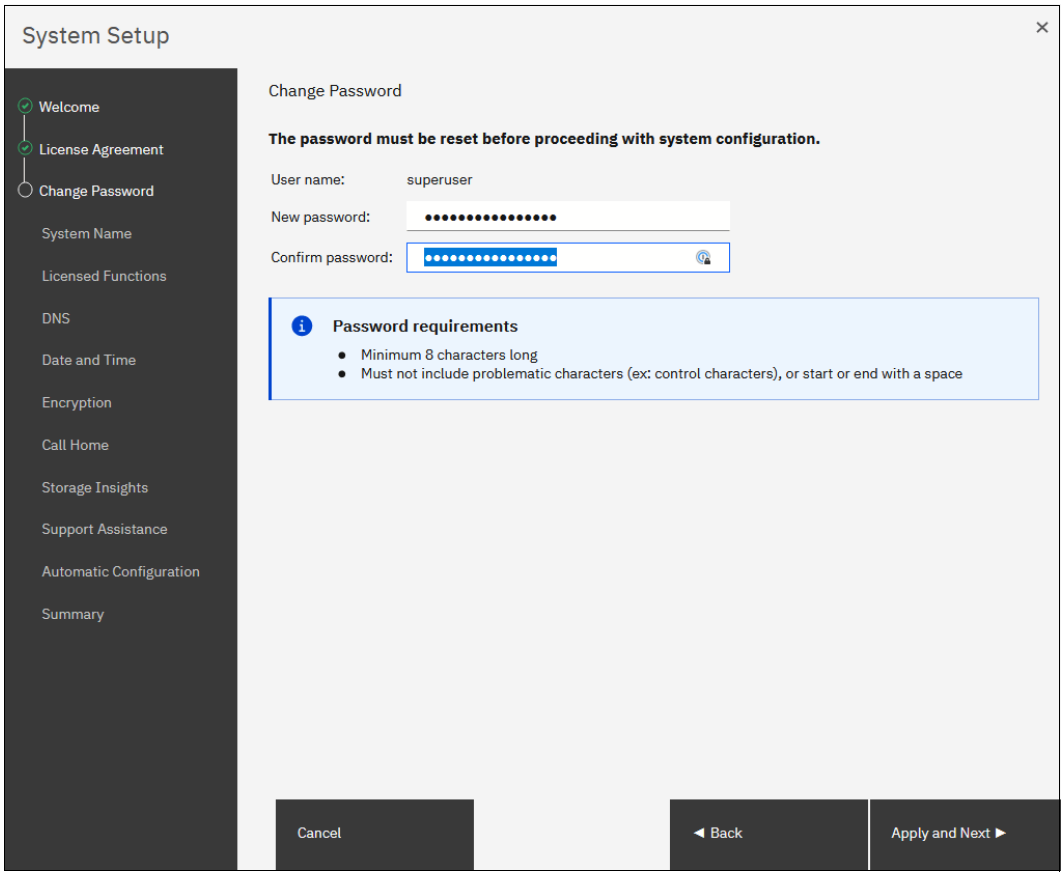


Figure 3-18 Changing the superuser password

5. Enter a name for the new system, as shown in Figure 3-19. Click **Apply and Next**.



Figure 3-19 Entering system name

Avoid the use of an underscore (_) in a system name. Although permitted here, it cannot be used in domain name server (DNS) shortnames and fully qualified domain names (FQDNs). Therefore, such naming might cause confusion and access issues. The following characters can be used: A - Z, a - z, 0 - 9, and - (hyphen).

Note: In a 3-Site Replication solution, ensure that the system name is unique for all three clusters to prepare the IBM Storage Virtualize clusters at Master, AuxNear, and AuxFar sites to work. The system names must remain different for the life of the 3-site configuration.

For more information about 3-Site Replication, see 10.14.1, “3-Site Replication Orchestrator” on page 892, or *IBM Spectrum Virtualize 3-Site Replication*, SG24-8504.

If required, the system name can be changed by running the **`chsystem -name <new_system_name>`** command. The system can also be renamed by using the management GUI by clicking **GUI** → **Monitoring** → **System Hardware** and selecting **System Actions** → **Rename System**.

6. Enter the licensed storage capacity units (SCU) and capacity for each function, as shown in Figure 3-20.

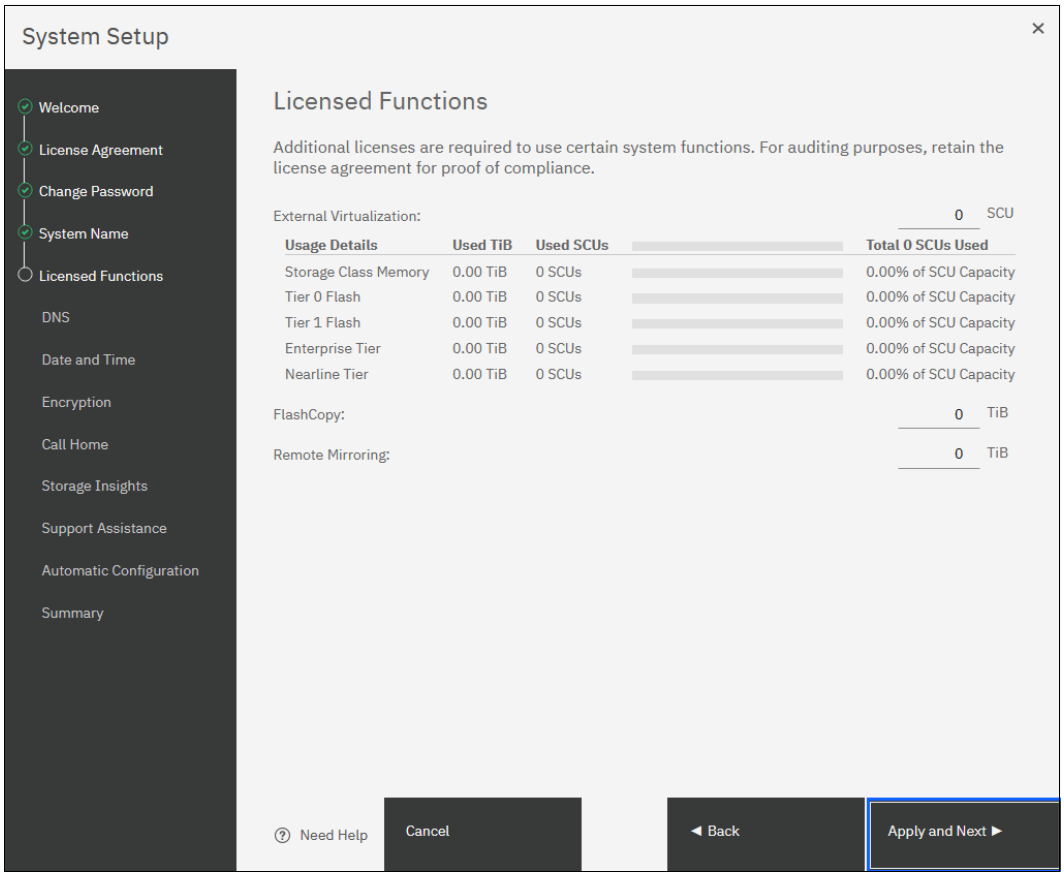


Figure 3-20 Licensed functions

The window for this step in the system setup might look different, depending on the systems that are used. Also, the way the license is enabled depends on the system that is used.

For more information about the available options, see 1.20, “Licensing” on page 115.

7. When done, click **Apply and Next**.

Note: Encryption uses a key-based licensing scheme.

8. DNS can be configured on the system, as shown in Figure 3-21. DNS helps the system to resolve the names of the computer resources that are in the external network if they are not indicated by an IP address. Click **Apply and Next** when done.

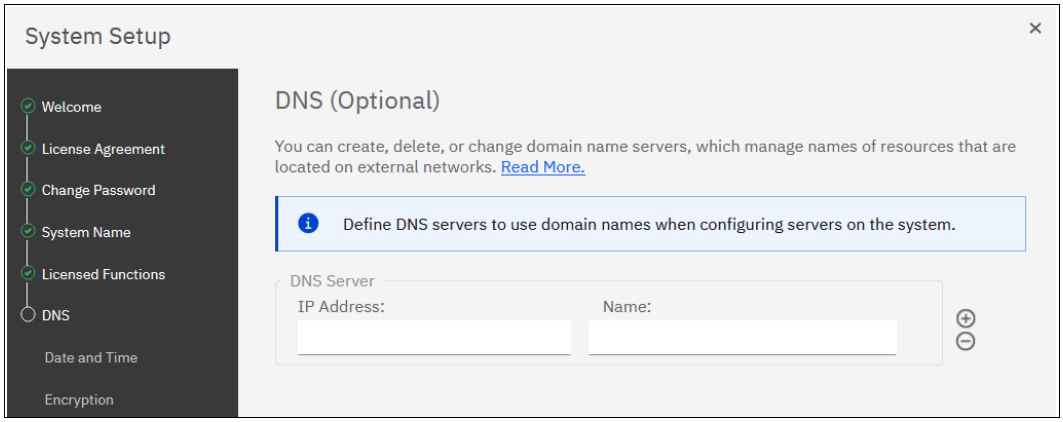


Figure 3-21 DNS server (optional)

9. Enter the date and time settings. In the example that is shown in Figure 3-22, the date and time are set by using an NTP server. Generally, use an NTP server so that all of your storage area network (SAN) and storage devices have a common timestamp. This practice facilitates troubleshooting and prevents time stamp-related errors.

   When done, click **Apply and Next**.

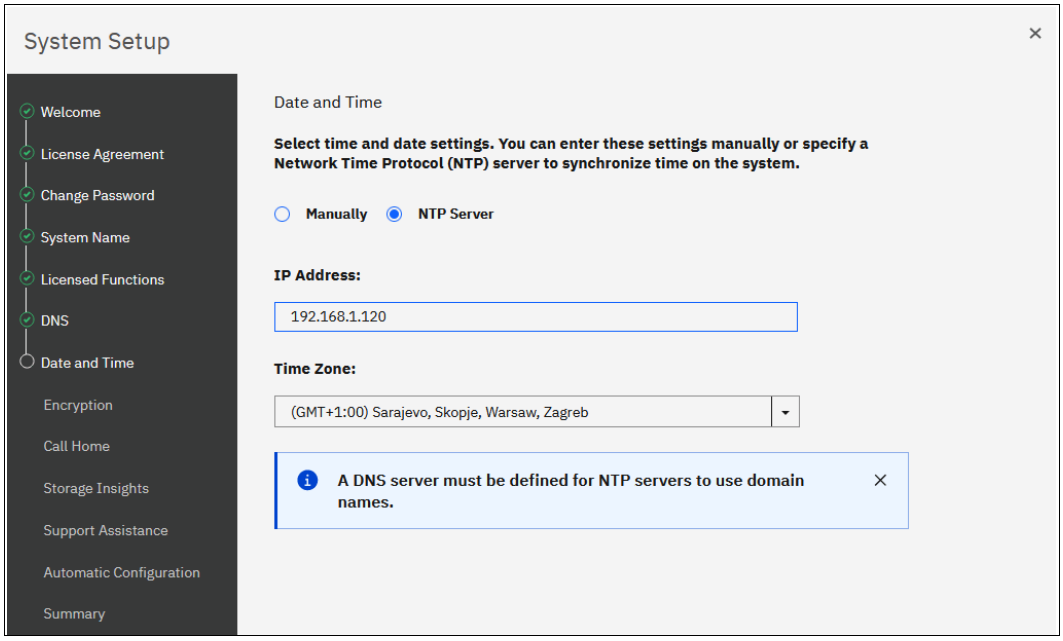


Figure 3-22 Data and time

10. The wizard prompts whether encryption was purchased. Make your selection, as shown in Figure 3-23.

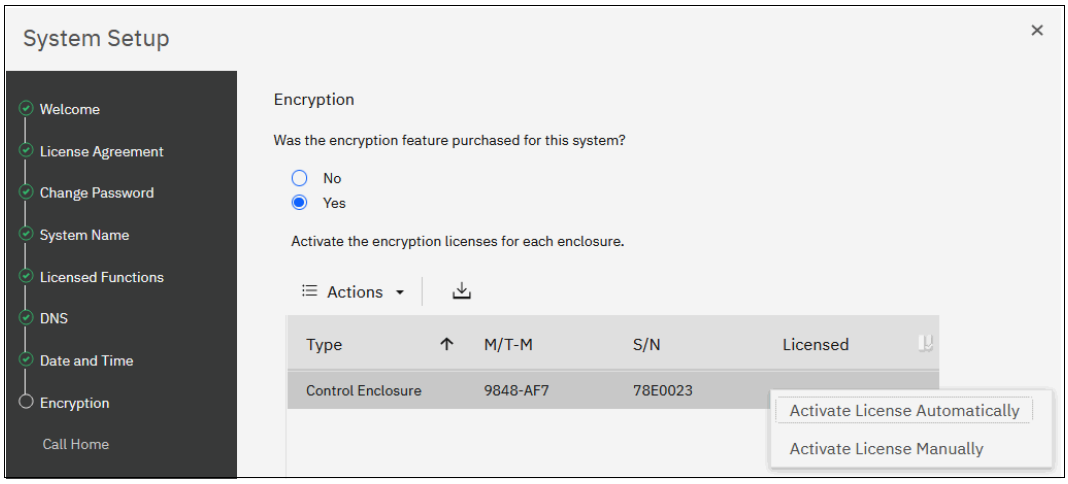


Figure 3-23 Encryption license activation

If encryption is not planned now, select **No** and then click **Next**. You can enable this feature later, as described in 12.8, “Activating and enabling encryption” on page 1103.

Note: When encryption is enabled on the system, encrypted storage pools can be created. If the system is a single control enclosure system where all FCM-drives should be in the same storage pool, encryption must be enabled before creating the storage pool. If a storage pool is created before encryption is enabled, any data in that pool must be migrated to an encrypted storage pool, if the data must be encrypted.

If you purchased the encryption feature, you are prompted to activate your license manually or automatically. The encryption license is key-based and required for each control enclosure.

You can use automatic activation if the workstation that you use to connect to the GUI and run the System Setup wizard has Internet access. If no Internet connection is available, use manual activation and follow the instructions.

After the encryption license is activated, you see a green checkmark for each enclosure, as shown in Figure 3-24. After all the control enclosures show that encryption is licensed, click **Next**.

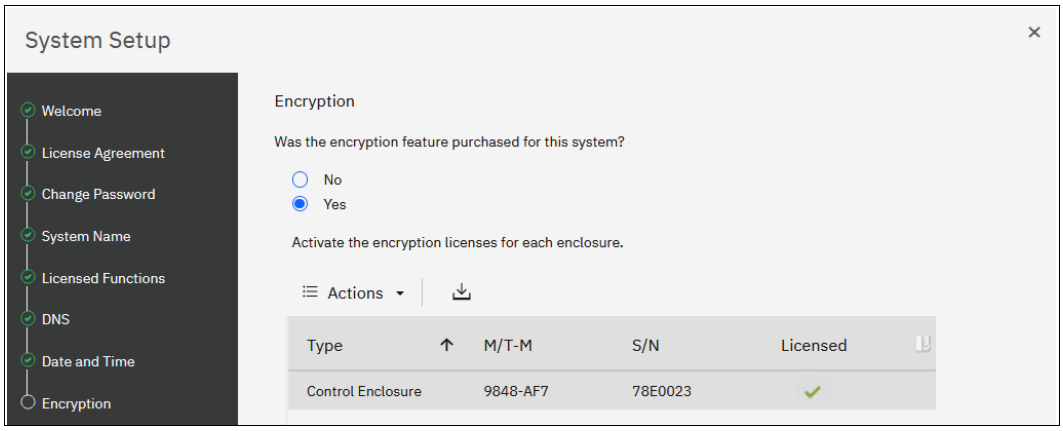


Figure 3-24 Encryption licensed

11. Call Home provides multiple benefits. It enables automatically creating tickets at IBM if errors occur, which improves the speed and efficiency by which calls are handled. Call Home also enables Storage Insights and Remote Support.

    Click **Next** at the Call Home information window, as shown in Figure 3-25.

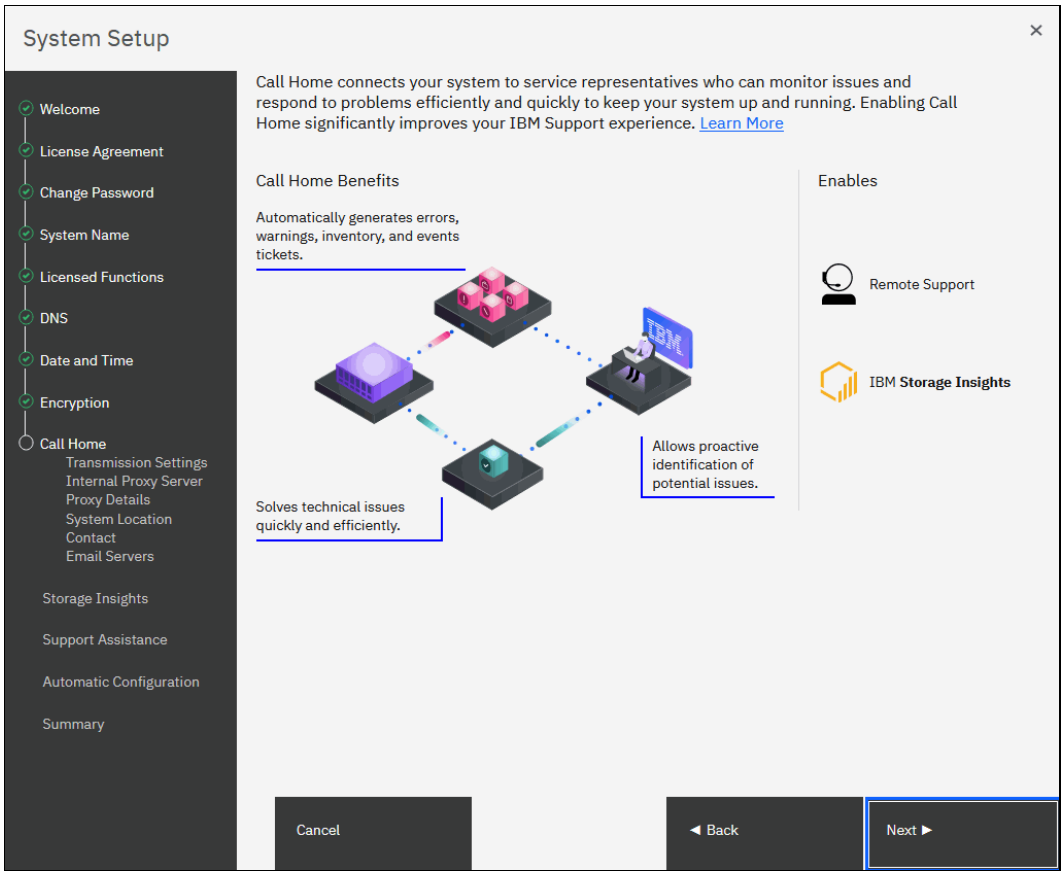


Figure 3-25 Call Home information window

On IBM FlashSystem 9500 systems and IBM SAN Volume Controller systems, an IBM SSR configures Call Home during installation. Verify that all the entered data is correct.

All IBM FlashSystem products and IBM SAN Volume Controller systems support the following methods of sending Call Home notifications to IBM:

- Cloud Call Home
- Call Home with email notifications

Cloud Call Home is the default and preferred option for a system to report event notifications to IBM Support. With this method, the system uses RESTful application programming interfaces (APIs) to connect to an IBM centralized file repository that contains troubleshooting information that is gathered from customers. This method requires no extra configuration.

The system can also be configured to use email notifications for this purpose. If this method is selected, you are prompted to enter the SMTP server IP address.

If both methods are enabled, Cloud Call Home is used, and the email notifications method is kept as a backup.

For more information about setting up Call Home, including Cloud Call Home, see 11.8, “Monitoring and Event Notification” on page 1006.

If either of these methods is selected, the system location and contact information must be entered. This information is used by IBM to provide technical support. All fields in the form must be completed. In this step, the system also verifies that it can contact the Cloud Call Home servers.

12. System Setup prompts to select which transmission types to be used for Call Home. As shown in the example in Figure 3-26 on page 196, both options were selected. Select the required options and click **Apply and Next**.

> **Note:** It is *not* recommended to select **I don’t want to use Call Home**.

13. Enter the proxy server details, if required, to enable Call Home connectivity (Figure 3-27 on page 196). Click **Next** to proceed.

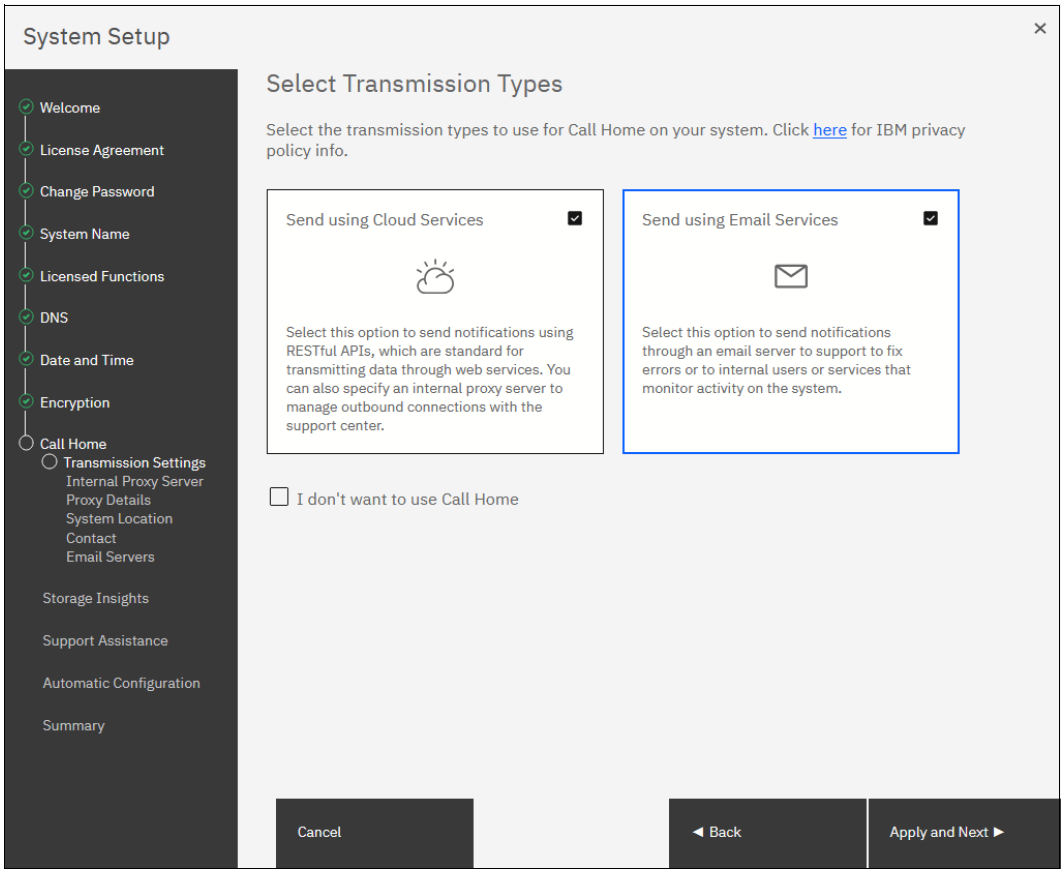


Figure 3-26 Call Home transmission type

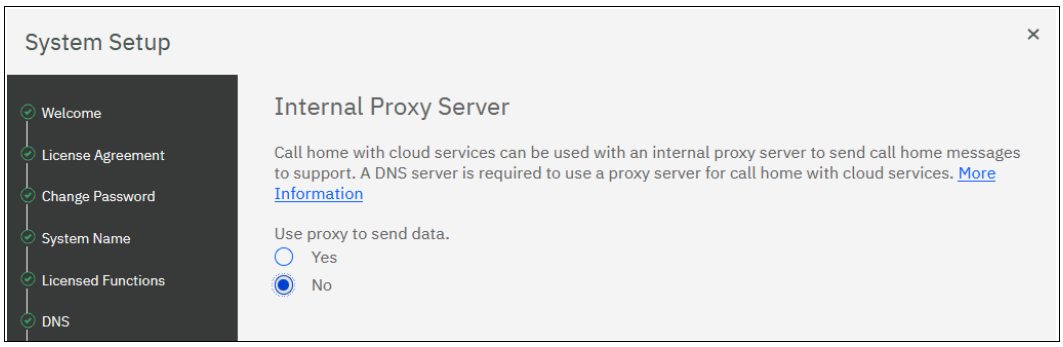


Figure 3-27 Proxy server (optional)

14. System Setup now prompts for the system location, as shown in Figure 3-28. Enter all of the information and click **Next**.

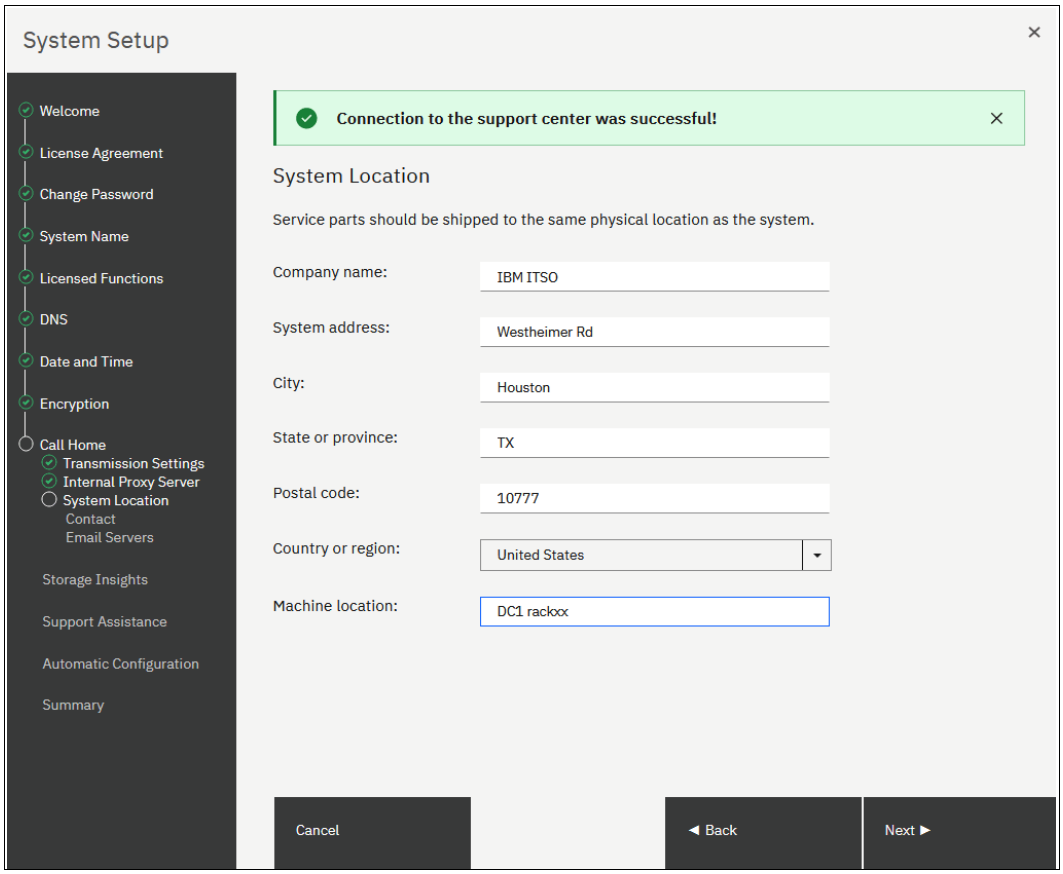


Figure 3-28 System location

15. Enter the contact information in the Contact window, as shown in Figure 3-29. Click **Apply and Next**.

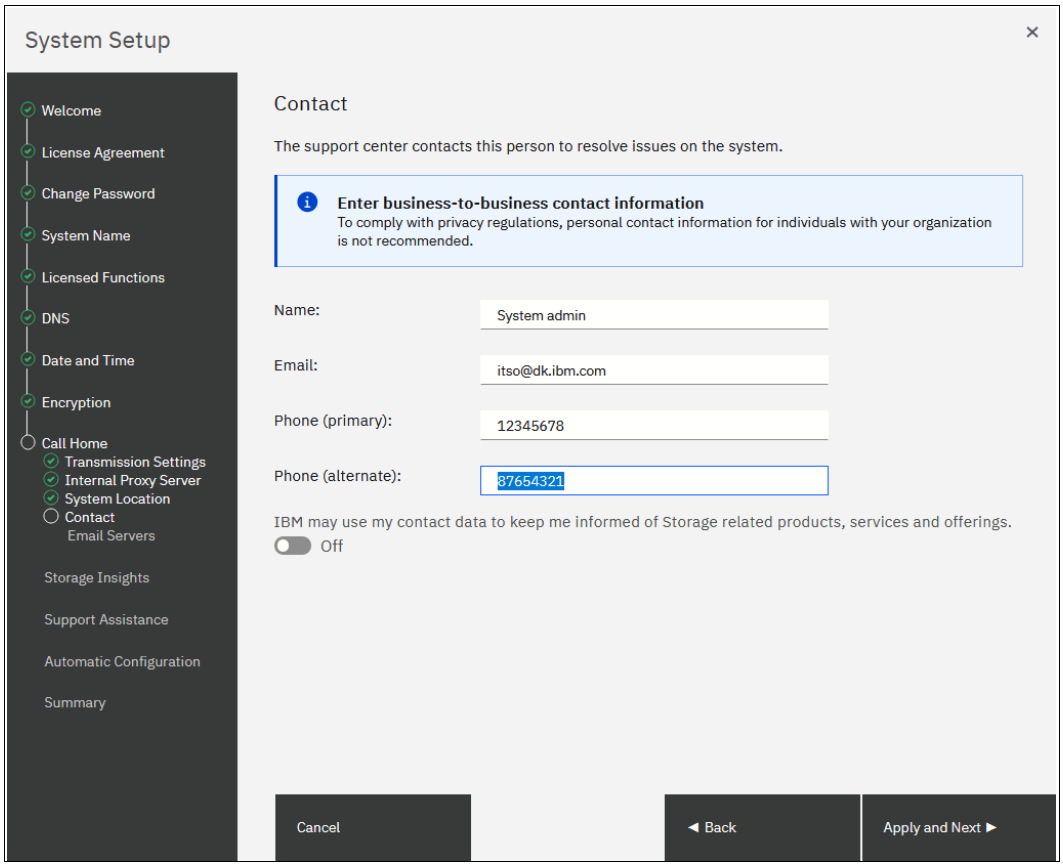


Figure 3-29 Setting up contact information for Call Home

16. If email was selected for Call Home, the user must provide the IP address of one or more email servers, as shown in Figure 3-30. Enabling email for Call Home also enables the possibility for the storage administrators to receive emails if warnings or errors occur. Enter the IP address of one or more servers and click **Apply and Next** to proceed.

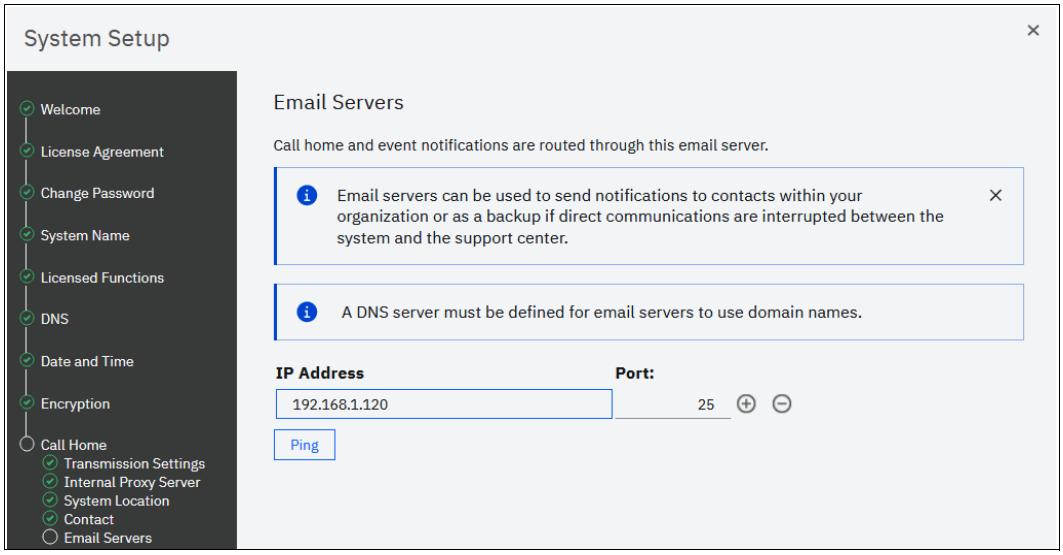


Figure 3-30 Email servers

17. IBM FlashSystem family systems and IBM SAN Volume Controller systems can be used with IBM Storage Insights, which is an IBM cloud storage monitoring and management tool. During this setup phase, the system attempts to contact the IBM Storage Insights web service. If it is available, you are prompted to sign up (see Figure 3-31). System Setup now provides information about how to enable Storage Insights. Click **Next** to continue.

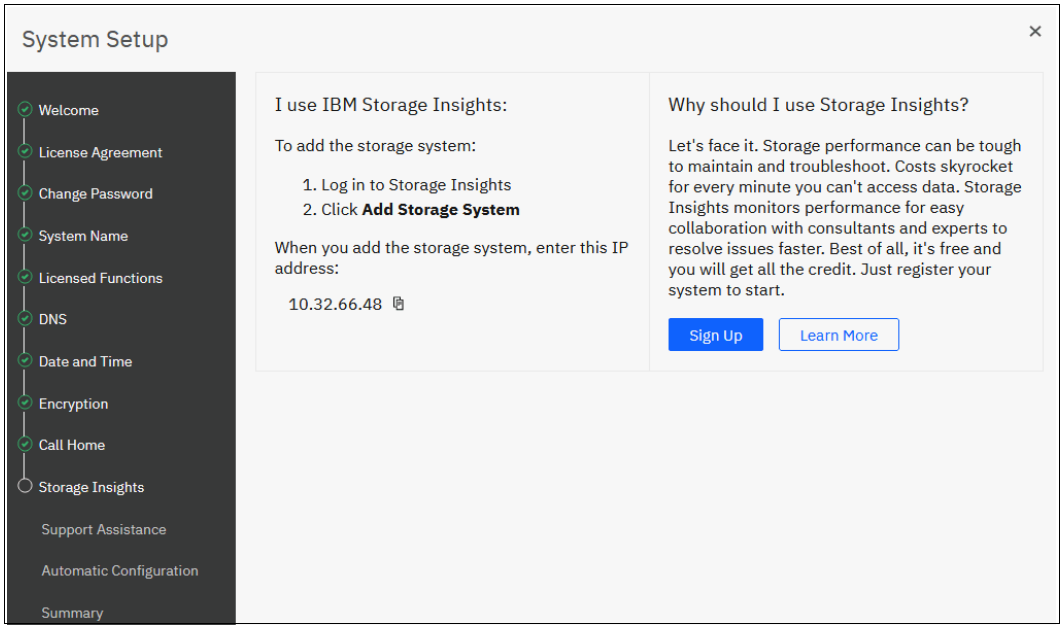


Figure 3-31 Storage Insights information window

18. If you enabled at least one Call Home method, the Support Assistance configuration window opens (see Figure 3-32). The Support Assistance function requires Call Home; therefore, if it is disabled, Support Assistance cannot be used. Click **Next** to continue.

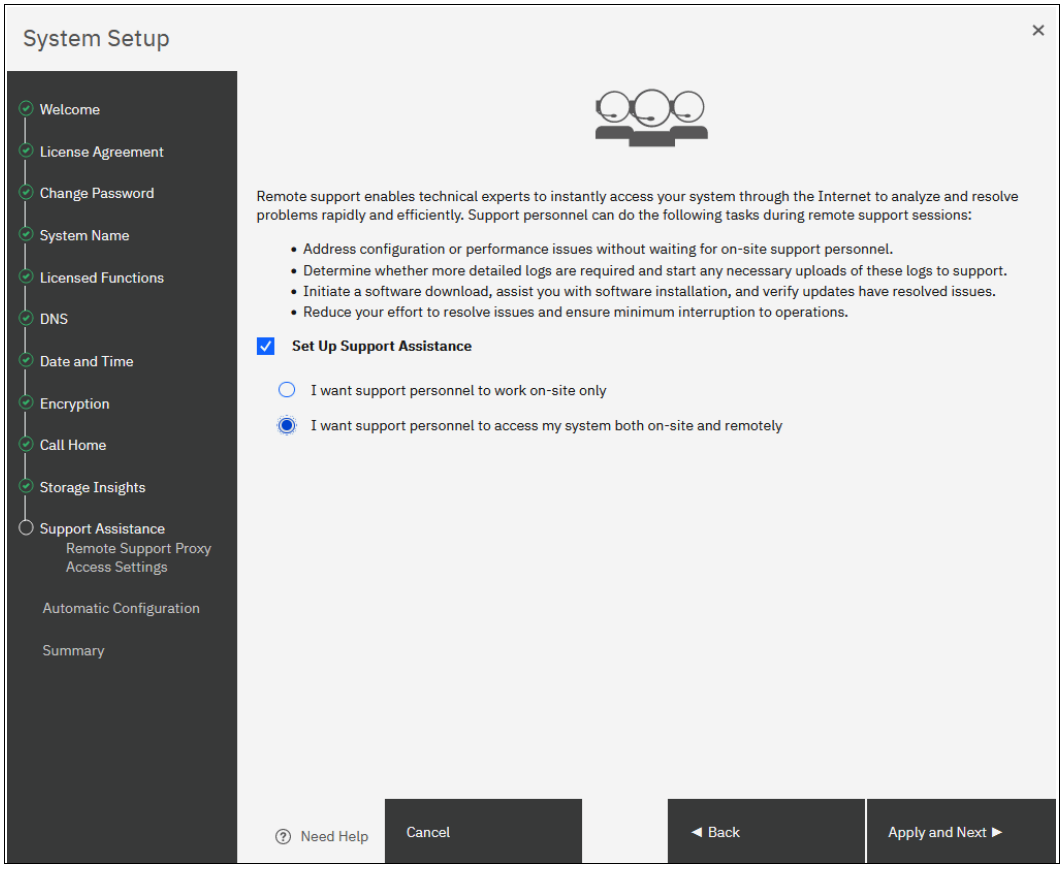


Figure 3-32 Support assistance

With the Support Assistance feature, you allow IBM Support to perform maintenance tasks on your system with support personnel onsite or remotely.

If an IBM SSR is onsite, the SSR can log in locally with your permission and a special user ID and password so that a superuser password does not need to be shared with the IBM SSR.

You can also enable Support Assistance with remote support to allow IBM Support personnel to log in remotely to the machine with your permission through a secure tunnel over the Internet.

For more information about the Support Assistance feature, see 11.8.2, “Remote Support Assistance” on page 1015.

19. If you allow remote support, you are provided with the IP addresses and ports of the remote support centers and an opportunity to provide proxy server details (if required) to allow the connectivity, as shown in Figure 3-33. Click **Apply and Next**.

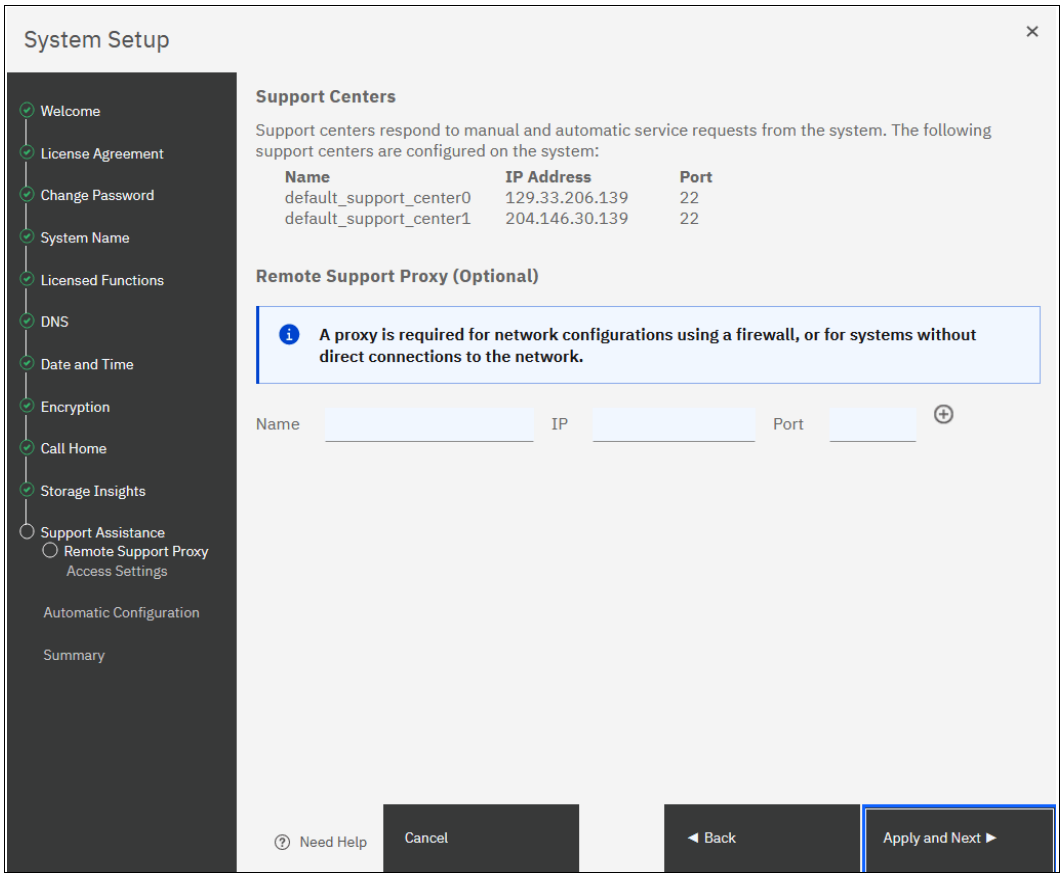


Figure 3-33 System communicating with named IBM Support servers

You can also allow remote connectivity at any time or only after obtaining permission from the storage administrator, as shown in Figure 3-34. Click **Apply and Next**.

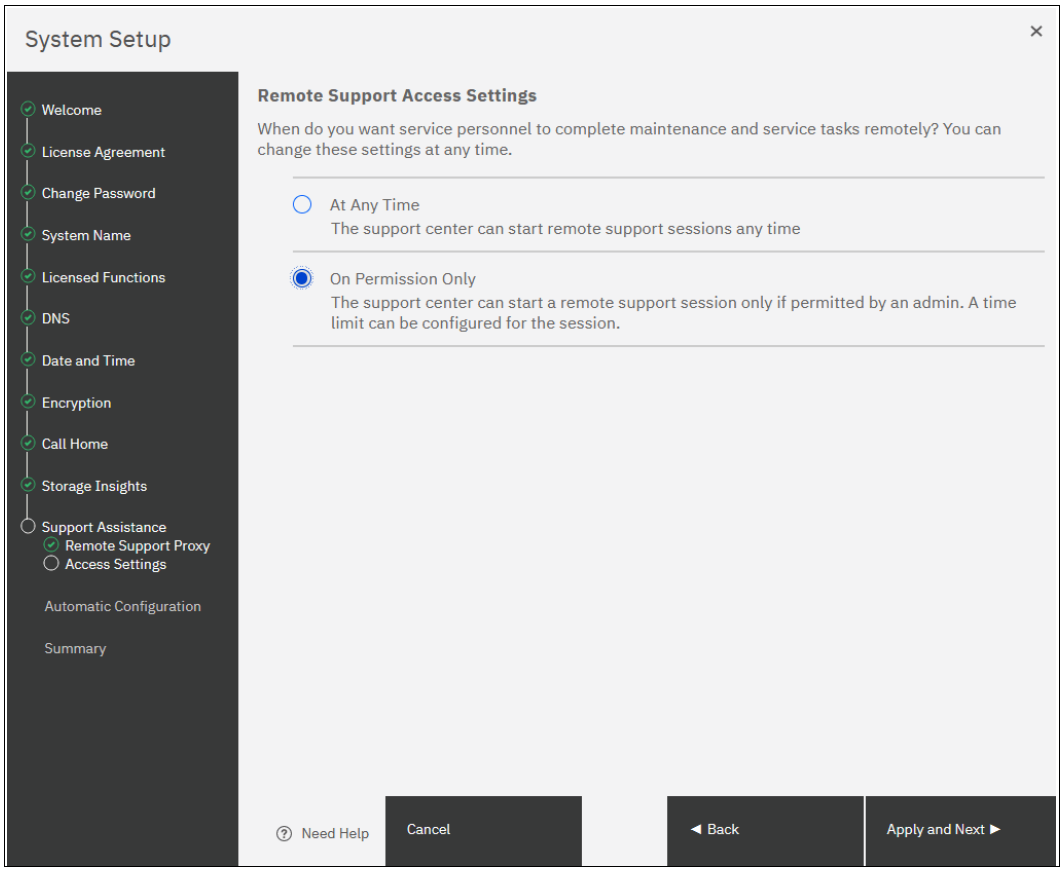


Figure 3-34 Remote support access settings

20. System Setup offers an option (IBM FlashSystem products only) to automatically configure the system if it is used as FC-attached, back-end storage for IBM SAN Volume Controller. If you plan to use the system in stand-alone mode (that is, not behind an IBM SAN Volume Controller), leave Automatic Configuration turned off, as shown in Figure 3-35. Click **Next** to continue.

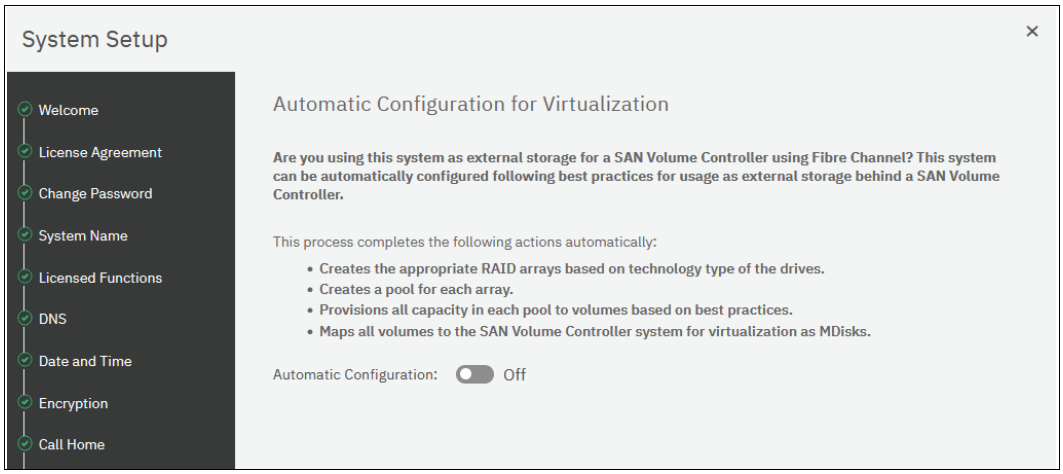


Figure 3-35 Automatic configuration for IBM SAN Volume Controller

For more information about how to enable Automatic configuration for IBM SAN Volume Controller on a running system after the System Setup wizard, see 3.3.8, “Automatic configuration for IBM SAN Volume Controller back-end storage” on page 226.

21. On the Summary page, the settings that were selected by the System Setup wizard are shown. If corrections are needed, you can return to a previous step by clicking **Back**. Otherwise, click **Finish** to complete the system setup wizard, as shown in Figure 3-36.

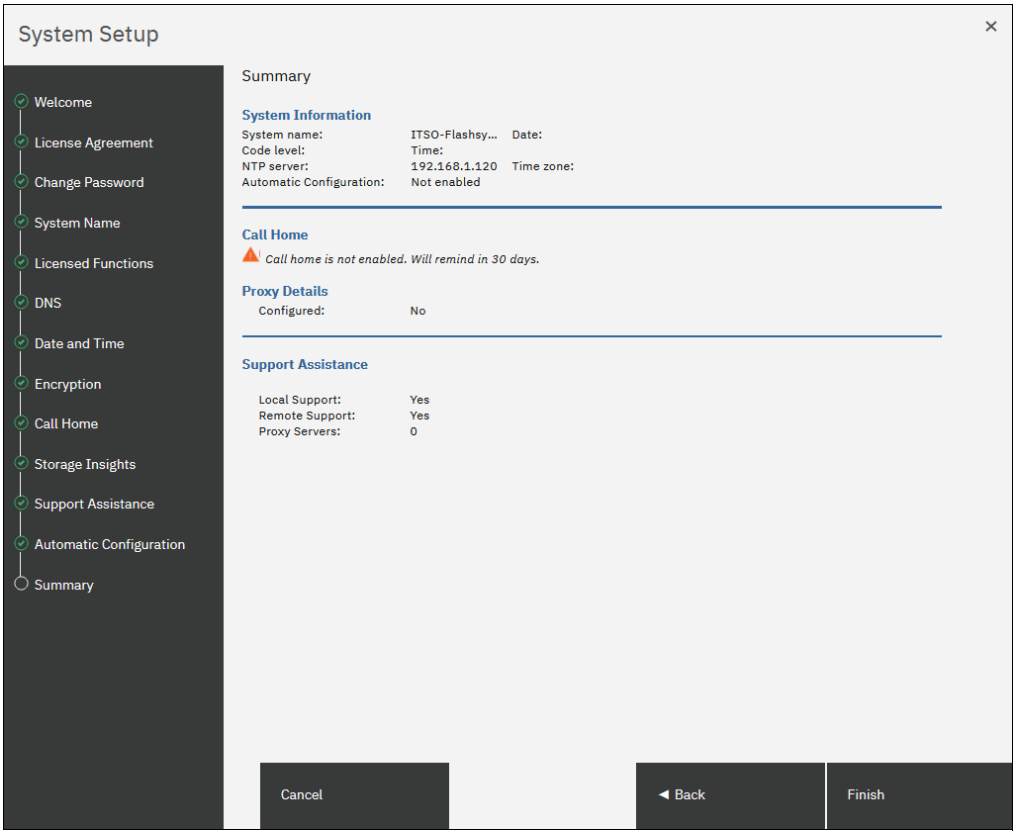


Figure 3-36 Summary page

Note: Call Home and Support Assistance cannot be enabled if the system cannot send data to IBM Support.

22. When the system setup wizard completes, your IBM FlashSystem consists only of the control enclosure that includes the node canister that you used to initialize the system and its partner, and the expansion enclosures that are attached to them.

    In the case of an IBM SAN Volume Controller, your system consists of only one node in the cluster, which might see other candidate nodes in the service GUI if they are connected to SAN and zoned together.

    If you have other control and expansion enclosures or IBM SAN Volume Controller nodes, you must add them to complete the System Setup.

    For more information about how to add a control or expansion enclosure, see 3.3.3, “Adding an enclosure in IBM FlashSystem” on page 210.

    For more information about how to add a node or hot spare node, see 3.3.4, “Adding a node or hot spare node in IBM SAN Volume Controller systems” on page 212.

If no other enclosures or nodes are to be added to this system, the System Setup process is complete and you can click **Finish** to be returned to the login window of the IBM FlashSystem.

All the required steps of the initial configuration are complete. If needed, you can configure other global functions, such as system topology, user authentication, or local port masking before configuring the volumes and provisioning them to hosts.

Click **Close** as shown in Figure 3-37.

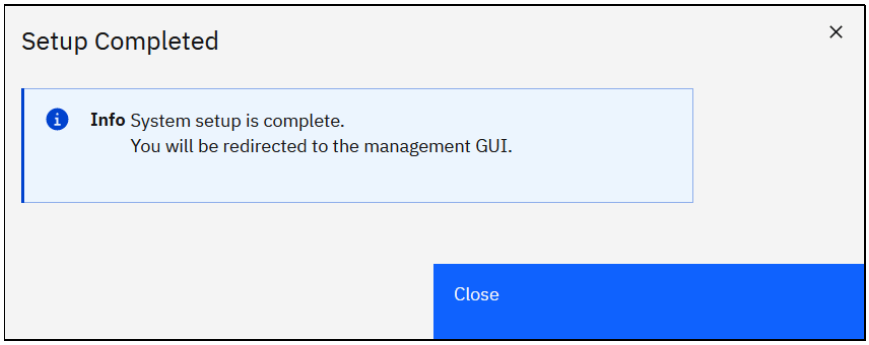


Figure 3-37 System Setup: Setup completed

23. After completing all steps of the System Setup wizard, the system GUI opens, as shown in Figure 3-38.

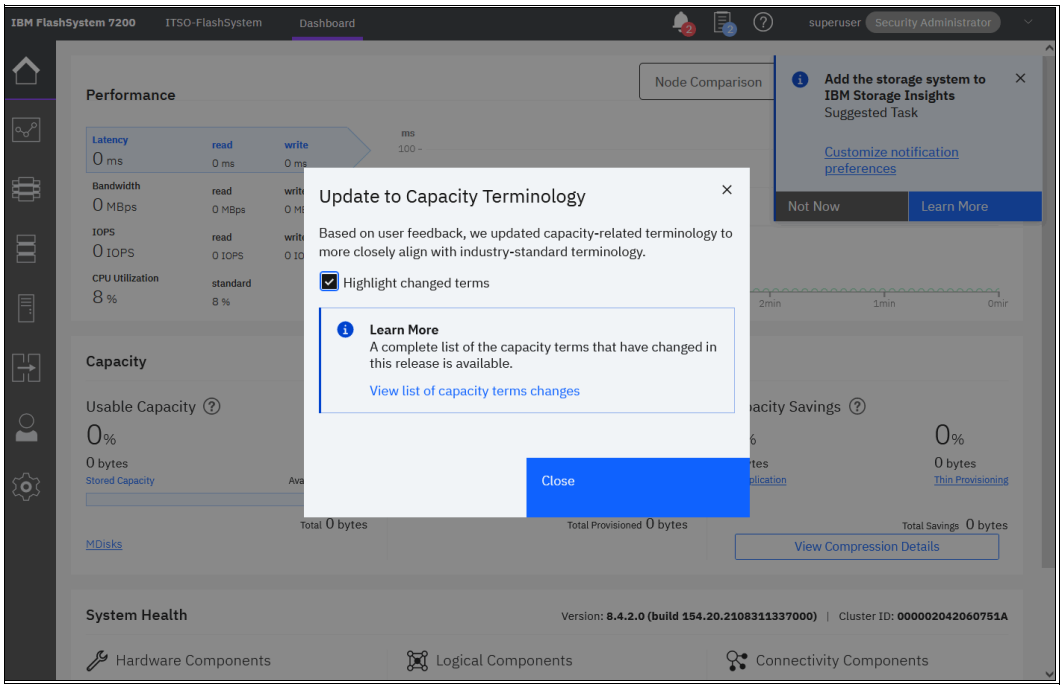


Figure 3-38 System Setup directs the user to the system GUI Base configuration

The tasks that are described next are used to define global system configuration settings. Often, they are performed during the System Setup process. However, they can also be performed later, such as when the system is expanded or the system environment is reconfigured.

3.3.2 Configuring clustering by using Ethernet connections

The system supports node-to-node connections that use Ethernet protocols that support remote direct memory access (RDMA) technology, such as RDMA over Converged Ethernet (RoCE) or iWARP. To use these protocols, the system requires that an RDMA-capable adapter is installed on each node, and dedicated RDMA-capable Ethernet ports are configured only for node-to-node communication.

RDMA technologies, such as RoCE and iWARP, enable the RDMA-capable adapter to transfer data directly between nodes, which bypasses CPU and caches and makes transfers faster. RDMA technologies provide faster connection and processing time than traditional iSCSI connections.

Up to four I/O groups can be joined in an IBM HyperSwap, a standard topology cluster, or an enhanced stretched cluster (IBM SAN Volume Controller only). This section describes the configuration steps that must be performed if a system is designed for IP-based RDMA node-to-node traffic.

For FC SAN clustering, no special configuration is required on the system. However, the SAN must be set up as described in Chapter 2, “Installation and configuration planning” on page 119.

Planning clustered systems over RDMA and TCP-based Ethernet

Before implementing, select one of the following inter-Switch link (ISL) configurations based on the need to support systems:

No ISL

In this network configuration, no inter-switch links are used between the switches across two sites. With no inter-switch links between systems, the recommended operational range of this configuration is limited to 300 meters. This configuration is best suitable for small-scale enterprises for achieving high availability in a cost-effective manner. A graphical representation of No ISL network connections is shown in Figure 3-39

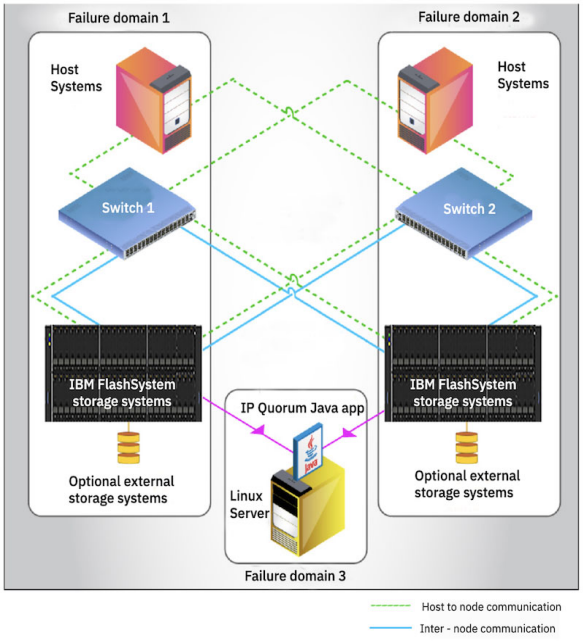


Figure 3-39 No ISL connectivity

Shared ISL

This network configuration uses shared inter-switch links, which carry inter-node and host-node traffic with other local area network (LAN) traffic. At times, the mixed traffic can cause congestion at ISL. However, configure priority flow control (PFC) for inter-node and host-node traffic to avoid congestion and system instability due to multiple traffics at ISL. Also, ensure that the bandwidth of ISL is adequate to support all the network traffic flowing through it. This configuration is used for the HyperSwap systems that involve multiple traffics from various subnet and VLANs in the network. A graphical representation of Shared ISL network connections is shown in Figure 3-40

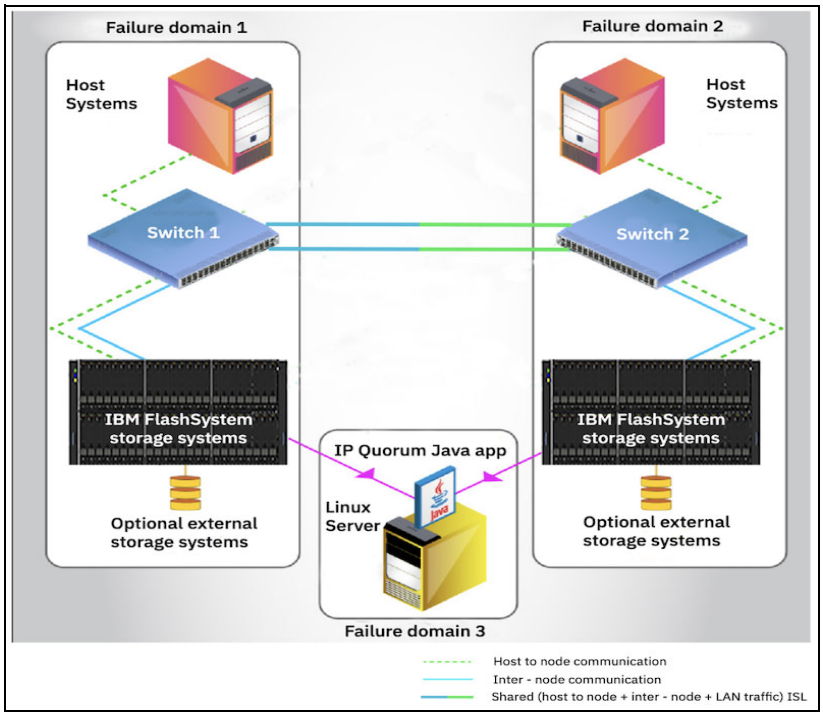


Figure 3-40 Shared ISL connectivity

Note: PFC (Priority Flow Control) is not supported when you use NVMe over TCP.

Dedicated ISL

This configuration uses dedicated inter-switch links, which carry only inter-node and host-node traffic for the HyperSwap systems. In this configuration, host-node and inter-node traffic are isolated, so ISL bandwidth is not affected. A dedicated ISL configuration can prevent decreased system performance and slower host response times. However, setting up this configuration incurs extra cost. A graphical representation of dedicated ISL network connections is shown in Figure 3-41 on page 207

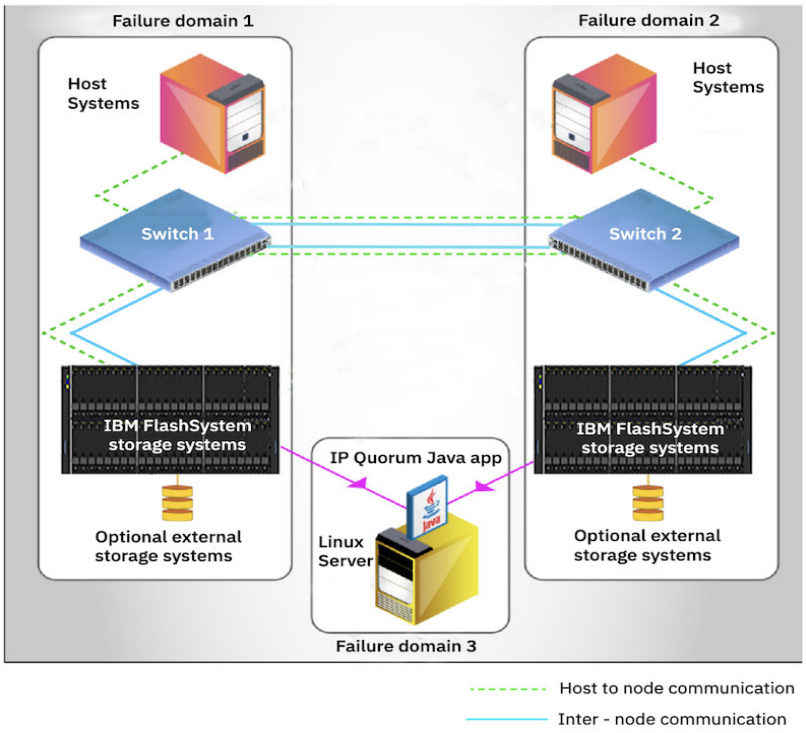


Figure 3-41 Dedicated ISL connectivity

Prerequisites

Before RDMA clustering is configured, ensure that the following prerequisites are met:

- 25 gigabits per second (Gbps) RDMA-capable Ethernet cards are installed in each node.
- RDMA-capable adapters in all nodes use the same technology, such as RDMA over Converged Ethernet (RoCE) or internet Wide Area RDMA Protocol (iWARP).
- RDMA-capable adapters are installed in the same slots across all the nodes of the system.
- Ethernet cables between each node are connected correctly.
- The network configuration does not contain more than two hops in the fabric of switches. The router must *not* be placed between nodes that use RDMA-capable Ethernet ports for node-to-node communication.
- The negotiated speeds on the local and remote adapters are the same.
- The local and remote port (RPORT) virtual local area network (VLAN) identifiers are the same. All the ports that are used for node-to-node communication must be assigned to one VLAN ID, and ports that are used for host attachment must have a different VLAN ID.

  If you plan to use VLAN to create this separation, you must configure VLAN support on all the Ethernet switches in your network before you define the RDMA-capable Ethernet ports on nodes in the system. On each switch in your network, set the VLAN to Trunk mode and specify the VLAN ID for the RDMA-ports that is to be in the same VLAN.
- A minimum of two dedicated RDMA-capable Ethernet ports are required for node-to-node communications to ensure best performance and reliability. These ports must be configured for inter-node traffic only and must not be used for host attachment, virtualization of Ethernet-attached external storage, or IP replication traffic.
- A maximum of four RDMA-capable Ethernet ports per node are allowed for node-to-node communications.

Configuring node port IP addresses

To enable RDMA clustering, IP addresses must be configured on each port of each node that is used for node-to-node communication. Complete the following steps:

1. Connect to a Service Assistant of a node by browsing to `https://<node_service_IP>/service`. Then, select a node and click **Change Node IP** (see Figure 3-42).

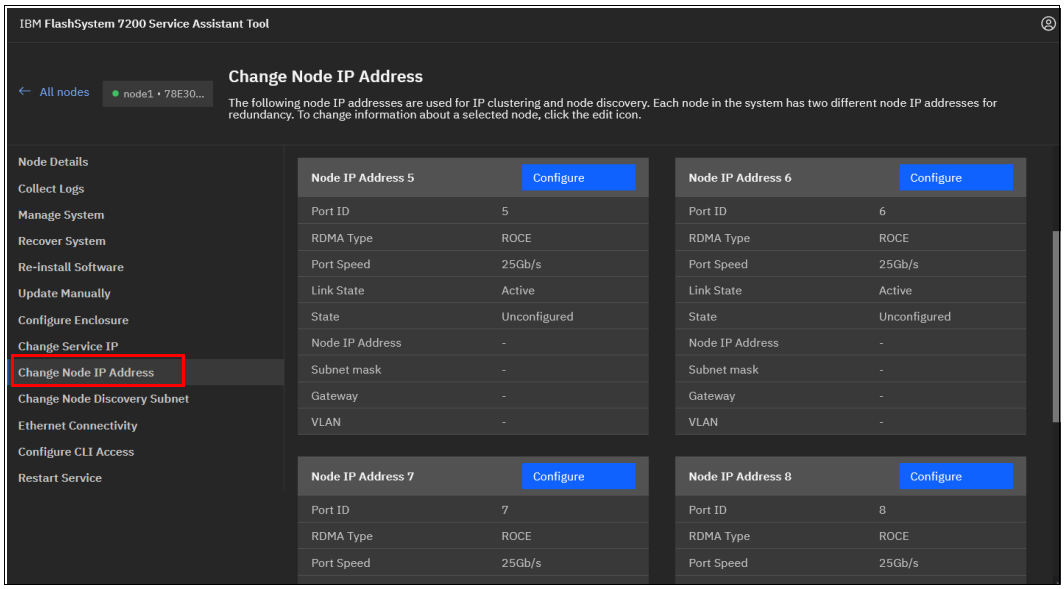


Figure 3-42 Node IP address setup for Remote Direct Memory Access clustering

2. Hover over a tile with a port and click **Configure** to set the IP address, netmask, gateway address, and VLAN ID for a port. The IP address for each port must be unique and cannot be used anywhere else on the system. The VLAN ID for ports that are used for node-to-node traffic must be the same on all nodes.

When the required information is entered, click **Save** and verify that the operation finished successfully, as shown in Figure 3-43. Repeat this step for all ports that you intend to use for node-to-node traffic, with a minimum of two and a maximum of four ports per node.

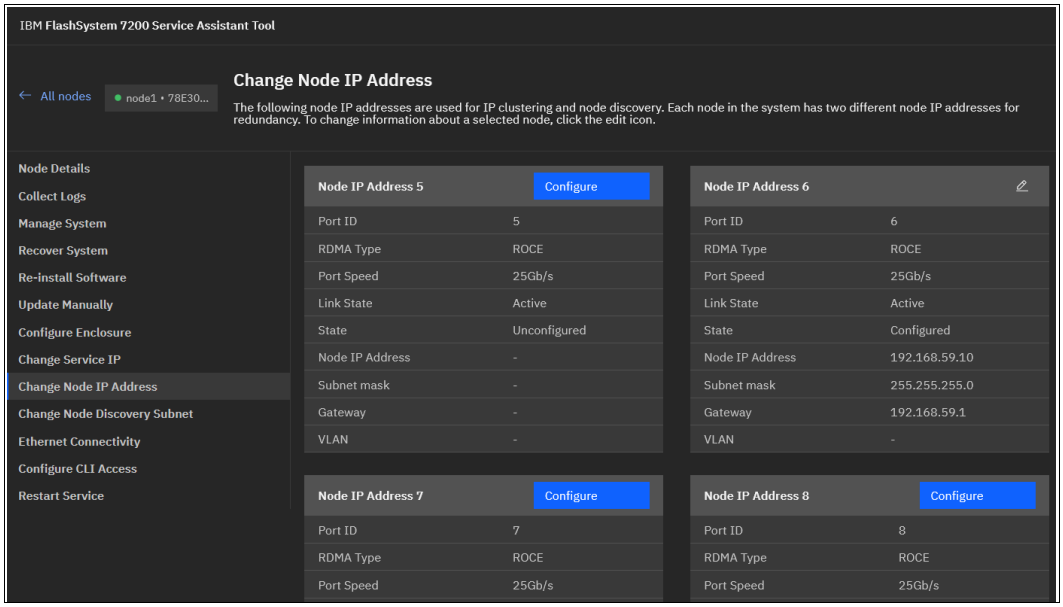


Figure 3-43 Node IP addresses configured

To list the node IP configuration by using the CLI, run the **`sainfo lsnodeip`** command, as shown in Example 3-2.

Example 3-2 Listing node IPs currently not set

```
ITSO-FlashSystem:superuser>sainfo lsnodeip
port_id    rdma_type port_speed vlan link_state state        node_IP_address gateway subnet_mask
1                                   inactive   unconfigured
2                                   inactive   unconfigured
3                                   inactive   unconfigured
4                                   inactive   unconfigured
5          RoCE      25Gb/s          active     unconfigured
6          RoCE      25Gb/s          active     unconfigured
7          RoCE      25Gb/s          active     unconfigured
8          RoCE      25Gb/s          active     unconfigured
9          RoCE      25Gb/s          active     unconfigured
10         RoCE      25Gb/s          active     unconfigured
```

Run the **`satask chnodeip`** commands to change node IP by using CLI, as shown in Example 3-3.

Example 3-3 Running commands to change node IP

```
superuser>satask chnodeip -ip 10.0.99.12 -gw 10.0.99.20 -mask 255.255.255.0 -port_id 5
superuser>satask chnodeip -ip 192.168.59.11 -gw 192.168.2.120 -mask 255.255.255.0 -port_id 6
```

To list the changed node IP, run the **`sainfo lsnodeip`** again, as shown in Example 3-4.

Example 3-4 Changed node IP (output shortened for clarity)

```
ITSO-FlashSystem:superuser>sainfo lsnodeip
port_id    rdma_type port_speed vlan link_state state        node_IP_address
1                                   inactive   unconfigured
2                                   inactive   unconfigured
3                                   inactive   unconfigured
4                                   inactive   unconfigured
```

```
5          RoCE      25Gb/s          active     configured   10.0.99.12
6          RoCE      25Gb/s          active     configured   192.168.59.11
7          RoCE      25Gb/s          active     unconfigured
8          RoCE      25Gb/s          active     unconfigured
9          RoCE      25Gb/s          active     unconfigured
10         RoCE      25Gb/s          active     unconfigured
```

3. Some environments might not include a stretched layer 2 subnet. In such scenarios, a layer 3 network (such as in standard topologies or long-distance RDMA node-to-node HyperSwap configurations) is applicable. To support the layer 3 Ethernet network, use the unicast discovery method for RDMA node-to-node communication. This method relies on unicast-based fabric discovery rather than multicast discovery.
4. To configure unicast discovery, use the **`satask addnodediscoverysubnet`**, **`satask rmnodediscoverysubnet`**, or **`sainfo lsnodediscoverysubnet`** commands. For more information, see Command-line interface.

   You can also configure discovery subnets by using the Service Assistant interface menu option **Change Node Discovery Subnet**, as shown in Figure 3-44.

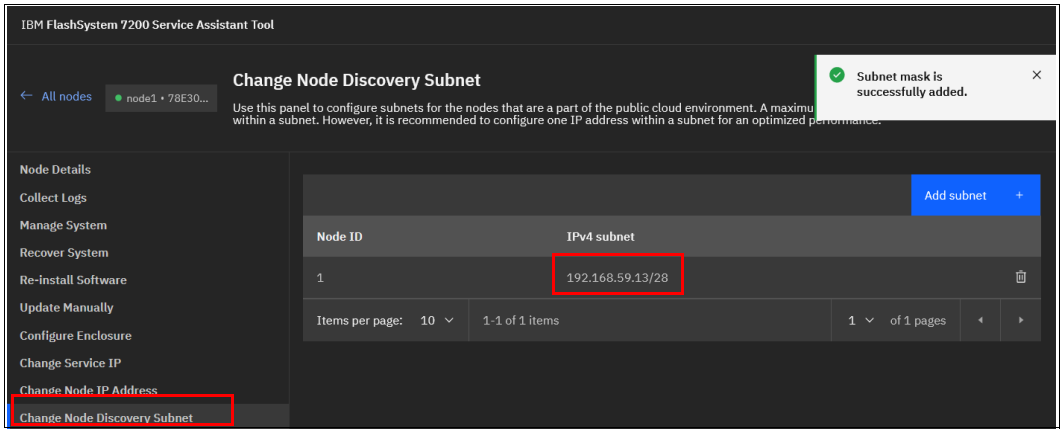


Figure 3-44 Setting the node discovery subnet

After the IP addresses are configured on all nodes in a system and the nodes to be partner nodes, from the Service Assistant GUI, navigate to **Ethernet Connectivity** to view which nodes are visible to the system.

Alternatively, run the **`sainfo lsnodeipconnectivity`** CLI command to verify that the partner nodes are visible on the IP network.

5. When all the nodes that are joined to the cluster are connected, add the enclosure to the cluster.

3.3.3 Adding an enclosure in IBM FlashSystem

This procedure is the same whether you are configuring the system for the first time or expanding it. When performed by using the system GUI, the same steps are used for adding expansion or control enclosures.

Before beginning this process, ensure that the new control enclosure is correctly installed and cabled to the system.

For FC node-to-node communication, verify that the correct SAN zoning is set.

For node-to-node communication over RDMA-capable Ethernet ports, ensure that the IP addresses are configured and a connection between nodes can be established.

To add an enclosure to the system, complete the following steps:

1. In the GUI, select **Monitoring** → **System Hardware**. When a new enclosure is detected by a system, the Add Enclosure button appears on the system next to System Actions, as shown in Figure 3-45.

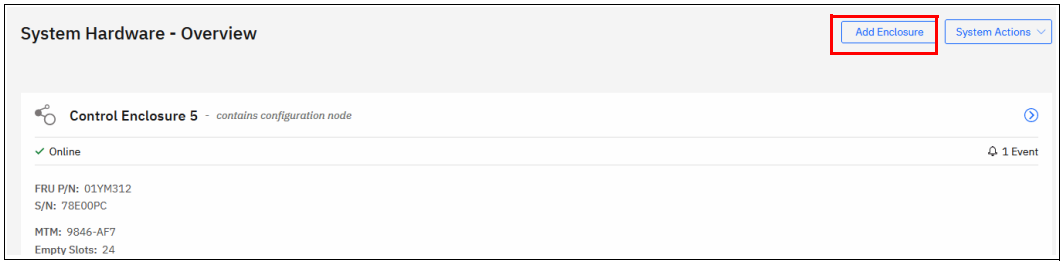


Figure 3-45 Add Enclosure button

> **Note:** If the Add Enclosure button does not appear, review the installation instructions to verify that the new enclosure is connected and set up correctly.

2. Click **Add Enclosure**, and a list of available candidate enclosures opens, as shown in Figure 3-46. To light the Identify light-emitting diode (LED) on a selected enclosure, select **Actions** → **Identify**. When the required enclosure (or enclosures) is chosen, click **Next**.

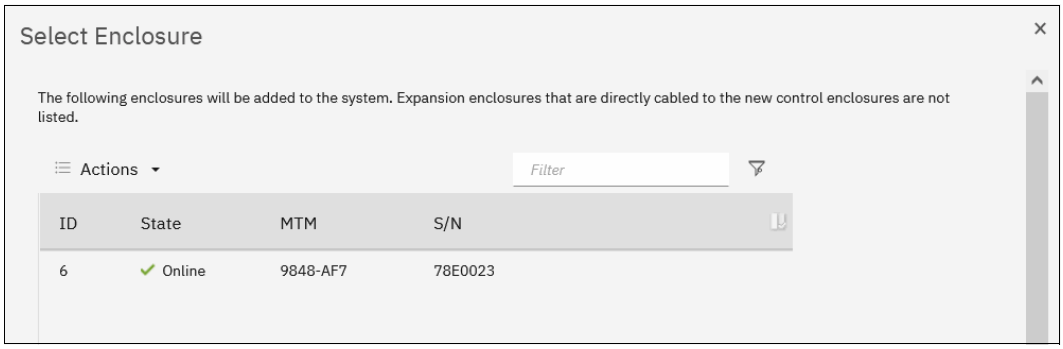


Figure 3-46 Selecting the control enclosure to add

3. Review the summary in the next window and click **Finish** to add the expansion enclosure or the control enclosure and all expansions that are attached to it to the system.

> **Note:** When a new control enclosure is added, the software version that is running on its nodes is upgraded or rolled back to match the system software version. This process can take up to 30 minutes or more, and the enclosure is added only when this process completes.

4. After the control enclosure is successfully added to the system, a success message appears. Click **Close** to return to the System Overview window and check that the new enclosure is visible and available for management.
5. To perform the same procedure by using a CLI, complete the following steps. For more information about the detailed syntax for each command, see Command-line interface.
   a. When you add control enclosures, check for unpopulated I/O groups by running the **`lsiogrp`** command. Because each control enclosure includes two nodes, it forms an I/O group.

Example 3-5 shows that only `io_grp0` has nodes. Therefore, a new control enclosure can be added to `io_grp1`.

Example 3-5 Listing the I/O groups

```
IBM_IBM FlashSystem:ITSO-FS9500:superuser>lsiogrp
id name            node_count vdisk_count host_count site_id site_name
0  io_grp0         2          0           0
1  io_grp1         0          0           0
2  io_grp2         0          0           0
3  io_grp3         0          0           0
4  recovery_io_grp 0          0           0
```

b. To list control enclosures that are available to add, run the **`lscontrolenclosurecandidate`** command, as shown in Example 3-6. To list the expansion enclosures, run the **`lsenclosure`** command. Expansions that have the **`managed`** parameter set to `no` can be added.

Example 3-6 Listing the candidate control enclosures

```
IBM_IBM FlashSystem:ITSO-FS9500:superuser>lscontrolenclosurecandidate
serial_number product_MTM machine_signature
78E005D       9848-AF8    4AD2-EA69-8B5E-D0C0
```

c. Add a control enclosure by running the **`addcontrolenclosure`** command, as shown in Example 3-7. The command triggers only the process, which starts in the background and can take 30 minutes or more.

Example 3-7 Adding a control enclosure

```
IBM_IBM FlashSystem:ITSO-FS9500:superuser>addcontrolenclosure -iogrp 1 -sernum
78E005D
```

d. To add an expansion enclosure, change its `managed` status to `yes` by running the **`chenclosure`** command, as shown in Example 3-8.

Example 3-8 Adding an expansion enclosure

```
IBM_IBM FlashSystem:ITSO-FS9500:superuser>lsenclosure
id status type      managed IO_group_id IO_group_name product_MTM serial_number
1  online control   yes     0           io_grp0       9848-AF8    78E006A
2  online expansion no      0           io_grp0       9848-AFF    78CBVF5

IBM_IBM FlashSystem:ITSO-FS9500:superuser>chenclosure -managed yes 2
```

3.3.4 Adding a node or hot spare node in IBM SAN Volume Controller systems

This procedure is the same whether you are configuring the system for the first time or expanding it later. The same process is used to add a node to an I/O group, or a hot spare node.

Before beginning this process, ensure that the new control enclosure is correctly installed and cabled to the system.

For FC node-to-node communication, verify that the correct SAN zoning is set.

For node-to-node communication over RDMA-capable Ethernet ports, ensure that the IP addresses are configured and a connection between nodes can be established.

To add a node to the system, complete the following steps:

1. In the GUI, select **Monitoring** → **System Hardware**. When a new enclosure is detected by a system, the **Add Node** button appears on the System - Overview window next to System Actions, as shown in Figure 3-47.

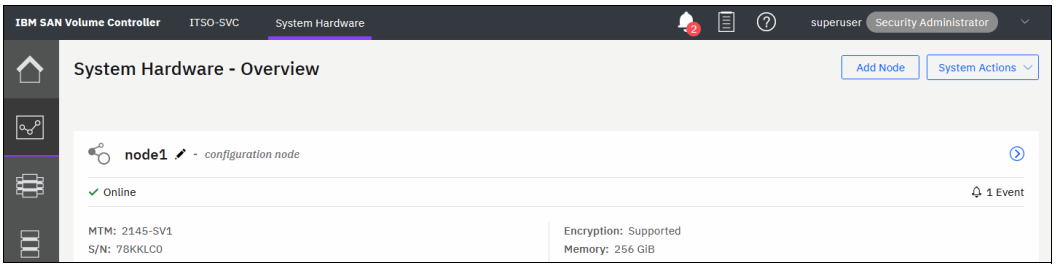


Figure 3-47 Add Node button

Note: If the Add Node button does not appear, review the installation instructions to verify that the new node is connected and set up correctly.

2. Click **Add Node**. A form that you can use to assign nodes to I/O groups opens, as shown in Figure 3-48. To light the Identify light-emitting diode (LED) on a node, click the LED icon that is next to a node name**.** When the required node (or nodes) is selected, click **Finish**.

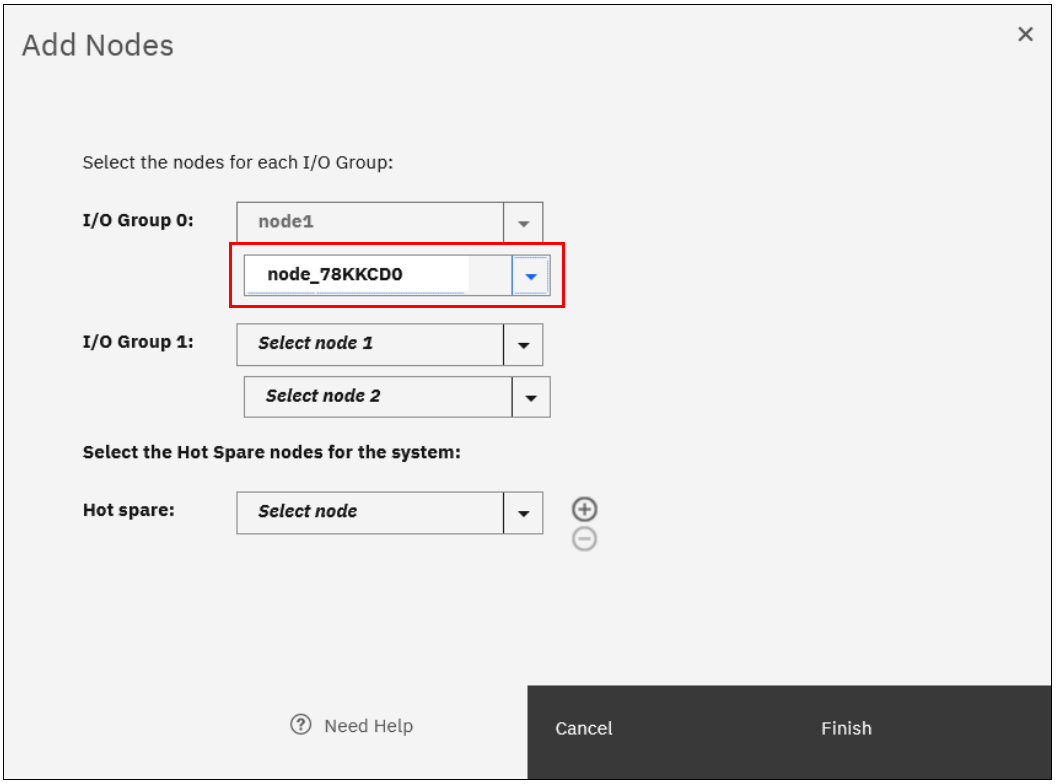


Figure 3-48 Adding a node

The Monitoring → Systems Hardware window now changes and shows that the node is added, as shown in Figure 3-49.

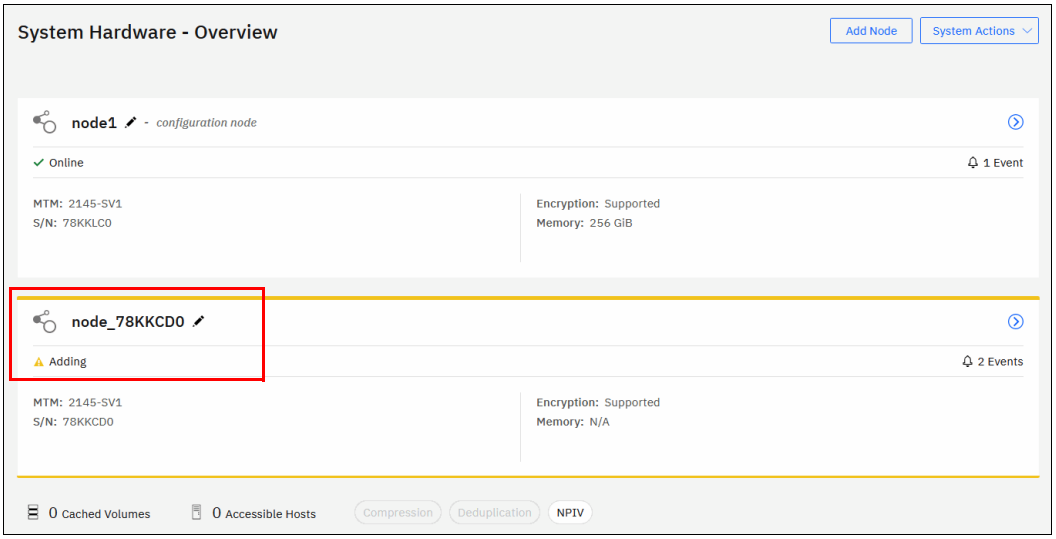


Figure 3-49 IBM SAN Volume Controller is adding node to the cluster

The node is added, as shown in Figure 3-50.

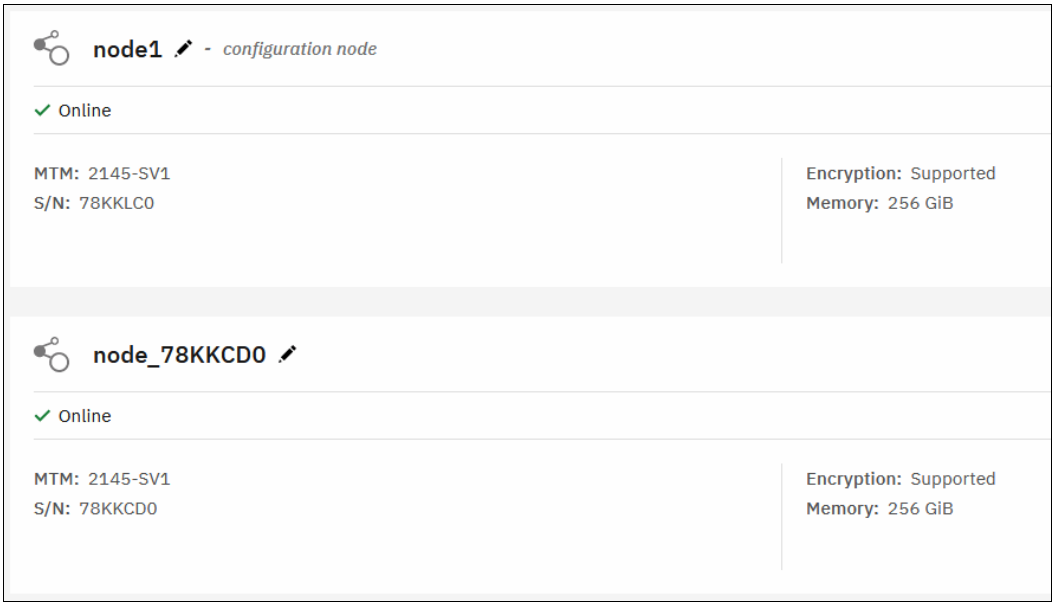


Figure 3-50 Node added

Note: When a node is added, the software version that is running is upgraded or rolled back to match the cluster software version. This process can take 30 minutes or more to complete. The node is added only after this process completes.

To perform the same procedure by using a CLI, complete the following steps. For more information about the detailed syntax for each command, see this IBM Documentation web page:

1. When adding nodes, check for unpopulated I/O groups by running **lsiogrp**. Each complete I/O group has two nodes. Example 3-9 shows that only io_grp0 has nodes; therefore, a new control enclosure can be added to io_grp1.

Example 3-9 Listing I/O groups

```
IBM_2145:ITSO-SVC:superuser>lsiogrp
id name            node_count vdisk_count host_count site_id site_name
0  io_grp0         2          0           0
1  io_grp1         0          0           0
2  io_grp2         0          0           0
3  io_grp3         0          0           0
4  recovery_io_grp 0          0           0
```

2. To list nodes that are available to add to the I/O group, run the **lscnodecandidate** command, as shown in Example 3-10.

Example 3-10 Listing the candidate nodes

```
BM_2145:ITSO-SVC:superuser>lsnodecandidate
id               panel_name UPS_serial_number UPS_unique_id    hardware
serial_number product_mtm machine_signature
500507680C00D98F 78KKLD0                    500507680C00D98F SV1     78KKLD0
2145-SV1    3F25-557E-21E6-2B7D
500507680C00D98A 78KKCH0                    500507680C00D98A SV1     78KKCH0
2145-SV1    702D-D5FE-76AA-4034
```

3. Add a node by running the **addnode** command, as shown in Example 3-11. The command triggers only the process, which starts in the background and can take 30 minutes or more.

Example 3-11 Adding a node as a spare

```
IBM_2145:ITSO-SVC:superuser>addnode -panelname 78KKLD0 -spare
Node, id [3], successfully added
```

Example 3-12 shows same command, but used to add a node to an I/O group io_grp1.

Example 3-12 Adding a node to an I/O group

```
IBM_2145:ITSO-SVC:superuser>addnode -panelname 78KKCH0 -name node3 -iogrp 1
Node, id [4], successfully added
```

4. Check the nodes in the system by using CLI. As shown in Example 3-13, the IBM SAN Volume Controller is configured with two nodes, which forms one IO-group. A spare node is configured for the IO-group.

Example 3-13 Single IO-group (two nodes) and one spare

```
IBM_2145:ITSO-SVC:superuser>lsnode
id name          UPS_serial_number WWNN             status IO_group_id IO_group_name config_node
UPS_unique_id hardware iscsi_name                                iscsi_alias panel_name
enclosure_id canister_id enclosure_serial_number site_id site_name
1  node1_78KKLC0               500507680C00D990 online 0         io_grp0      yes
SV1      iqn.1986-03.com.ibm:2145.itso-svc.node178kklc0             78KKLC0
2  node2_78KKCD0               500507680C00D982 online 0         io_grp0      no
SV1      iqn.1986-03.com.ibm:2145.itso-svc.node278kkcd0             78KKCD0
```

```
3  spare1                    500507680C00D98F spare                      no
SV1                                                                  78KKLD0
```

A two-IO-group system with no spare is shown in Example 3-14.

Example 3-14 Two IO-groups (four nodes) configured- no spare

```
IBM_2145:ITSO-SVC:superuser>lsnode
id name          UPS_serial_number WWNN             status IO_group_id IO_group_name config_node
UPS_unique_id hardware iscsi_name                                  iscsi_alias panel_name
enclosure_id canister_id enclosure_serial_number site_id site_name
1 node1_78KKLC0               500507680C00D990 online 0         io_grp0      yes
SV1      iqn.1986-03.com.ibm:2145.itso-svc.node178kklc0            78KKLC0
2 node2_78KKCD0               500507680C00D982 online 0         io_grp0      no
SV1      iqn.1986-03.com.ibm:2145.itso-svc.node278kkcd0            78KKCD0
3 node3_78KKCH0               500507680C00D98A online 1         io_grp1      no
SV1      iqn.1986-03.com.ibm:2145.itso-svc.node378kkch0            78KKCH0
4 node4_78KKLD0               500507680C00D98F online 1         io_grp1      no
SV1      iqn.1986-03.com.ibm:2145.itso-svc.node478kkld0            78KKLD0
```

The administrator might want to rename the nodes to feature consistent names. This process can be done by clicking **Monitoring** → **System Hardware** → **Node Actions** → **Rename**.

3.3.5 Changing the system topology

Note: HyperSwap is *not* supported by IBM FlashSystem 5015.

The HyperSwap function is a high availability (HA) feature that provides dual-site, active-active access to a volume. You can create an HyperSwap topology system configuration in which each I/O group in the system is physically on a different site. When these configurations are used with HyperSwap volumes, they can be used to maintain access to data on the system if site-wide outages occur.

IBM SAN Volume Controller supports a second multi-site topology: Enhanced Stretched Cluster (ESC). Here each I/O group in the system has one node on one site and one node on the other site.

Both topologies enable full configuration of the highly available volumes through a single point of configuration.

Note: Only the IBM SAN Volume Controller can use Enhanced Stretched Cluster (ESC).

If your solution is designed for HyperSwap or ESC, use the guidance in this section to configure your topology for either solution.

For a list of requirements for a HyperSwap or ESC configuration, see the following IBM Documentation web pages:

- Stretched system configuration details (for IBM SAN Volume Controller)
- HyperSwap system configuration details (for IBM SAN Volume Controller)
- HyperSwap system configuration details (for IBM FlashSystem)

To change the system topology, complete the following steps:

1. In the GUI, click **Monitoring** → **System Hardware** to open the System - Overview window. Click **System Actions** and expand **Modify System Topology**, as shown in Figure 3-51.

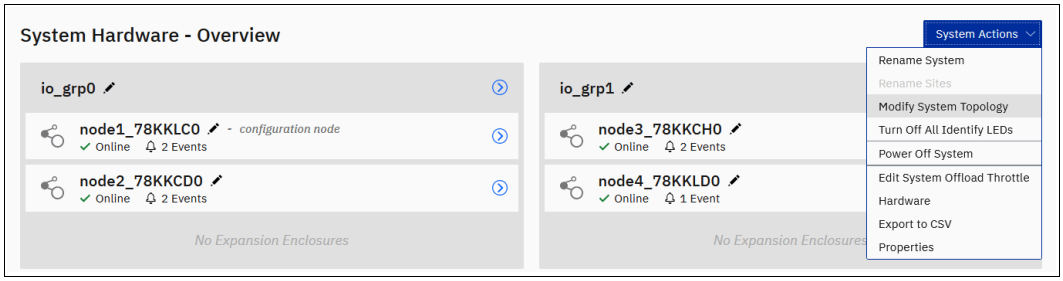


Figure 3-51 Starting the Modify System Topology wizard

2. The Modify Topology wizard welcome window opens. You are prompted to change the default site names, as shown in Figure 3-52. The site names can indicate, for example, building locations for each site, or other descriptive information. Click **Next**.

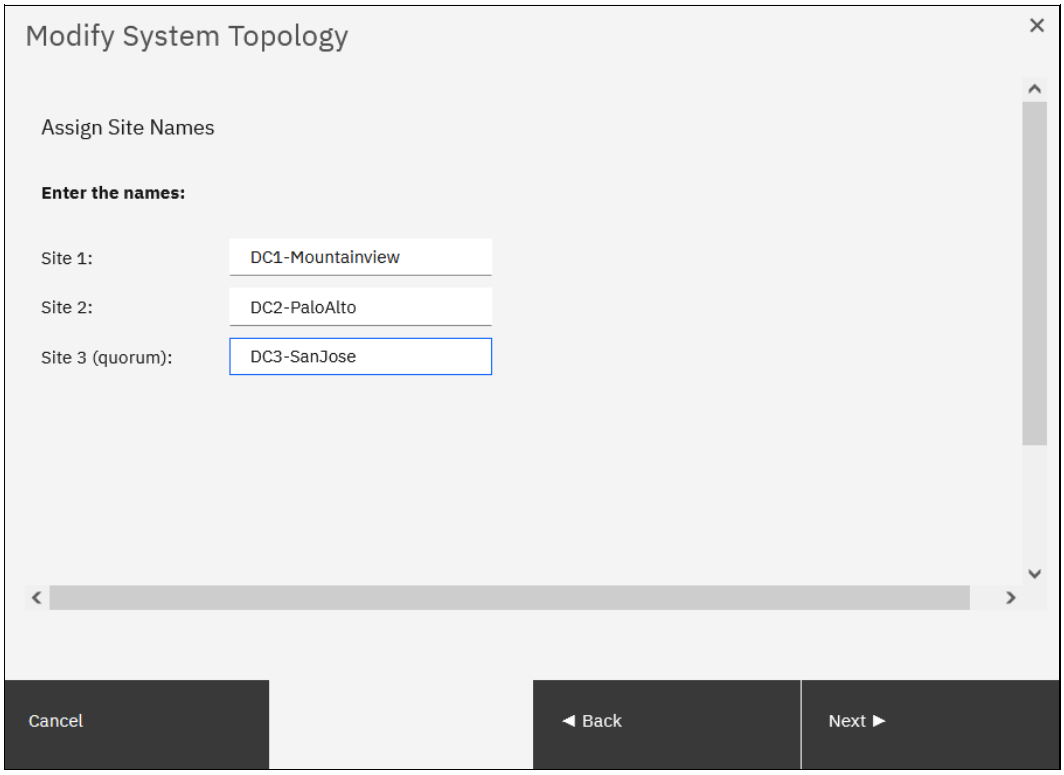


Figure 3-52 Assigning site names

3. Select **HyperSwap System** or **Stretched System** for the topology and assign I/O groups to sites. For IBM FlashSystem products, HyperSwap System is the only available topology. Click the marked icons in the center of the window to swap site assignments Figure 3-53. Click **Next**.

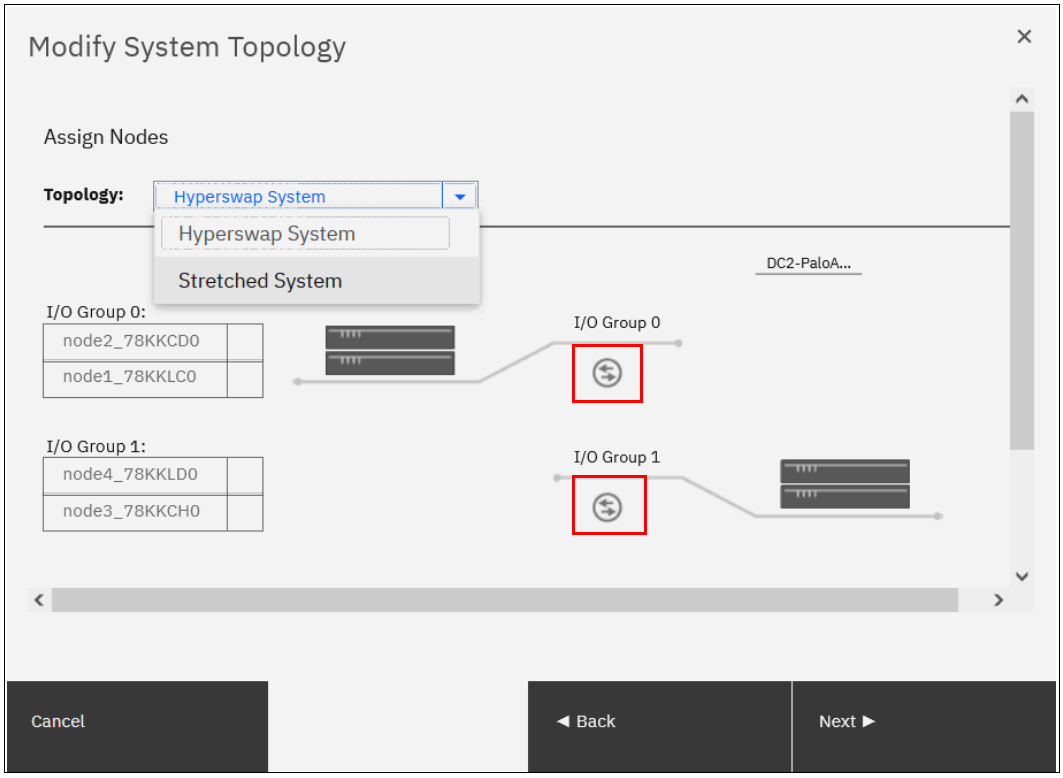


Figure 3-53 Specifying the system topology

4. If any host objects or back-end storage controllers are configured, you must assign a site for each of them. Right-click the object and click **Modify Site**. When done, click **Next,** as shown in Figure 3-54 on page 219.

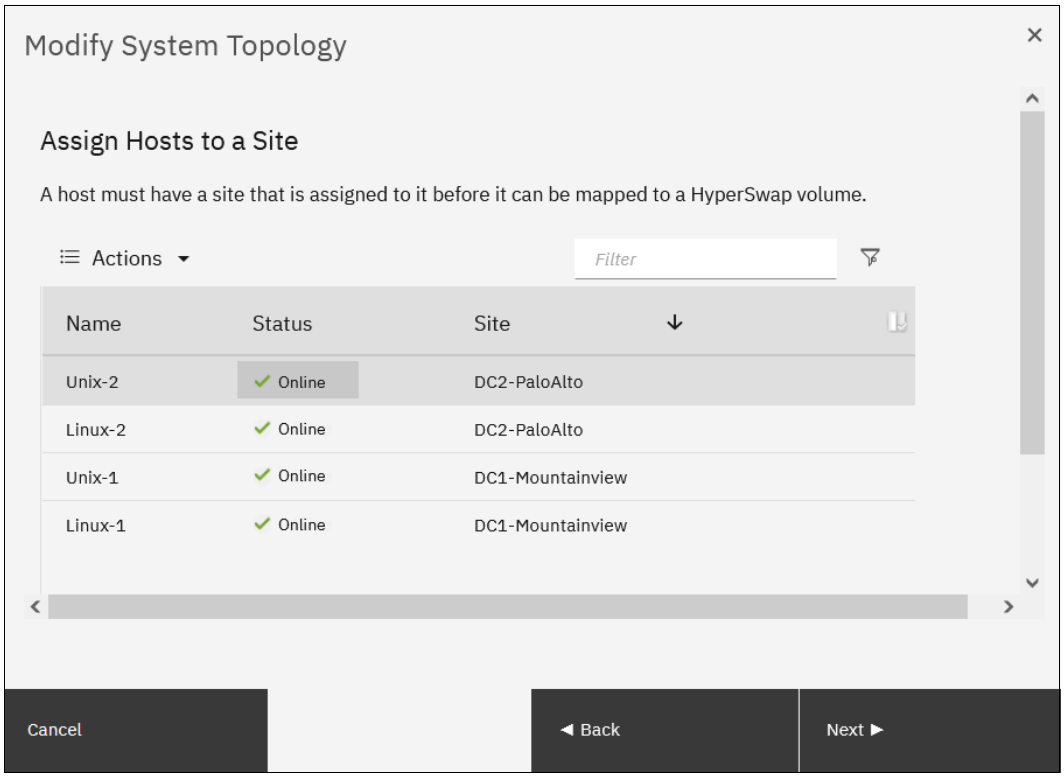


Figure 3-54 Assigning hosts to sites

5. If you are configuring HyperSwap topology, set the maximum background copy operations bandwidth between the sites. A background copy is the initial synchronization and any subsequent resynchronization traffic for HyperSwap volumes. Use this setting to limit the effect of volume synchronization to host operations. You can also set it higher during the initial setup when no host operations are on the volumes yet, and set it lower when the system is in production.

   As shown in Figure 3-55 on page 220, you must specify the total bandwidth between the sites in megabits per second (Mbps) and what percentage of this bandwidth that can be used for background copying. Click **Next**.

   Note: An ESC topology system does not require this setting.

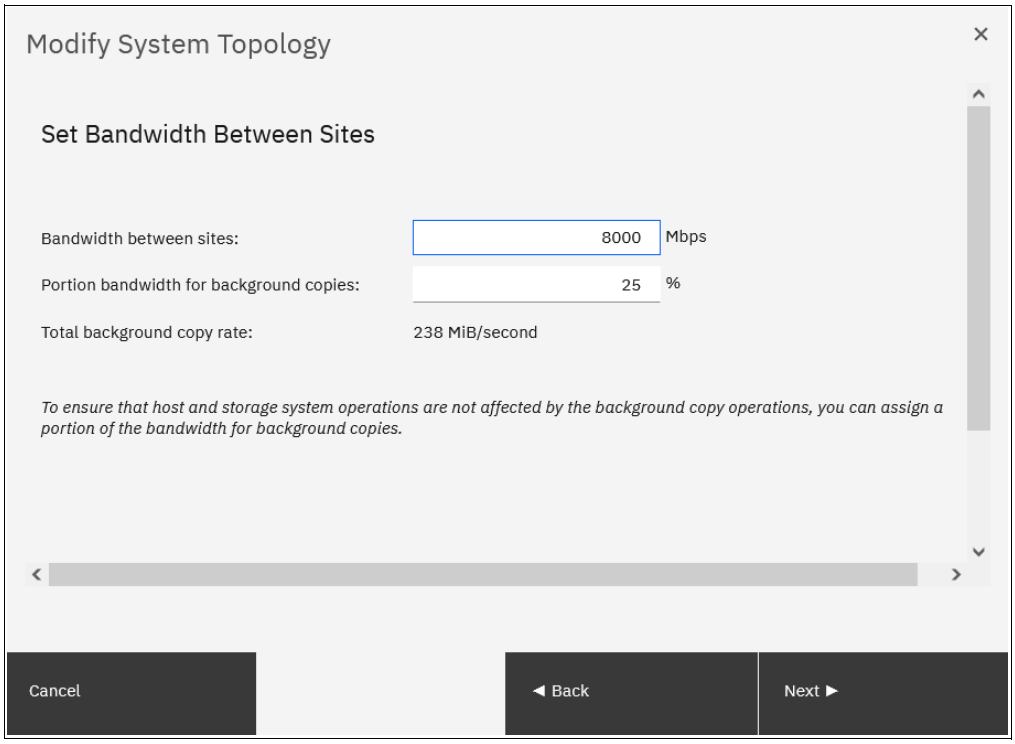


Figure 3-55 Setting the bandwidth between the sites

6. Review the summary and click **Finish**. The wizard starts implementing changes to migrate the system to the HyperSwap solution.

When you later add a host or back-end storage controller objects, the GUI prompts you to set an object site during the creation process.

3.3.6 Configuring quorum disks or applications

Quorum devices are required for a system to hold a copy of important system configuration data. An internal drive of an IBM FlashSystem, a managed disk (MDisk) from FC-attached external back-end storage, or a special application that is connected over an IP network can work as a quorum device.

One of these items is selected for the *active quorum* role, which is used to resolve failure scenarios where half the nodes on the system become unavailable or a link between enclosures is disrupted. The active quorum determines which nodes can continue processing host operations. It also avoids a “split brain” condition, which occurs when both halves of the system continue I/O processing independently of each other.

For IBM FlashSystem products with a single control enclosure, quorum devices are selected automatically from the internal drives. For IBM SAN Volume Controller systems with a standard topology, quorum devices are automatically assigned from an MDisk. No special configuration actions are required. This function also applies for IBM FlashSystem products with multiple control enclosures, a standard topology, and virtualizing external storage.

For HyperSwap and ESC topology systems, install an active quorum device on a third, independent site. Because of the costs that are associated with deploying a separate FC-attached storage device on a third site, an IP-based quorum device can be used for this purpose.

If no third site exists, the quorum must be configured to select a site to always win a tie breaker. If connectivity is lost between the sites, the site that is configured as the winner continues operating and processing I/O requests and the other site stops until the fault is fixed. On HyperSwap configurations, IP quorum settings help to choose the preferred site to control which site continues to run during connectivity issues between sites.

If a site outage occurs at the winning site, the system stops processing I/O requests until this site is recovered or the manual quorum override procedure is used.

On IBM FlashSystem products in a standard topology system with two or more control enclosures and no external storage, none of the internal drives can be the active quorum device. For such configurations, it is a best practice to deploy an IP-based quorum application to avoid a “split brain” condition.

Creating and installing an IP quorum application

To create and install an IP quorum application, complete the following steps:

1. Select **System** → **Settings** → **IP Quorum** to download the IP quorum application, as shown in Figure 3-56. If you use IPv6 for management IP addresses, the Download IPv6 Application button is available and the IPv4 option is disabled. In our example, we select **Download IPv4 Application**.

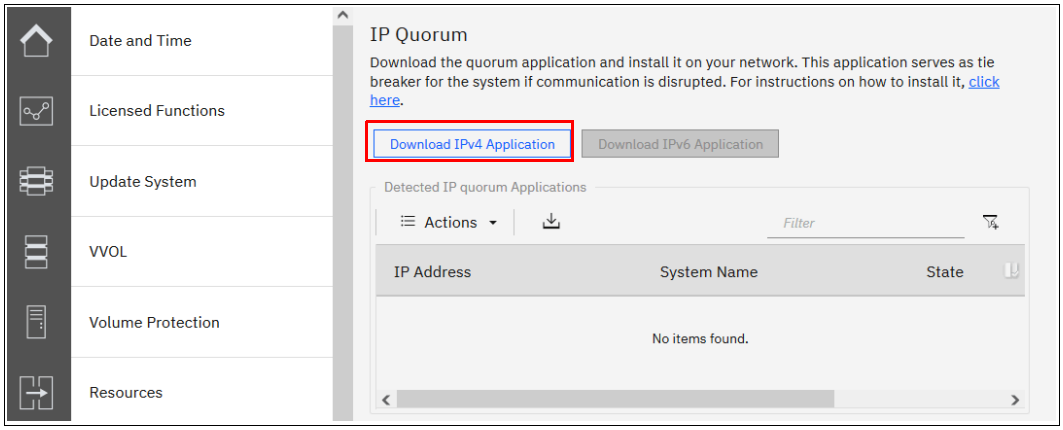


Figure 3-56 Download IPv4 quorum button

2. Click **Download...** and a window opens, as shown in Figure 3-57. It provides an option to create an IP application that is used for tie-breaking only, or an application that can be used as a tie-breaker and to store recovery metadata.

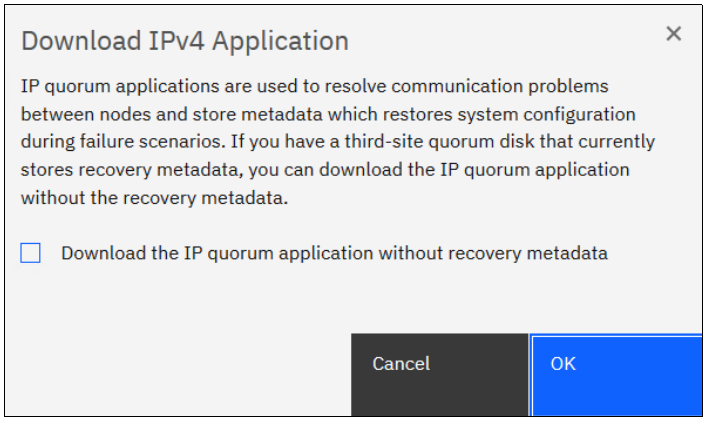


Figure 3-57 Download IP quorum application window

An application that does not store recovery metadata requires less channel bandwidth for a link between the system and the quorum app, which might be a decision-making factor for using a multi-site HyperSwap system.

For a full list of IP quorum app requirements, see IP quorum application configuration.

3. Click **OK**. The `ip_quorum.jar` file is created. Save the file and transfer it to a supported AIX, Linux, or Windows host that can establish an IP connection to the service IP address of each system node. Move it to a separate directory and start the application, as shown in Example 3-15.

Example 3-15 Starting the IP quorum application on the Windows operating system

```
C:\IPQuorum>java -jar ip_quorum.jar
=== IP quorum ===
Name set to null.
Successfully parsed the configuration, found 2 nodes.
Trying to open socket
Trying to open socket
Handshaking
Handshaking
Waiting for UID
Creating UID
*Connecting
Connected to 10.0.0.42
Connected to 10.0.0.41
```

Note: Add the IP quorum application to the list of auto-started applications at each start or restart or configure your operating system to run it as an auto-started service in the background. The server that runs the IP quorum must be in the same subnet as the IBM FlashSystem. A total of five IP quorums can be used.

The IP quorum log file and recovery metadata are stored in the same directory with the `ip_quorum.jar` file.

4. Check that the IP quorum application is successfully connected and running by verifying its Online status by selecting **System** → **Settings** → **IP Quorum**, as shown in Figure 3-58.

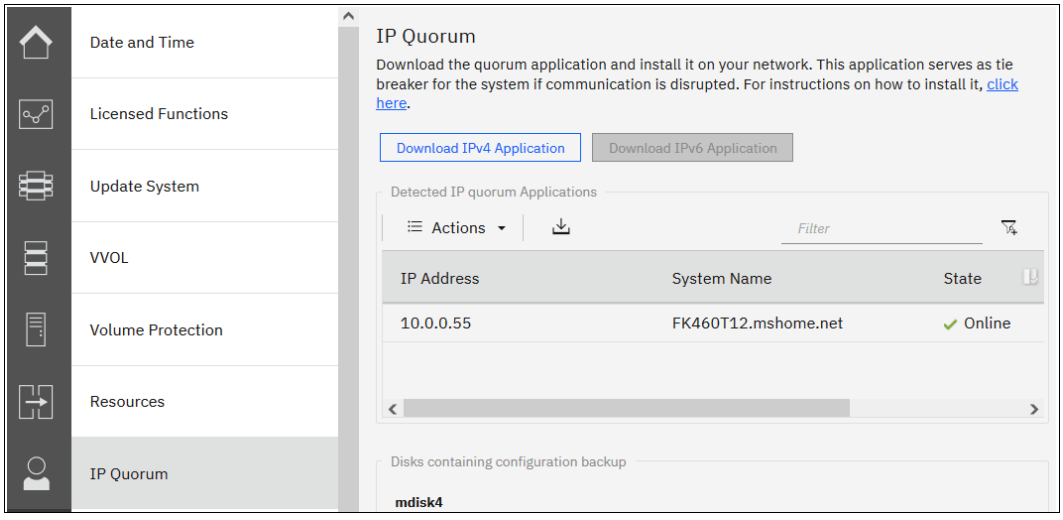


Figure 3-58 IP quorum application that is deployed and connected

Configuring the IP quorum mode

On a standard topology system, only the Standard quorum mode is supported. No other configuration is required. On a HyperSwap topology, you can configure the following tie-breaker scenarios (a tie occurs when half of the nodes that were a member of the system are present):

- If the quorum mode is set to `Standard`, both sites have an equal chance to continue working after the tie breaker.
- If the quorum mode is set `Preferred`, during a disruption, the system delays processing tie-breaker operations on non-preferred sites, which leaves more time for the preferred site to win. If during an extended period a preferred site cannot contact the IP quorum application (for example, if it is destroyed), a non-preferred site continues working.
- If the quorum mode is set to `Winner`, the selected site is always the tie-breaker winner. If the winner site is destroyed, the remaining site can continue operating only after manual intervention.

The `Preferred` quorum mode is supported by an IP quorum only.

To set a quorum mode, select **System** → **Settings** → **IP Quorum** and then click **Quorum Setting**. The Quorum Setting window opens, as shown in Figure 3-59.

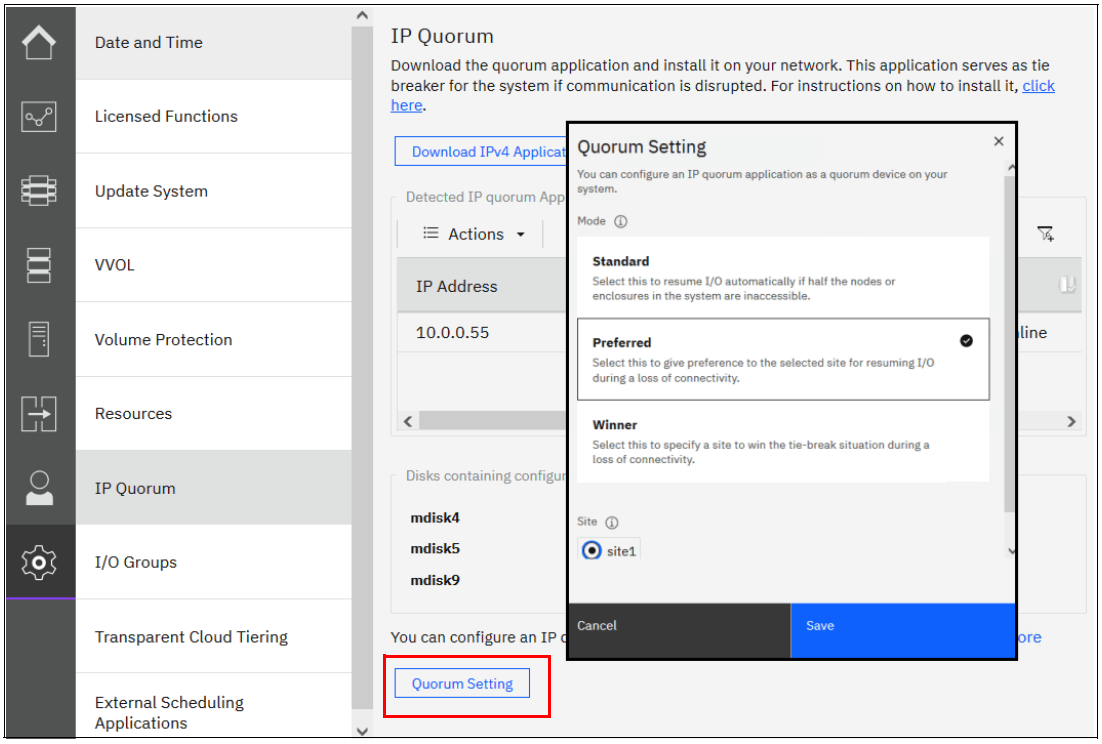


Figure 3-59 Changing the quorum mode

3.3.7 Configuring the local Fibre Channel port masking

With FC port masking, you control the use of FC ports. By applying a mask, you restrict node-to-node communication or replication traffic on selected ports.

To decide whether your system must have port masks configured, see 2.7.8, “Port designation recommendations” on page 135.

To set the FC port mask by using the GUI, complete the following steps:

1. Select **Settings** → **Network** → **Fibre Channel Ports**. In a displayed list of FC ports, the ports are grouped by a system port ID. Each port is configured identically across all nodes in the system. You can click the arrow next to the port ID to expand a list and see which node ports (N_Port) belong to the selected system port ID and their worldwide port names (WWPNs).
2. Right-click a system port ID that you want to change and select **Modify Connection**, as shown in Figure 3-60.

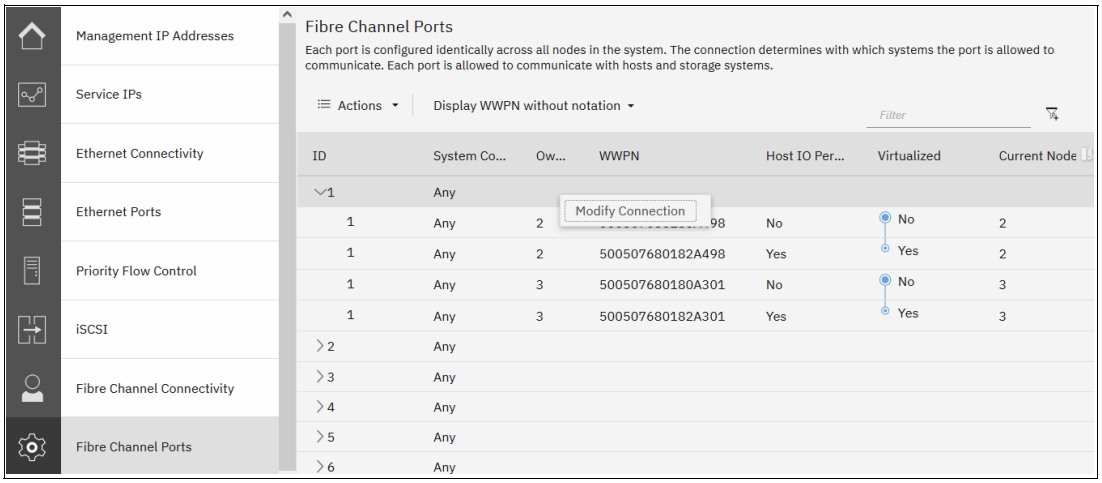


Figure 3-60 Applying a port mask by using a GUI

By default, all system ports can send and receive traffic of any kind, including the following examples:

- Host traffic
- Traffic to virtualized back-end storage systems
- Local system traffic (node to node)
- Partner system (remote replication) traffic

The first two types are always allowed, and you can control them only with SAN zoning. The other two types can be blocked by port masking.

3. In the Modify Connection dialog box (see Figure 3-61 on page 225), you can choose which type of traffic a port can send; for example, Remote if the port is dedicated to Remote Replication traffic. Click **Modify** when done.

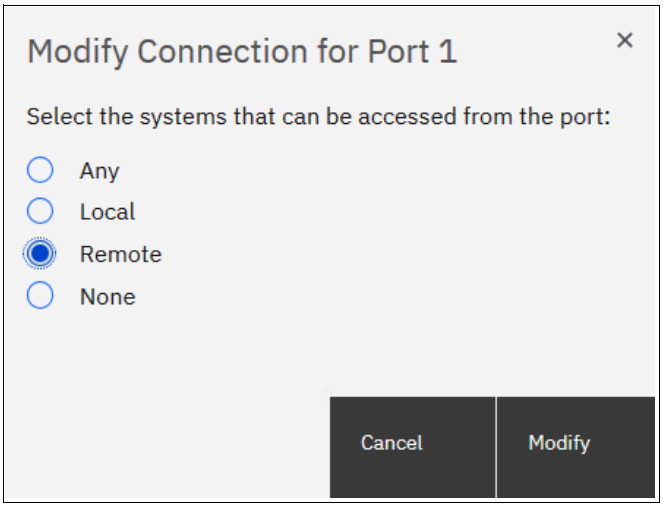


Figure 3-61 Modify Connection dialog box

The following types of traffic are allowed for each choice:

- Any: A port can work with all types of traffic.
- Local: Remote replication traffic is blocked on this port.
- Remote: Local node-to-node traffic is blocked.
- None: Local and remote systems traffic is allowed, but system-to-host and system-to-back-end storage communication still exists.

Port masks can also be set by using the CLI. Local and remote partner port masks are internally represented as a string of zeros and ones. The last digit in the string represents port one. The previous digits represent ports two, three, and so on.

If the digit for a port is set to `1`, the port is enabled for the specific type of communication. If it is set to `0`, the system does not send or receive traffic that is controlled by a mask on the port.

To view the current port mask settings, run the **`lssystem`** command, as shown in Example 3-16. The output shows that all system ports allow all types of traffic.

Example 3-16 Viewing the local port mask

```
IBM_IBM FlashSystem:ITSO-FS9500:superuser>lssystem |grep mask
local_fc_port_mask 1111111111111111111111111111111111111111111111111111111111111111
partner_fc_port_mask 1111111111111111111111111111111111111111111111111111111111111111
```

To set the local or RPORT mask, run the **`chsystem`** command. Example 3-17 shows the mask setting for a system with four FC ports on each node and that has RC relationships. Masks are applied to allow local node-to-node traffic on ports 1 and 2, and replication traffic on ports 3 and 4.

Example 3-17 Setting a local port mask by running the chsystem command

```
IBM_IBM FlashSystem:ITSO-FS9500:superuser>chsystem -localfcportmask 0011
IBM_IBM FlashSystem:ITSO-FS9500:superuser>chsystem -partnerfcportmask 1100
IBM_IBM FlashSystem:ITSO-FS9500:superuser>lssystem |grep mask
local_fc_port_mask 0000000000000000000000000000000000000000000000000000000000000011
partner_fc_port_mask 0000000000000000000000000000000000000000000000000000000000001100
```

The mask is extended with zeros, and all ports that are not set in a mask have the selected type of traffic blocked.

Note: When replacing or upgrading your node hardware, consider that the number of FC ports and their arrangement might be changed. If so, make sure that any configured port masks are still valid for the new configuration.

3.3.8 Automatic configuration for IBM SAN Volume Controller back-end storage

If a system is supposed to work as FC-attached back-end storage for IBM SAN Volume Controller, you can enable Automatic Configuration for Virtualization during the initial System Setup or anytime later by selecting **Settings** → **System** → **Automatic Configuration**.

Automatic Configuration for Virtualization is intended for a new system. If host, pool, or volume objects are configured, all the user data must be migrated out of the system and those objects must be deleted.

The Automatic Configuration for Virtualization wizard starts immediately after you complete the initial setup wizard if you set Automatic Configuration to `On`.

Complete the following steps:

1. Click **Settings** → **System** → **Automatic Configuration**. Then, select Automatic Configuration **ON** and click **Save**, as shown in Figure 3-62.

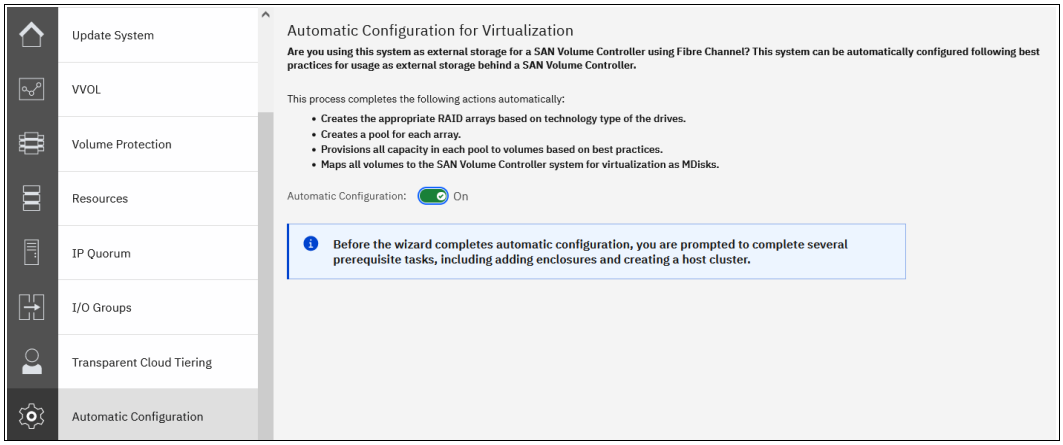


Figure 3-62 Automatic Configuration wizard enablement

2. If any control or expansion enclosures must be included as part of the external storage to be virtualized, you can add them. If you do not have more enclosures to add, this part of the prerequisite steps can be skipped.

   Click **Add Enclosure** to start the adding process, or click **Skip** to move to the next step (see Figure 3-63).

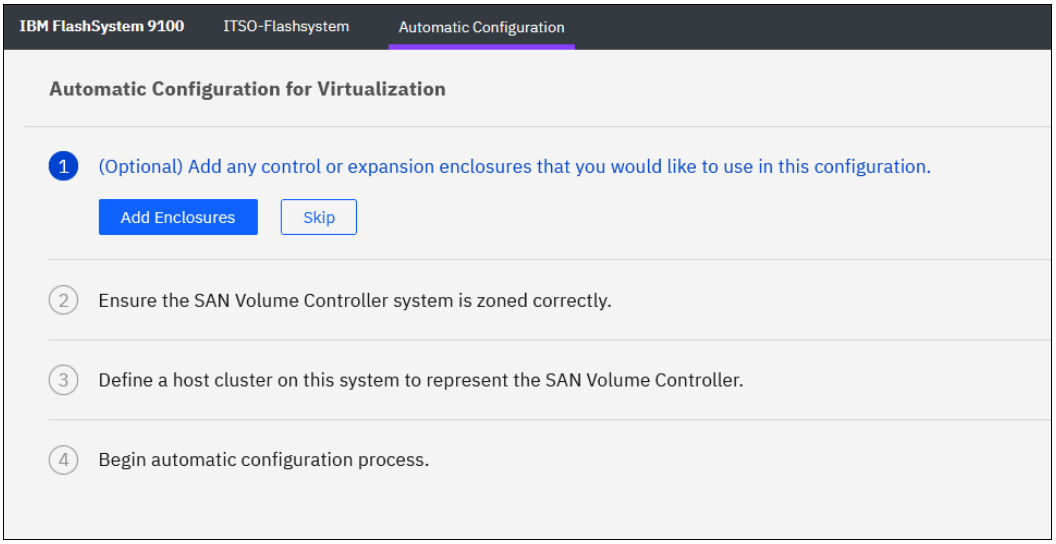


Figure 3-63 Automatic configuration: Add Enclosure

Note: You can turn off the Automatic Configuration for Virtualization wizard at any step by clicking the dotted symbol in the upper right corner.

3. The wizard checks whether the IBM SAN Volume Controller is correctly zoned to the system. By default, newly installed systems run in N_Port ID Virtualization (NPIV) mode (Target Port Mode). The system's virtual (host) WWPNs must be zoned for IBM SAN Volume Controller. On the IBM SAN Volume Controller side, physical WWPNs must be zoned to a back-end system independently of the NPIV mode setting.
4. Create a host cluster object for IBM SAN Volume Controller. Each IBM SAN Volume Controller node has its own worldwide node name (WWNN). Make sure to select all WWNNs that belong to nodes of the same IBM SAN Volume Controller cluster.

   Figure 3-64 shows that because the system detected an IBM SAN Volume Controller cluster with dual I/O groups, four WWNNs are selected.

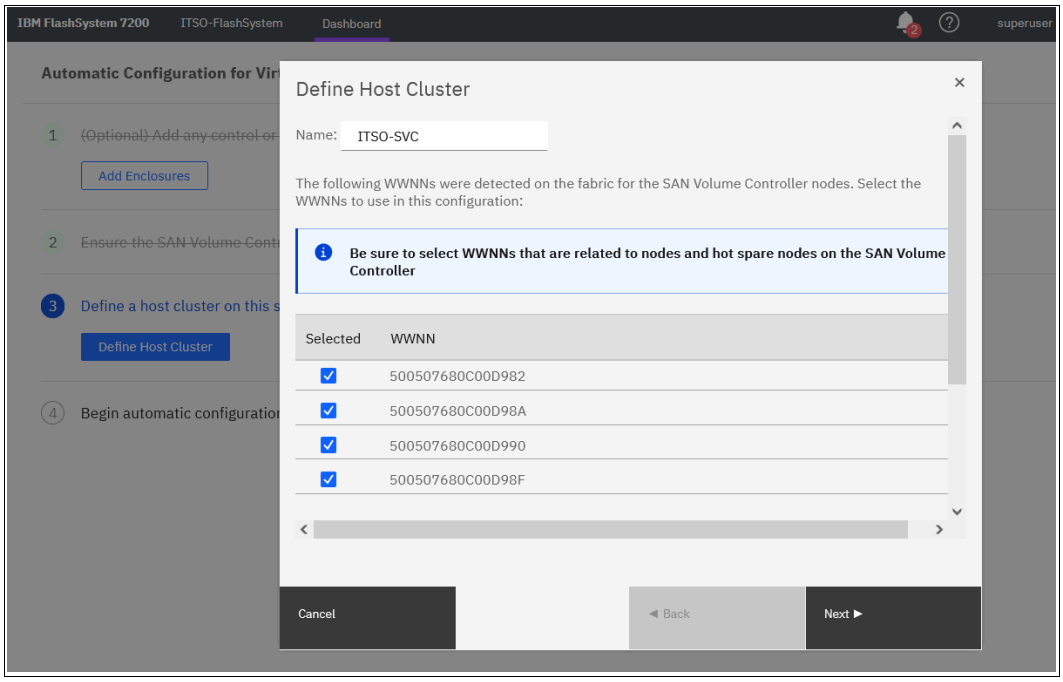


Figure 3-64 Defining a host cluster

5. When all nodes of an IBM SAN Volume Controller cluster (including the spare cluster) are selected, you can change the host object name for each one, as shown in Figure 3-65 on page 229. For convenience, name the host objects to match the IBM SAN Volume Controller node names or serial numbers.

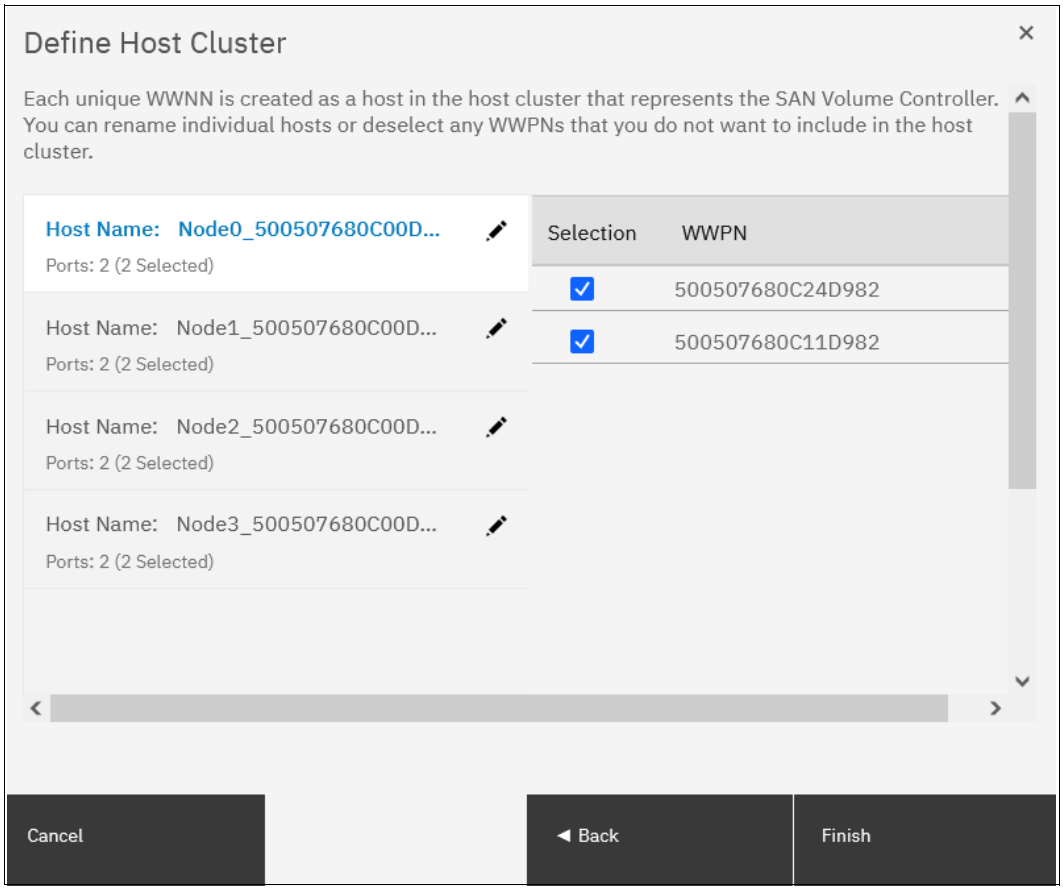


Figure 3-65 Hosts inside an IBM SAN Volume Controller host cluster

6. Click **Automatic Configuration** and check the list of internal resources that are used, as shown in Figure 3-66.

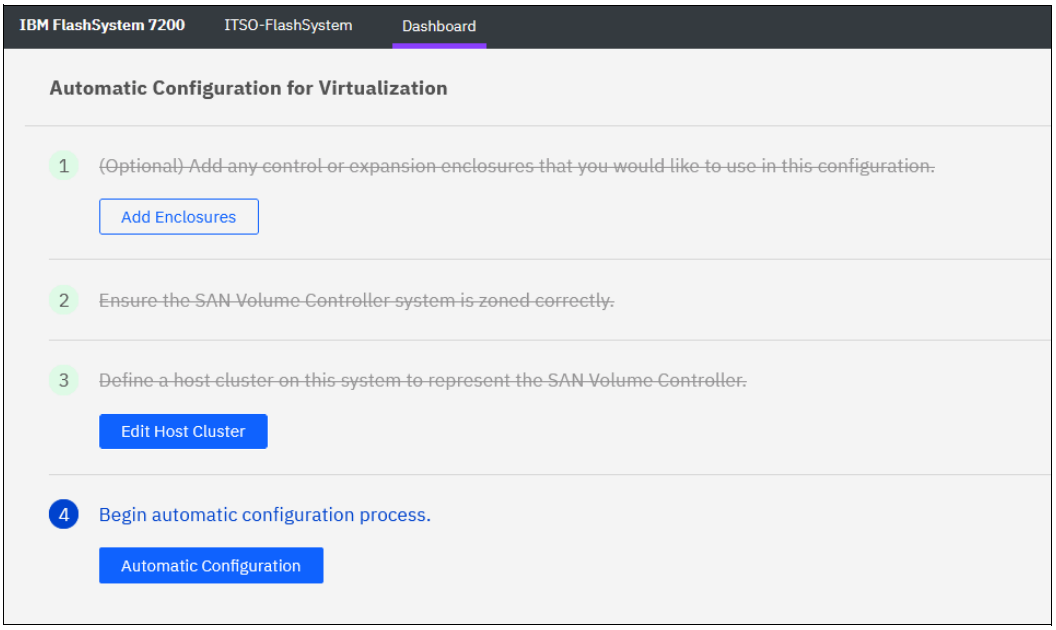


Figure 3-66 Begin the automatic configuration process

7. If the system uses compressed drives (FCM drives), you are prompted to enter your expected compression ratio (or total capacity that is to be provisioned to IBM SAN Volume Controller), as shown in Figure 3-67. If IBM SAN Volume Controller uses encryption or writes data that is not compressible, set the ratio to 1:1 and then, click **Next**.

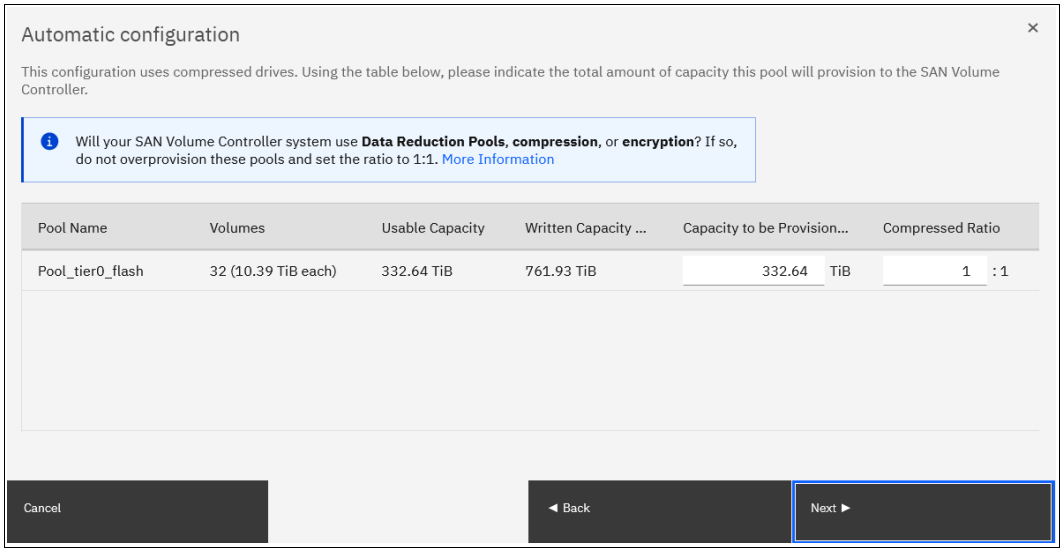


Figure 3-67 Automatic pool configuration

8. Review the pool (or pools) configuration, as shown in Figure 3-68, and click **Proceed** to trigger the commands that applies it.

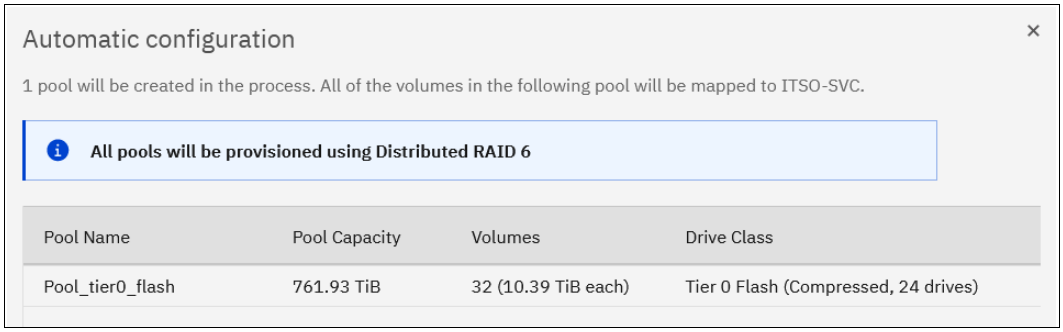


Figure 3-68 Pools configuration

9. When the Automatic Configuration for Virtualization wizard completes, you see the window that is shown in Figure 3-69. After clicking **Close**, you can proceed to the IBM SAN Volume Controller GUI and configure a new provisioned storage, as shown in Figure 3-69.

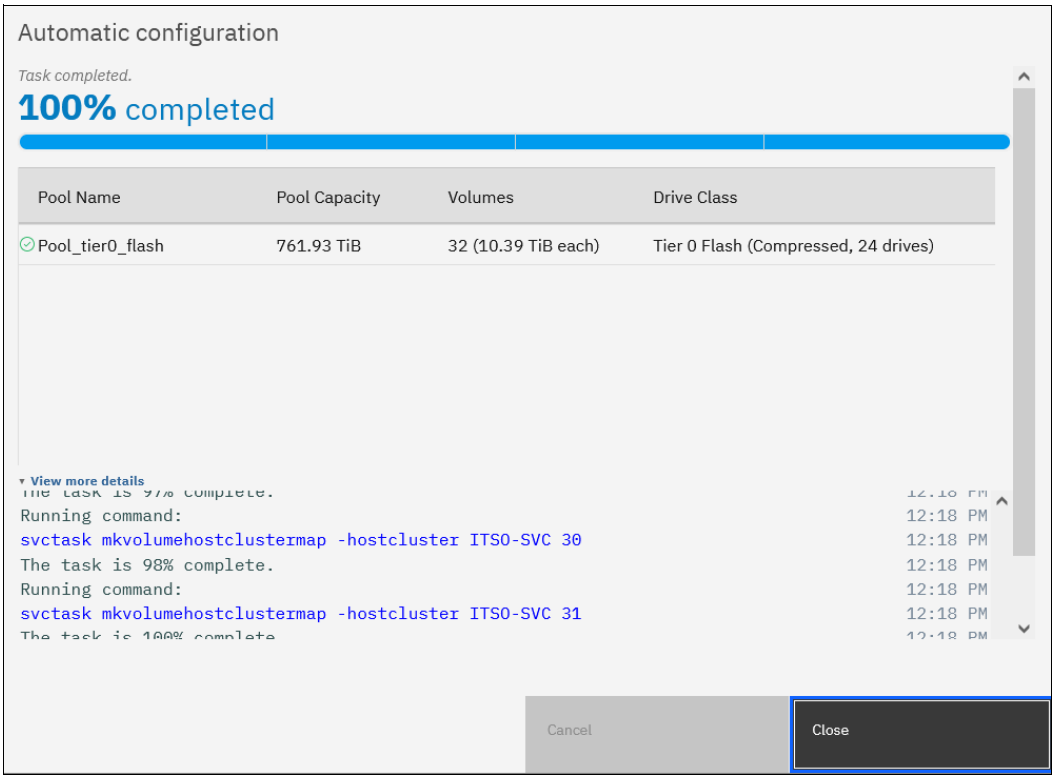


Figure 3-69 Automatic configuration running commands

10. You can export the system volume configuration data in `.csv` format by using this window or anytime later by selecting **Settings** → **System** → **Automatic Configuration**.

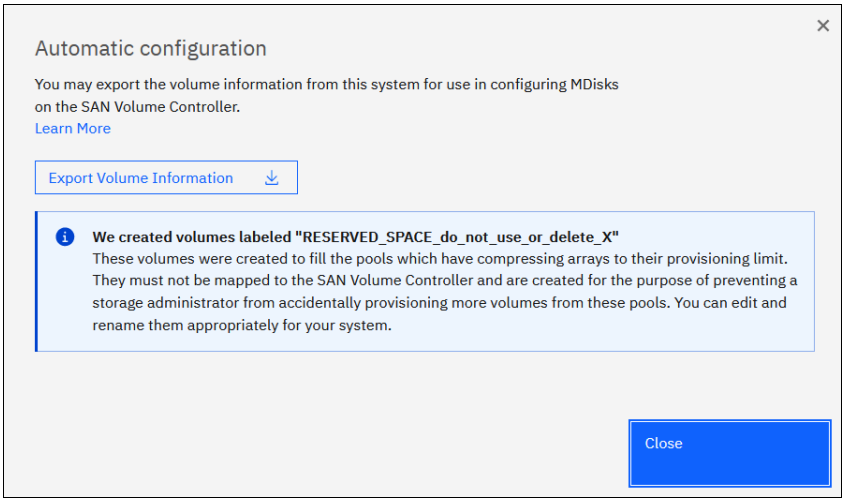


Figure 3-70 Automatic configuration complete

3.3.9 Configuring management access

The system can be managed by using the GUI and the CLI. Access to the system management interfaces requires user authentication.

For more information about user authentication and the secure communication implementation steps, see Chapter 12, “Security and encryption” on page 1045.

4

IBM Storage Virtualize GUI

This chapter provides an overview of the IBM Storage Virtualize GUI. The management GUI is a tool that is enabled and provided by IBM Storage Virtualize that helps you to monitor, manage, and configure your system.

This chapter explains the basic view and the configuration procedures that are required to get your system environment running as quickly as possible by using the GUI. This chapter does not describe advanced troubleshooting or problem determination and some of the complex operations (compression and encryption).

For more information, see Chapter 11, “Reliability, availability, and serviceability” on page 963.

Throughout this chapter, all GUI menu items are introduced in a systematic, logical order as they appear in the GUI. However, topics that are described in more detail in other chapters of the book are only referred to here. For example, Storage pools (Chapter 5, “Using storage pools” on page 365), Volumes (Chapter 6, “Volumes” on page 419), Hosts (Chapter 7, “Hosts” on page 517), and Copy Services (Chapter 10, “Advanced Copy Services” on page 717) are described in separate chapters.

This chapter includes the following topics:

- 4.1, “Performing operations by using the GUI” on page 234
- 4.2, “Introduction to the GUI” on page 239
- 4.3, “System Hardware - Overview window” on page 247
- 4.4, “Monitoring menu” on page 250
- 4.5, “Using the menus” on page 265
- 4.6, “Ownership groups” on page 268
- 4.7, “Settings” on page 285
- 4.8, “Other frequent tasks in the GUI” on page 356

© Copyright IBM Corp. 2024. All rights reserved.

4.1 Performing operations by using the GUI

This section describes useful tasks that use the GUI that help administrators to manage, monitor, and configure the system as quickly as possible. For the example in this chapter, the configured system is in a standard topology and not in an Enhanced Stretched or HyperSwap topology.

Demonstration video: You can view a demonstration video IBM Storage Spectrum Virtualize 8.6 GUI, including Volume Group Snapshots (with Safeguarded Copy) and Policy Based Replication.

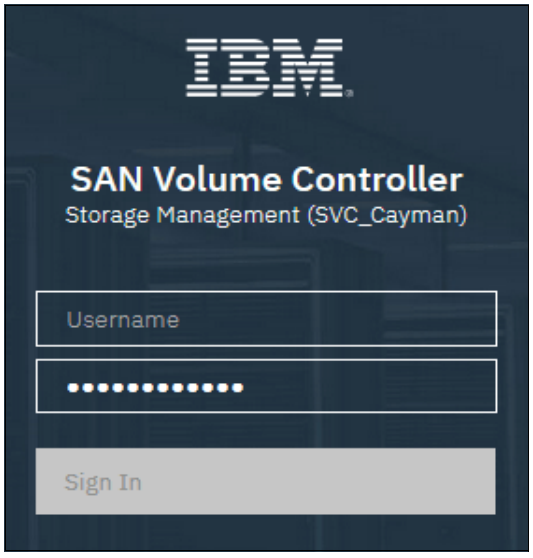
The GUI is a built-in software component within the IBM Storage Virtualize Software. Multiple users can be logged in to the GUI. However, because no locking mechanism exists, be aware that the last action that is entered from the GUI is the action that takes effect if two users change the same object simultaneously.

Important: Data entries that are made through the GUI are case-sensitive.

You must enable Java Script in your browser. For Mozilla Firefox, JavaScript is enabled by default and requires no other configuration steps. For more Information, see any FlashSystem or the IBM SAN Volume Controller Documentation. For more information, see, Flash9x00 and the IBM SAN Volume Controller Documentation.

4.1.1 Accessing the GUI

To access the IBM GUI, enter the IP address that was set during the initial setup process into your web browser. You can connect from any workstation that can communicate with the system. The login window opens (Figure 4-1).



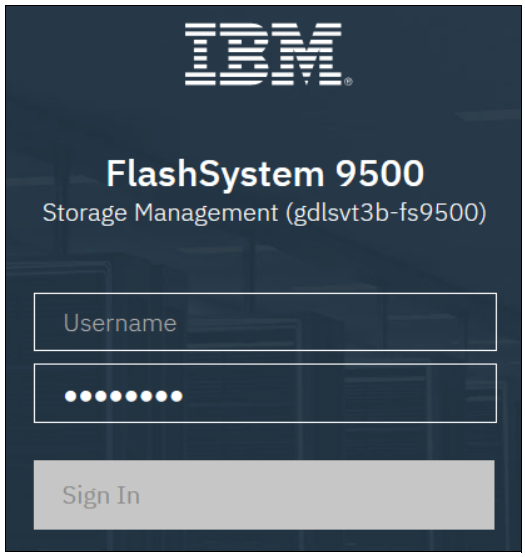


Figure 4-1 Login window of the GUI

Note: If you log in to the GUI by using the configuration node, you receive another option: Service Assistant Tool (SAT). Clicking this option takes you to the service assistant instead of the cluster GUI, as shown in Figure 4-2.

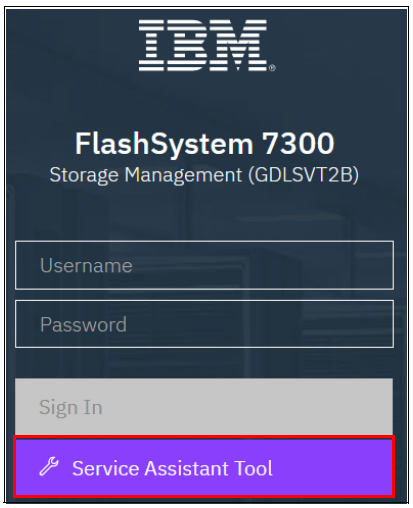


Figure 4-2 Login window of the storage system when it is connected to the configuration node

With IBM Storage Virtualize Version 8.5 and later, you can configure a Single sign On (SSO) login. (Figure 4-3). Setup can be done under the Security section (see “Single Sign-on” on page 310).

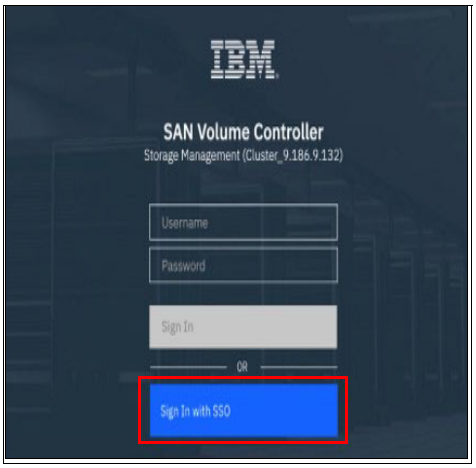


Figure 4-3 Single Sign On

It is a best practice for each user to have their own unique account. The default user accounts must be disabled for use or their passwords changed and kept secured for emergency purposes only. This approach helps to identify any personnel who are working on the systems and track all important changes that are done by them. The *superuser* account is used for initial configuration only.

After a successful login, the Version 8.6 Welcome window opens and displays the system dashboard (see Figure 4-4).

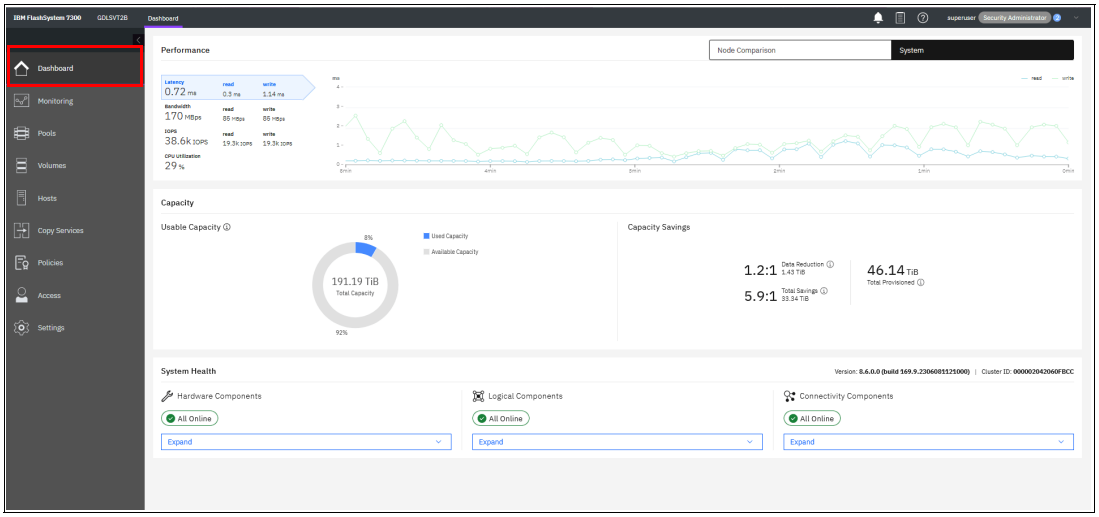


Figure 4-4 Welcome page with the dashboard

The Dashboard is divided into three sections:

- Performance

  This section provides important information about latency, bandwidth, input/output operations per second (IOPS), and CPU utilization. All this information can be viewed at the system or canister levels. A “Node comparison” view shows the differences in characteristics of each node (see Figure 4-5). The performance graph is updated with new data every 5 seconds. Measurement of time now includes microseconds support.

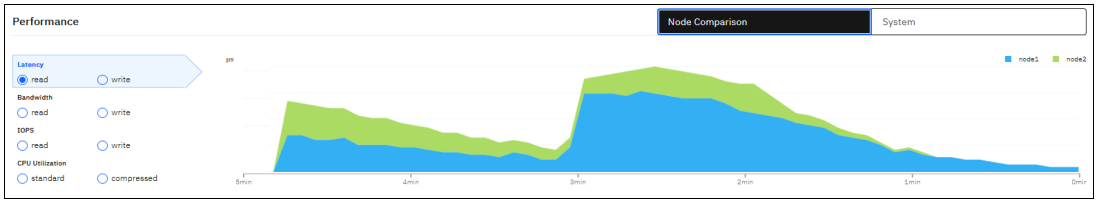


Figure 4-5 Performance statistics

- Capacity

  This section (see Figure 4-6) shows the current utilization of attached storage. It also shows provisioned capacity and capacity savings. To display a complete list of the options refer to the Volumes tab in the GUI.

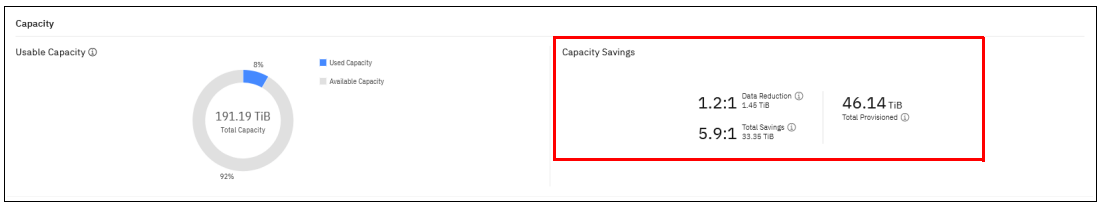


Figure 4-6 Capacity overview

Figure 4-7 shows the overview of compressed volumes

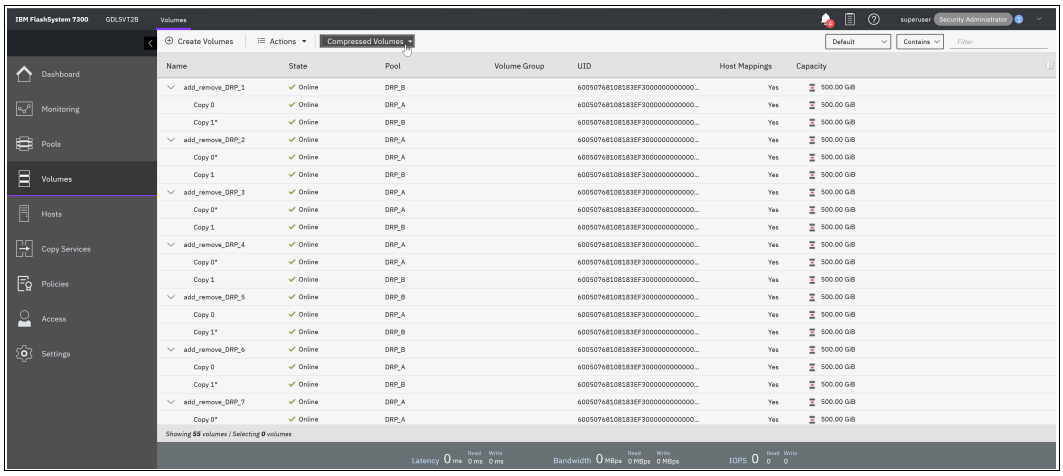


Figure 4-7 Overview of compressed volumes

Figure 4-8 shows a view of the Compression Details.

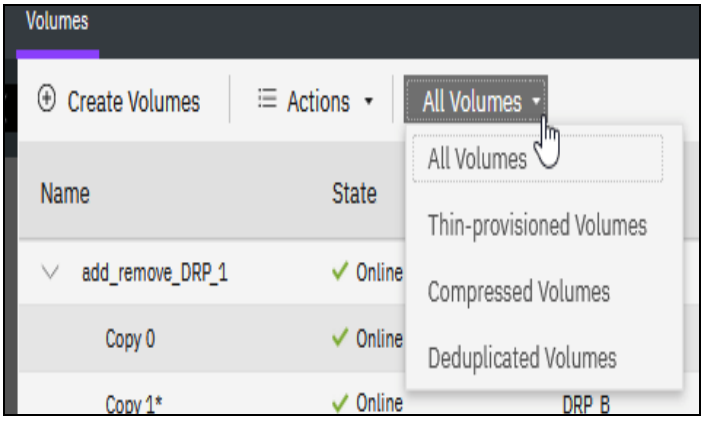


Figure 4-8 Compression Details

If the Overprovisioned External Systems section appears, you can click the section to see a list of related MDisks and pools, as shown in Figure 4-9 on page 238.

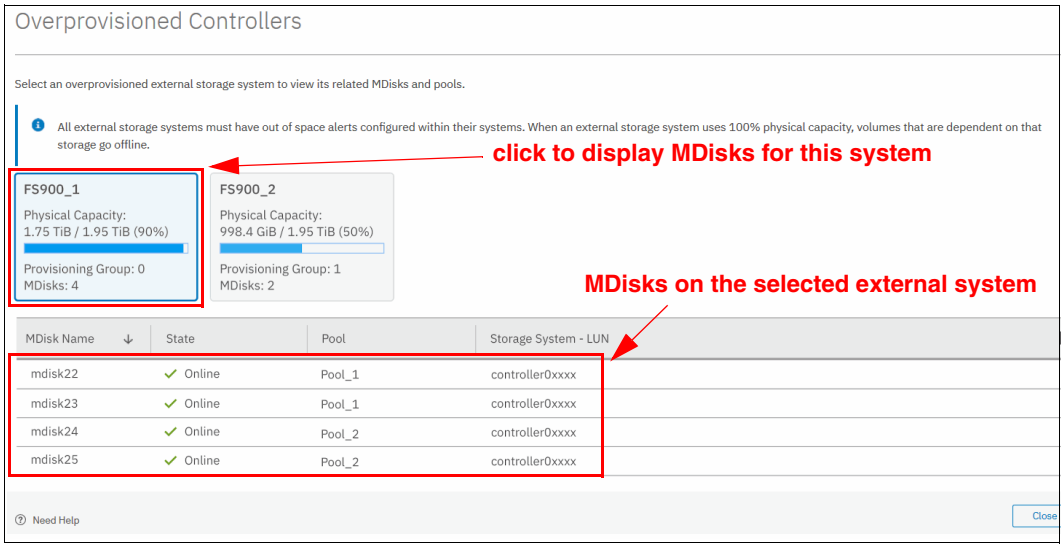


Figure 4-9 List that shows overprovisioned external storage

You also see a warning when assigning MDisks to pools if the MDisk is on an overprovisioned external storage controller.

- System Health

  This section indicates the status of all critical system components, which are grouped in three categories: Hardware, logical, and connectivity components, as shown in Figure 4-10. When you click **Expand**, each component is listed as a subgroup. You can then go directly to the section of GUI where the component that you are interested in is managed.

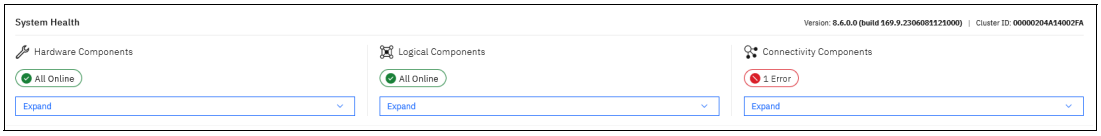


Figure 4-10 System Health overview window

Figure 4-11 shows the expanded view.

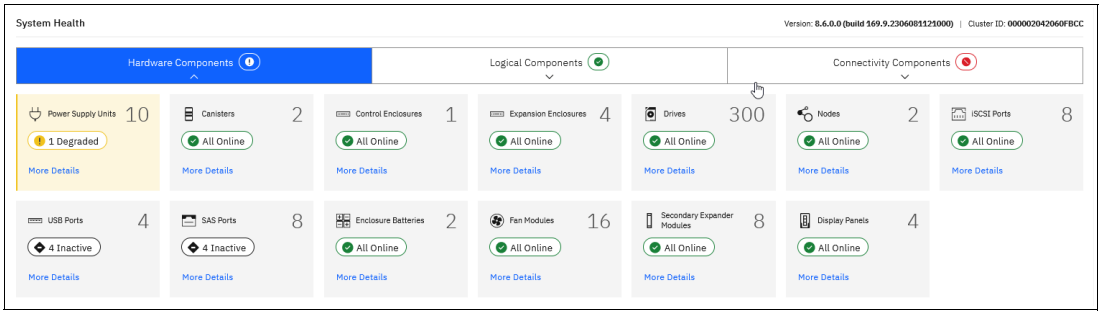


Figure 4-11 Expanded System health view

The dashboard displays as a welcome page instead of the system window as in previous versions. This system overview was moved to the **Monitoring** → **System Hardware** menu.

Although the Dashboard window provides key information about system behavior, the System menu is a preferred starting point to obtain the necessary details about your system components.

4.2 Introduction to the GUI

As shown in Figure 4-12, the former GUI System window is moved to **Monitoring** → **System Hardware**.

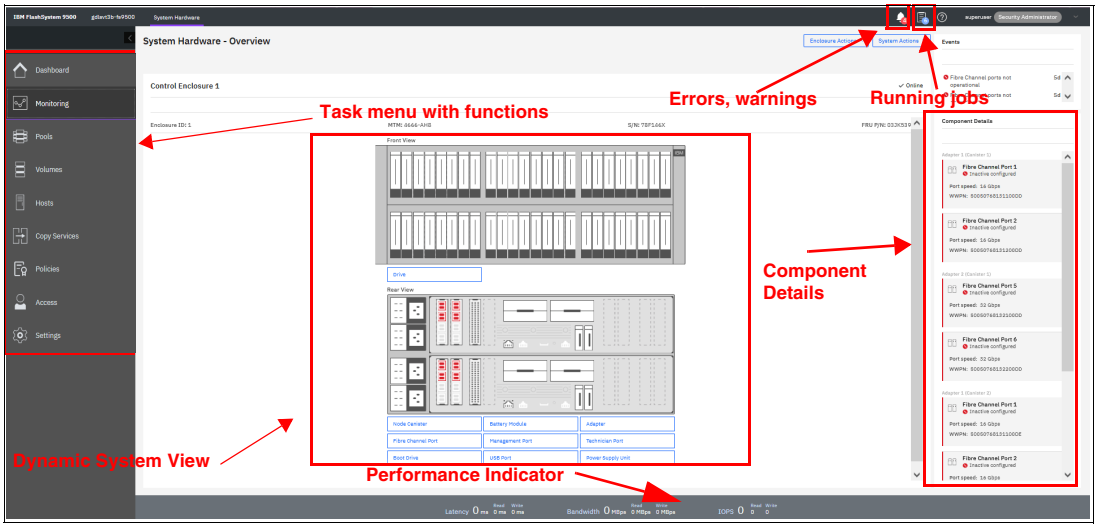


Figure 4-12 IBM Storage System Hardware window

4.2.1 Task menu

The IBM Storage Virtualize GUI task menu is always available on the left side of the GUI window. To browse by using this menu, click the action and choose a task that you want to display, as shown in Figure 4-13 on page 240.

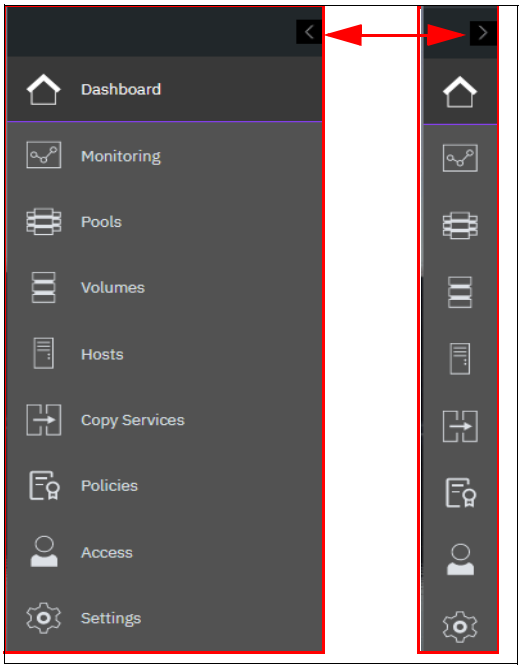


Figure 4-13 The task menu on the left side of the GUI

If you reduce the horizontal size of your browser window by using the arrows on the top, the wide task menu shrinks to the icons only.

4.2.2 Background tasks

After the initial configuration process is complete, IBM Storage Virtualize shows the information about tasks, which are running in the background, to notify the administrator that several key functions are not yet finished. If necessary, this indicator can be closed and these tasks can be reviewed later.

Figure 4-14 shows the background tasks in the System window.

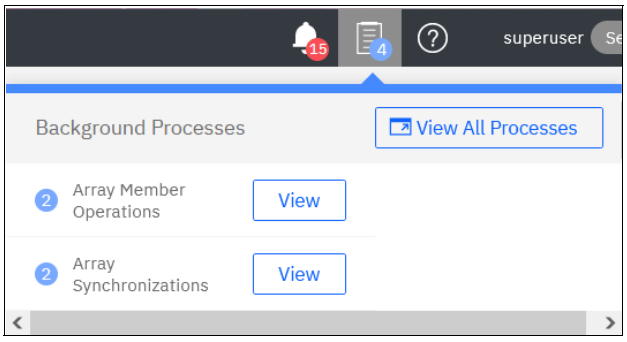


Figure 4-14 Background tasks

In this case, the GUI has two background processes that help with the general administration of the system. You can directly view the tasks from this window. Other suggested tasks that typically appear after the initial system configuration are to create a volume and configure a storage pool.

The dynamic IBM Storage Virtualize menu contains the following windows:

- Dashboard
- Monitoring
- Pools
- Volumes
- Hosts
- Copy Services
- Policies
- Access
- Settings

4.2.3 Notification icons and help

Three notification icons are in the upper navigation area of the GUI (see Figure 4-15):

- The left icon indicates warning and error alerts that were recorded in the Event log.
- The middle icon shows running jobs and suggested tasks.
- The third rightmost icon offers a help menu with content that is associated with the current tasks and the currently opened GUI menu.



Figure 4-15 Notification area

Alerts indication

The left icon in the notification area informs administrators about important alerts in the systems. Click the icon to list warning messages in yellow and errors in red (see Figure 4-16).

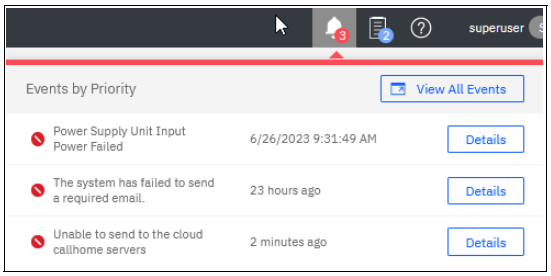


Figure 4-16 System alerts

You can go directly to the Events menu by clicking the **View All Events** option, as shown in Figure 4-17.

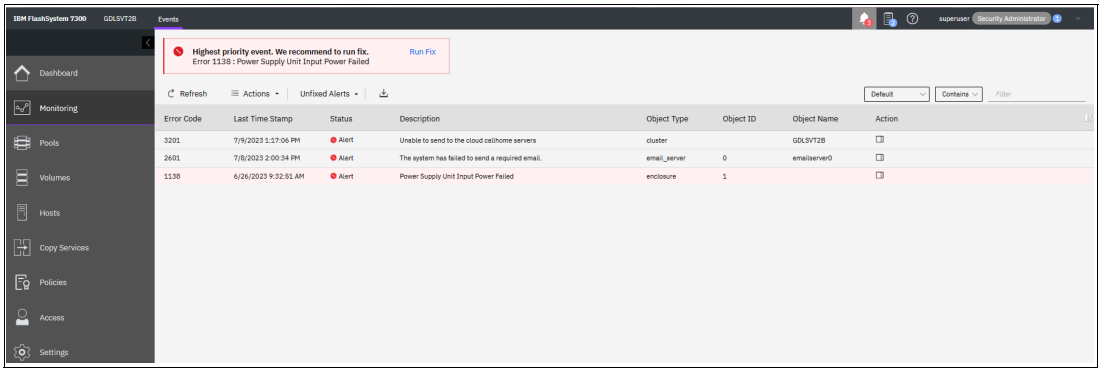


Figure 4-17 View all Events

You can see each event message separately by right-clicking the event and selecting **Properties** for the specific message. Then you can analyze the content and eventually run the suggested fix procedure, as shown in Figure 4-18.

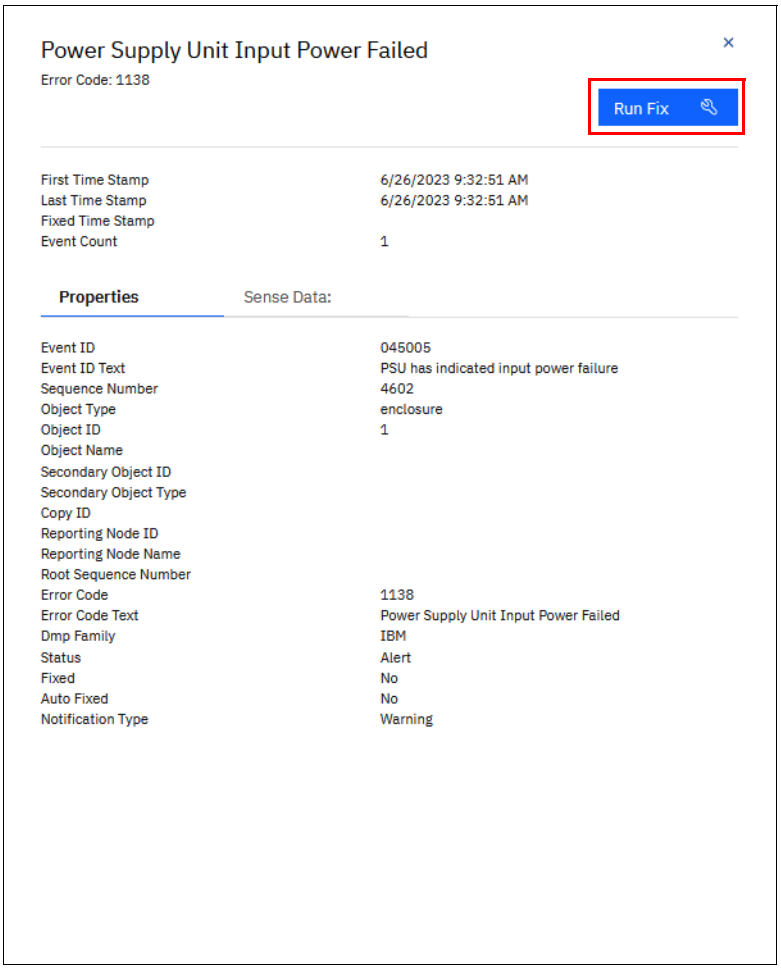


Figure 4-18 Power Supply Unit Input Power Failed

Running tasks and suggested tasks

The middle icon in the notification area provides an overview of running tasks that are triggered by the administrator.

Note: The button for suggested tasks was moved to the User Area in the upper right of the GUI window.

Running tasks

Similarly, you can analyze the details of running tasks (all of them together in one window or of a single task). Figure 4-19 shows the summary of all running background processes.

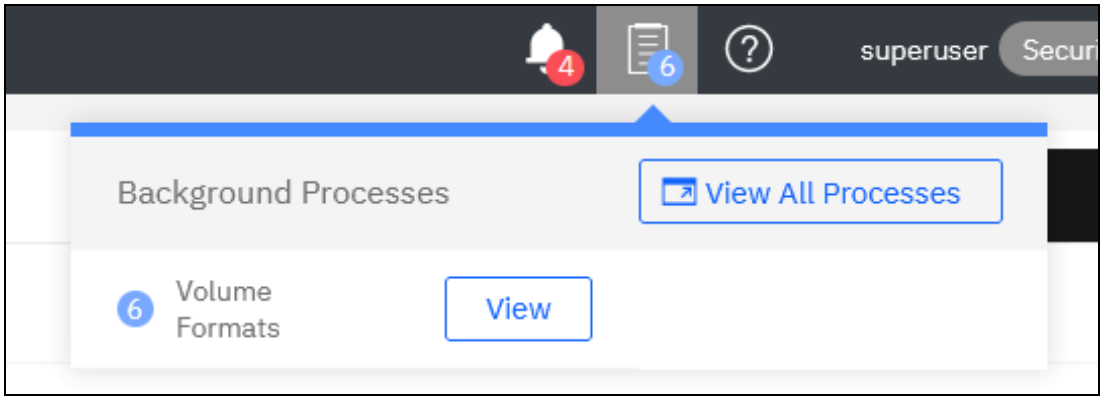


Figure 4-19 Background Processes

Click **View** to open the volume format job, as shown in Figure 4-20.

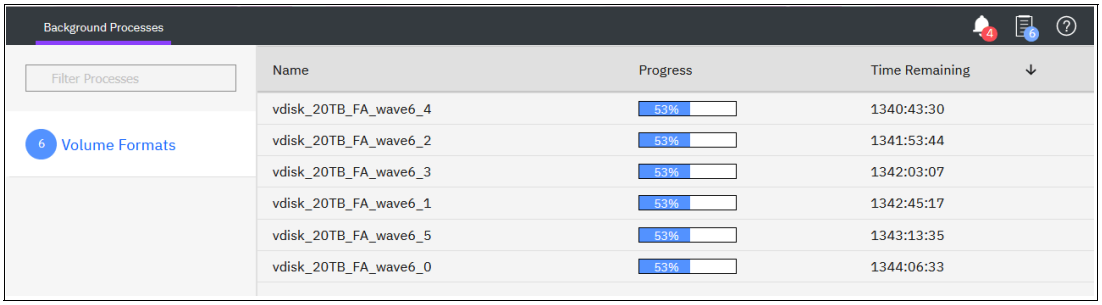


Figure 4-20 Details of a running task

The following information can be displayed as part of the running tasks:

- Volume migration
- MDisk removal
- Image mode migration
- Extent migration
- IBM FlashCopy
- Metro Mirror (MM) and Global Mirror (GM)
- Volume formatting
- Space-efficient copy repair
- Volume copy verification and synchronization
- Estimated time for the task completion

Suggested tasks

The suggested task is shown on the right side of the User Task button (see Figure 4-21).

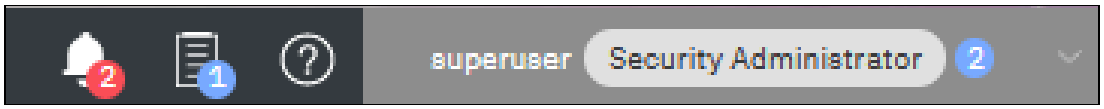


Figure 4-21 User Task button

To open the suggested task window, click the down arrow on the right side of the user action field, as indicated in Figure 4-22.

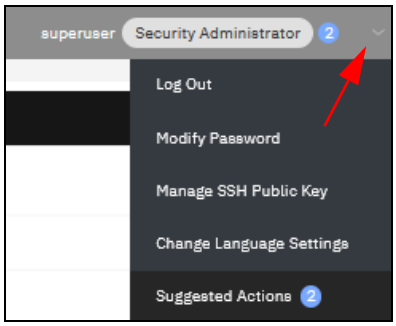


Figure 4-22 Opening the Suggested Actions window

Now, select **Suggested Actions**. The window that is shown in Figure 4-23 opens.

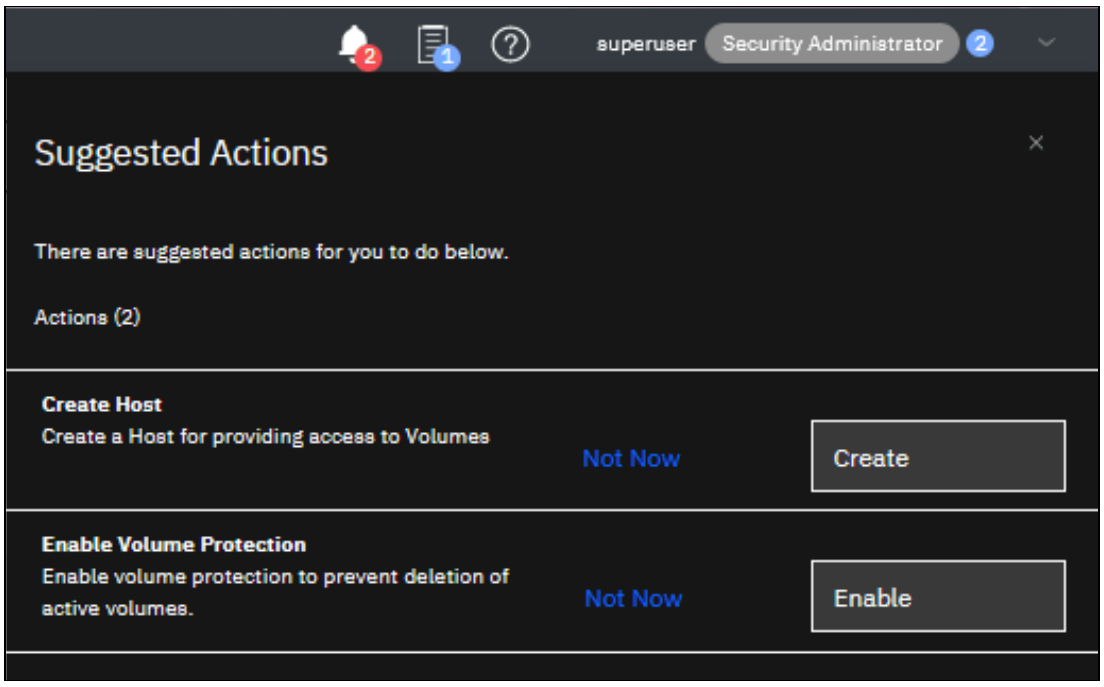


Figure 4-23 Suggested Actions window

In the example that is shown in Figure 4-23, we have not yet defined a host. Therefore, the system suggests that we do so and offers us direct access to the associated Create Host menu. If you do not want to create a host now, click **Not Now** and the suggestion message disappears.

Making selections

Recent updates to the GUI brought improved selection making. You can now select multiple items more easily. Go to a wanted window, press and hold the **Shift** or **Ctrl** key, and make your selection.

Press and hold the **Shift** key and select the first item in your list that you want, and then select the last item. All items between the two that you choose are also selected, as shown in Figure 4-24.

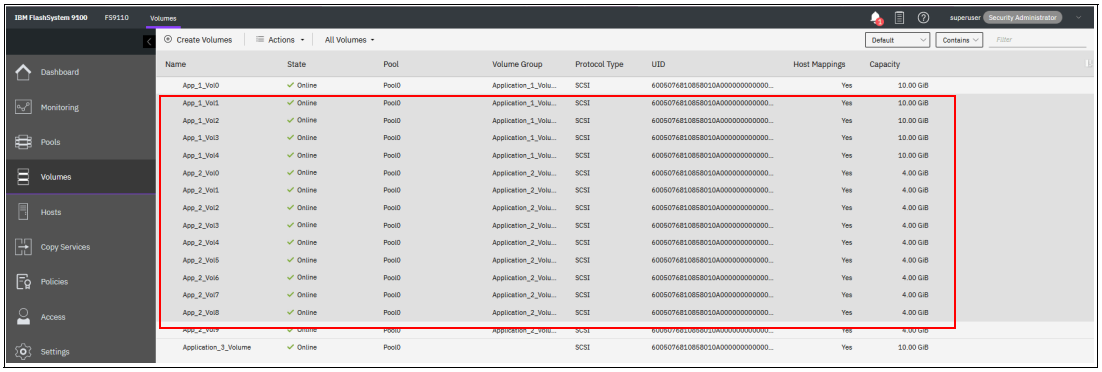


Figure 4-24 Selecting items by using the Shift key

Press and hold the **Ctrl** key and select any items from the entire list. You can select items that do not appear in sequential order, as shown in Figure 4-25.

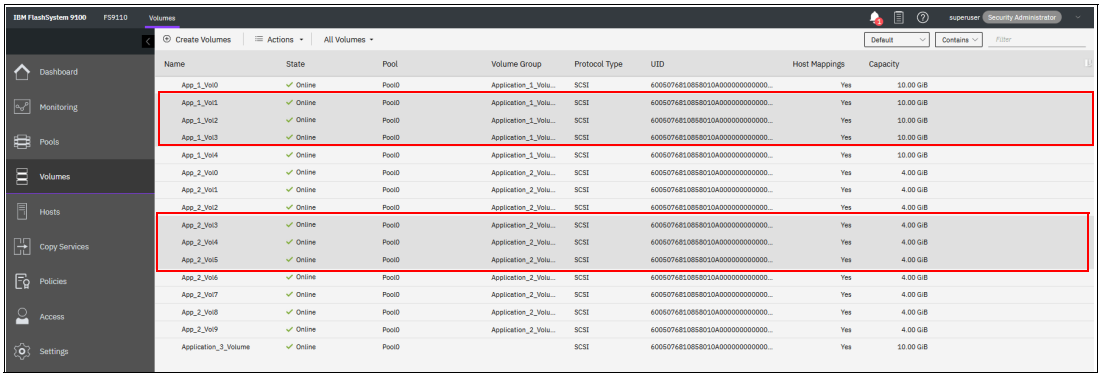


Figure 4-25 Selecting items by using the Ctrl key

You can also select items by using the built-in filtering function. For more information, see 4.3.1, “Content-based organization” on page 247.

Help

If you need help, you can select the (**?**) button, as shown in Figure 4-26.

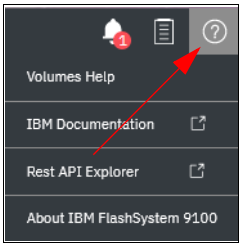


Figure 4-26 Access help menu

The following options are available:

- The first option opens a new tab with plain text information about the window that you are in and its contents.
- The second option shows the same information in IBM Documentation. This option requires an internet connection, but the first option does not because the information is stored locally on the system.
- The third option guides you to the IBM Storage Virtualize REST API

For example, in the Dashboard window, you can open help information that is related to the dashboard-provided information, as shown in Figure 4-27.

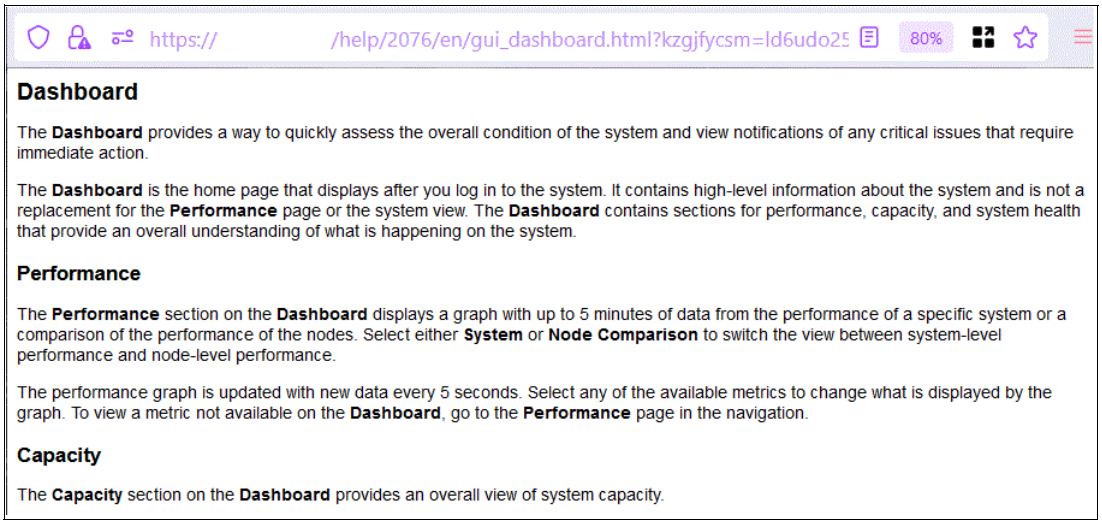


Figure 4-27 Example of Dashboard help content

Or as shown in Figure 4-28, the Help function guides you to the web page overview of the IBM Storage Virtualize REST API section.



Figure 4-28 IBM Storage Virtualize REST API

4.3 System Hardware - Overview window

The System Hardware - Overview window is shown in Figure 4-29.

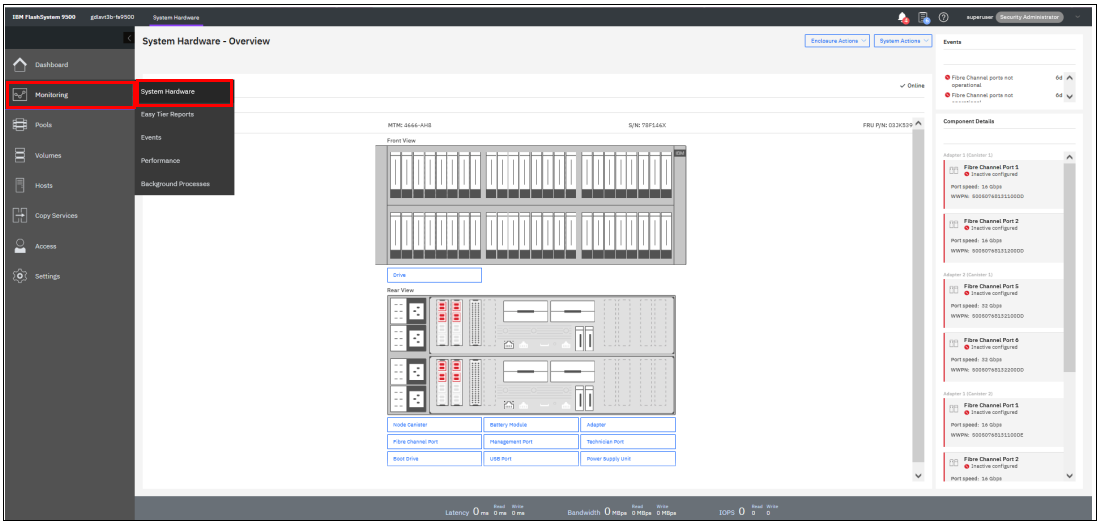


Figure 4-29 System Hardware - Overview window

The next section describes the structure of the window and how to go to various system components to manage them more efficiently and quickly.

4.3.1 Content-based organization

The following sections describe several view options within the GUI in which you can filter (to minimize the amount of data that is shown on the window), sort, and reorganize the content of the window.

Table filtering

On most pages, a Filter box is available at the upper right of the window. Use this option if the list of object entries is too long and you want to search for something specific.

To use search filtering, complete the following steps:

1. In the Filter box that is shown in Figure 4-30 on page 248, enter a search term by which you want to filter. You can also use the drop-down menus to modify what the system searches for.

   For example, if you want an exact match to your filter, select **=** instead of **Contains**. The first drop-down list limits your filter to search through a specific column only; for example, Name and State.

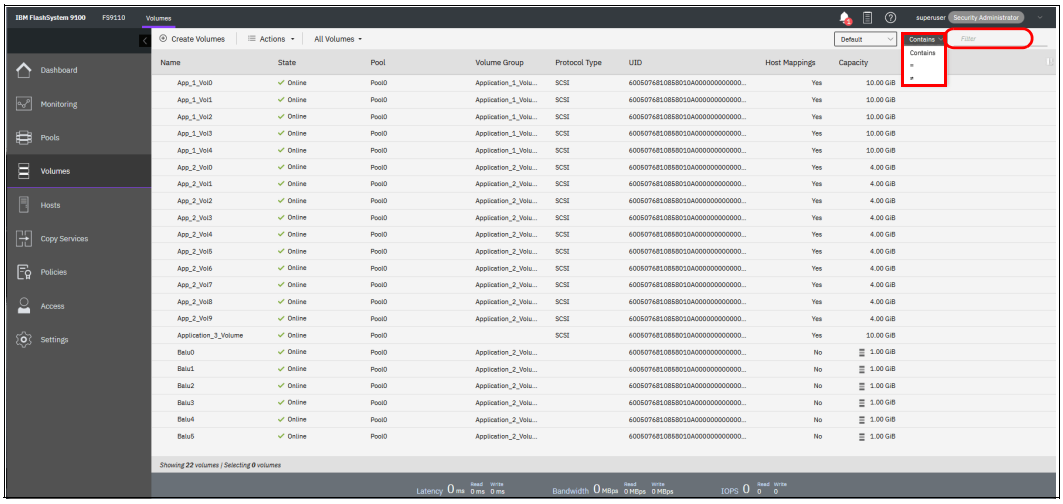


Figure 4-30 Filter search box

2. Enter the text string that you want to filter and press **Enter**.

   By using this function, you can filter your table based on column names. In our example, a volume list is displayed that contains the names that include *Balu* somewhere in the name. *Balu* is highlighted in turquoise, as are any columns that contain this information, as shown in Figure 4-31. The search option is not case-sensitive.

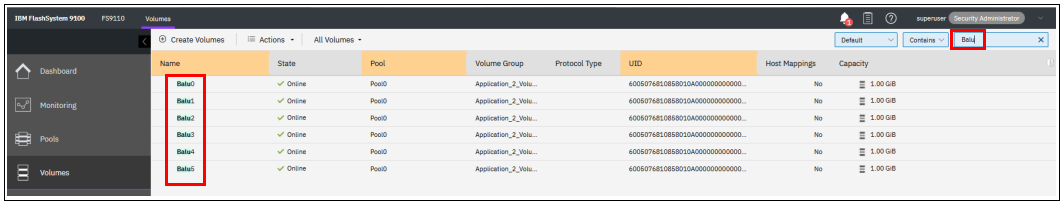


Figure 4-31 Showing filtered rows

3. Remove this filtered view by clicking the **X** icon that displays in the Filter box or by deleting what you searched for and pressing Enter, as shown in Figure 4-32.

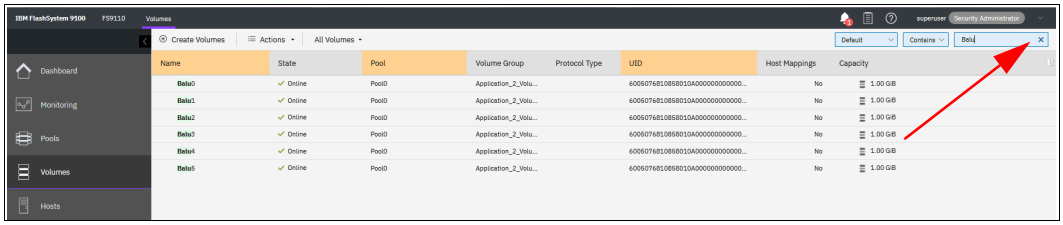


Figure 4-32 Removing the filtered view

Filtering: This filtering option is available in most menu options of the GUI.

Table information

In the table view, you can add or remove the information in the tables on most pages.

For example, on the Volumes window, complete the following steps to add a column to the table:

1. Right-click any column headers of the table or select the icon in the upper left of the table header. A list of all of the available columns displays, as shown in Figure 4-33.

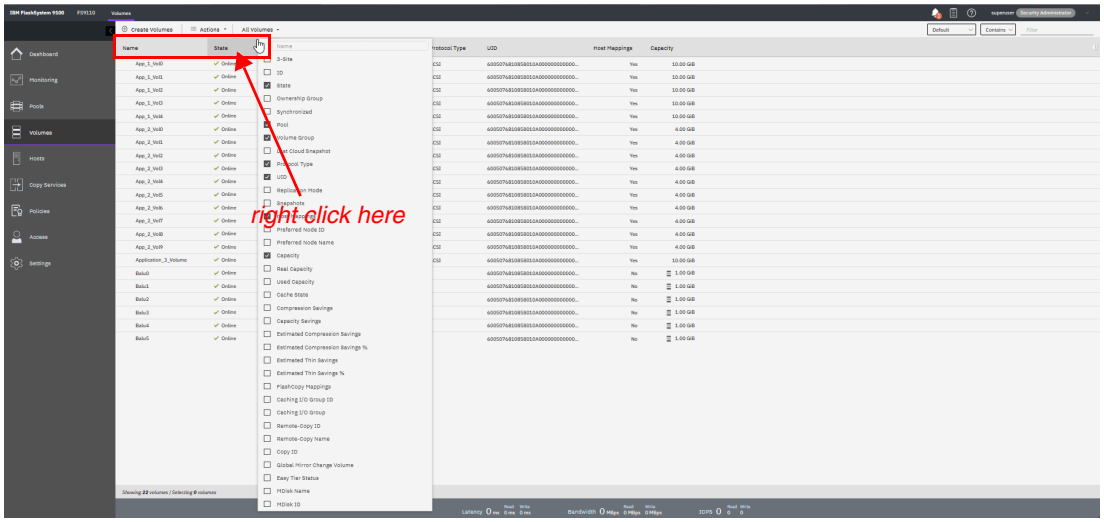


Figure 4-33 Adding or removing details in a table

2. Select the column that you want to add or remove from this table. In this example, the volume ID column was added and the content was sorted by ID, as shown on the left in Figure 4-34.

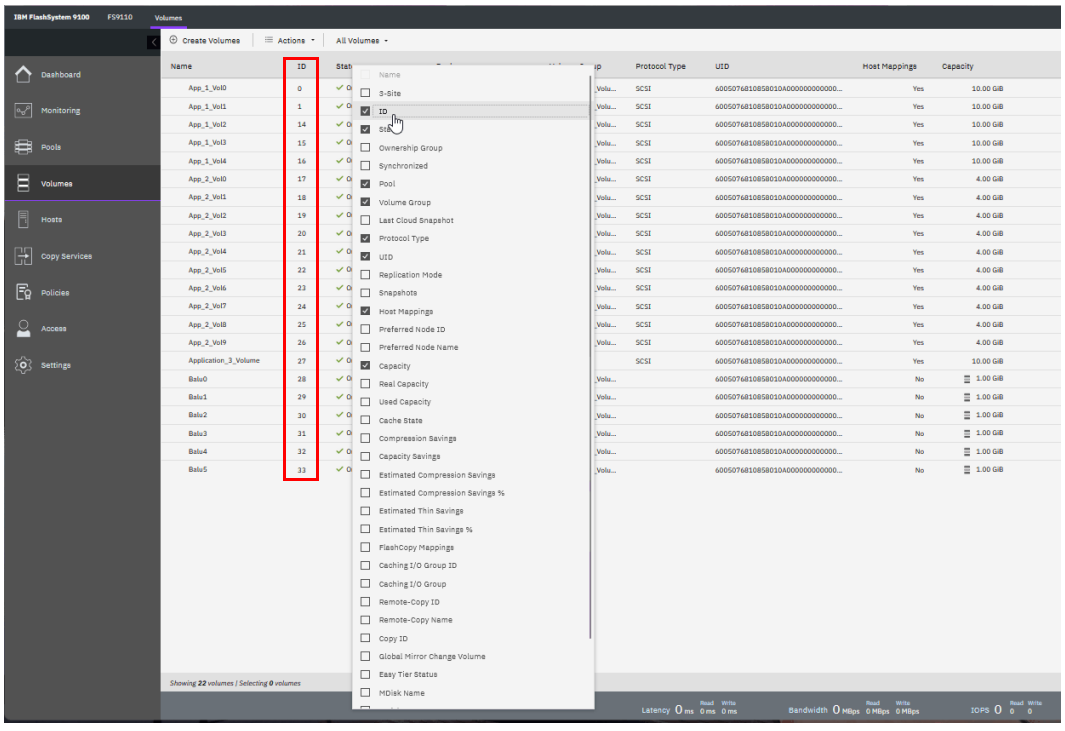


Figure 4-34 Table with an added ID column

3. You can repeat this process several times to create custom tables to meet your requirements.
4. Return to the default table view by selecting **Restore Default View** (the last entry) in the column selection menu, as shown in Figure 4-35.

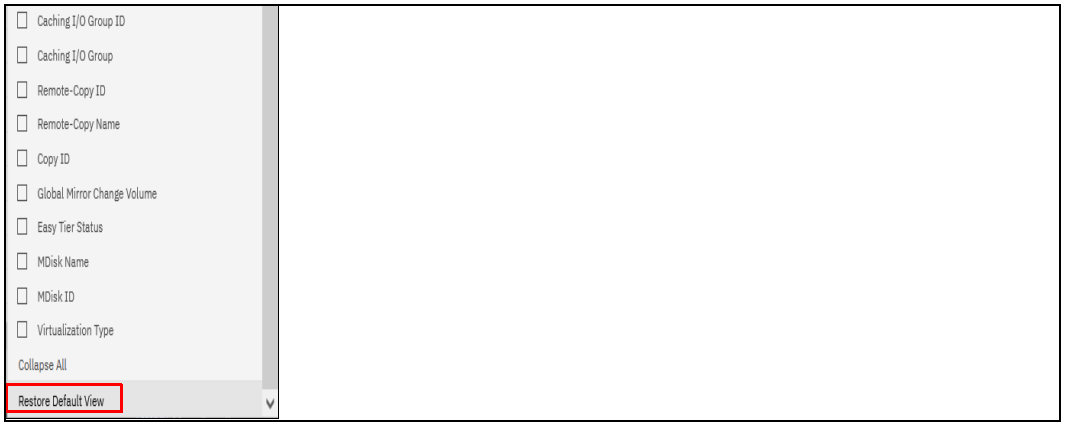


Figure 4-35 Table Restore Default View

Sorting: By clicking a column, you can sort a table based on that column in ascending or descending order.

Shifting columns in tables

You can move columns by clicking and moving the column right or left, as shown in Figure 4-36. In this example, the Capacity column is moved before the Pool column.

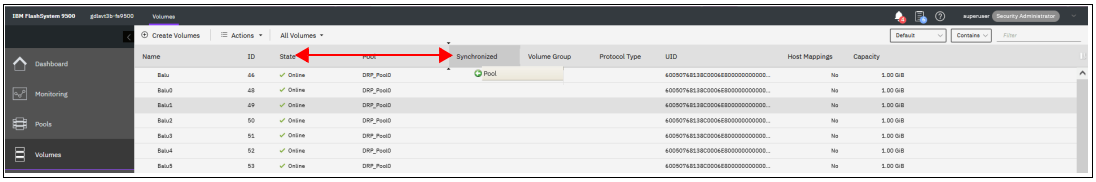


Figure 4-36 Reorganizing table columns

4.4 Monitoring menu

Click the **Monitoring** icon in the left window to open the Monitoring menu (see Figure 4-37).

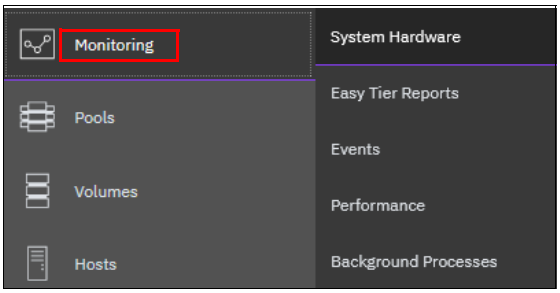


Figure 4-37 Monitoring menu

The Monitoring menu offers the following navigation options:

- System Hardware

  This option opens an overview of the system. It shows all control enclosures and groups them into I/O groups if more than one control enclosure is present. Useful information about each enclosure is displayed, including status, number of events against each enclosure, and key enclosure information, such as ID and serial number.

  For more information, see 4.4.1, “System Hardware Overview window” on page 252.

- Easy Tier Reports

  This option gives you an overview of the Easy Tier Activities in your system. Easy Tier eliminates manual intervention when you assign highly active data on volumes to faster responding storage. In this dynamically tiered environment, data movement is seamless to the host application, regardless of the storage tier in which the data belongs.

  However, you can manually change the default behavior. For example, you can turn off Easy Tier on pools that have any combination of the four types of MDisks. If a pool contains one type of MDisk, Easy Tier enters balancing mode. When the pool contains multiple types of MDisks, Easy Tier is automatically turned on.

  For more information, see 4.4.2, “Easy Tier reports” on page 261.

- Events

  This option tracks all informational, warning, and error messages that occurred in the system. You can apply various filters to sort the messages according to your needs or export the messages to an external comma-separated value (CSV) file.

  For more information, see 4.4.3, “Events option” on page 262.

- Performance

  This option reports the general system statistics that relate to the processor (CPU) utilization, host and internal interfaces, volumes, and MDisks. With this option, you can switch between megabytes per second (MBps) or IOPS.

  For more information, see 4.4.4, “Performance window” on page 264.

- Background Processes

  The option shows the progress of all tasks running in the background, as described in 4.4.5, “Background Processes” on page 265.

4.4.1 System Hardware Overview window

This section describes the System Hardware Overview window for an IBM SAN Volume Controller Cluster and a FlashSystem.

System Hardware Overview window for an IBM SAN Volume Controller Cluster

The System Hardware Overview for an IBM SAN Volume Controller Cluster is shown in Figure 4-38.

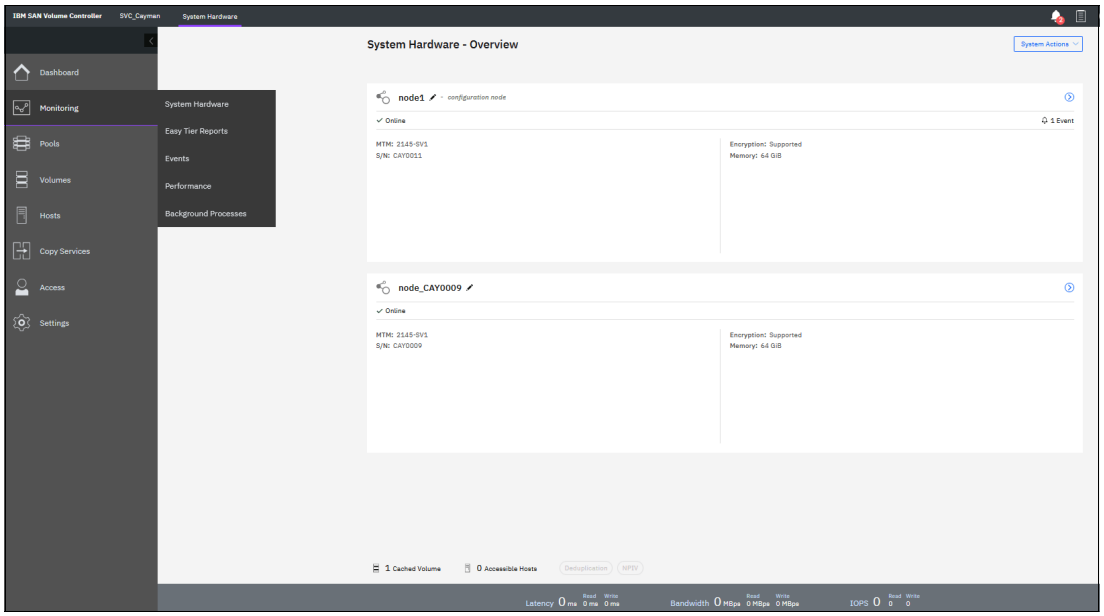


Figure 4-38 System Hardware Overview for an IBM SAN Volume Controller Cluster

To select the **Detailed Monitoring** view, select one node, as shown in Figure 4-39.

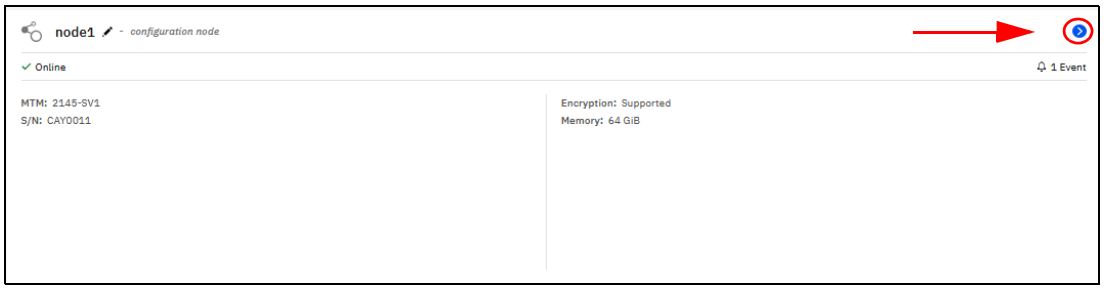


Figure 4-39 Select the IBM SAN Volume Controller node

Now, you can select different components to view detail; for example, FRU P/N or the health status of the batteries (see Figure 4-40 on page 253).

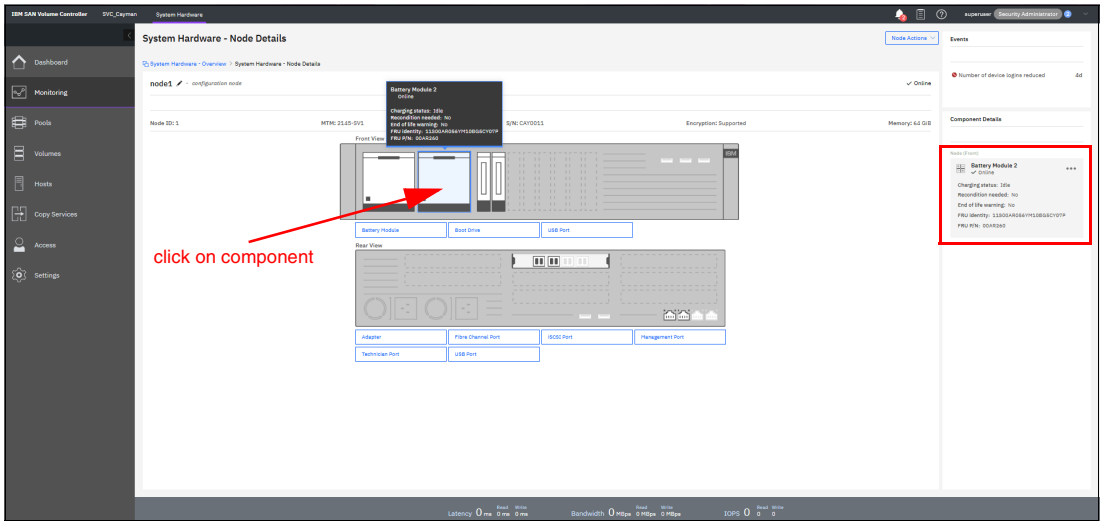


Figure 4-40 Health status of components

To view detailed properties, right-click the three dots that are in the upper right of the window (see Figure 4-41).

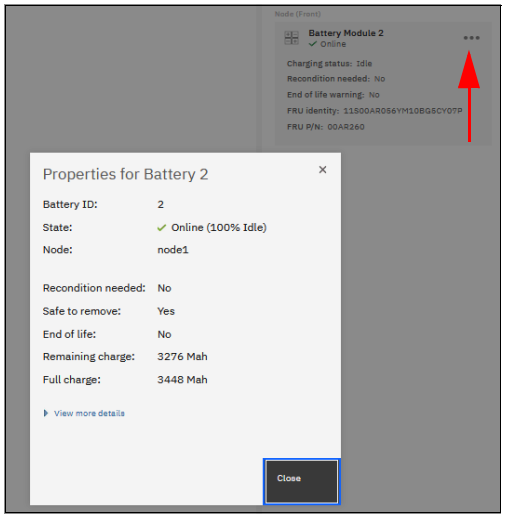


Figure 4-41 Detailed properties

In an environment with several IBM SAN Volume Controller clusters, you can easily direct the onsite personnel or technician to the correct device by enabling the identification LED on the front panel by completing the following steps:

1. Select the suitable Node Actions and click **Identify**, as shown in Figure 4-42

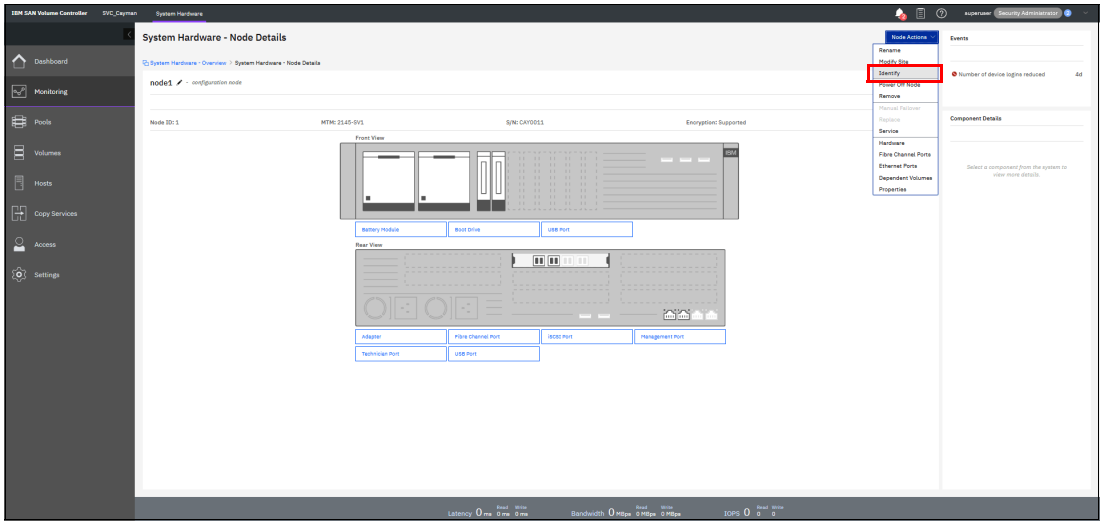


Figure 4-42 Identify node

2. Wait for confirmation from the technician that the device in the data center was correctly identified. In the GUI, you see an Info field, which indicates that the Identify LED was turned on.
3. After the confirmation, click **Turn LED Off** (see Figure 4-43).

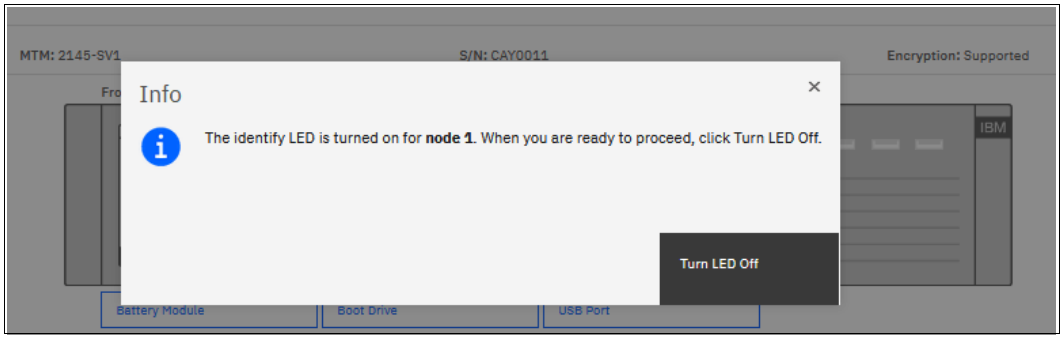


Figure 4-43 Turn LED Turn off LED

Alternatively, you can use the command line interface (CLI) to get the same results. Enter the following commands in the sequence that is shown:

1. **`chnode -identify yes node1`**
2. **`chnode -identify no node1`**

You can use the same CLI to obtain results for a specific controller or drive.

Figure 4-44 shows the backward view of a SAN Volume Controller SV3.

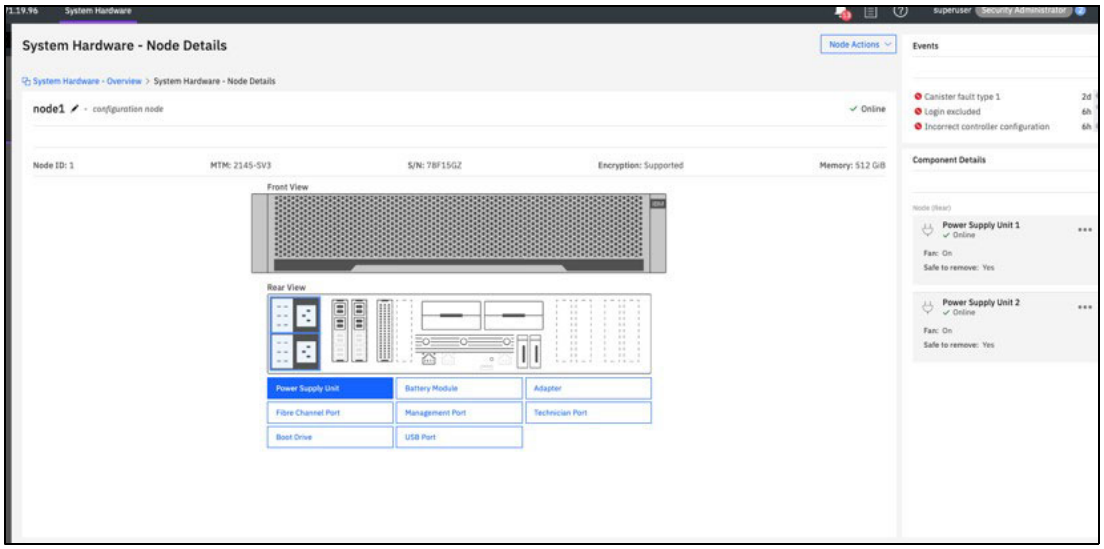


Figure 4-44 Backward view of a SAN Volume Controller SV3

System Hardware Overview window for a FlashSystem

The System Hardware option of the Monitoring menu provides a general overview. If more than one control enclosure exists in a cluster, each enclosure has its own I/O group section (see Figure 4-45).

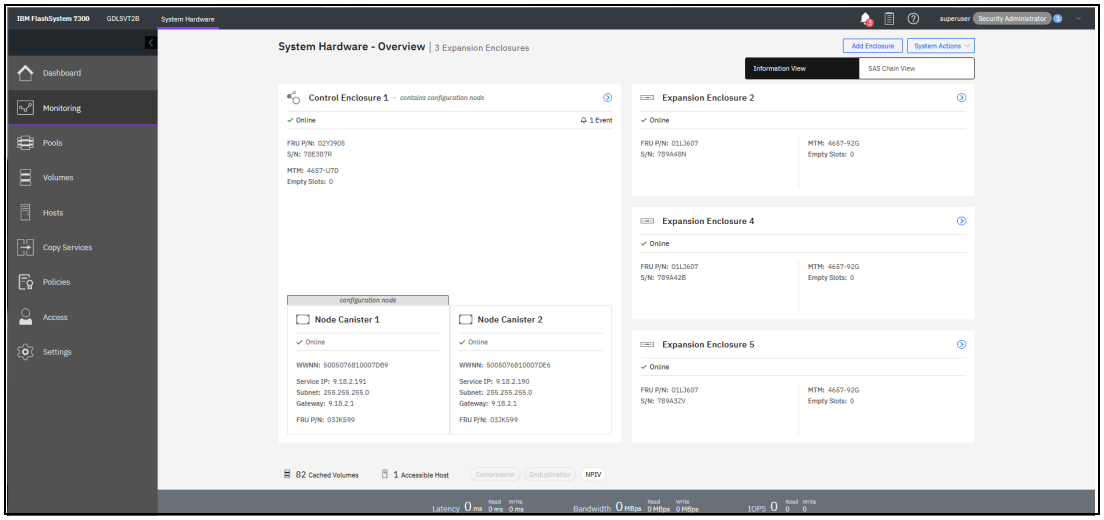


Figure 4-45 System Hardware Overview for a FlashSystem 7300

Figure 4-46 shows how to display more information about the system hardware enclosure.

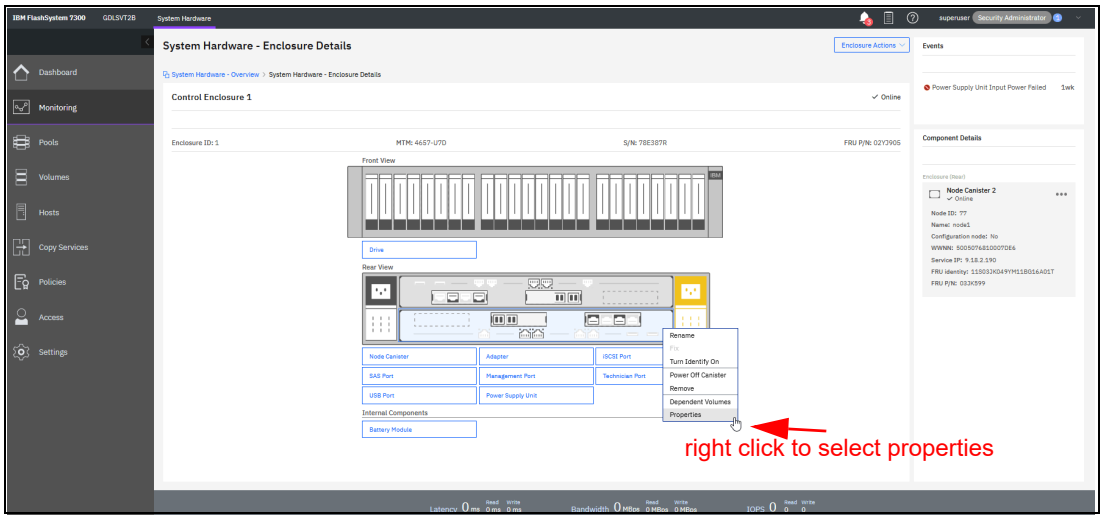


Figure 4-46 Selecting more system hardware enclosure details

This view shows all external components in real time. The properties of Canister 2 are shown in Figure 4-47. You can click any component in the graphic view or the list view at the bottom to view details.

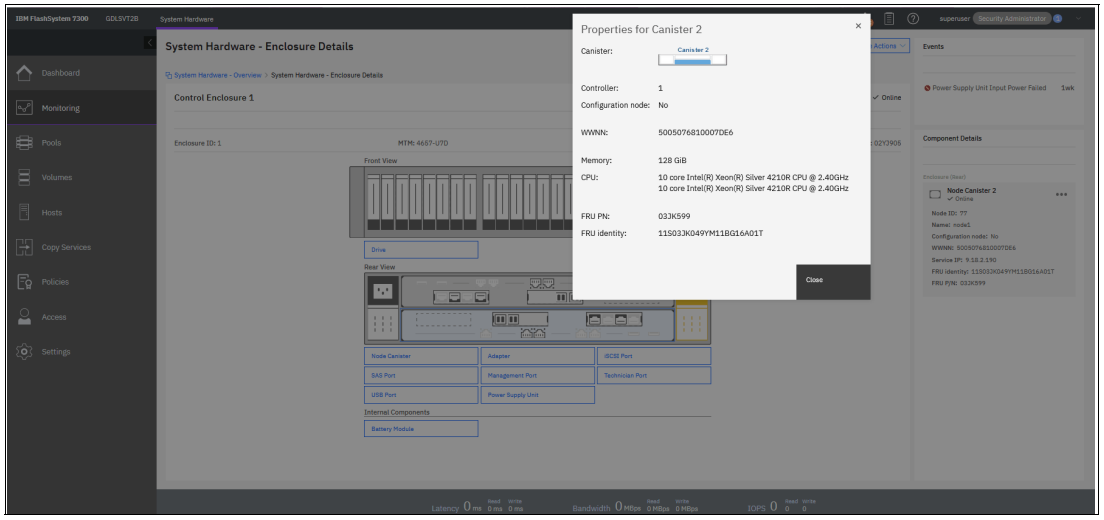


Figure 4-47 System hardware enclosure details

For example, clicking the **Adapter** provides details, such as whether the Adapter is online and the IP Address of each port, as shown in Figure 4-48. For more information about the component, see the right side of the window under the Component Details section that shows when a component is selected.

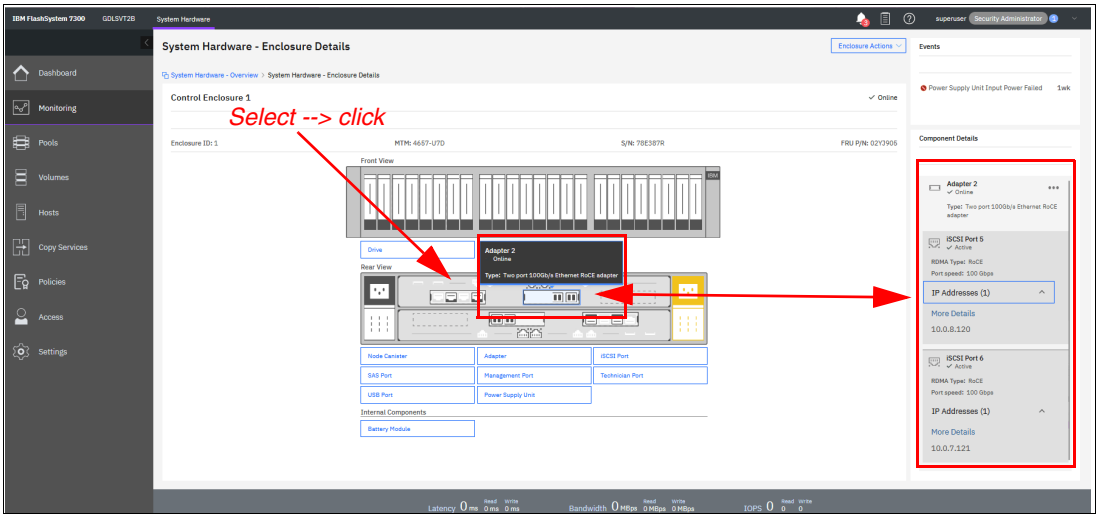


Figure 4-48 Showing the Adapter details

By right-clicking and selecting **Properties**, you see detailed technical parameters, such as Written Capacity Limit, Raw Capacity, Type, and field-replaceable unit (FRU) number, as shown in Figure 4-49.

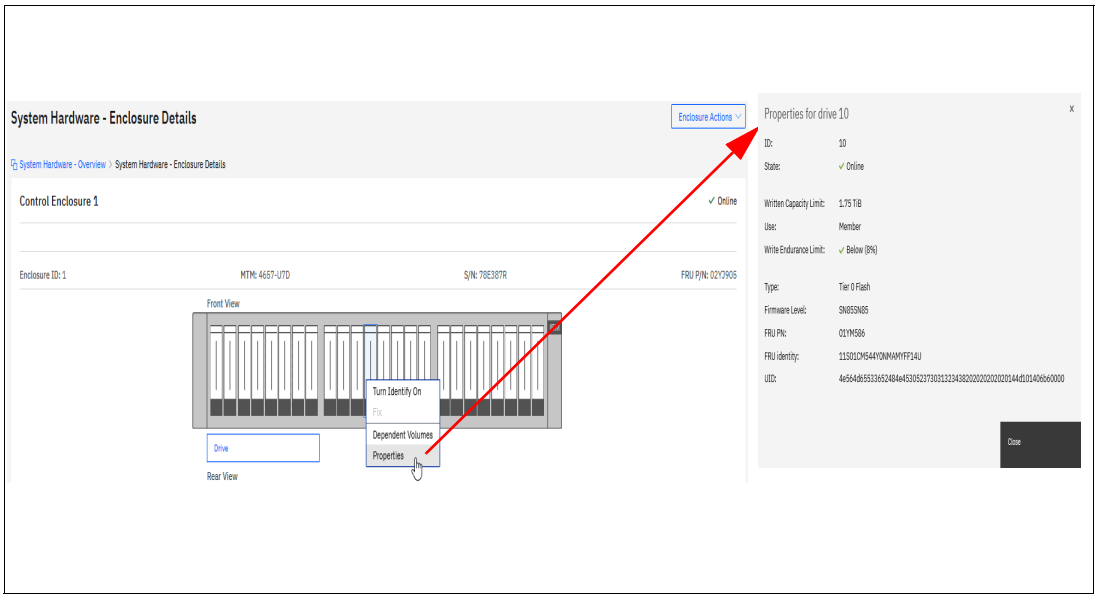


Figure 4-49 Drive information

In an environment with multiple IBM FlashSystem storage system clusters, you can easily direct the onsite personnel or technician to the correct device by enabling the identification LED on the front window by completing the following steps:

1. Select the suitable drive and click **Turn Identify On**, as shown in Figure 4-50.

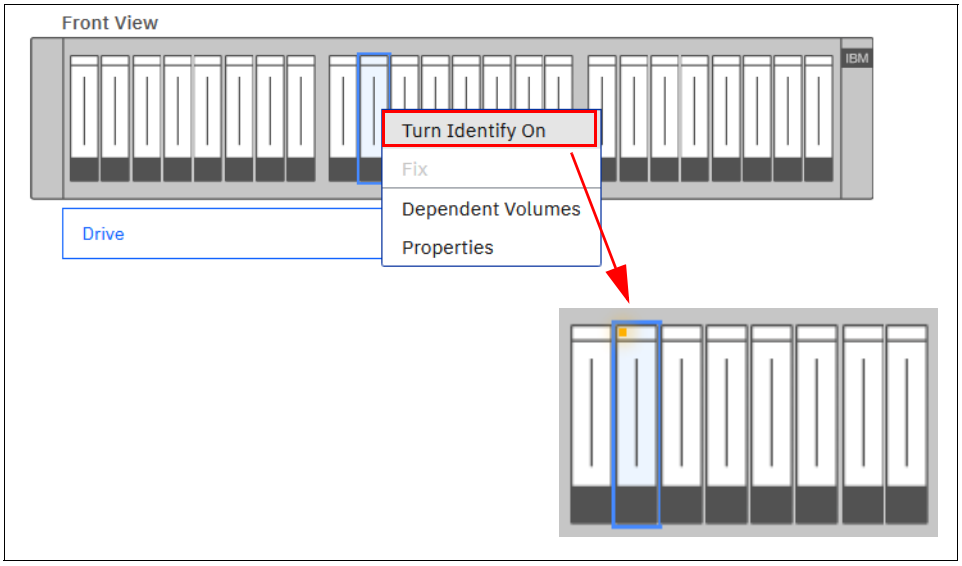


Figure 4-50 Turning on the Identify LED

2. Wait for confirmation from the technician that the device in the data center was correctly identified. In the GUI, you see a flashing light, which indicates that the Identify LED was turned on.
3. After the confirmation, click **Turn Identify Off** (see Figure 4-51).

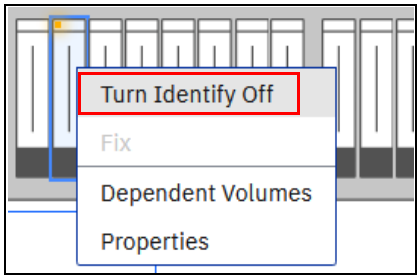


Figure 4-51 Turning off the Identify LED

Alternatively, you can use the command line interface (CLI) to get the same results. Use the following commands in the sequence that is shown:

1. Enter **`svctask chenclosure -identify yes 1`** or **`chenclosure -identify yes 1`**)
2. Enter **`svctask chenclosure -identify no 1`** or **`chenclosure -identify no 1`**)

You can use the same CLI to obtain results for a specific controller or drive.

To view internal components (that is, internal components that are hidden from view), review the bottom of the GUI underneath where the list of external components is displayed. You can select any of these components and the details are displayed in the right window, as with the external components.

Figure 4-52 shows the backside of the enclosure.

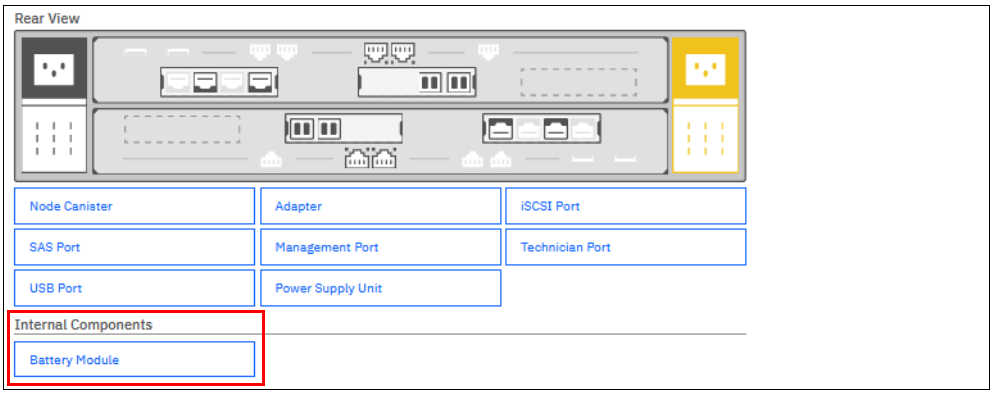


Figure 4-52 Viewing the internal components

You can also choose **SAS Chain View** to view directly attached expansion enclosures, as shown in Figure 4-53. A useful view of the entire serial-attached Small Computer System Interface (SCSI) (SAS) chain is displayed, with selectable components that show port numbers and canister numbers, along with a cable diagram for easy cable tracking.

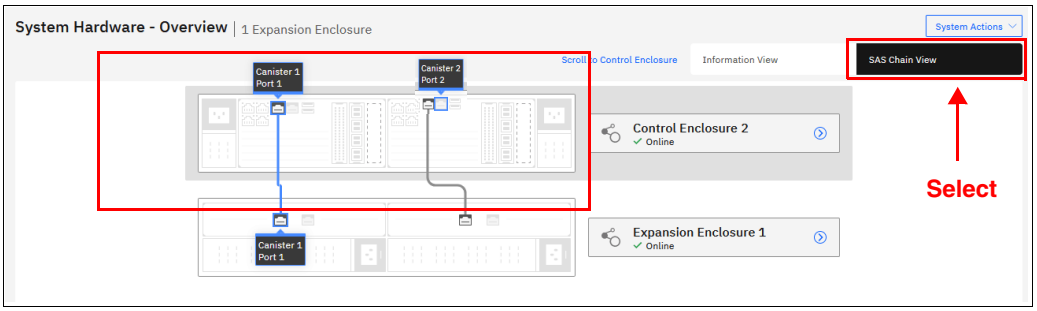


Figure 4-53 SAS Chain View

You can select any enclosure to get more information, including serial number and model type, as shown in Figure 4-54 on page 260, where expansion enclosure 2 is selected. You can also see the Events and Component Details areas at the right side of the window, which shows information that relates to the enclosure or component that you select.

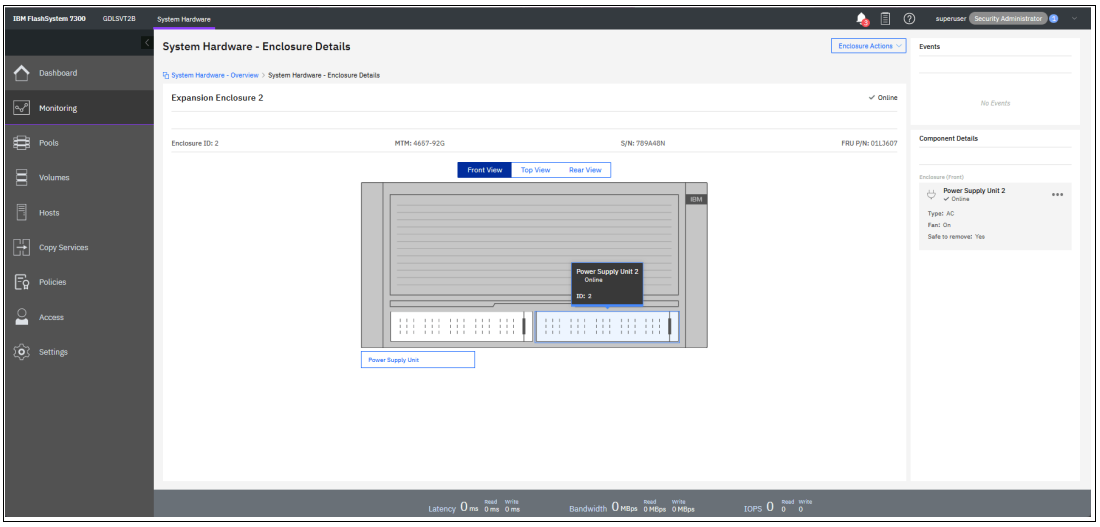


Figure 4-54 Enclosure Details window

This example includes a 4657-92G Expansion. Figure 4-55 provides a Top View.

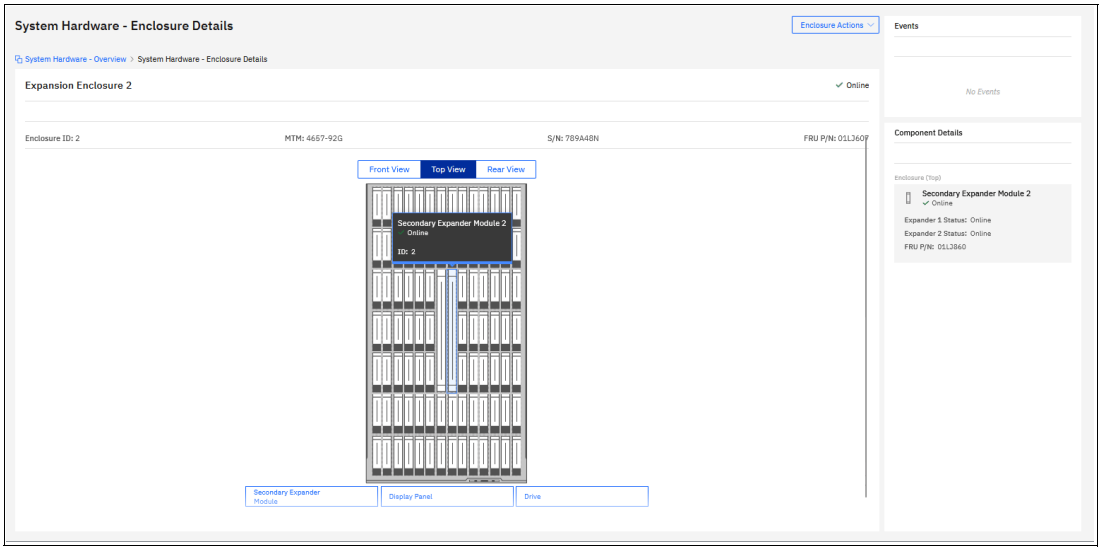


Figure 4-55 Top View of a 4657-92G Expansion Enclosure

With directly attached expansion enclosures, the view is condensed to show all expansion enclosures on the right side, as shown in Figure 4-56. The number of events against each enclosure and the enclosure status are displayed for quick reference. Each enclosure is selectable, which brings you to the Expansion Enclosure View window.

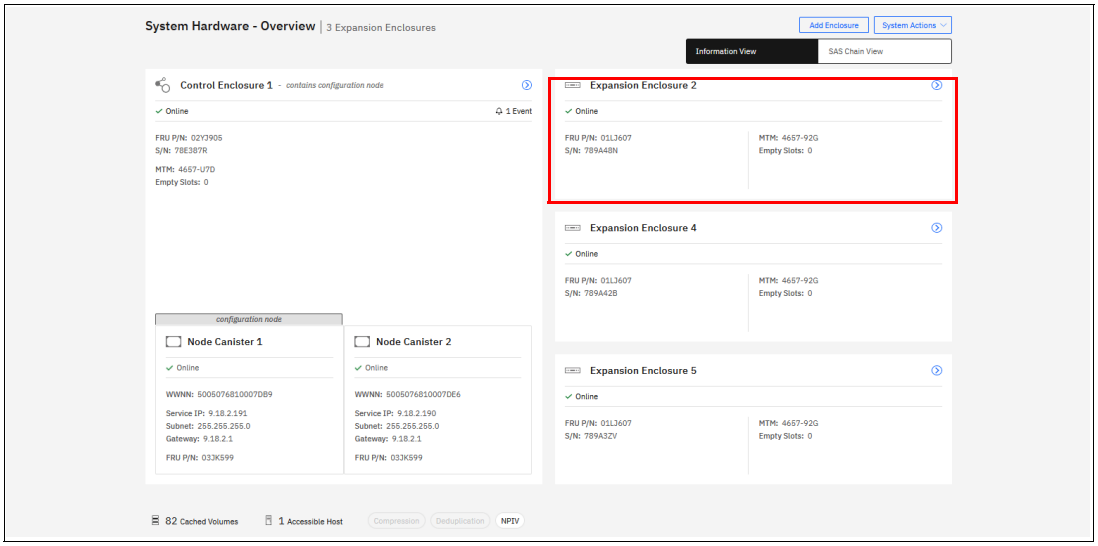


Figure 4-56 System Overview with attached enclosures

4.4.2 Easy Tier reports

The management GUI supports monitoring Easy Tier data movement in graphical reports to help you understand the performance of your storage device. Charts for data movement, tier composition, and workload skew comparison can be viewed as web-generated HTML files in a browser, or downloaded as `.csv` files.

Data is collected by the IBM Storage Tier Advisor Tool (IBM STAT) in 5-minute increments. When data that is displayed in increments that are larger than 5 minutes (for example, 1 hour), the data that is displayed for that 1 hour is the sum of all the data points that were received for that 1-hour time span.

To view Easy Tier data and reports in the management GUI, from the management GUI, select **Monitoring** → **Easy Tier Reports** (see Figure 4-57). A second way is to view Easy Tier Statistics over the Pools view: **Pools** → **View Easy Tier Reports**.

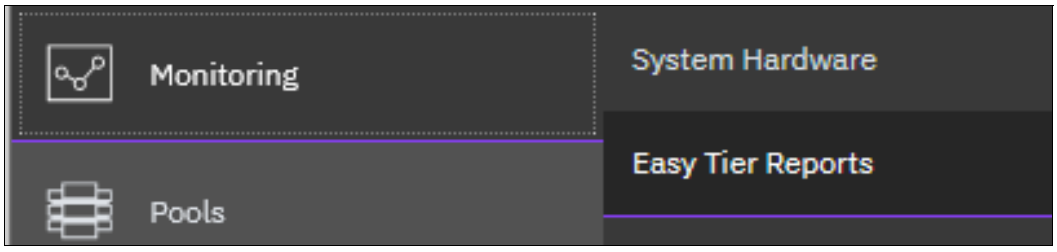


Figure 4-57 Selecting Easy Tier Reports

The following charts are available for every pool:

- Data Movement statistics

  This chart displays the migration actions that are triggered by Easy Tier.

- Tier composition statistics

  This chart displays the distributed workload between the top tier, middle tier, and bottom tier. Each tier is composed of one or more tier types.

- Workload Skew Comparison

  The percentage of I/O workload compared to the total capacity is displayed in this chart.

Figure 4-58 shows how to export the Easy Tier Reports.

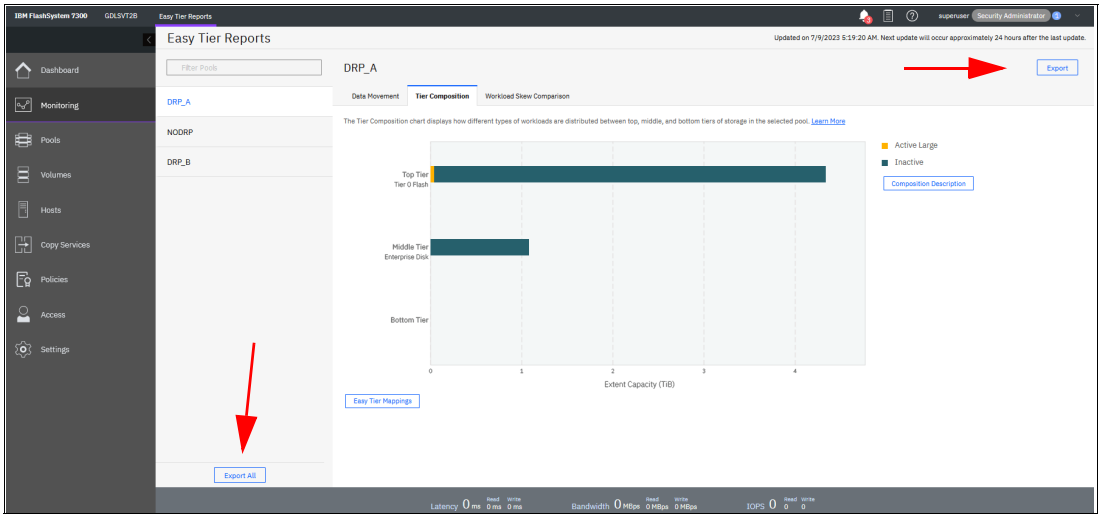


Figure 4-58 Easy Tier Reports

You can export your Easy Tier stats to a `.csv` file for further analysis. For more information about Easy Tier Reports, see Chapter 9, “Advanced features for storage efficiency” on page 669.

4.4.3 Events option

The Events option, which is available in the Monitoring menu, tracks all informational, warning, and error messages that occur in the system. You can apply various filters to sort them, or export them to an external `.csv` file. A `.csv` file can be created from the information that is shown here. Figure 4-59 on page 263 provides an example of records in the system Event log. Another comment field is available when you select a special event. For the error messages with the highest internal priority, perform corrective actions by running fix procedures. Click **Run Fix** (see Figure 4-59 on page 263).

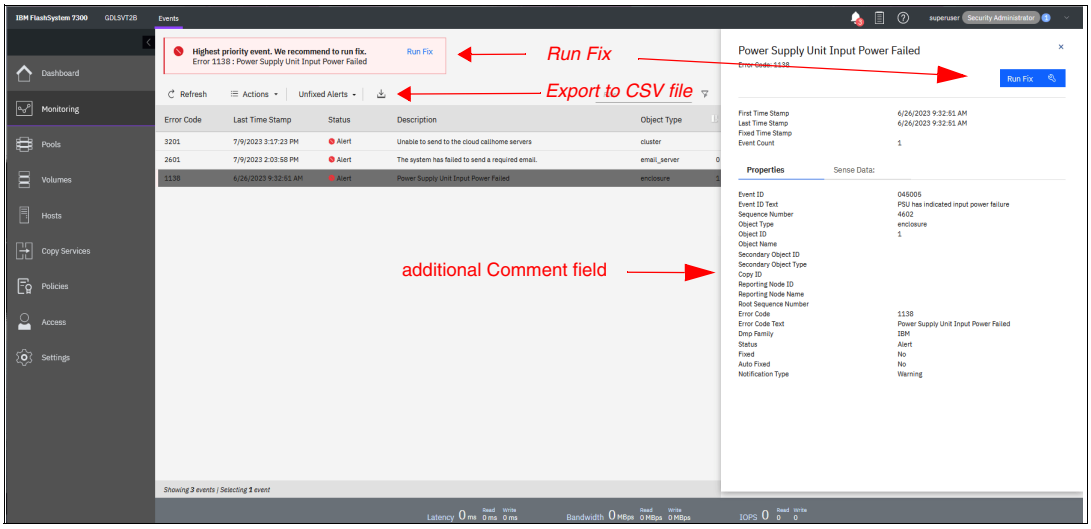


Figure 4-59 Event log list

The fix procedure wizard opens, as shown in Figure 4-60.

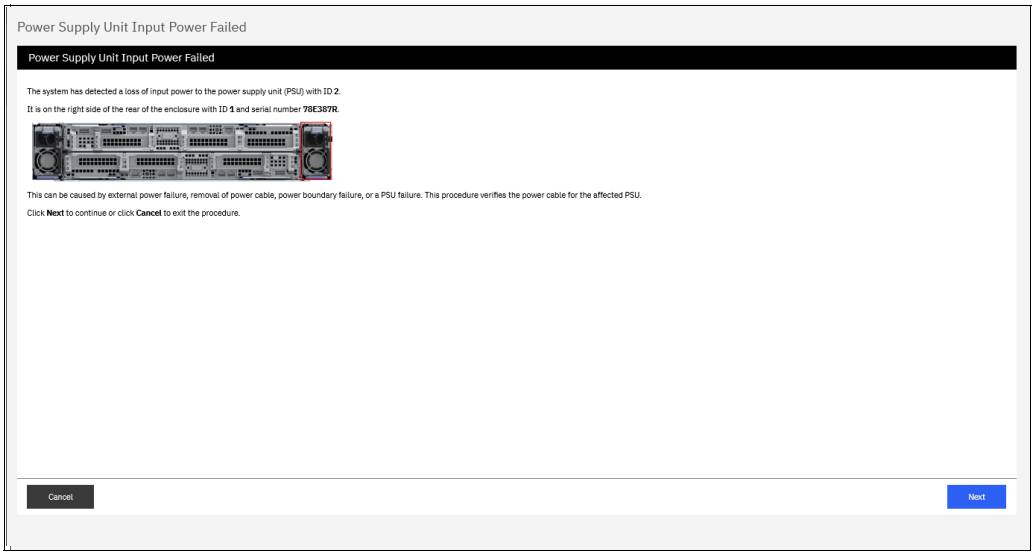


Figure 4-60 Performing a fix procedure

The wizard guides you through the troubleshooting and fixing process from a hardware or software perspective. If you determine that the problem cannot be fixed without a technician's intervention, you can cancel the procedure execution at any time.

For more information about fix procedures, see Chapter 11, "Reliability, availability, and serviceability" on page 963.

4.4.4 Performance window

The Performance window reports the general system statistics that relate to processor (CPU) utilization, host and internal interfaces, volumes, and MDisks. You can switch between MBps or IOPS, and drill down in the statistics to the node level. This capability might be useful when you compare the performance of each control canister in the system if problems exist after a node failover occurs (see Figure 4-61).



Figure 4-61 Performance statistics of the IBM FlashSystem storage system

By default, the performance statistics in the GUI show the latest 5 minutes of data. To see details of each sample, click the graph and select the timestamp, as shown in Figure 4-62.

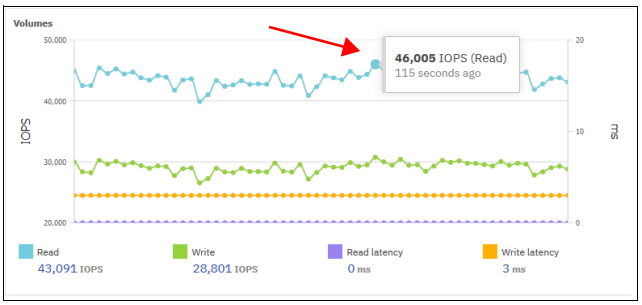


Figure 4-62 Sample details

The charts that are shown in Figure 4-62 represent 5 minutes of the data stream. For in-depth storage monitoring and performance statistics with historical data about your system, use IBM Storage Control or IBM Storage Insights.

Note: The management GUI supports latency metrics that are displayed in microseconds and milliseconds. Measurements for metrics are rendered based on whether current values are less than 1.000 milliseconds. Values over 1.000 milliseconds are displayed as milliseconds and metrics under that value display as microseconds. Whereas in command line interface (CLI) the values are always displayed in microseconds.

You can also obtain a no-charge unsupported version of the Quick Performance Overview (**`qperf`**). For more information, see How to schedule Quick Performance Overview (qperf) tool via cron.

4.4.5 Background Processes

Use the Background Processes window to view and manage current tasks that are running on the system (see Figure 4-63).

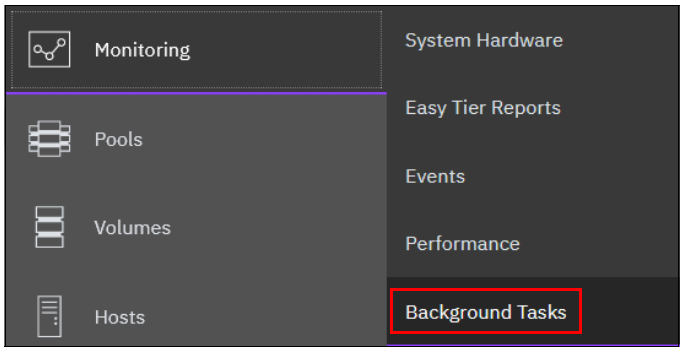


Figure 4-63 Selecting Background Processes

This menu provides an overview of currently running tasks that are triggered by the administrator. The overview provides more details than the indicator, as shown in Figure 4-64.

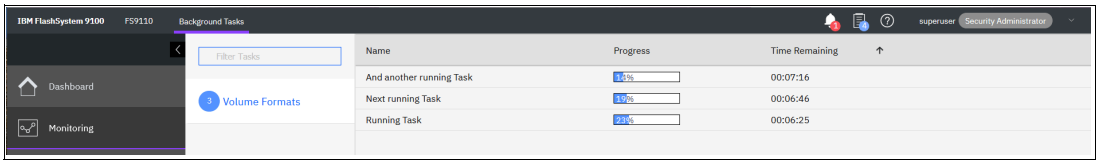


Figure 4-64 List of running tasks

4.5 Using the menus

Several menus are available in the GUI to help you configure and manage various components.

4.5.1 Pools

The Pools menu option is used to configure and manage storage pools, internal, and external storage, MDisks, and to migrate old attached storage to the system.

The Pools menu contains the following items accessible from the GUI (Figure 4-65 on page 266):

- Pools
- Volumes by Pool
- Internal Storage
- External Storage
- MDisks by Pool
- Import External Storage

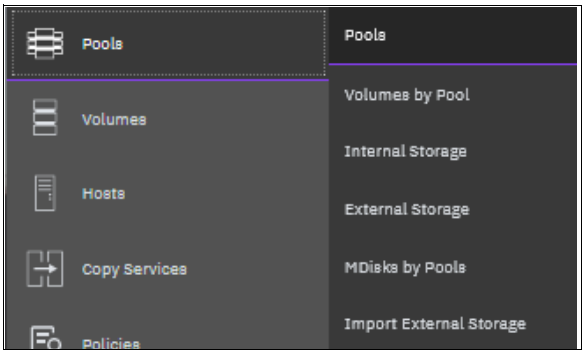


Figure 4-65 Pools menu

For more information about storage pool configuration and management, see Chapter 5, “Using storage pools” on page 365.

4.5.2 Volumes

A volume is a logical disk that the system presents to attached hosts. By using GUI operations, you can create different types of volumes depending on the type of topology that is configured on your system.

The Volumes menu contains the following options, as shown in Figure 4-66:

- Volumes
- Volumes by Pool
- Volumes by Host and Cluster
- Cloud Volumes
- Volume Groups

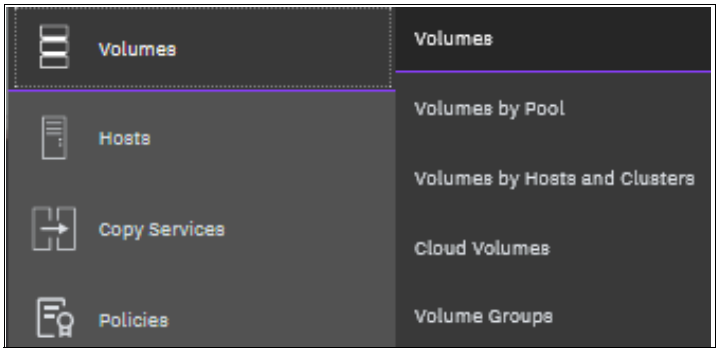


Figure 4-66 Volumes menu

For more information about these tasks and configuration and management process guidance, see Chapter 6, “Volumes” on page 419.

4.5.3 Hosts

A *host system* is a computer that is connected to the system through a Fibre Channel (FC) interface or an IP network. It is a logical object that represents a list of worldwide port names (WWPNs) that identify the interfaces that the host uses to communicate with your System. FC and SAS connections use WWPNs to identify the host interfaces to the systems.

The Hosts menu consists of the following choices, as shown in Figure 4-67:

- Hosts
- Host Clusters
- Mappings
- Volumes by Host and Clusters

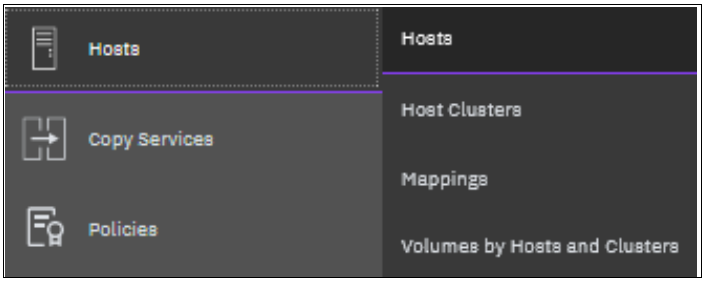


Figure 4-67 Hosts menu

For more information about configuration and management of hosts by using the GUI, see Chapter 7, “Hosts” on page 517.

4.5.4 Copy Services

The IBM Storage Virtualize Copy Services consist of the 2-Site Partnerships. Two-site partnerships replicate volume data that is on one system to a remote system. Two-site partnerships are required for policy-based replication. Partnerships can be used for migration, 3-site replication, and disaster recovery situations. Before you can configure policy-based replication or configure remote-copy objects, a partnership between all the systems must be established. If relationships or consistency groups exist between different systems, those systems must maintain their partnership. Each system can maintain up to three partnerships. The system can either support a maximum of three IP partnerships, three Fibre Channel partnerships, or a combination of these two types of partnerships. As many as four systems can be directly associated with each other.

The Copy Services menu offers the following operations in the GUI, as shown in Figure 4-68:

- Partnerships and Remote Copy



Figure 4-68 Copy Services in GUI

For more information about Copy Services (which is one of the most important features for resiliency solutions), see Chapter 10, “Advanced Copy Services” on page 717.

4.5.5 Access menu

The Access menu in the GUI is used to maintain who can log in to the system, define the access rights to the user, and track what was done by each privileged user to the system. It is divided into the following categories:

- Ownership groups
- Users by group
- Audit log

This section includes explanations of how to create, modify, or remove a user, and how to see records in the audit log.

The Access menu is available from the left window, as shown in Figure 4-69.

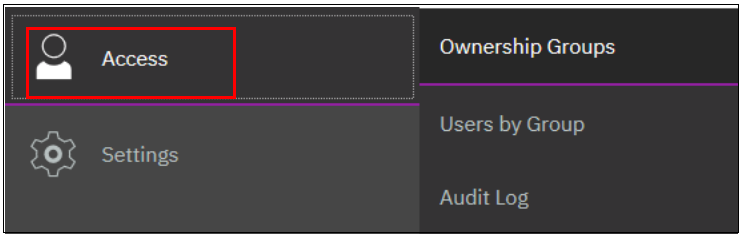


Figure 4-69 Access menu

4.6 Ownership groups

An *ownership group* defines a subset of users and objects within the system. You can create ownership groups to further restrict access to specific resources that are defined in the ownership group. Only users with Administrator or Security Administrator roles can configure and manage ownership groups.

Ownership groups restrict access to only those objects that are defined within that ownership group. An owned object can belong to one ownership group.

An *owner* is a user with an ownership group that can view and manipulate objects within that group.

The first time that you start the Ownership Group task, you see the window that is shown in Figure 4-70 on page 269.

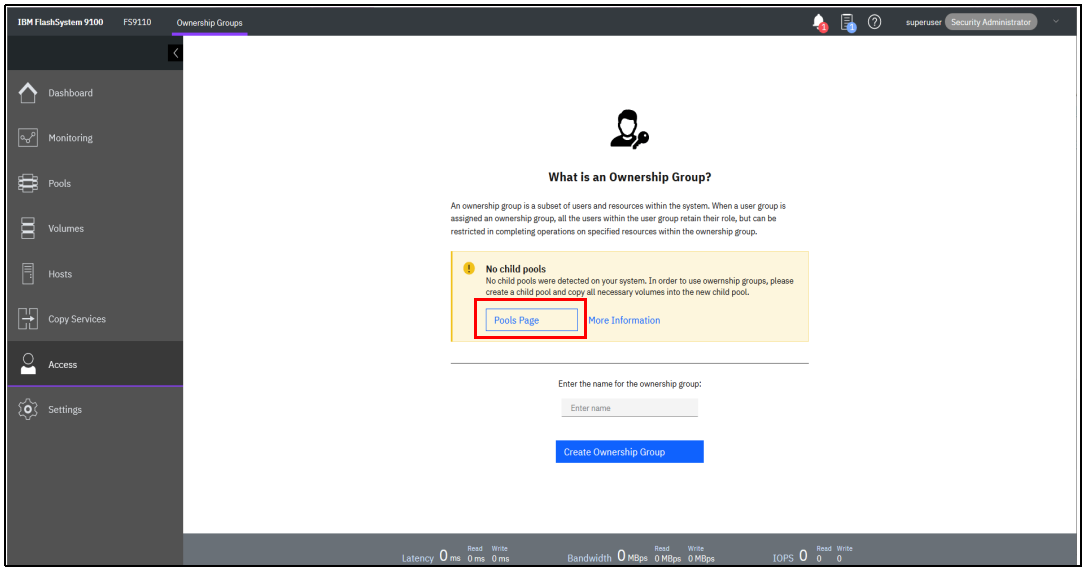


Figure 4-70 Start of an Ownership Group

In our example, no child pool exists; therefore, the GUI guides you to the Pools page to create child pools.

The system supports the following resources that you assign to ownership groups:

- Child pools
- Volumes
- Volume groups
- Hosts
- Host clusters
- Host mappings
- FlashCopy mappings
- FlashCopy consistency groups
- User groups
- Portsets

The following basic use cases can be applied to the use of ownership groups on the system:

- Objects are created within the ownership group. Other objects can be on the system that are not in the ownership group.
- On a system where these supported objects are configured, and you want to migrate these objects to use ownership groups.

When a user group is assigned to an ownership group, the users in that user group retain their role, but are restricted to only those resources within the same ownership group. User groups can define the access to operations on the system, and the ownership group can further limit access to individual resources.

For example, you can configure a user group with the Copy Operator role, which limits access of the user to Copy Services functions, such as FlashCopy and Remote Copy operations. Access to individual resources, such as a specific FlashCopy consistency group, can be further restricted by adding it to an ownership group.

When the user logs on to the management GUI, only resources that they can access through the ownership group are displayed. Also, only events and commands that are related to the ownership group in which a user belongs are viewable by those users.

4.6.1 Inheriting ownership

Depending on the type of resource, ownership can be defined explicitly or inherited from the user, user group, or other parent resources. Objects inherit their ownership group from other objects whenever possible, as shown in the following examples:

- Volumes inherit the ownership group from the child pool that provides capacity for the volumes.
- FlashCopy mappings inherit the ownership group from the volumes that are configured in the mapping.
- Hosts inherit the ownership group from the host cluster that they belong to, if applicable.
- Host mappings inherit the ownership group from the host and volume to which the host is mapped.

These objects cannot be moved to a different ownership group without creating inconsistent ownership.

Ownership groups are also inherited from the user. Objects that are created by an owner inherit the ownership group of the owner. If the owner is in more than one ownership group (only possible for remote users), the owner must choose the group when the object is created.

Figure 4-71 shows how different objects inherit ownership from ownership groups.

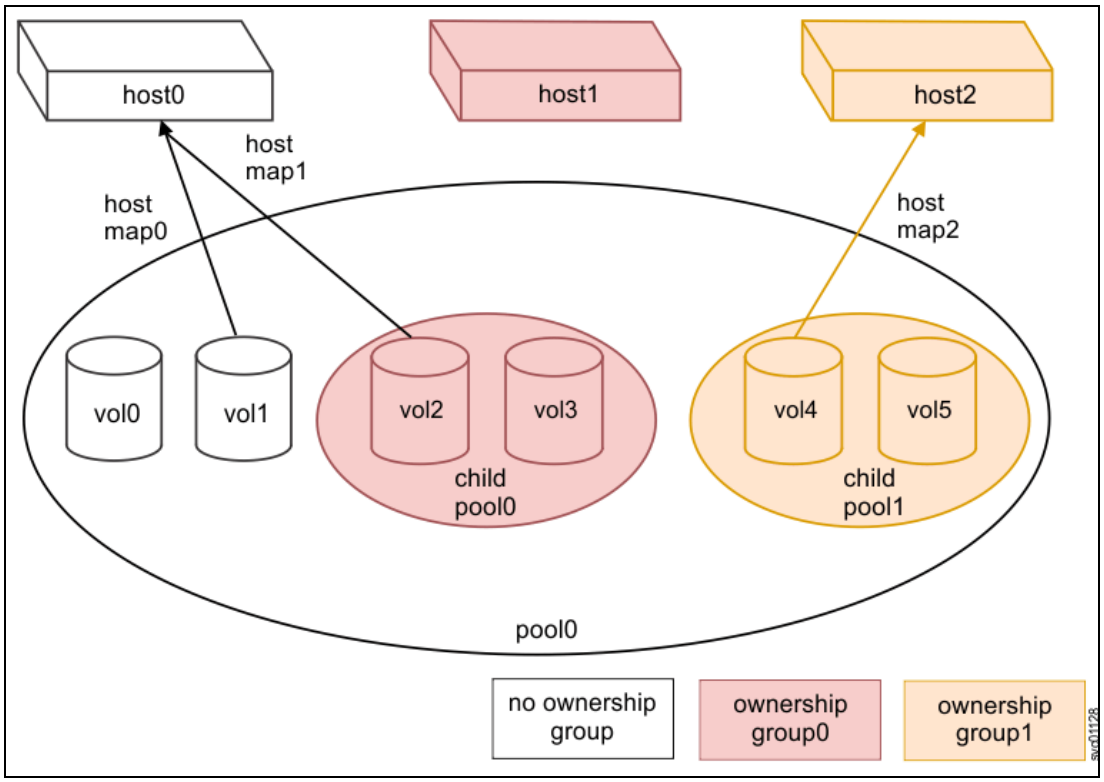


Figure 4-71 Ownership group inheritance

The following objects have ownership that is assigned explicitly and do not inherit ownership from other parent resources:

- Child pools
- Host clusters
- Hosts that are not part of a host cluster
- Volume groups
- FlashCopy consistency groups
- User groups
- Portsets
- Hosts that are a part of a host cluster
- Volumes
- Users
- Volume-to-host mappings
- FlashCopy mappings
- Configuring ownership groups
- Migrating to ownership groups

Child pools

The following rules apply to child pools that are defined in ownership groups:

- Child pools can be assigned to an ownership group when you create a pool or change a pool.
- Users who assign the child pool to the ownership group cannot be defined within that ownership group.
- Resources that are within the child pool inherit the ownership group that is assigned for the child pool.

Host clusters

The following rules apply to host clusters that are defined in ownership groups:

- If the user who is creating the host cluster is defined in only one ownership group, the host cluster inherits the ownership group of that user.
- If the user is defined in an ownership group but is also defined in multiple user groups, the host cluster inherits the ownership group. The system uses the lowest role that the user has from the user group. For example, if a user is defined in two user groups with the roles of Monitor and Copy Operator, the host cluster inherits the Monitor role.
- Only users that are not within an ownership group can assign ownership groups when a host cluster is created or changed.

Hosts that are not part of a host cluster

The following rules apply to a host that are not part of a host cluster that is defined in ownership groups:

- If the user who is creating the host is in only one ownership group, the host cluster inherits the ownership group of that user.
- If the user is defined in an ownership group but is also defined in multiple user groups, the host inherits the ownership group. The system uses the lowest role that the user has from the user group. For example, if a user is defined in two user groups with the roles of Monitor and Copy Operator, the host inherits the Monitor role.
- Only users not within an ownership group can assign ownership groups when you create a new host or change a host.

Volume groups

Volume groups can be created to manage multiple volumes that are used with Transparent Cloud Tiering (TCT) support. The following rules apply to volume groups that are defined in ownership groups:

- If the user that is creating the volume group is defined in only one ownership group, the volume group inherits the ownership group of that user.
- If the user is defined in an ownership group but is also defined in multiple user groups, the volume group inherits the ownership group. The system uses the lowest role that the user has from the user group. For example, if a user is defined in two user groups with the roles of Monitor and Copy Operator, the host inherits the Monitor role.
- Only users not within an ownership group can assign ownership groups when you create a volume group or change a volume group.
- Volumes can be added to a volume group if the volume and volume group are within the same ownership group or if both are not in an ownership group. Situations exist in which a volume group and its volumes can belong to different ownership groups. Volume ownership can be inherited from the ownership group or from one or more child pools.
- The ownership of a volume group does not affect the ownership of the volumes it contains. If a volume group and its volumes are owned by different ownership groups, the owner of the child pool that contains the volumes can change the volume directly.

  For example, the owner of the child pool can change the name of a volume within it. The owner of the volume group can change the volume group and indirectly change the volume, such as deleting a volume from the volume group. The ownership group of the child pools or the owner of the volume group cannot directly manipulate the resources that are not defined in their ownership group.

FlashCopy consistency groups

FlashCopy consistency groups can be created to manage multiple FlashCopy mappings. The following rules apply to FlashCopy consistency groups that are defined in ownership groups:

- If the user that is creating the FlashCopy consistency group is in only one ownership group, the FlashCopy consistency group inherits the ownership group of that user.
- If the user is defined in an ownership group but is also defined in multiple user groups, the FlashCopy consistency group inherits the ownership group. The system uses the lowest role that the user has from the user group.
- Only users that are not within an ownership group can assign ownership groups when a FlashCopy consistency is created or changed.
- FlashCopy mappings can be added to a consistency group if the volumes in the mapping and the consistency group are within the same ownership group. You can also add a FlashCopy mapping to a consistency group if it and all of its dependent resources are not in an ownership group.
- Situations exist in which a FlashCopy consistency group and its resources can belong to different ownership groups.
- As with volume groups and volumes, the ownership of the consistency group has no effect on the ownership of the mappings it contains.

User groups

The following rules apply to user groups that are defined in ownership groups:

- If the user that is creating the user group is in only one ownership group, the user group inherits the ownership group of that user.
- If the user is with multiple user groups, the user group inherits the ownership group of the user group with the lowest role.
- Only users not within an ownership group can assign an ownership group when a user group is created or changed.

These resources inherit ownership from the parent resource. Although the user cannot change the ownership group of the resource, they can change the ownership group of the parent object.

Portsets

Actions in portsets are performed by restricted and unrestricted users.

Restricted users

Restricted users can perform the following tasks:

- View and create IP addresses or assign hosts to portsets that are in their ownership group
- Assign hosts to portsets that are not assigned to an ownership group
- View all IP addresses in portsets that are owned by them and portsets not assigned to an ownership group

When a restricted user creates a portset, it is automatically assigned to the same ownership group as that restricted user.

Restricted users are unable to perform the following tasks:

- Modify portsets that are associated with other restricted users
- View portsets that were assigned to a different ownership group
- View IP addresses that are associated with a portset that is assigned to a different ownership group
- Create or delete IP addresses from a portset that is part of a different ownership group

Unrestricted users

Only unrestricted users can own remote copy and storage portsets.

Unrestricted users can perform the following tasks:

- View, create, and manage all portsets on the system
- View and create IP addresses to any portsets on the system
- Assign hosts in an ownership group to portsets that are not assigned to an ownership group
- Assign hosts to portsets that are assigned the same ownership group

Unrestricted users are unable to perform the following tasks:

- Assign hosts to portsets if the host and portsets are in different ownership groups
- Change the ownership of remote copy and storage portsets

Other objects inherit the ownership group from these assigned objects. Usually, the ownership group is defined when an object is created. Ownership can also be changed when

the defined object from which the ownership is established is assigned to a different ownership group.

Hosts that are a part of a host cluster

The following rules apply to hosts that are defined in ownership groups:

- The host inherits the ownership group of the host cluster to which it belongs.
- If a host is removed from a host cluster within an ownership group, the host inherits the ownership group of the host cluster to which it used to belong.
- If a host is removed from a host cluster that is not within an ownership group, the host inherits no ownership groups.
- Hosts can be added to a host cluster if the host and host cluster have the same ownership group.
- Changing the ownership group of a host cluster automatically changes the ownership group of all the hosts inside the host cluster.

Volumes

The following rules apply to volumes that are defined in ownership groups:

- The volume inherits the ownership group of the child pools that provide capacity for the volume and its copies.
- If the child pool that provides capacity for the volume or its copies is defined in different ownership groups, the volume cannot be created in an ownership group.
- When creating a volume copy or migrating a volume in the CLI, use the **`-inconsistentownershipgroup`** flag to allow for inconsistent ownership groups. However, do not leave volumes or volume copies in different ownership groups.

  After the migration, the user with the Security Administrator role must ensure that all volumes or copies are within the same ownership group as the users who need access.
- With volume groups, the volume group and its volumes can belong to different ownership groups. However, the ownership of a volume group does not affect the ownership of the volumes that it contains.

Users

The following rules apply to users that are defined in ownership groups:

- A user inherits the ownership group of the user group to which it belongs.
- Users that use Lightweight Directory Access Protocol (LDAP) for remote authentication can belong to multiple user groups and multiple ownership groups.

Volume-to-host mappings

The following rules apply to volume-to-host mappings that are defined in ownership groups:

- These mappings inherit the ownership group of the host or host cluster and the volume in the mapping.
- If host or host cluster and volume are within different ownership groups, the mapping cannot be assigned an ownership group.

FlashCopy mappings

The following rules apply to FlashCopy mappings that are defined in ownership groups:

- FlashCopy mappings inherit the ownership group of both volumes that are defined in the mapping.

- If the volumes are within different ownership groups, the mapping cannot be assigned to an ownership group.
- As with FlashCopy consistency groups, it is possible for a consistency group and its mappings to belong to different ownership groups. However, the ownership of the consistency group does not affect the ownership of the mappings that it contains.

Configuring ownership groups

You can configure ownership groups to manage access to resources on the system. You can create ownership groups to further restrict access to specific resources that are defined in the ownership group. Only users with Administrator or Security Administrator roles can configure and manage ownership groups.

Migrating to ownership groups

If you updated your system to a software level that supports ownership groups, you must reconfigure specific resources if you want to configure ownership groups.

To create an ownership group, select **Create Ownership Group** (see Figure 4-72).

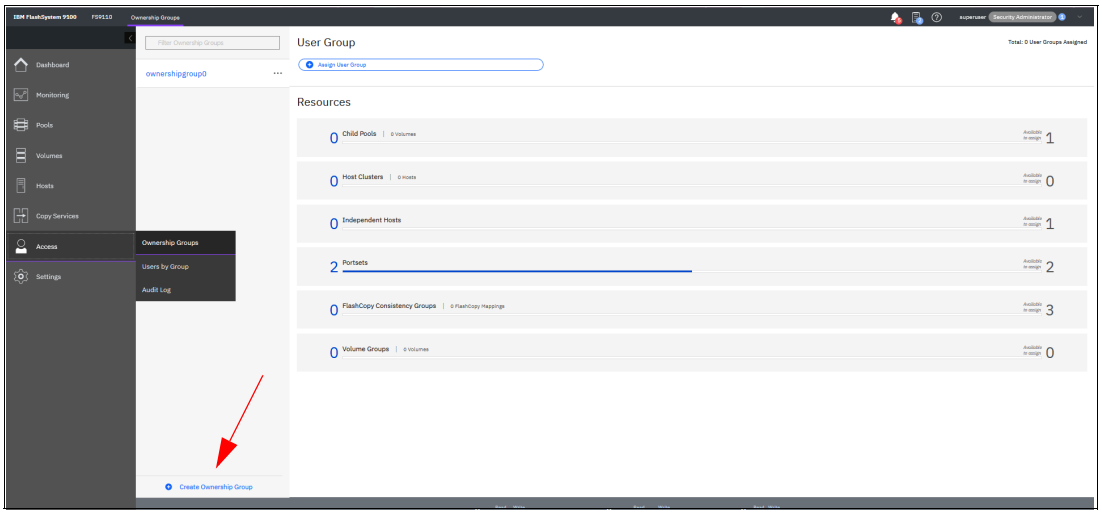


Figure 4-72 Ownership by Groups

4.6.2 Users by groups

You can create local users who can access the system. These user types are defined based on the administrative privileges that they have on the system.

Local users must provide a password, Secure Shell (SSH) key, or both. Local users are authenticated through the authentication methods that are configured on the system. If the local user needs access to the management GUI, a password is needed for the user. If the user requires access to the CLI through SSH, a password or a valid SSH key file is necessary.

Local users must be part of a user group that is defined on the system. User groups define roles that authorize the users within that group to a specific set of operations on the system.

Complete the following steps to define your user group in your system:

1. Select **Access** → **Users by Group**, as shown in Figure 4-73.

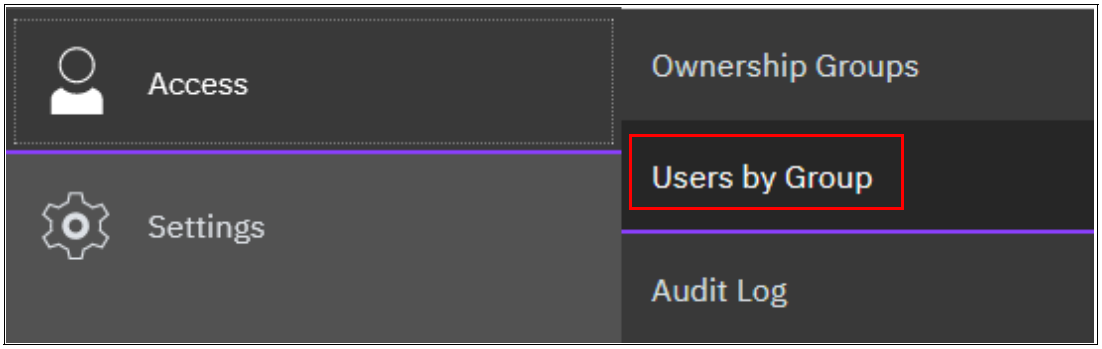


Figure 4-73 Accessing Users by Group

2. Select **Create User Group**, as shown in Figure 4-74.

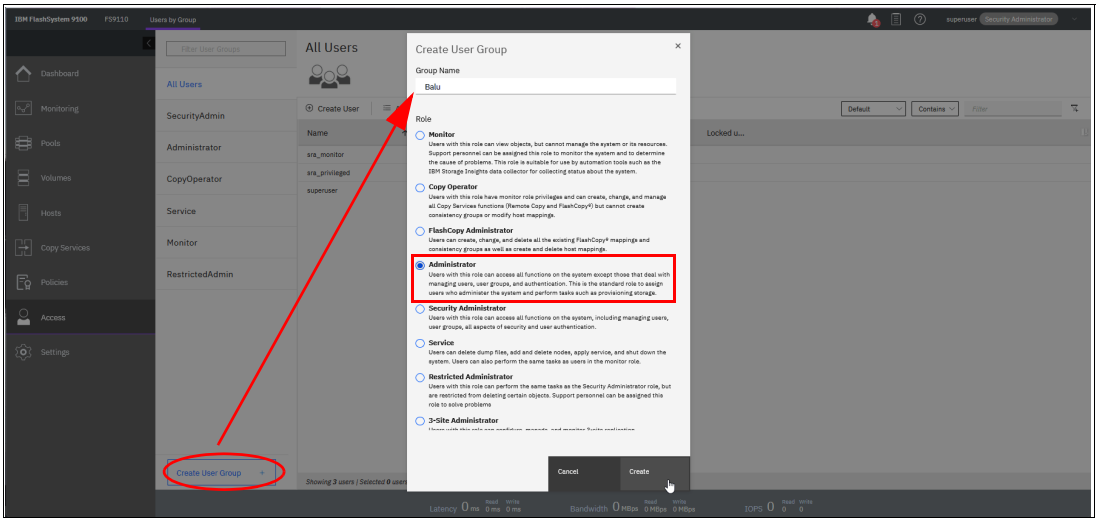


Figure 4-74 Defining a User Group

Figure 4-75 shows the newly created User Group.

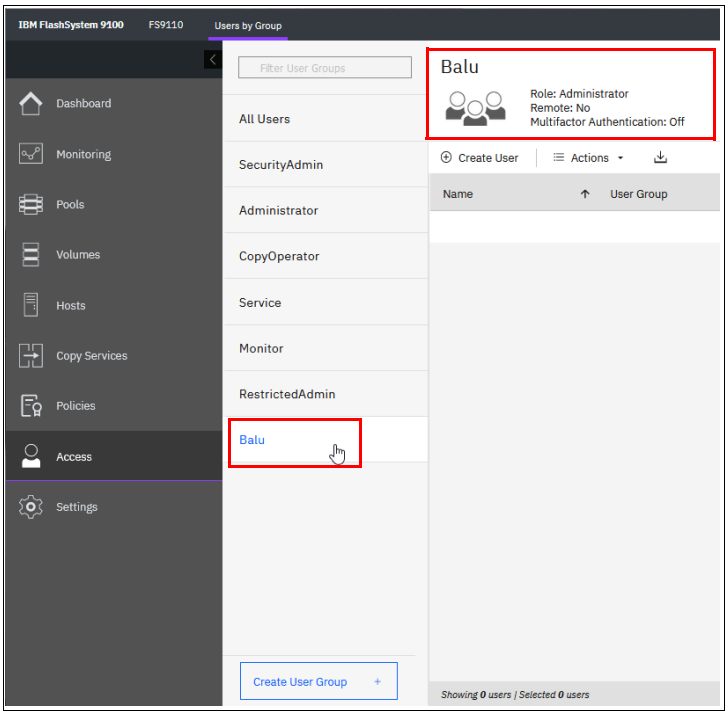


Figure 4-75 User Group

The following privileged user group roles are available in IBM Storage Virtualize:

- **Monitor**

  These users can access all system viewing actions. Monitor role users cannot change the state of the system or the resources that the system manages. Monitor role users can access all information-related GUI functions and commands, back up configuration data, and change their own passwords.

- **Copy Operator**

  These users can start and stop all FlashCopy, MM, and GM relationships. Copy Operator role users can run the system commands that Administrator role users can run that deal with FlashCopy, MM, and GM relationships.

- **FlashCopy Administrator**

  These users can create, change, and delete all the existing FlashCopy mappings and consistency groups and create and delete host mappings.

- **Administrator**

  These users can manage all functions of the system except for those functions that manage users, user groups, and authentication. Administrator role users can run the system commands that the Security Administrator role users can run from the CLI, except for commands that deal with users, user groups, and authentication.

- **Security Administrator**

  These users can manage all functions of the system, including managing users, user groups, user authentication, and configuring encryption. Security Administrator role users can run any system commands from the CLI. However, they cannot run the **`sainfo`** and **`satask`** commands from the CLI. Only the superuser ID can run those commands.

- **Service**

  These users can delete dump files, add and delete nodes, apply service, and shut down the system. Users can also perform the same tasks as users in the monitor role.

- **Restricted Administrator**

  These users can perform the same tasks and run most of the same commands as Administrator role users. However, users with the Restricted Administrator role are not authorized to run the **rmvdisk**, **rmvdiskhostmap**, **rmhost**, or **rmmdiskgrp** commands. Support personnel can be assigned this role to help resolve errors and fix problems.

- **3-Site Administrator**

  These users can configure, manage, and monitor 3-site replication configurations by using specific command operations that are available only on the 3-Site Orchestrator. Before you can work with 3-Site Orchestrator, a user profile must be created.

- **VASA Provider**

  A user with the vSphere application programming interfaces (APIs) for Storage Awareness (VASA) Provider role can manage virtual volumes or VMware vSphere virtual volumes (VVOLs) that are used by VMware vSphere and managed through IBM Storage Control software.

4.6.3 Users

Each user of the management GUI must provide a username and a password to sign on. Each user also has an associated role, such as monitor or security administrator. These roles are defined at the system level. For example, a user can be the administrator for one system, but the security administrator for another system.

The system supports local users and remote users who are authenticated to the system by using a remote authentication service. You can create local users who can access the system. These user types are defined based on the administrative privileges that they have on the system.

Local users

Local users must provide a password, a Secure Shell (SSH) key, or both. Local users are authenticated by using the authentication methods that are configured on the system. If the local user needs access to the management GUI, a password is needed for the user. If the user requires access to the command line interface (CLI) through SSH, a password or a valid SSH key file is necessary.

Local users must be part of a user group that is defined on the system. User groups define roles that authorize the users within that group to a specific set of operations on the system.

In addition to these first factor authentication methods for local users, you can enable the system to require multifactor authentication for all local users defined in a user group.

To use multifactor authentication on the system, a supported authentication service must be configured and multifactor authentication must be enabled on the system and user groups.

Currently, the system integrates with IBM Security Verify that provides second factor authentication. Local users must also be manually added to the supported authentication service and set up their second factors.

For more information, see configuring multifactor authentication.

Remote users

A remote user is authenticated on a remote LDAP server. A remote user does not need to be added to the list of users on the system, although they can be added to configure optional SSH keys.

For remote users, an equivalent user group must be created on the system with the same name and role as the group on the remote LDAP server. Remote users cannot access the system when the remote LDAP server is down. In that case, a local user account must be used until the LDAP service is restored. Remote users have their groups that are defined by the remote authentication server.

The system also supports requiring multifactor authentication for remote users. As with local users, you must configure a supported authentication service and enable the feature on the system and remote user groups.

However, remote users can be automatically managed by using IBM Security Verify Bridge for Directory Sync.

For remote users that authenticate with LDAP servers, install and configure IBM Security Verify Bridge for Directory Sync on your LDAP server, such as Windows Active Directory. IBM Security Verify Bridge for Directory Sync duplicates any users and groups that are defined on the source LDAP server into the Cloud Directory in IBM Security Verify. Any subsequent changes that are made to the source LDAP server are copied automatically to the IBM Security Verify directory.

In addition to multifactor authentication, the system supports single sign-on for remote users only. Remote users are authenticated to all applications through a single set of credentials. Single sign-on requires that the feature is enabled at the system level and on user groups.

For more information about configuring single sign-on for remote users, see Single Sign-on.

Registering a user

After you define your group, you can register a user within this group by clicking **Create User** (Figure 4-76).

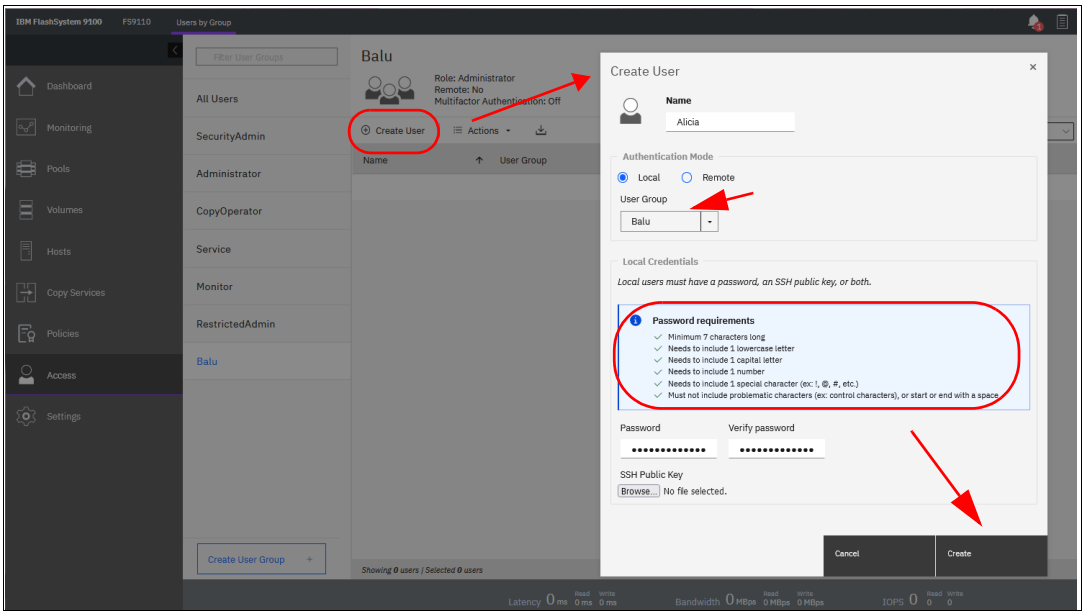


Figure 4-76 Registering a user account

Deleting a user

To remove a user account, right-click the user in the **All Users** list and select **Delete**, as shown in Figure 4-77.

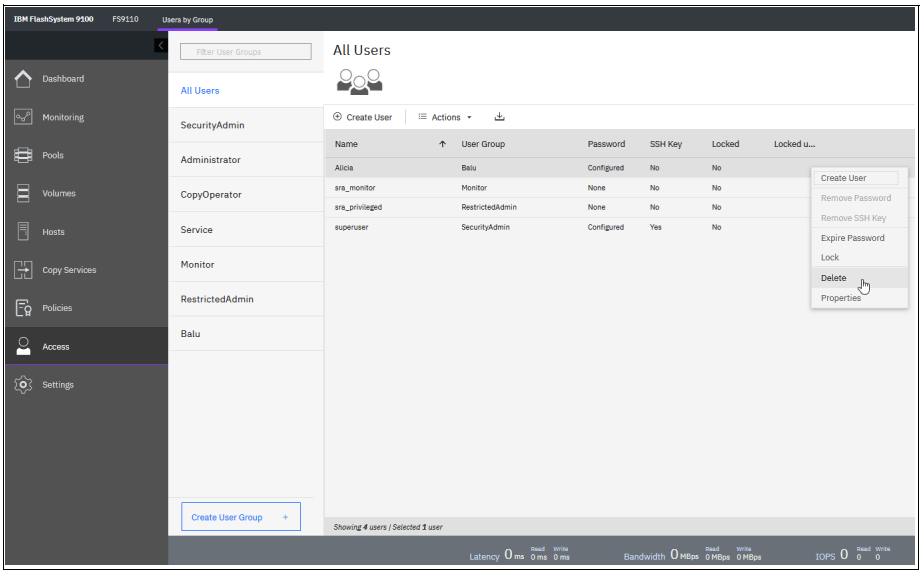


Figure 4-77 Deleting a user account

Attention: When you click **Delete**, the user account is deleted immediately. No other confirmation request is presented.

Password actions

You can use the GUI to set a password, lock and unlock a user, and expire a password.

Setting a password

To set a new password for the user, right-click the user (or click **Actions**) and select **Properties**. In this window, you can assign the user to a different group or reset their password (see Figure 4-78).

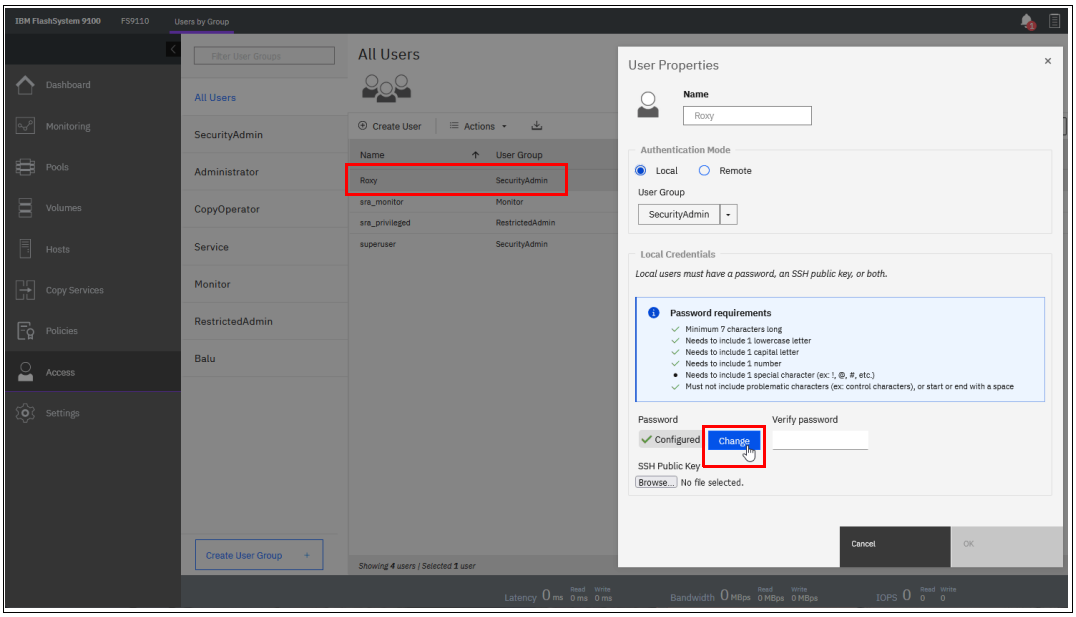


Figure 4-78 Setting a new password

New functions Lock User and Expire Password, which were available with IBM Storage Virtualize Version are described next.

Locking a password

To lock a user, right-click the username and then select **Lock**, as shown in Figure 4-79.

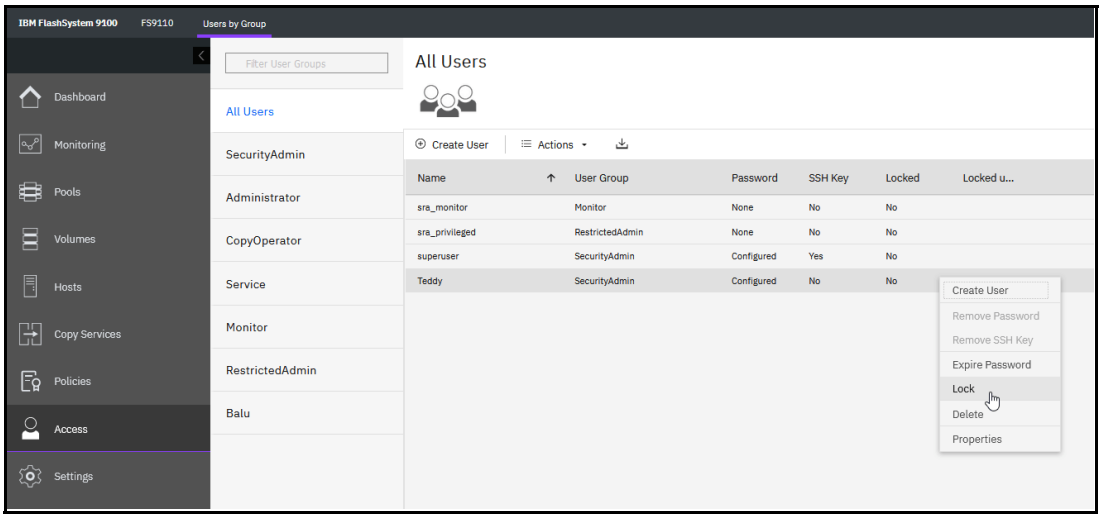


Figure 4-79 Locking a user

The verification window opens (Figure 4-80 on page 282).

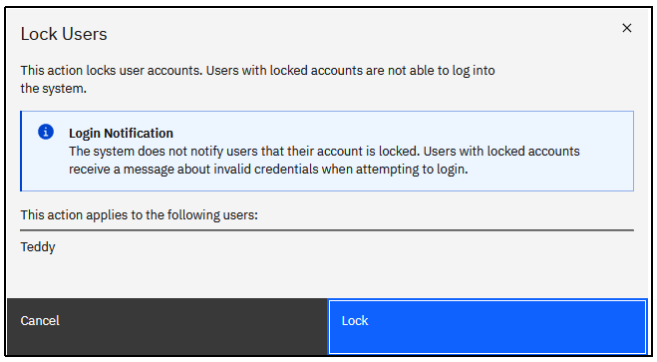


Figure 4-80 Lock verification

Note: Users do not receive a notification that their account was locked. Users with locked accounts receive a message about invalid credentials only when attempting to log in.

Figure 4-81 shows a locked user.

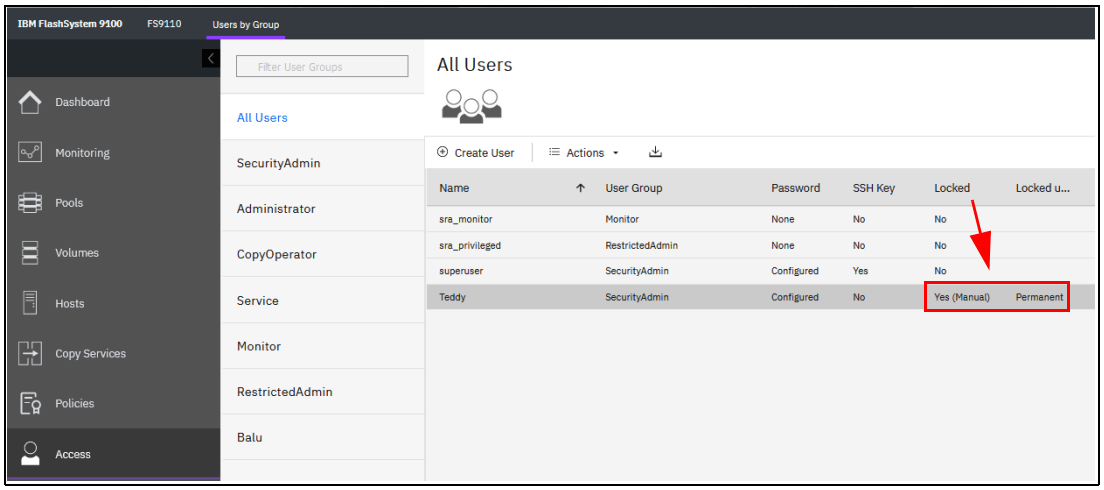


Figure 4-81 Locked user

To unlock the user, right-click the user and select **Unlock**, as shown in Figure 4-82.

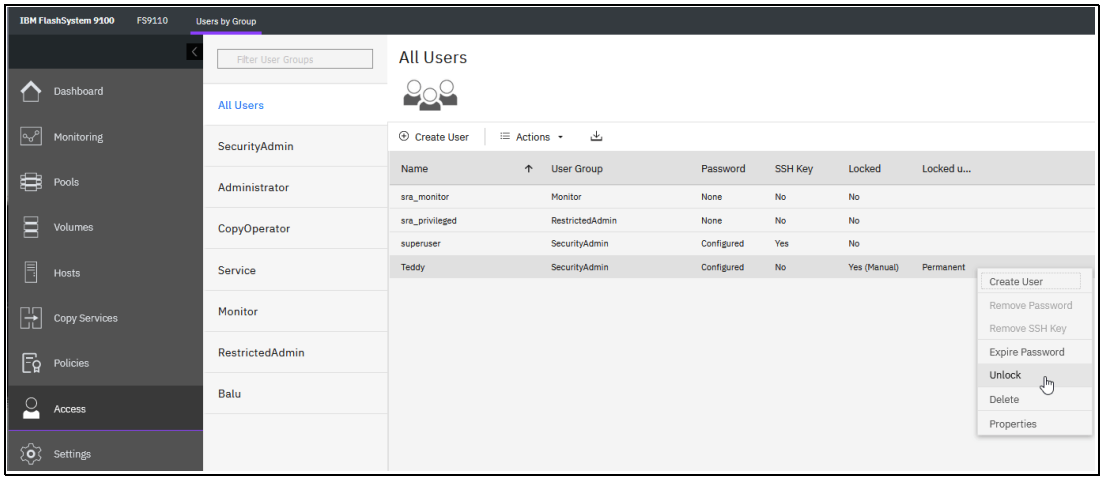


Figure 4-82 Unlock user

A confirmation window opens, as shown in Figure 4-83.

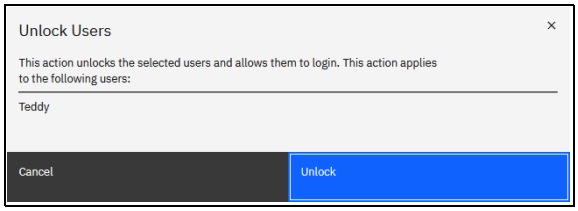


Figure 4-83 Unlock confirmation

Expiring a password

To force a password change for a dedicated user, you can use the Expire Password function (see Figure 4-84).

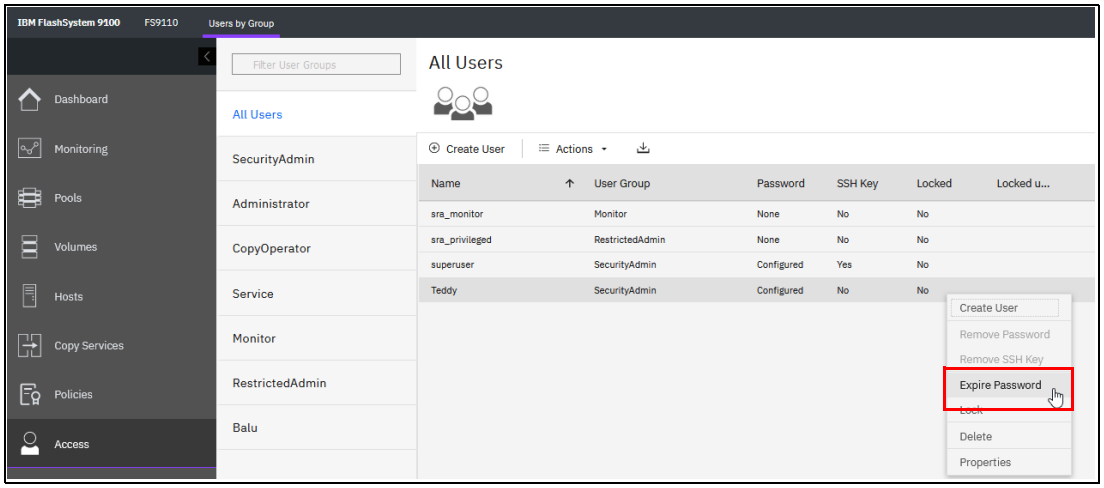


Figure 4-84 Expire Password

Select **Expire Password** and the verification windows opens, as shown in Figure 4-85.

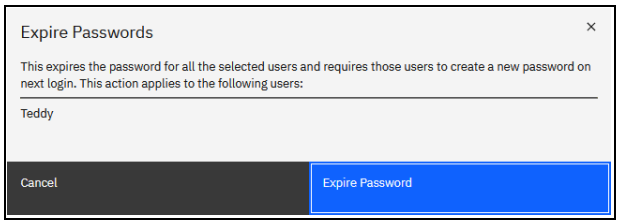


Figure 4-85 Expire Password verification

Note: This process expires the password for all the selected users and requires those users to create a password on the next login. This action applies to the selected users.

4.6.4 Audit log

An *audit log* documents action that is submitted through the management GUI or the CLI. You can use the audit log to monitor user activity on your system.

The audit log entries provide the following information:

- Time and date when the action or command was submitted.

  Note: When logs are displayed in the CLI, the timestamps for the logs in the CLI are the system time. However, when logs are displayed in the management GUI, the timestamps are converted to the local time where the web browser is running.

- Name of the user who completed the action or command.
- IP address of the system where the action or command was submitted.
- Name of source and target node on which the command was submitted.
- Parameters that were submitted with the command, excluding confidential information.
- Results of the command or action that completed successfully.
- Sequence number and the object identifier that is associated with the command or action.
- The origin of the command or action (possible values are: GUI, CLI, or REST).
- For service-related actions that are completed by support personnel, the audit log displays the Challenge information.

  This information indicates the challenge with which the support user who is authenticated to the system. Support users connect to the system through a challenge-response authentication. If authentication succeeds, a session is opened and Support users can conduct service-related actions on the system. This field identifies the support user at the support center. This field is blank for other users. New with Version 8.5 a Custom Range Filter is added, which allows you to specify a specific time range for searching.

An example of the audit log is shown in Figure 4-86 on page 285.

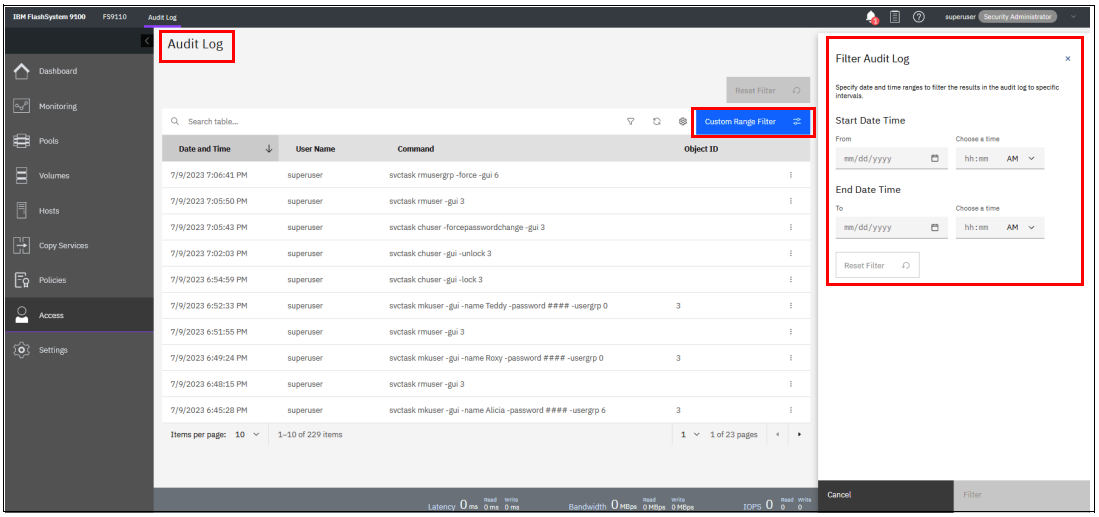


Figure 4-86 Audit log with Custom Range Filter

The following commands are not documented in the audit log:

- `dumpconfig`
- `cpdumps`
- `finderr`
- `dumperrlog`

The following items are also not documented in the audit log:

- Failed commands
- A result code of `0` (success) or `1` (success in progress)
- Result object ID of node type (for the **`addnode`** command)
- Views

Important: Failed commands are not recorded in the audit log. Commands that are triggered by IBM Support personnel are recorded with the flag **`Challenge`** because they use challenge-response authentication.

4.7 Settings

Use the Settings menu to configure system options for notifications, security, IP addresses, and preferences that are related to display options in the management GUI (see Figure 4-87).

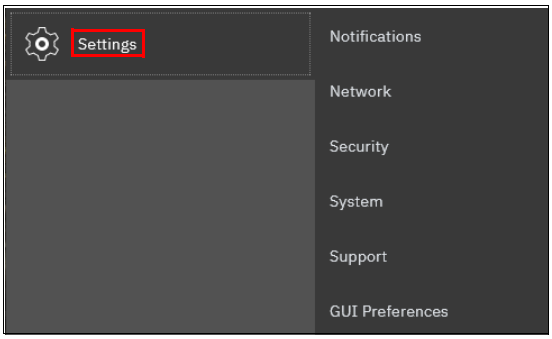


Figure 4-87 Settings menu

The following options are available for configuration from the Settings menu:

- Notifications: The system can use Simple Network Management Protocol (SNMP) traps, syslog messages, and Call Home emails to notify you and IBM Support Center when significant events are detected. Any combination of these notification methods can be used simultaneously.
- Network: Use the Network window to manage the management IP addresses for the system, service IP addresses for the nodes, and internet Small Computer Systems Interface (iSCSI) and FC configurations. The system must support FC or Fibre Channel over Ethernet (FCoE) connections to your storage area network (SAN).
- Security: Use the Security window to configure and manage remote authentication services.
- System: Use the System menu to manage overall system configuration options, such as licenses, updates, and date and time settings.
- Support: Use this option to configure and manage connections, and upload support packages to the support center.
- GUI Preferences: Configure the welcome message that appears after you log in, and refresh internals and GUI logout timeouts.

These options are described next.

4.7.1 Notifications

Your IBM FlashSystem storage system can use SNMP traps, syslog messages, and Call Home email to notify you and the IBM Support Center when significant events are detected. Any combination of these notification methods can be used simultaneously.

Notifications are normally sent immediately after an event is raised. However, events can occur because of service actions that are performed. If a recommended service action is active, notifications about these events are sent only if the events are still unfixed when the service action completes.

SNMP notifications

Simple Network Management Protocol (SNMP) is a standard protocol for managing networks and exchanging messages. The system can send SNMP messages that notify personnel about an event. You can use an SNMP manager to view the SNMP messages that the system sends. The system supports both SNMP version 2 and version 3.

To view the SNMP configuration, click the **Settings** icon and then select **Notifications** → **SNMP** (see Figure 4-88).

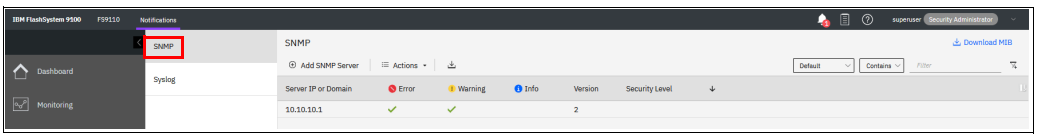


Figure 4-88 Setting the SNMP server and traps

In Figure 4-88, you can view and configure an SNMP server to receive various informational, error, or warning notifications by setting the following information:

- IP Address or Domain

  The address for the SNMP server.

- Community

  SNMP Community strings are used only by devices that support the SNMPv1 and SNMPv2c protocols. SNMPv3 uses username and password authentication and an encryption key. By convention, most SNMPv1 to v2c equipment ships from the factory with a read-only community string set to `public`.

- Server Port

  The remote port (RPORT) number for the SNMP server. The RPORT number must be a value of 1 - 65535.

- Event Notifications

  Consider the following points about event notifications:

  - Select **Error** if you want the user to receive messages about problems, such as hardware failures that must be resolved immediately.

    Important: Browse to **Recommended Actions** to run the fix procedures on these notifications.

  - Select **Warning** if you want the user to receive messages about problems and unexpected conditions. Investigate the cause immediately to determine any corrective action.

  - Select **Info** if you want the user to receive messages about expected events. No action is required for these events.

    Note: To remove an SNMP server, right-click the SNMP server and click **Remove**. To add an SNMP server, click the plus sign (**+**).

- Engine ID

  Enter a hexadecimal value that identifies an SNMP server instance. Each SNMPv3 server instance requires an engine ID that uniquely identifies the server to the administrative domain. Engine IDs are composed of a series of octets (bytes). They must be 5 - 32 octets. Each octet can contain a value of 0 - 255 or hexadecimal values from ‘00’H up to ‘FF’H. The engine ID translates to even number of hexadecimal 10 - 64 characters.

- Security Name

  The username must not exceed 31 characters.

- Authentication Protocol:

  - MD5 message digest algorithm in HMAC:
    - Directly provides data integrity checks
    - Indirectly provides data origin authentication
    - Uses private key that is known by sender and receiver
    - 16-byte key
    - 128-bit digest (truncates to 96 bits)
  - SHA, an optional alternative algorithm
  - Loosely synchronized monotonically increasing time indicator values defend against specific message stream modification attacks

- Privacy Protocol

  User-based Privacy Mechanism is based on:

  - Data Encryption Standard (DES) Cipher Block Chaining (CBC) mode:
    - Provides data confidentiality
    - Uses encryption
    - Subject to export and use restrictions in many jurisdictions
  - Uses 16-byte key (56-bit DES key, 8-byte DES initialization vector) that is known by sender and receiver
  - DES faces limitations for international compliance due to known security weaknesses.
  - Triple Data Encryption Standard (Triple DES)
  - Advanced Encryption Standard (128, 192, and 256, bit keys)

- Privacy Passphrase

  Enter a passphrase that is used to encrypt and decrypt messages between the system and the SNMP server. If you specify a privacy protocol, you must enter a corresponding passphrase.

Syslog notifications

The syslog protocol is a standard protocol for forwarding log messages from a sender to a receiver on an IP network. The IP network can be IPv4 or IPv6. The system can send syslog messages that notify personnel about an event.

You can use the Syslog window to view the syslog messages that are sent by your Storage System. To view the Syslog configuration, go to the System window and click **Settings**, and then select **Notifications** → **Syslog** (see Figure 4-89). A domain name server (DNS) server is required to use domain names in syslog.

To create another Syslog Server, select **Create Syslog Server**, as shown in Figure 4-89.

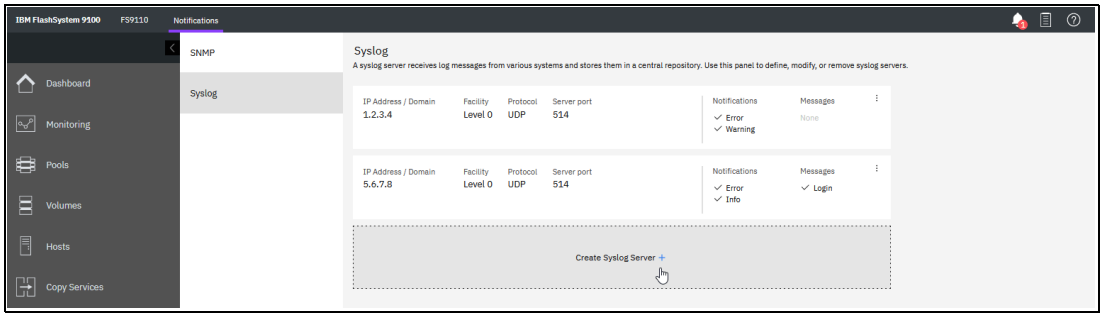


Figure 4-89 Creating Syslog Server

Figure 4-90 on page 289 shows that which input is needed to create a Syslog Server.

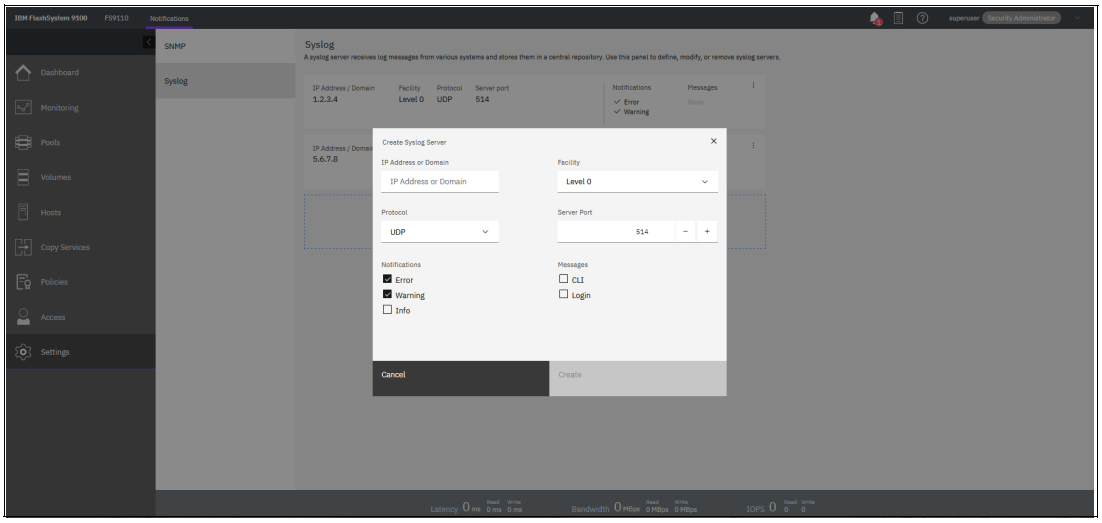


Figure 4-90 Setting the syslog messages

From this window, you can view and configure a syslog server to receive log messages from various systems and store them in a central repository by entering the following information:

- IP Address or Domain

  The IP address for the syslog server.

- Facility

  The facility determines the format for the syslog messages. The facility can be used to determine the source of the message.

- Protocol of the transmission protocol

  Select `UDP` or `TCP`.

- Server Port

  Port number of the syslog server.

- Event Notifications

  Consider the following points about event notifications:

  – Select **Error** if you want the user to receive messages about problems, such as hardware failures that must be resolved immediately.

  > **Important:** Browse to **Recommended Actions** to run the fix procedures on these notifications.

  – Select **Warning** if you want the user to receive messages about problems and unexpected conditions. Investigate the cause immediately to determine whether any corrective action is necessary.

  – Select **Info** if you want the user to receive messages about expected events. No action is required for these events.

- Message Format

  The message format depends on the facility. The system can transmit syslog messages in the following formats:

  – Concise message format provides standard details about the event
  – Expanded format provides more details about the event

To remove a syslog server, select **Delete** from the drop-down menu, as shown in Figure 4-91.

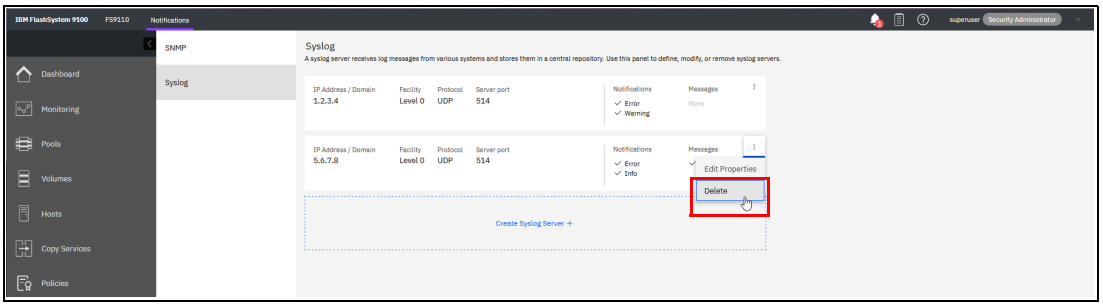


Figure 4-91 Delete Syslog Server

To create another Syslog Server, select **Create Syslog Server**, as shown in Figure 4-89 on page 288.

The syslog messages can be sent in concise message format or expanded message format.

Example 4-1 shows a compact format syslog message.

Example 4-1 Compact syslog message example

```
IBM2145 #NotificationType=Error #ErrorID=077001 #ErrorCode=1070 #Description=Node
CPU fan failed #ClusterName=FS9110 #Timestamp=Wed Jun16 08:00:59 2023 BST
#ObjectType=Node #ObjectName=Node1 #CopyID=0 #ErrorSequenceNumber=100
```

Example 4-2 shows an expanded format syslog message.

Example 4-2 Full format syslog message example

```
IBM2145 #NotificationType=Error #ErrorID=077001 #ErrorCode=1070 #Description=Node
CPU fan failed #ClusterName=FS9110 #Timestamp=Wed Jun16 08:00:59 2023 BST
#ObjectType=Node #ObjectName=Node1 #CopyID=0 #ErrorSequenceNumber=100 #ObjectID=2
#NodeID=2 #MachineType=9846AF7#SerialNumber=1234567 #SoftwareVersion=8.6.0.1
(build 169.9.2306081121000)#FRU=fan 24P1118, system board 24P1234
#AdditionalData(0->63)=0000000021000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000#Additional
Data(64-127)=0000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000
```

4.7.2 Network

This section describes how to view the network properties of the storage system.

The network information can be obtained by clicking **Network**, as shown in Figure 4-92 on page 291.

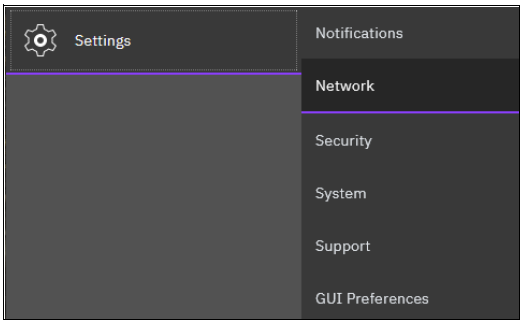


Figure 4-92 Accessing network information

Configuring the network

The procedure to set up and configure an IBM FlashSystem storage system network interface is described in Chapter 3, “Initial configuration” on page 177.

Management IP addresses

To view the management IP addresses of IBM Storage Virtualize, select **Settings** → **Network.** Click **Management IP Addresses**. The GUI shows the management IP address by pointing to the network ports, as shown in Figure 4-93.

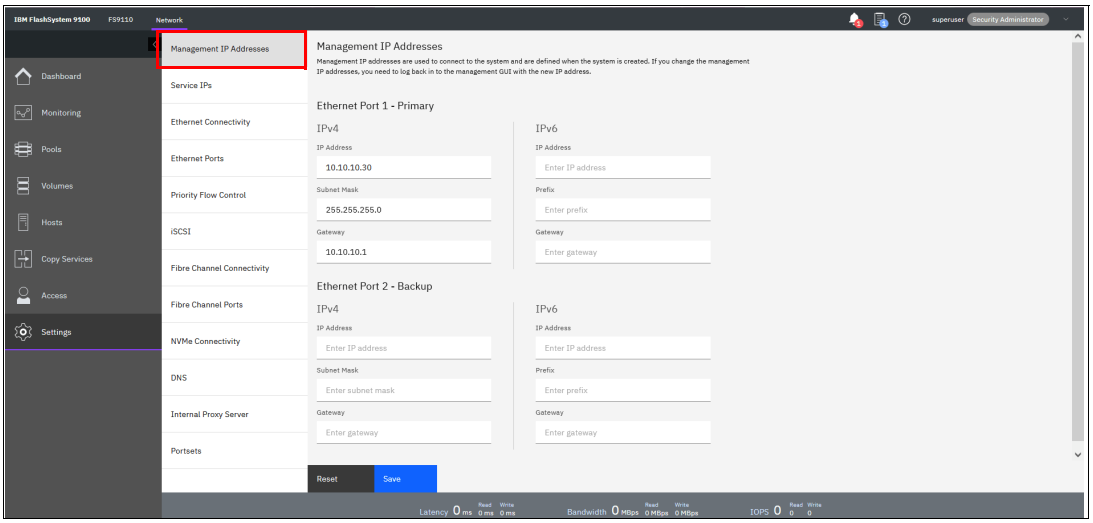


Figure 4-93 Viewing the management IP addresses

Service IP information

To view the Service IP information of your IBM Storage Virtualize installation, select **Settings** → **Network**, as shown in Figure 4-92. Click the **Service IPs** option to view the properties, as shown in Figure 4-94 on page 292. You can choose the upper or lower node canister.

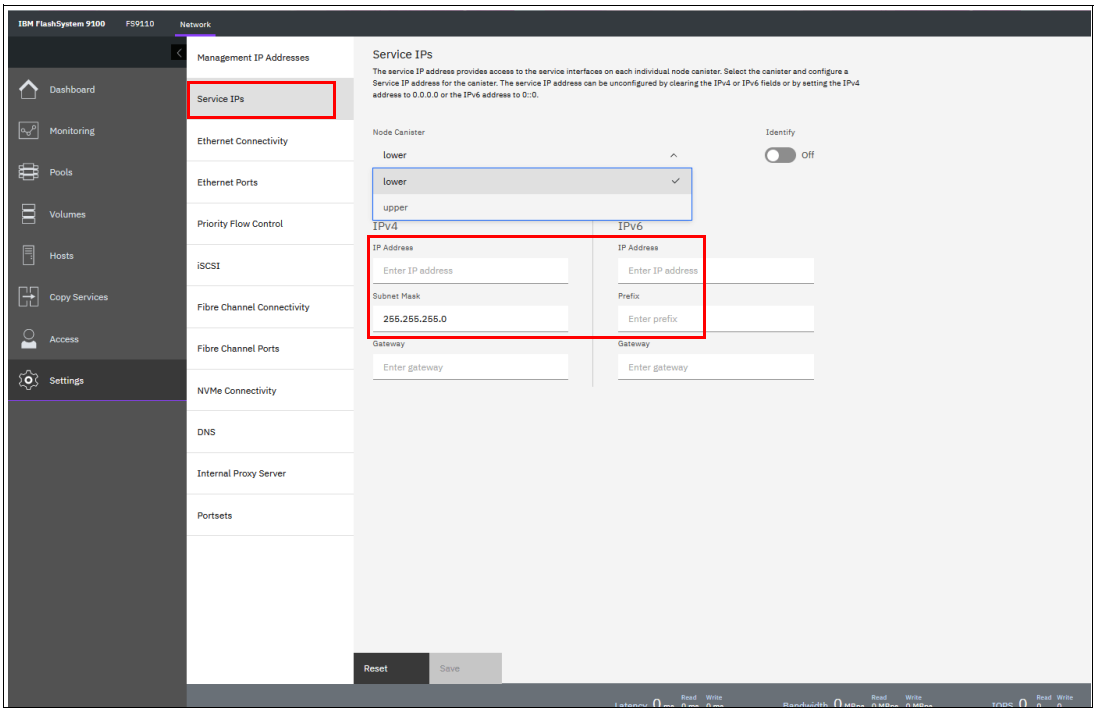


Figure 4-94 Viewing service IP addresses

The service IP address is commonly used to provide access to the network interfaces on the individual nodes of the control enclosure.

Instead of reaching the management IP address, the service IP address directly connects to each node canister for service operations. You can select a node canister of the control enclosure from the drop-down list and then click any of the ports that are shown in the GUI. The service IP address can be configured to support IPv4 or IPv6.

Ethernet Connectivity

Use the Ethernet Connectivity page to view node-to-node connections that use Ethernet protocols that support remote direct memory access (RDMA) technology, such as RDMA over Converged Ethernet (RoCE) or internet Wide Area RDMA Protocol (iWARP). To use these protocols, the system requires that an RDMA-capable adapter is installed on each node, and dedicated RDMA-capable Ethernet ports are configured only for node-to-node communication (see Figure 4-95).

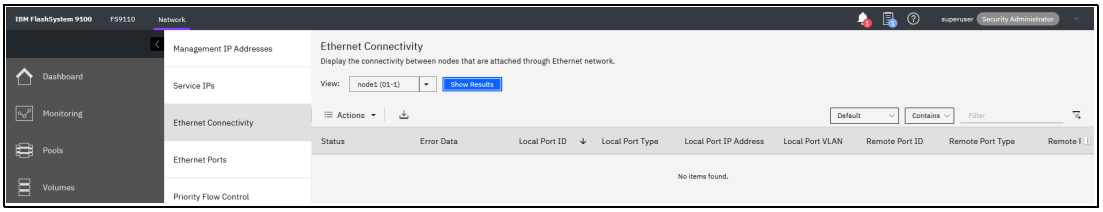


Figure 4-95 Ethernet Connectivity

RDMA technologies, such as RoCE and iWARP, enable the RDMA-capable adapter to transfer data directly between nodes and bypass CPU and caches, which make transfers faster. RDMA technologies provide faster connection and processing time than traditional iSCSI connections.

Select the node from the list to view status, error data, and other details about connections for that node. The status field displays information that is related to the connection for the selected node. The following statuses are possible:

- `Connected`

  Indicates that Ethernet ports are configured and a connection is established.

- `Discovered`

  Indicates that an Ethernet port on the selected node is configured, but a connection cannot be established. This status indicates a potential problem that must be resolved. The Error Data column indicates the reason for the Discovered status.

  The following values are possible in the Error Data column:

  - `Protocol mismatch`

    Indicates that the protocol on the source and destination adapters is not the same. This error occurs when one node in the system does not have the 25 Gbps Ethernet adapter installed.

  - `Unreachable`

    Indicates that the local and remote IP addresses cannot be reached. This error can occur if one of the nodes in the system is offline. Select **Monitoring** → **Events** to view errors and run any necessary fix procedures to return the node to an online status.

  - `Duplicate IP addresses`

    Indicates that one or more IP addresses are used in the network. Each node IP address must be unique. To fix this error, you can use the Service Assistant interface to change the node IP address.

  - `Degraded`

    Indicates that the negotiated speed on the local and remote adapters is not the same. Degraded status occurs when one or both adapters are configured at a lower speed rather than the maximum speed that the adapters support. To fix this issue, ensure that adapters on both nodes are configured at the maximum speed.

  - `VLAN ID Mismatch`

    Indicates that the local and remote port virtual LAN identifiers are not the same. To fix this error, ensure that the local and remote nodes belong to the same VLAN. If you use VLAN in your network, you must configure VLAN on switches by setting VLAN to `Trunk` mode and specifying the VLAN ID on the switch before you configure IP addresses and other settings for the RDMA-capable Ethernet ports on the nodes in the system.

    For more information, see Configuring VLAN for hosts.

Ethernet ports

Ethernet ports for each node are at the rear of the system and used to connect the system to hosts, external storage systems, and to other systems that are part of RC partnerships. Depending on the model of your system, supported connection types include FC, when the ports are FCoE-capable, iSCSI, and iSCSI Extensions for Remote Direct Memory Access (RDMA) (iSER). iSER connections use the RDMA over Converged Ethernet (RoCE) protocol or the internet-Wide Area RDMA Protocol (iWARP). The window indicates whether a specific port is being used for a specific purpose and traffic.

You can modify how the port is used by selecting **Actions**. Then select one of the following options to change the use of the port:

- Manage IP Addresses
- Modify Remote Copy

- Modify iSCSI Hosts
- Modify Storage Ports
- Modify Maximum Transmission Unit

You can also display the login information for each host that is logged in to a selected node.

To display this information, select **Settings** → **Network** → **Ethernet Ports** and then right-click the node and select **IP Login Information**. This information can be used to detect connectivity issues between the system and hosts and improve the configuration of iSCSI host to optimize performance.

Select **Ethernet Ports** for an overview from the menu, as shown in Figure 4-96. For more information about planning, see Chapter 2, “Installation and configuration planning” on page 119.

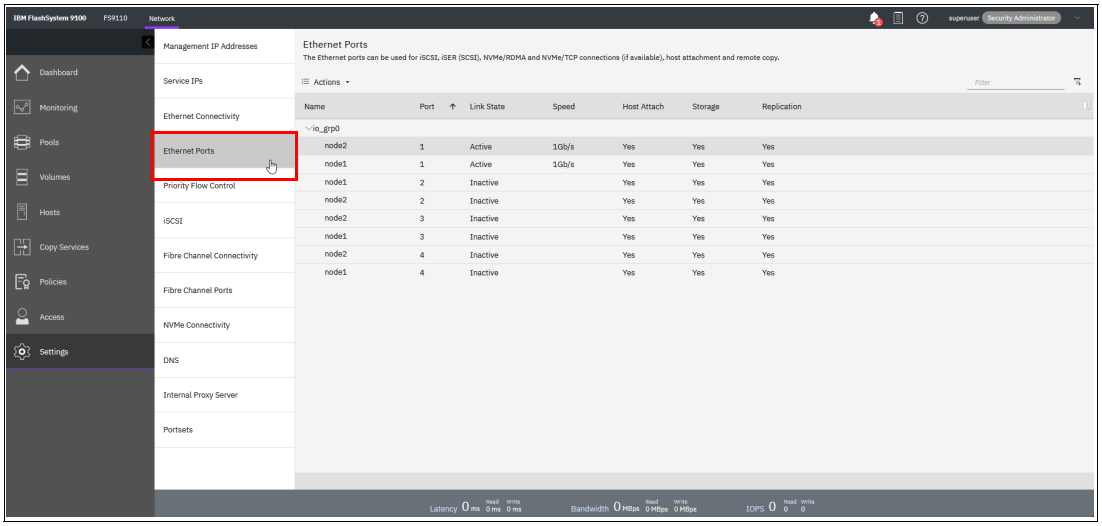


Figure 4-96 Ethernet Ports

The Ethernet Port page displays whether a specific port has PFC configured for a specific traffic type. For each supported connection type, PFC-related settings are displayed in the table. If On is displayed, it means that PFC is fully configured in the network.

A full configuration for PFC requires that PFC settings are defined on the Ethernet switch and the system. In addition, each Ethernet port must also be configured for the traffic type and VLAN settings for each port must be configured for a full configuration of PFC capabilities.

For more information about configuring PFC settings on the switch and system, see Configuring priority flow control.

Priority flow control

Priority flow control (PFC) is an Ethernet protocol that you can use to select the priority of different types of traffic within the network. With PFC, administrators can reduce network congestion by slowing or pausing specific classes of traffic on ports, which provides better bandwidth for more important traffic.

The system supports PFC on various supported Ethernet-based protocols on three types of traffic classes: system, host attachment, and storage traffic. You can configure a priority tag for each of these traffic classes.

The priority tag can be any value 0 - 7. You can set identical or different priority tag values to all these traffic classes. You can also set bandwidth limits to ensure quality of service (QoS) for these traffic classes by using the Enhanced Transmission Selection (ETS) setting in the network. When you plan to configure PFC, follow these guidelines and examples.

To use PFC and ETS, ensure that the following tasks are completed:

- Ensure that ports support 10 Gb or higher bandwidth to use PFC settings.
- Configure a virtual local area network (VLAN) on the system to use PFC capabilities for the configured IP version.
- Ensure that the same VLAN settings are configured on the all entities, including all switches between the communicating end points.
- Configure the QoS values (priority tag values) for host attachment, storage, or system traffic by running the **chsystemethernet** command.
- To enable priority flow for host attachment traffic on a port, make sure that the host flag is set to `yes` on the configured IP on that port.
- To enable priority flow for storage traffic on a port, make sure that storage flag is set to `yes` on the configured IP on that port.
- On the switch, enable the Data Center Bridging Exchange (DCBx). DCBx enables switch and adapter ports to exchange parameters that describe traffic classes and PFC capabilities. For these steps, check your switch documentation for more information.
- For each supported traffic class, configure the same priority tag on the switch. For example, if you plan to have a priority tag setting of 3 for storage traffic, ensure that the priority also is set to 3 on the switch for that traffic type.
- If you are planning to use the same port for different types of traffic, ensure that the ETS settings are configured in the network.

Note: PFC is not supported when you use NVMe over TCP connections.

4.7.3 Using the management GUI

To set PFC on the system, complete the following steps:

1. In the management GUI, select **Settings** → **Network** → **Priority Flow Control**, as shown in Figure 4-97.

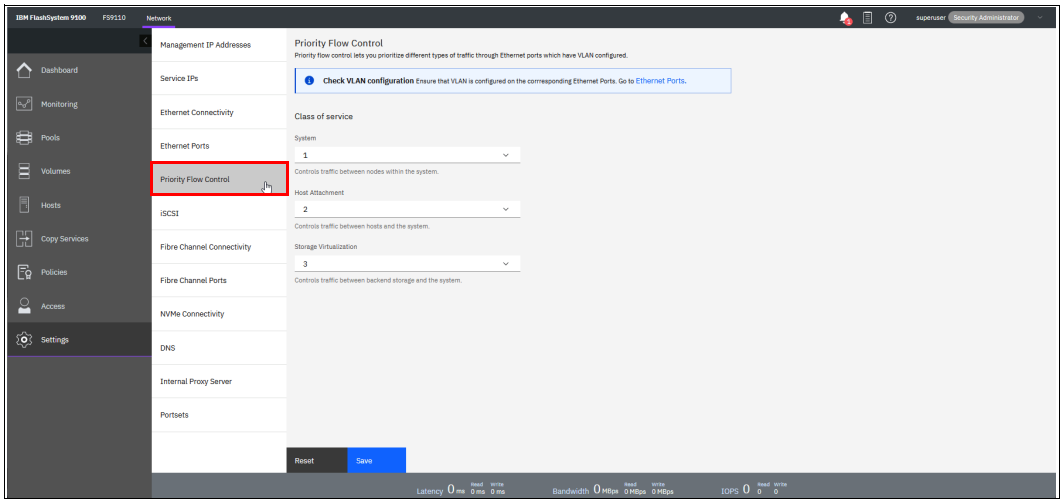


Figure 4-97 Priority flow control

2. For each of the following classes of service, select the priority setting for that traffic type:
   - System

     Set a value 0 - 7 for the system traffic, which includes communication between nodes within the system. The system priority tag is supported on iSCSI connections and systems that support RDMA over Ethernet connections between nodes. Ensure that you set the same priority tag on the switch to use PFC capabilities.
   - Host attachment

     Set the priority tag 0 - 7 for system to host traffic. The host attachment priority tag is supported on iSCSI connections and systems that support RDMA over Ethernet connections. Ensure that you set the same priority tag on the switch to use PFC capabilities.
   - Storage virtualization

     Set the priority tag 0 - 7 for system to external storage traffic. The storage virtualization priority tag is supported on storage traffic over iSCSI connections. Ensure that you set the same priority tag on the switch to use PFC capabilities.

     Finally, make sure that IP is configured with VLAN.

iSCSI information

From the iSCSI window in the **Settings** menu, you can display and configure parameters for the system to connect to iSCSI-attached hosts, as shown in Figure 4-98.

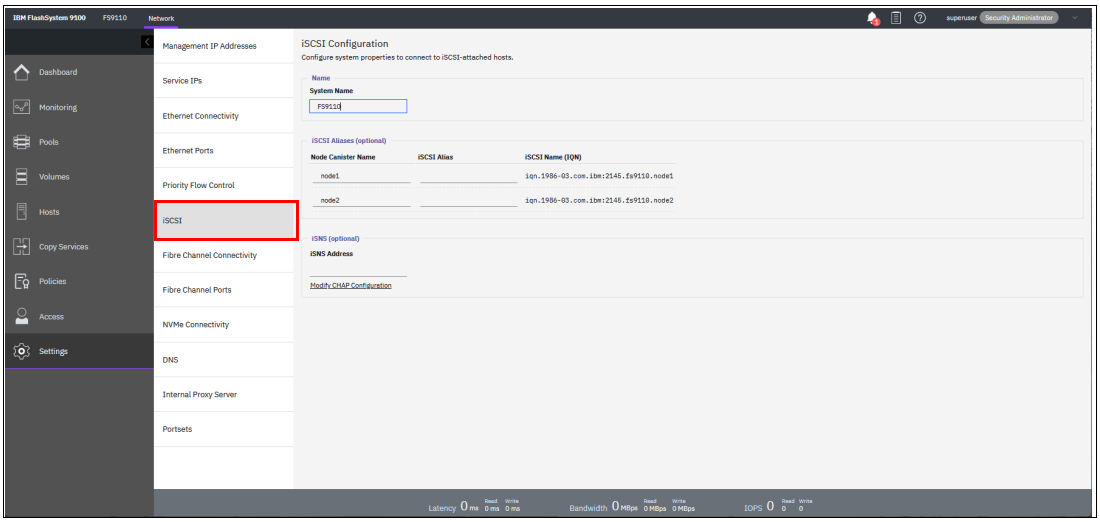


Figure 4-98 iSCSI Configuration window

The following parameters can be updated:

- System Name

  It is important to set the system name correctly because it is part of the iSCSI Qualified Name (IQN) for the node.

  > **Important:** If you change the name of the system after iSCSI is configured, you might need to reconfigure the iSCSI hosts.

  To change the system name, click the system name and specify the new name.

  > **System name**: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The name can be 1 - 63 characters.

- iSCSI aliases (optional)

  An *iSCSI alias* is a user-defined name that identifies the node to the host. To change an iSCSI alias, click the iSCSI alias and specify a name for it.

  Each node has a unique iSCSI name that is associated with two IP addresses. After the host starts the iSCSI connection to a target node, this IQN from the target node is visible in the iSCSI configuration tool on the host.

- Internet Storage Name Service (iSNS) and Challenge Handshake Authentication Protocol (CHAP)

  You can specify the IP address for the iSNS. Host systems use the iSNS server to manage iSCSI targets and for iSCSI discovery.

  You can also enable CHAP to authenticate the system and iSCSI-attached hosts with the specified shared secret.

  The CHAP secret is the authentication method that is used to restrict access for other iSCSI hosts that use the same connection. You can set the CHAP for the entire system under the system properties or for each host definition. The CHAP must be identical on

the server and the system and host definition. You can create an iSCSI host definition without the use of CHAP.

Fibre Channel information

As shown in Figure 4-99, you can use the Fibre Channel Connectivity window to display the Fibre Channel connection between nodes and other storage systems and hosts that attach through the FC network. You can filter by selecting one of the following fields:

- All nodes, storage systems, and hosts
- Systems
- Nodes
- Storage systems
- Hosts

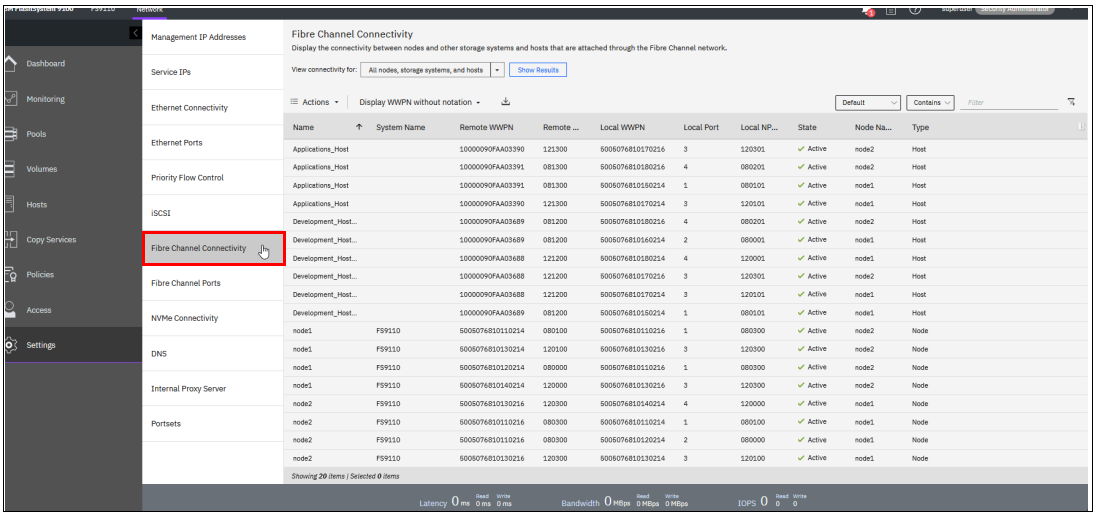


Figure 4-99 Fibre Channel Connectivity

In the Fibre Channel Ports window, you can use this view to display how the FC port is configured across all control node canisters in the system. This view helps, for example, to determine which other clusters and hosts the port can communicate with, and which ports are virtualized. `No` indicates that this port cannot be online on any node other than the owning node (see Figure 4-100).

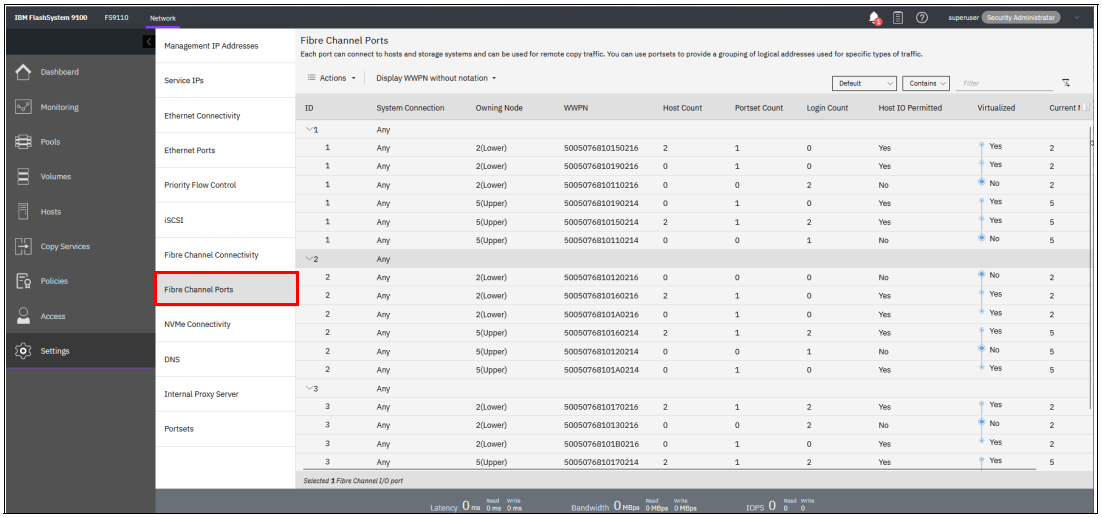


Figure 4-100 Viewing Fibre Channel Port properties

You can use the Fibre Channel ports window in addition to SAN fabric zoning to restrict node-to-node communication. You can specify specific ports to prevent communication between nodes in the local system or between nodes in a remote-copy partnership. This port specification is called *Fibre Channel port masking*.

The Fibre Channel port masking is used to filter out the data that is not intended on a specific Fibre Channel port. In this way, you can choose the type of the traffic (service) that is required on each port. The following services normally use Fibre Channel for data communication in the system:

- Hosts
- External storage systems
- Nodes in the local system (node-to-node communications)
- Nodes in partnered systems (remote node communications or remote-copy operations)

Host traffic is configured by using storage area network (SAN) zone configuration and the port mask parameter on **mkhost** or **chhost** commands. External storage system communication is also managed by the SAN zoning. However, to zone traffic between nodes, you must create separate fabrics for each port per node, which can lead to large and complex fabric zones.

To reduce the number of zones that are required, use the Fibre Channel panel to prevent a specific type of node-to-node traffic. SAN zoning is still required to reduce latency and prevent congestion on the SAN. Use the following configuration guidelines when you configure ports for Fibre Channel traffic:

- For redundancy, each node must have at least two logins to every other node in the same system.
- A node can have up to 16 active logins to any other node.

Note: Hosts can be zoned for up to eight paths per volume.

Non-Volatile Memory Express (NVMe) connectivity

A Non-Volatile Memory Express (NVMe) over FC host can be attached to the system. For more information about Non-Volatile Memory Express over Fibre Channel (FC-NVMe), such as interoperability requirements, see this IBM SAN Volume Controller Support web page or this IBM FlashSystem Support web page.

If your system supports an FC-NVMe connection between nodes and hosts, you can display more information about each side of the connection. To display node information, select the node from the drop-down menu and then select **Show Results**.

You can also display the host details for the connection or for all hosts and nodes. Use this window to troubleshoot issues between nodes and hosts that use FC-NVMe connections. For these connections, the Status column displays the current state of the connection.

The following states for the connection are possible:

- Active

  Indicates that the connection between the node and host is being used.

- Inactive

  Indicates that the connection between the node and host is configured, but no FC-NVMe operations occurred in the last 5 minutes.

  Because the system sends a periodic heartbeat message to keep the connection open between the node and the host, it is unusual to see an inactive state for the connection. However, it can take up to 5 minutes for the state to change from inactive to active.

If the inactive state remains beyond the 5-minute refresh interval, it can indicate a connection problem between the host and the node. Verify these values in the management GUI and view the messages by selecting **Monitoring** → **Events**.

Figure 4-101 shows the NVMe Connectivity menu.

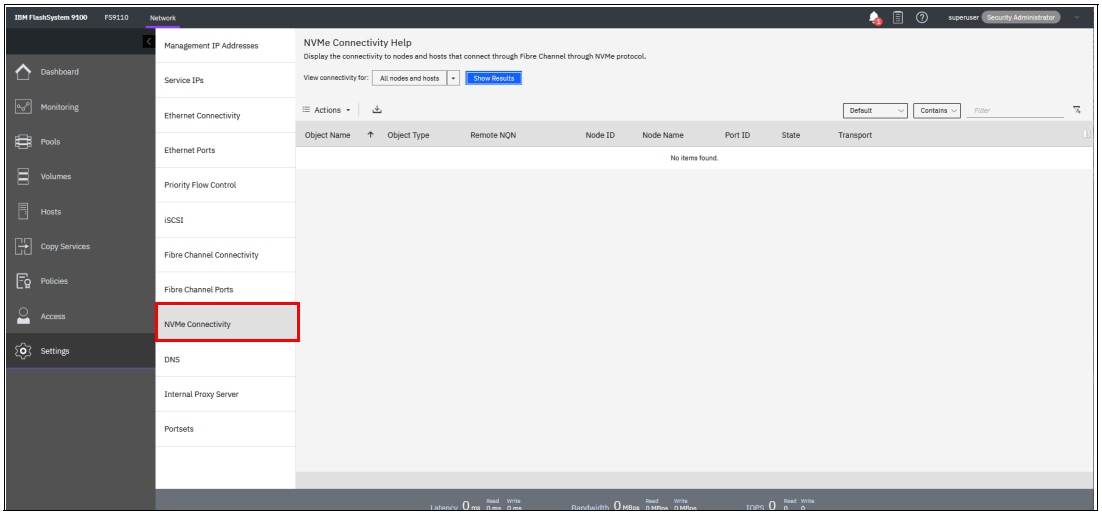


Figure 4-101 NVMe Connectivity window

Consider FC-NVMe target limits when you plan and configure the hosts. Include the following points in your plan:

- An NVMe host can connect to four NVMe controllers on each system node. The maximum per node is four with an extra four in failover.
- Zone up to four ports in a single host to detect up to four ports on a node. To allow failover and avoid outages, zone the same or extra host ports to detect an extra four ports on the second node in the I/O group.
- A single I/O group can contain up to 256 FC-NVMe I/O controllers. The maximum number of I/O controllers per node is 128 plus an extra 128 in failover. Zone a total maximum of 16 hosts to detect a single I/O group. Also, consider that a single system target port can have up to 16 NVMe I/O controllers.

When you install and configure attachments between the system and a host that runs the Linux operating system, follow specific guidelines. For more information about these guidelines, see NVMe hosts that run the Linux operating system.

Domain name server

IBM Storage Virtualize allows DNS entries to be manually set up in the system. The information about the DNS helps the system to access the DNS and resolve the names of the computer resources that are in the external network.

To view and configure DNS information in IBM Storage Virtualize, complete the following steps:

1. In the left panel, click **DNS** and select **Add DNS server**. Enter the IP address and the name of each DNS server. IBM Storage Virtualize supports a maximum of two DNS servers for IPv4 or IPv6 (see Figure 4-102 on page 301).

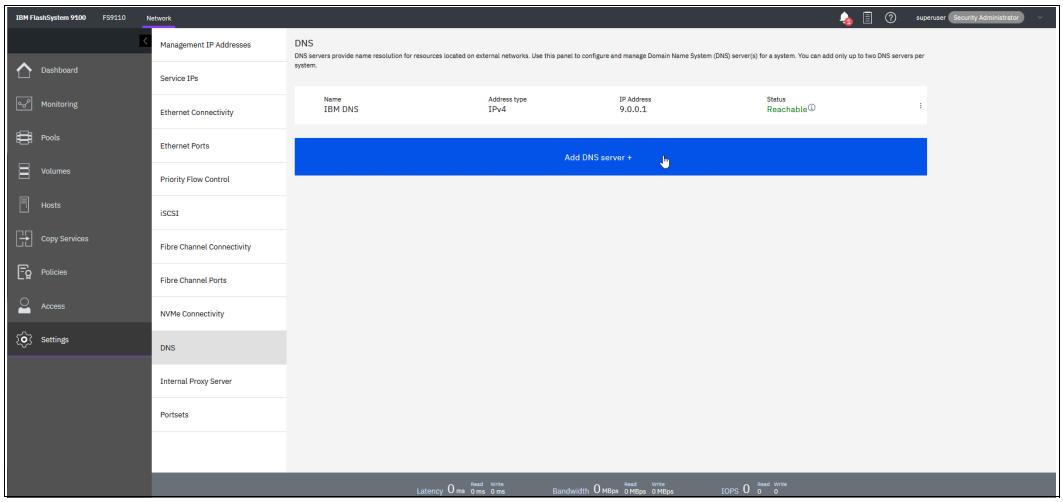


Figure 4-102 DNS information

2. Click **Save** after you finish entering the DNS information.

Internal Proxy Server

You can configure an internal proxy server to manage incoming and outgoing connections to the system, as shown in Figure 4-103.

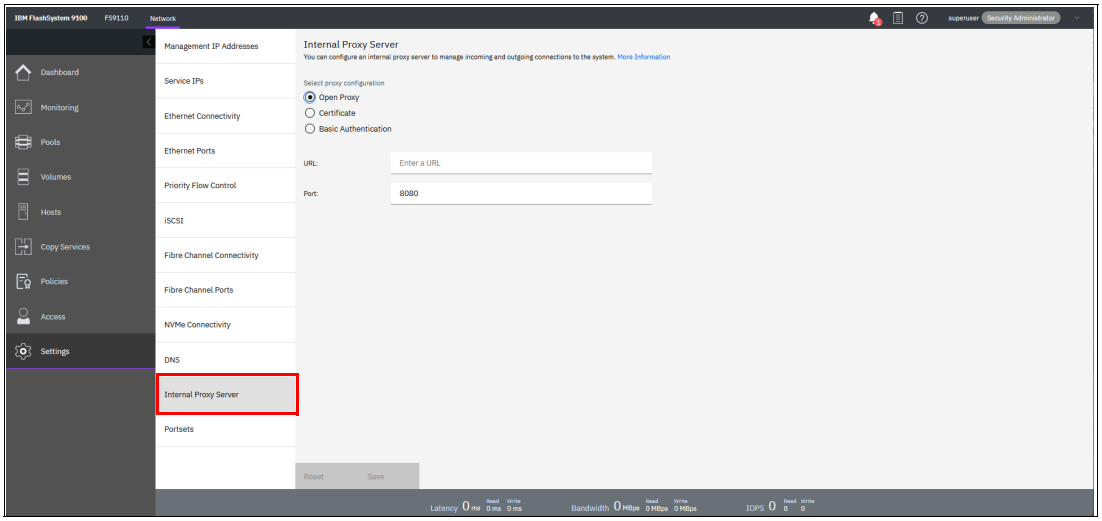


Figure 4-103 Internal Proxy Server

As part of Call Home with cloud services configuration, you can choose to use your proxy server within a network to manage connections between the system and the support center. A proxy server manages connections between your internal network and any entity that is outside of your network.

Call Home with cloud services or remote support assistance requires connections to the support center for quick and efficient problem resolution for your system. Instead of opening ports for these external connections through your firewall, you can specify an internal proxy server for better security. To use an internal proxy server, a DNS server is also required to resolve hostnames to IP addresses.

You can specify a hostname and port to define the proxy configuration and optionally add security-related settings. To define an internal proxy server, complete the following steps:

1. In the management GUI, select **Settings** → **Network** → **Internal Proxy Server**. For more information, see this Creating a DNS server.
2. Select one of the following types of proxy configuration and complete the required fields:
   - Open Proxy

     Select this option to define an internal proxy server with a URL and port only. This configuration does not include any security controls.
   - Certificate

     Select this option to import the certificate for the internal proxy server, which allows the system to verify the authenticity of the internal proxy server.
   - Basic Authentication

     Select this option to require authentication with a username and password for outbound connections through the internal proxy server.
3. Click **Save**.

Portsets

Portsets are groupings of logical addresses that are associated with the specific traffic types. The system supports portsets for host, storage, and remote-copy traffic. An example of a portset is shown in Figure 4-104.

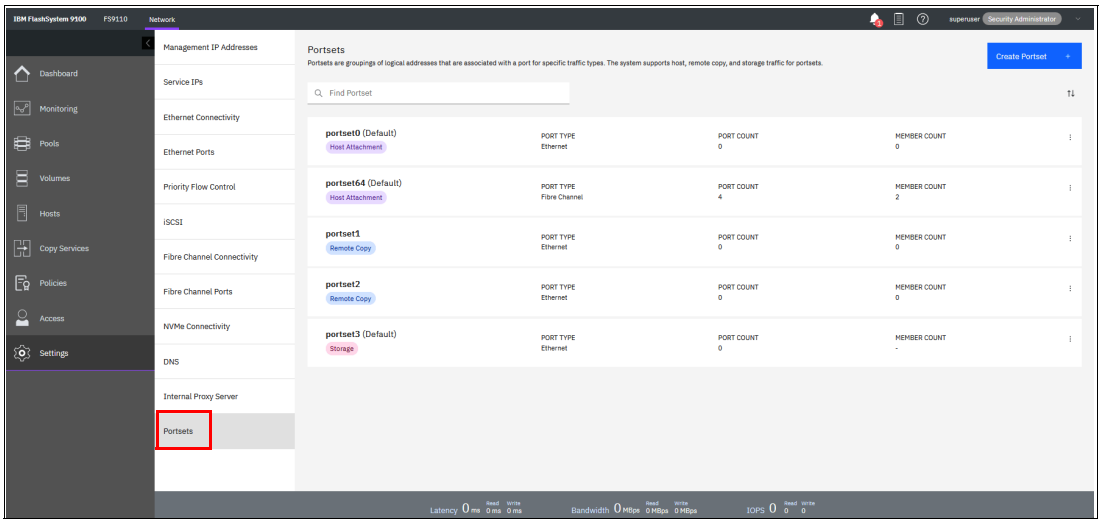


Figure 4-104 portsets

Each port supports one IP address per portset and each port can have 64 IP addresses, which associate with different portsets to allow multiple tenants to attach to the nodes through Ethernet host attachment protocols, such as iSCSI and iSER. For cloud environments, Ethernet ports support different IP addresses and VLANs for network isolation for multiple clients that share storage resources on the system.

Requirements for portsets

In general, portsets have the following characteristics:

- Portsets are system-wide objects where IP addresses from all nodes must be included in the portset to host, storage, and replication.

- Each IP address in a portset must be configured on a separate Ethernet port.
- The same ports can share IP addresses across different portsets that allow the same IP address to be used for host, storage, and remote-copy traffic. All shared IP addresses must use the same port and have the same VLAN, gateway, and prefix.

  When IP addresses are shared among multiple portsets, the system creates a logical copy of the IP address and its attributes, rather than a new IP address.
- Portsets that are owned by different ownership groups can share an IP address.
- The system supports a maximum of 64 portsets. However, restrictions on the number of IP addresses per port can indicate that not all of these supported IP addresses can be used.
- A port can have 64 unique IP addresses. All 64 IP addresses must be IPv4 or IPv6, or a mix of IPv4 and IPv6.
- Each port can be configured with only one unique routable IP address (gateway specified). The routable IP can be shared among multiple portsets.
- As part of portset and multi-tenancy support, you must configure VLAN settings on the switches and hosts before you can create portsets and IP addresses. For more information, see /Configuring VLAN.
- Portset 0 is a default portset that is automatically configured when the system is updated or created. Portset 0 is a host portset by default and cannot be deleted, even if it is empty.

  Portset 0 serves as the default portset for any IP addresses and host objects that are configured without a specified portset. Portset 0 allows administrators to continue with an original configuration that does not require multi-tenancy. After an update, all configured host objects are automatically mapped to portset 0.
- Portset 3 is another default portset that is automatically configured when the system is updated or created, but it is reserved for storage virtualization traffic only. Portset 3 cannot be deleted, even when it is empty. When an IP address is configured for storage traffic, in previous releases, it is automatically mapped to portset 3 after upgrade.

Host portset requirements

In addition to portset 0, you can create portsets for host traffic. Host traffic includes traffic to individual hosts and host clusters. The following requirements are specific to host portsets:

- Portsets can have a maximum of 4 IP addresses per node.
- A single portset can contain IPv4, IPv6 IP, or a mix of IPv4 and IPv6 addresses.
- Portsets can have IP addresses that are configured on ports with different capabilities, such as those that support iWARP+RoCE.
- For a host to log in to nodes on the system, the host must be mapped to a portset that contains at least one IP address from any of the nodes on that system.

Replication portset requirements

You can create portsets for IP partnerships. The following characteristics are specific to replication portsets:

- Replication portset can have a maximum of 1 IP address per node.
- All IP addresses in replication portsets must be IPv4 or IPv6 addresses. You cannot mix IP protocol versions on replication portsets. The protocol for both the IP partnership and the portset must be the same.

- Each IP partnership can be mapped to two portsets: one for each link between systems. For a partnership with a single link, a single portset can be defined in the Portset Link 1 field on the Create Partnership page. For a partnership with dual links, a second portset must be defined in the Portset 2 field in the GUI.
- Portsets replace the requirement for creating remote-copy groups for IP partnerships. During software updates, any IP addresses that are assigned to remote-copy groups with an IP partnership are automatically moved to a corresponding portset.

  For example, if remote-copy group 1 is defined on the system before the update, IP addresses from that remote-copy group are in portset 1 after the update. Any IP address in remote-copy group 2 is placed in portset 2.

Creating portsets

In the management GUI, complete the following steps to create a portset:

1. Select **Settings** → **Network** → **Portsets**.
2. Select **Create Portset**.
3. On the Create Portset page, enter a name of the portset, and select the one of the following options for the portset type:
   - Host attachment

     Indicates that the IP addresses that are added to the portset are used for host attachment only
   - Remote Copy

     Indicates that the IP addresses that are added to the portset are used for IP partnerships only
4. For the Port Type, select **Ethernet**.
5. Select the ownership group for the portset. An ownership group defines a subset of users and objects within the system. You can create ownership groups to further restrict access to specific resources that are defined in the ownership group.

   Only users with Security Administrator roles can configure and manage ownership groups. Restricted users are those users who are defined to a specific ownership group and can view or manage only specific resources that are assigned to that ownership group.

   Unrestricted users are not defined to an ownership group and can manage any objects on the system based on their role on the system.

   For more information, see 12.5, “Ownership groups principles of operations” on page 1083.

   When you define an ownership group for portsets, you can limit and restrict users to view and manage only specific portsets (see Figure 4-105 on page 305).

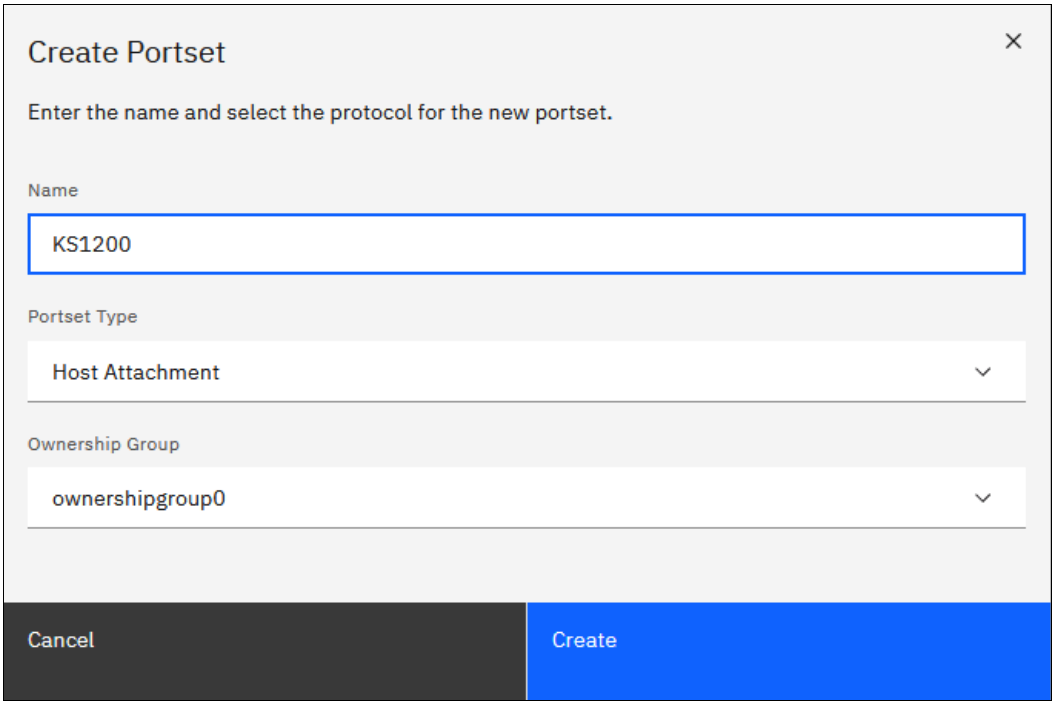


Figure 4-105 Create Portset window

6. Click **Create**.

After you create the portset, complete the following steps to assign IP addresses to that portset:

1. Select **Settings** → **Network** → **Ethernet Ports**.
2. Right-click the port and select **Manage IP addresses**.
3. On the **Manage IP Addresses** page, select **Add IP Address**.

   Enter the following information for the IP address that you are adding to the selected port:

   – IP address

     IP address to associate with the selected port

   – Type

     The IP protocol version of the IP address

   – Subnet Mask or Prefix

     The subnet mask for the IPv4 addresses or the prefix for IPv6 addresses

   – VLAN

     The corresponding VLAN tag to which this IP address belongs

   – Portset

     The name of the portset. Ensure that the portset type matches the traffic type that is assigned to the port.
4. Click **Back** to return to the Ethernet Ports page. Verify that the port displays the **Configured** state. Select another port and add IP addresses to the corresponding portsets.

5. Right-click the port and select **Modify Remote Copy**, **Modify iSCSI hosts**, or **Modify Storage Ports**. The traffic type for the port must match the traffic for the portset that you created.

After you create portsets and assign IP addresses, you can assign hosts, host clusters, and IP partnerships to the portset for those traffic types.

4.7.4 Security

Select **Settings** → **Security** (see Figure 4-106) to view and change security settings, authenticate users, and manage secure connections.

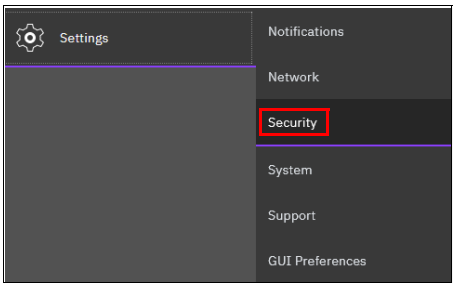


Figure 4-106 Security menu

With IBM Storage Virtualize version 8.6, some new options are available in **Settings** → **Security**: System Certificates, User access, SSH Rules and Security Levels (Figure 4-107).

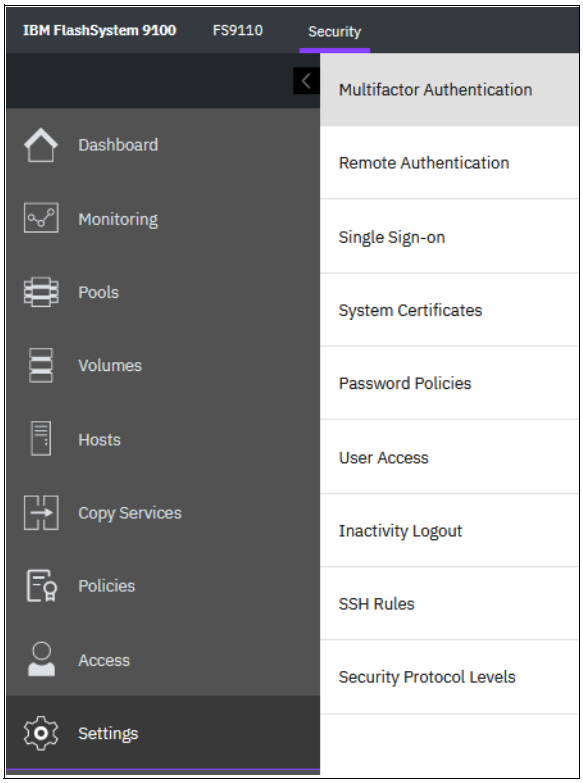


Figure 4-107 Security Settings

For more information about all of the options, see Chapter 12, “Security and encryption” on page 1045.

Multifactor authentication

Use the Multifactor Authentication (MFA) page to enable multifactor authentication for the system. Multifactor authentication requires users to provide several pieces of information when they log in to the system to prove their identity.

Multifactor authentication uses any combination of two or more methods, which are referred to as *factors*, to authenticate users to your resources and protect those resources from unauthorized access.

One of the key concepts of multifactor authentication is that each factor comes from the following categories

- Something a user knows

  By using this first factor, the users authenticate with information that only each user knows, such as a password or IBM PIN®.

- Something a user has

  By using this second factor, the users prove their identity with information that is provided to the user by a trusted authentication service, such as one-time pass codes that are generated by an application or mobile device.

- Something a user is

  By using this third factor, users prove their identity with biometrics, such as a fingerprint or retinal scan.

With the adoption of cloud-based services, multifactor authentication increases the control over user access and security settings. First-factor authentication methods alone, such as username and password combinations, do not provide the level of protection and security that is required in cloud and hybrid-cloud environments.

With multifactor authentication support, security administrators can reinforce account protection, create granular access for users and user groups, and monitor access more efficiently at a system level.

The system integrates with IBM Security Verify or Duo Security, which are cloud-based identity and access management (IAM) service providers. These services provide different factors to validate and verify users who access the system.

To configure multifactor authentication, you must complete the prerequisite and configuring steps on IBM Security Verify and in the management GUI. Figure 4-108 on page 308 shows the Multifactor Authentication window with IBM Security Verify. The system integrates with IBM Security Verify to provide multifactor authentication for system users.

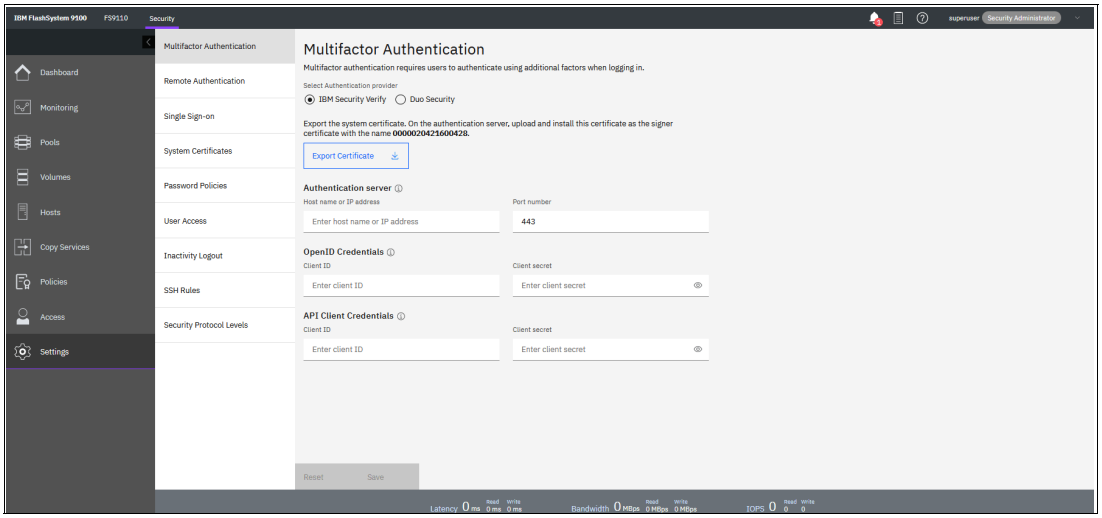


Figure 4-108 Multifactor Authentication with IBM Security Verify

- Figure 4-109 shows the Multifactor Authentication window with Duo Security. The system integrates with Duo Security to provide multifactor authentication for system users. For more information, see 12.4, “Multifactor authentication” on page 1066 and Confguring multifactor authentication.

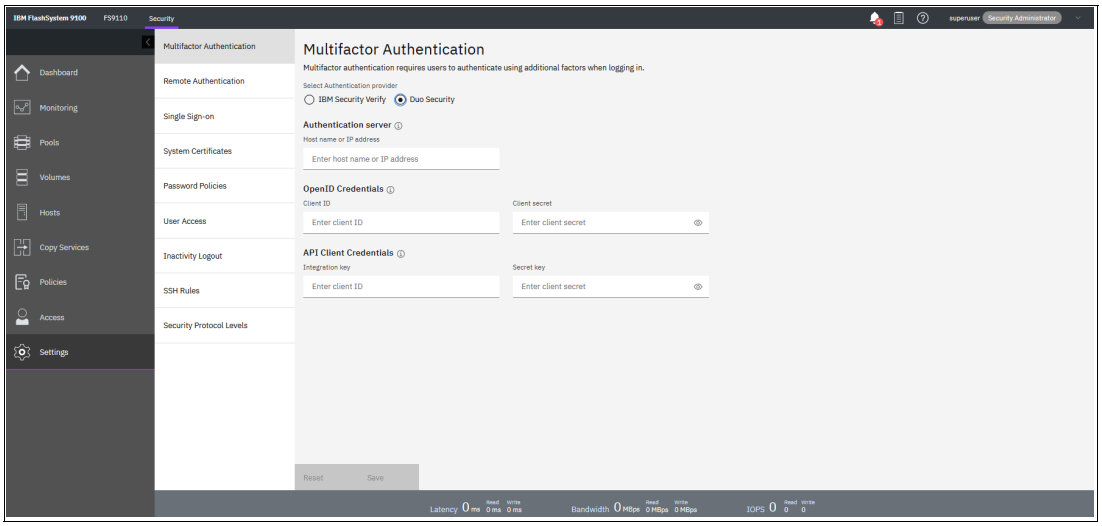


Figure 4-109 Multifactor Authentication with Duo Security

Remote Authentication

Figure 4-110 shows the Remote Authentication window. Here, you can refresh the authentication cache or configure the remote authentication.

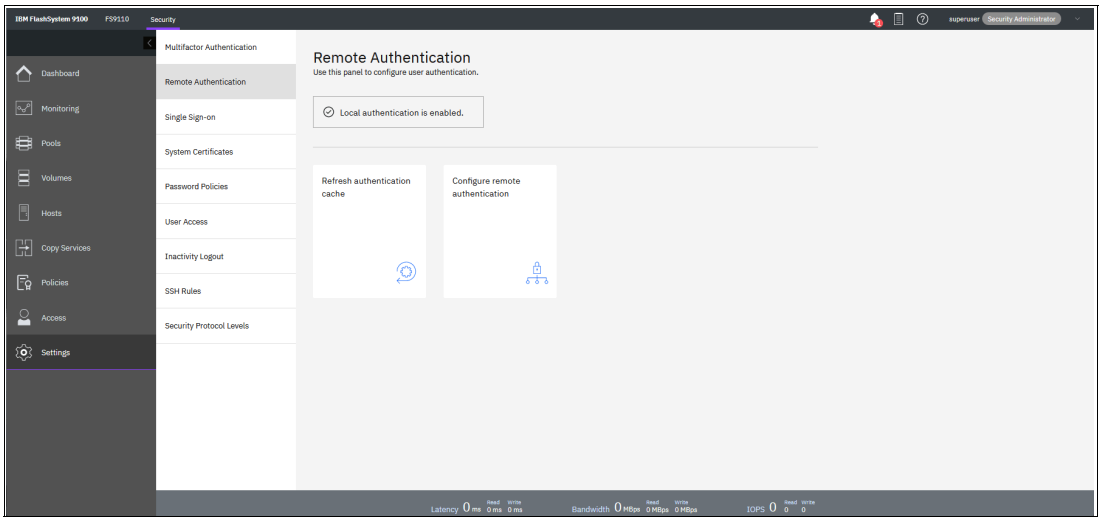


Figure 4-110 Remote Authentication

Note: If you refresh the authentication cache, you might lose access to the system if you logged in as a remote user (see Figure 4-111).

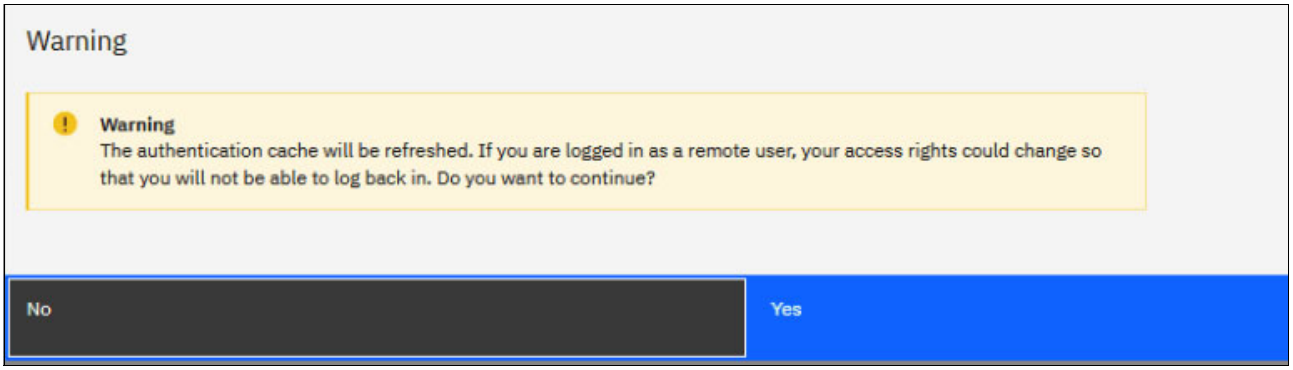


Figure 4-111 Refresh Authentication

In the Configure remote authentication window, you can configure remote authentication by using LDAP, as shown in Figure 4-112.

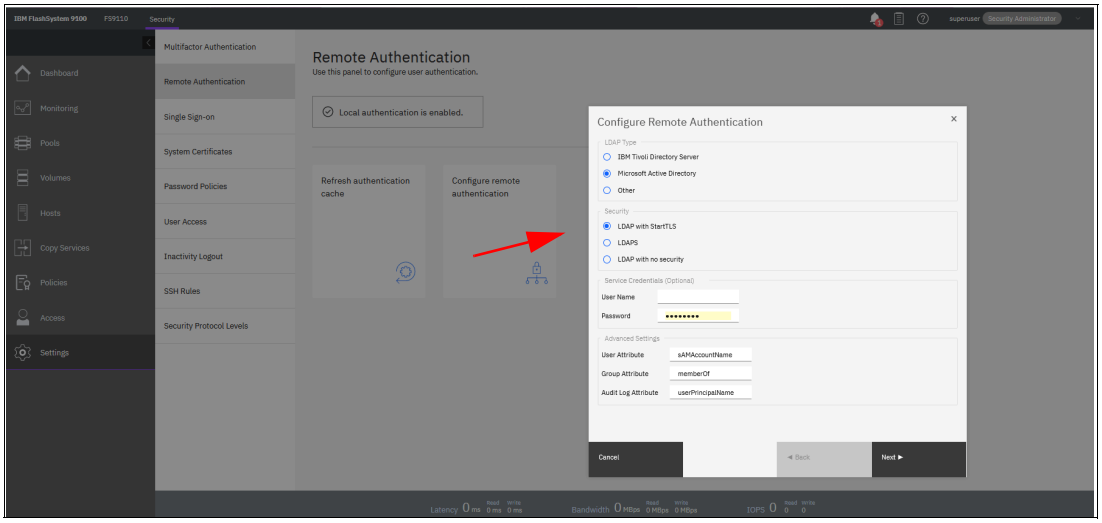


Figure 4-112 Configuring Remote Authentication

By default, the system has local authentication that is enabled. When you configure remote authentication, you do not need to configure users on the system or assign more passwords. Instead, you can use your passwords and user groups that are defined on the remote service to simplify user management and access, enforce password policies more efficiently, and separate user management from storage management.

For more information, see Configuring remote authentication.

Single Sign-on

Use the Single Sign-On page to configure single-sign on for the entire system (see Figure 4-113).

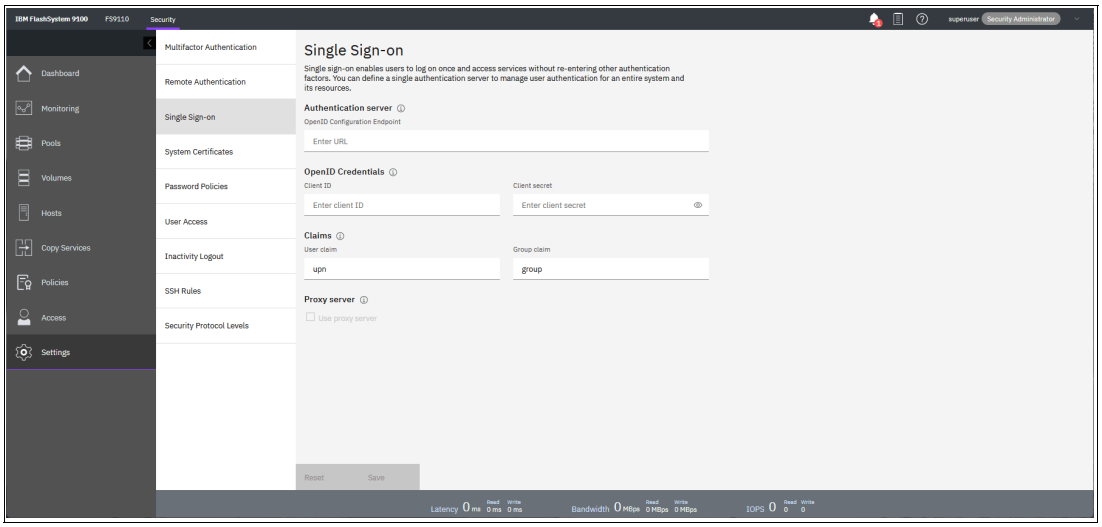


Figure 4-113 Single Sign-on

Single sign-on delegates all authentication to a trusted Identity Provider (IdP). With single sign-on, users must provide their credentials only once when they log in to an application or system, rather than repeatedly providing the credentials for every individual application or system. Each individual IBM Storage Virtualize system is considered a separate application and must be added to the IdP.

When single sign-on is enabled on the system, first- and second-factor authentication are delegated to the IdP. The system supports Microsoft Active Directory Federation Services (AD FS) to provide single sign-on.

In the management GUI, users select Sign In with SSO at the log in prompt and are redirected to complete authentication through the configured IdP. You can configure the IdP to provide first factor authentication to your users.

Several multifactor authentication cloud-based providers can be added to the single sign-on configuration to require more user authentication, if necessary.

When authentication is completed successfully, the users are redirected to the management GUI. The single Sign-on login window is shown in Figure 4-114.

For more information about prerequisites and how to configure single sign-on, see Chapter 12, “Security and encryption” on page 1045 and Configuring single sign-on.

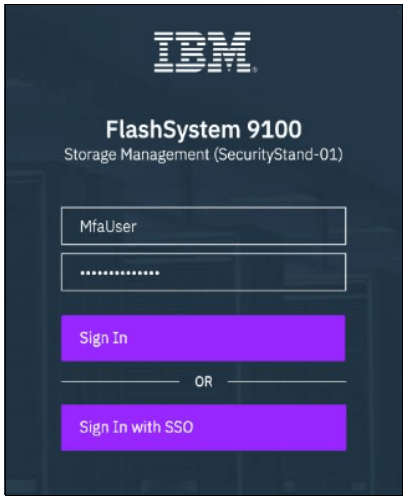


Figure 4-114 Single Sign-on login

System Certificates

To enable or manage secure communications, select the **System Certificates** window, as shown in Figure 4-115 on page 312. Before you create a request for either type of certificate, ensure that your browser does not restrict the type of keys that can be used for certificates.

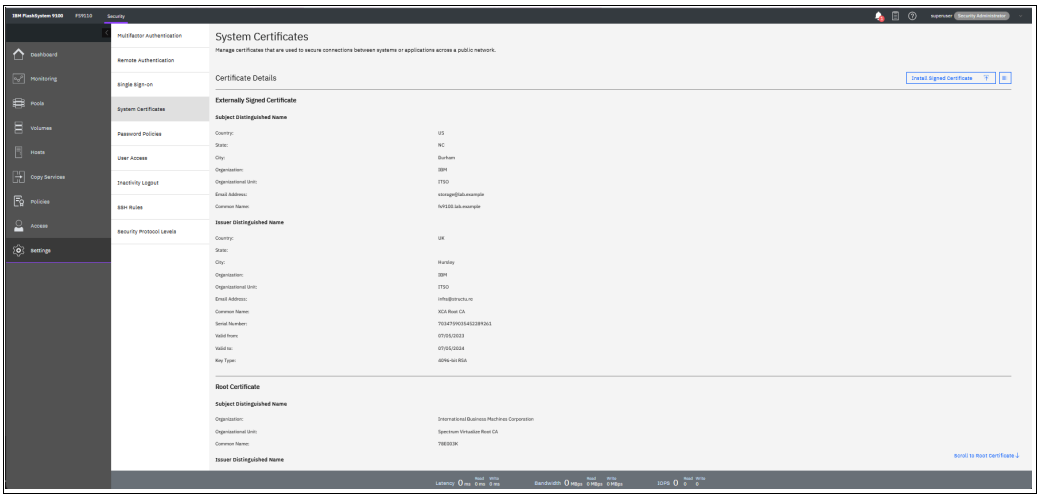


Figure 4-115 Configuring secure communications and updating certificates

Use the System Certificates page to enable and manage secure connections between systems or applications across networks. Before you create system certificates, determine how certificates are signed based on your current security requirements. The system supports several ways to sign system certificates.

Self-signed certificate

In 8.5.3.0, self-signed certificates are no longer created by default. After an update to 8.5.3.0 or later, the current self-signed certificate is usable until it expires. Before the self-signed certificate expires, you must update the system certificate to either an internally signed certificate or an externally signed certificate.

Internally signed certificate

The system has a root certificate authority (CA) that can be used to create internally signed system certificates. In 8.5.3.0 or later, system setup creates a certificate that is signed by the root CA to secure connections between the management GUI and the browser. The root certificate can be exported from the system and added to truststores on other systems, browsers, or devices to establish trust. Internally signed certificates can be renewed automatically before they expire. Automatic renewal simplifies the certificate renewal process and prevents security warnings from expired certificates. Automatic renewal is only supported by using an internally signed certificate.

Externally signed certificate

Externally signed certificates are issued and signed by a trusted third-party provider of certificates, called an external certificate authority (CA). This CA can be a public CA or your own organization's CA. Most web browsers trust well-known public CAs and include the root certificate for these CAs in the device or application. Externally signed certificates cannot be renewed automatically because they must be issued by the external CA. Externally signed certificates must be manually updated before they expire by creating a new certificate signing request (CSR) on the system and supplying it to the CA. The CA signs the request and issues a certificate that must be installed on the system. The system raises a warning in the event log 30 days before the certificate expires.

Additional information about certificates

This section describes root and system certificates.

Root certificate

When running version 8.5.3.0 or later, IBM Storage Virtualize has a root certificate authority that can be used to generate system certificates. The root CA is generated during the first boot on 8.5.3.0 and is shared between all nodes in the clustered system. The root CA cannot be modified by the user. The root certificate can be exported from the system and added to trust stores for web browsers, Thales CipherTrust Manager key servers, and other devices and applications that support chain of trust checking. If chain of trust checking is supported, then the signed system certificate can be renewed and does not need to be exported.

The root certificate can be identified by its subject and issuer distinguished name:

- O represents International Business Machines Corporation.
- OU represents IBM Storage Virtualize Root CA.
- CN represents a serial number, for example, 1234567. The common name (CN) field is the enclosure serial number of the configuration node that generated the certificate.

The root certificate is valid for 20 years and uses a 4096-bit RSA key pair. After the first system upgrade to 8.5.3.0 or later, the system continues to use its current certificate that is either self-signed or externally signed until a new certificate is generated.

System certificate

The system certificate is signed by the system's root CA, a trusted third-party CA, or is self-signed. This certificate is presented to other devices such as web browsers, key servers, or partner systems in IP partnerships, to authenticate the system and create a secure connection.

The system certificate supports the following key types:

- RSA 4096
- RSA 2048
- ECDSA 521
- ECDSA 384

SSL protocol level 4 restricts the use of RSA key exchange cipher suites. Therefore, you cannot enable SSL protocol level 4 when using a certificate with an RSA key. Likewise, you cannot generate a certificate with an RSA key if SSL protocol level 4 is set.

Note: Multifactor Authentication with IBM Security Verify uses the system certificate to sign JSON Web Tokens (JWTs). The system certificate must be exported and added as a signer certificate in the IBM Security Verify interface. If the system certificate expires or is renewed, access to the management GUI is unavailable until the new system certificate is installed in IBM Security Verify. A user must use the CLI to log in to the system to export the new system certificate and install it in IBM Security Verify.

Note: IBM Security Guardium Key Lifecycle Manager key servers do not currently support chain of trust checking with IBM Storage Virtualize. The system certificate must be installed on the IBM Security Guardium Key Lifecycle Manager key servers in order to establish a connection. If the system certificate is expired or renewed, then the new system certificate must be installed on the key servers so that IBM Storage Virtualize can establish a connection with the key servers. Access to encrypted storage is not disrupted while the connection to the key servers is unavailable unless all nodes in the clustered system are powered off and on at the same time.

During system setup, an initial certificate is created to use for secure connections between web browsers. This certificate is signed by the system's root CA. A new certificate should be generated which includes the relevant DNS or IP entries for the system in the Subject Alternate Name field. The System Certificates page in the management GUI suggests the DNS names if a DNS server is added to the system. If a DNS server is not added, then the management GUI suggests the IP addresses.

Note: If a certificate is changed, the certificate must also be updated on *all* configured key servers or access to encrypted data can be lost.

Password Policies

In this window, you can define policies for password management and expiration, as shown in Figure 4-116.

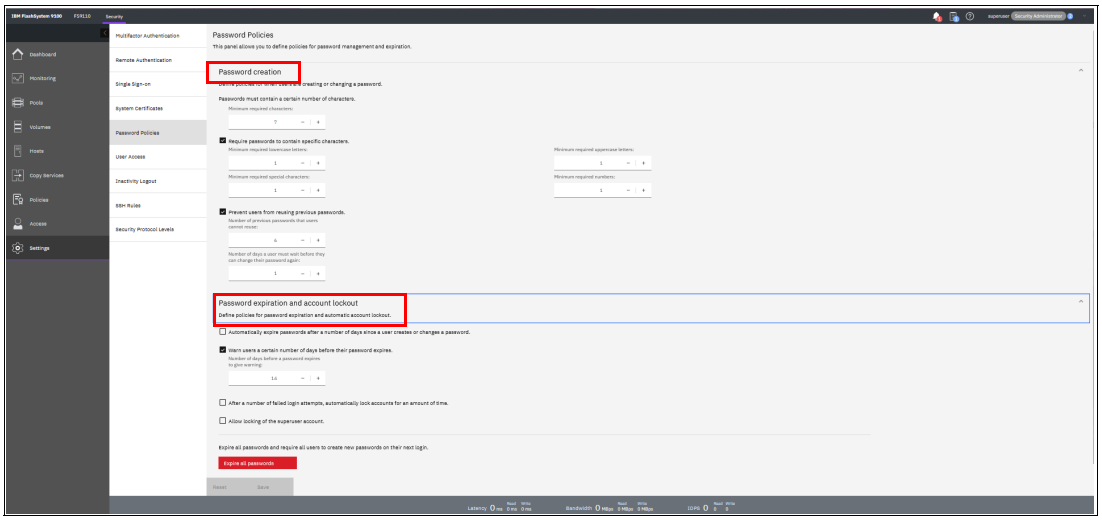


Figure 4-116 Password Policies window

With password policy support, system administrators can set security requirements that are related to password creation and expiration, timeout for inactivity, and actions after failed login attempts. Password policy support allows administrators to set security rules that are based on their organization's security guidelines and restrictions.

The system supports the following password and security-related rules:

- Password creation rules:

  An administrator can set and manage the following rules for all passwords that are created on the system:

  - Password length requirements for all users
  - Password content requirements:
    - Use of uppercase and lowercase characters
    - Use of numbers
    - Use of special characters
  - Rules for reuse of recent passwords
  - Requirement to change a password on the next login under any of the following conditions:
    - Expired password
    - Temporary password for a created account that has a temporary password

- Password expiration and account locking rules:
  - The administrator can set the following rules for password expiration:
    - Password expiration limit
    - Password to expire immediately
    - Number of failed login attempts before the account is locked
    - Time for locked accounts
  - Locking superuser account access.
- Automatic logout for inactivity.

Note: Systems that support a dedicated technician port can lock the superuser account. The superuser account is the default user that can complete installation, initial configuration, and other service-related actions on the system. If the superuser account is locked, service tasks cannot be completed.

User Access

Use the User Access page to control access of the system for the user. The user can access all system resources and objects, and also perform service assistant actions. Depending on your security requirements, you might need to limit access to the system interfaces for the user. To update the user access settings in the management GUI, select **Settings** → **Security** → **User Access**.

Superuser access

To update the Superuser access, complete these steps:

- Update the following properties that affect superuser access to the system:
  - Require both password and SSH key for superuser authentication. Indicates whether the superuser must provide both password and SSH key during login. Select **On** to enable both password and SSH key authentication for superuser.
  - Prevent GUI access for superuser. Indicates whether the management GUI access is allowed or denied for the superuser. Select **On** to deny the superuser access to the management GUI.
  - Prevent REST API access for superuser. Indicates whether REST API access is allowed or denied for the superuser. Select **On** to deny the superuser access to the REST API.
- Click **Save**.

Two person integrity(TPI)

Use the TPI settings to manage security controls of security administrators and prohibit certain tasks in the system by requiring the involvement of two security administrators. When you enable TPI, the users that belong to user groups of security administrator role are assigned the restricted security administrator role instead. However, their user groups retain their security administrator role. Once TPI is enabled, a role elevation request and approval process is required to perform certain sensitive tasks:

- The restricted security administrator can issue a role elevation request on its own behalf to complete certain tasks in the system.
- Another restricted security administrator or a security administrator must approve the role elevation request.
- For example, this role elevation request and approval process is required to remove a Safeguarded snapshot.

- The restricted security administrators or security administrators can approve or deny role elevation requests, cancel role elevation requests, or revoke a role elevation request that was approved.

Restricted security administrator with an approved role elevation request

A restricted security administrator with an approved role elevation request can perform the following tasks:

- Create, change, or remove security administrator user groups.
- Change the non-security administrator user group attribute on an existing local user to a security administrator user group.
- Modify attributes on existing local users that are members of the security administrator user groups.
- Change the role of existing non-security administrator user groups to the security administrator role.
- Change the security administrator role of an existing user group to a non-security administrator role.
- Remove and change Safeguarded backups and Safeguarded backup locations.
- Delete Safeguarded snapshots.
- Use a provisioning policy to define a set of rules that are applied when volumes are created within a storage pool or child pool.
- Change the single sign-on credentials that are used for the system.
- Remove the Safeguarded snapshot policy association from a volume group.

Enabling TPI

To enable TPI, complete the following steps:

1. In the management GUI, select **Settings** → **Security** → **User Access** → **Two person integrity**.
2. Select **Enabled**.
3. Click **Save**.
4. To apply the TPI changes, click **Log out**.
5. After you sign in, the page displays that the current role is updated to restricted security administrator, and Manage Role Elevation Requests and Request Elevated role are displayed in the user menu list
6. Click **Snooze for 5 minutes** to stay on the page for 5 more minutes to review the settings.
7. To disable TPI, click **Disabled**. After TPI is enabled, a user with an approved TPI request can disable TPI.

   See Figure 4-117 on page 317 for the User Access settings

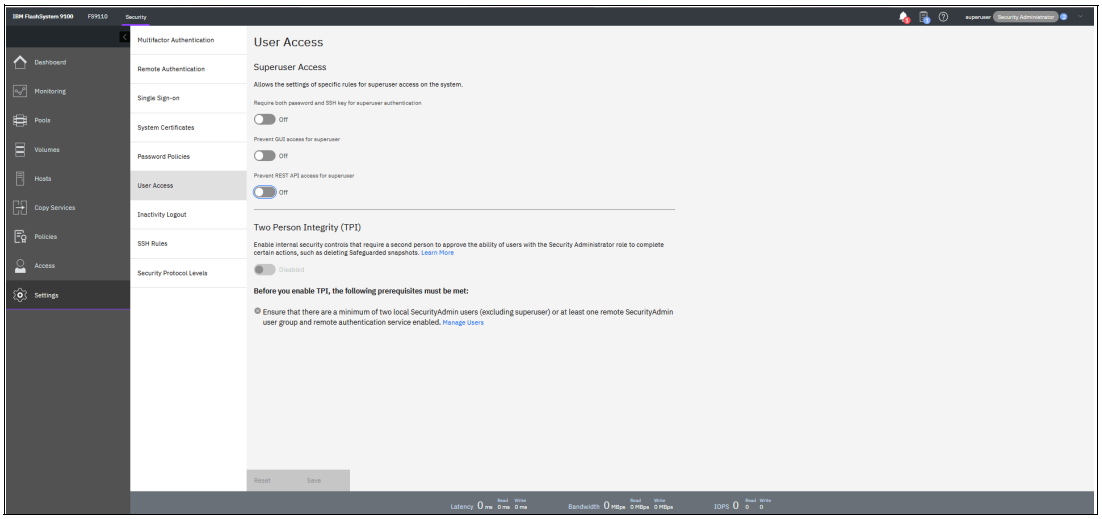


Figure 4-117 User Access

Inactivity Logout window

By using this window, you can set the inactivity time that is allowed before the system logs out users (see Figure 4-118). You can set values for command line and management GUI access.

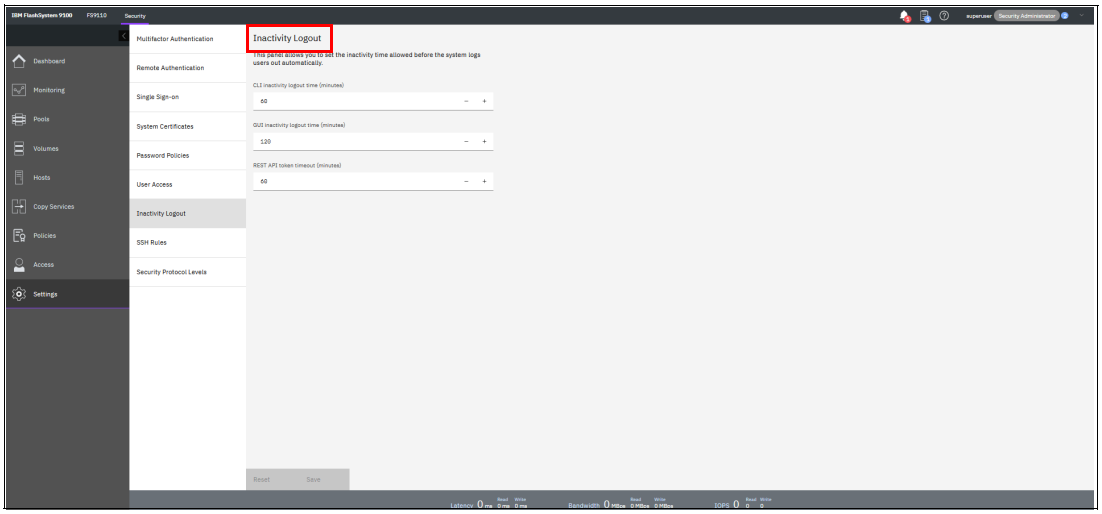


Figure 4-118 Inactivity Logout

SSH Rules

Use the SSH Rules page to customize settings related to Secure Shell (SSH) access. SSH is used to authenticate users to the command line interface.

You can update any of the following details:

SSH login grace period (seconds)

Indicates the amount of time in seconds to log in before SSH times out. The range is 15–1800.

Maximum login attempts (SSH)

Indicates the total number of login attempts allowed per single SSH connection. The range is 1–10.

See Figure 4-119 for the settings of SSH Rules

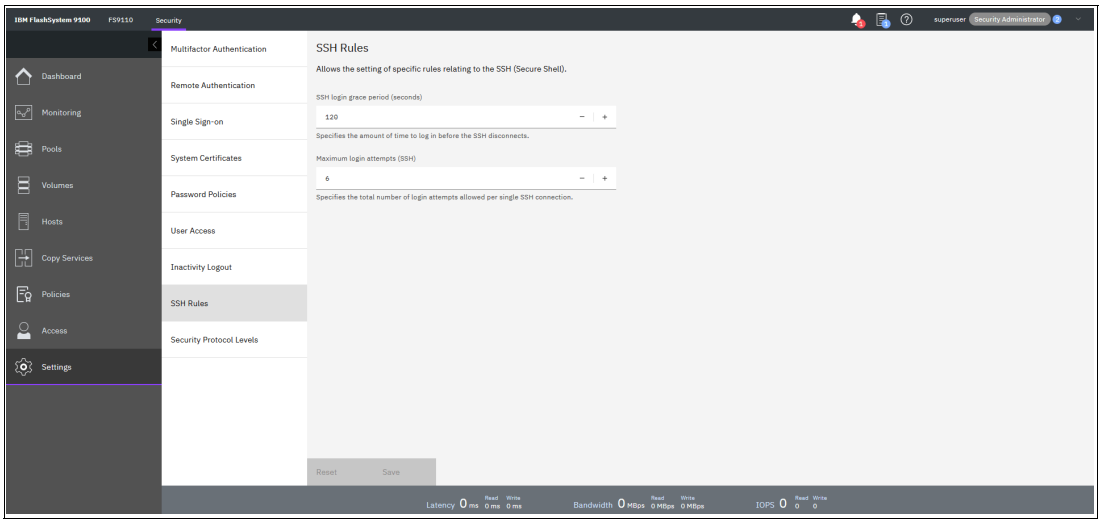


Figure 4-119 SSH Rules

Security protocol levels

The system supports different SSL and SSH protocol levels that control which security ciphers are used to encrypt connections between system and devices. Different SSL and SSH protocol levels support a range of security ciphers that provide differing strengths of encryption.

SSL protocol levels

The SSL protocol levels are described below:

- Level 1: Support TLS 1.2 and allow SSL 3.0 and TLS (1.0, 1.1, 1.2) ciphers. For security reasons, level 1 is deprecated and no longer supported from 8.6.0.0.
- Level 2: Support TLS 1.2 and allow legacy TLS ciphers.
- Level 3: Support TLS 1.2 and allow TLS 1.2 ciphers.
- Level 4: Support TLS 1.2, allow TLS 1.2 ciphers, and disallow RSA and static key exchange ciphers.
- Level 5: Support TLS 1.2 and TLS 1.3 but disallow static key exchange ciphers. This is the compatible mode.
- Level 6: Support TLS 1.3 and allow TLS 1.3 ciphers.
- Level 7: Support TLS 1.3 and a single FIPS cipher.

SSH protocol levels

The *SSH protocol level* supports the following levels:

- 1 - Indicates to allow block ciphers.
- 2 - Indicates to disallow block ciphers.
- 3 - Indicates to disallow SHA1.

Changing the SSL protocol level causes services that use the protocol level to restart. Restarting the services terminates the existing sessions and ensures that there are no active sessions on the old security level. After the services restart, it might take a few minutes for the services to become usable again.

Note: By default, the SSL protocol level is set to 5, and the SSH protocol level is set to 3.

Figure 4-120 shows you the possible setting for the different security protocol levels

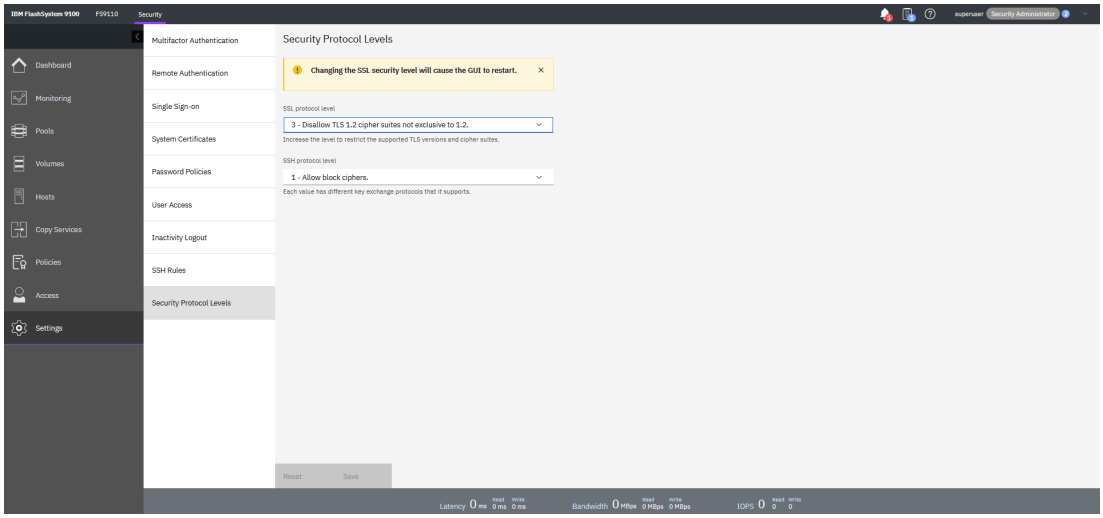


Figure 4-120 Security protocol levels

4.7.5 System menus

Click **Settings** → **System** (Figure 4-121) to view and change the date and time settings, work with licensing options, download configuration settings, work with VVOLs and IP Quorum, or download software upgrade packages.

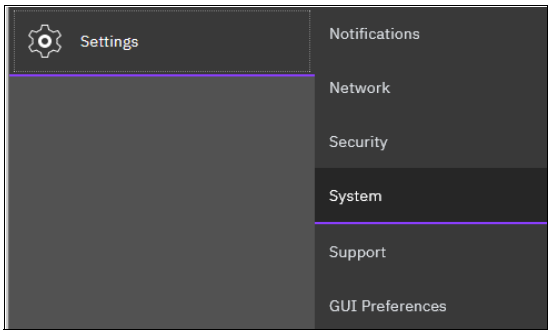


Figure 4-121 System option

Date and time

To view or configure the date and time settings, complete the following steps:

1. From the main System window, click **Settings** → **System**.
2. In the left column, select **Date and Time**, as shown in Figure 4-122.

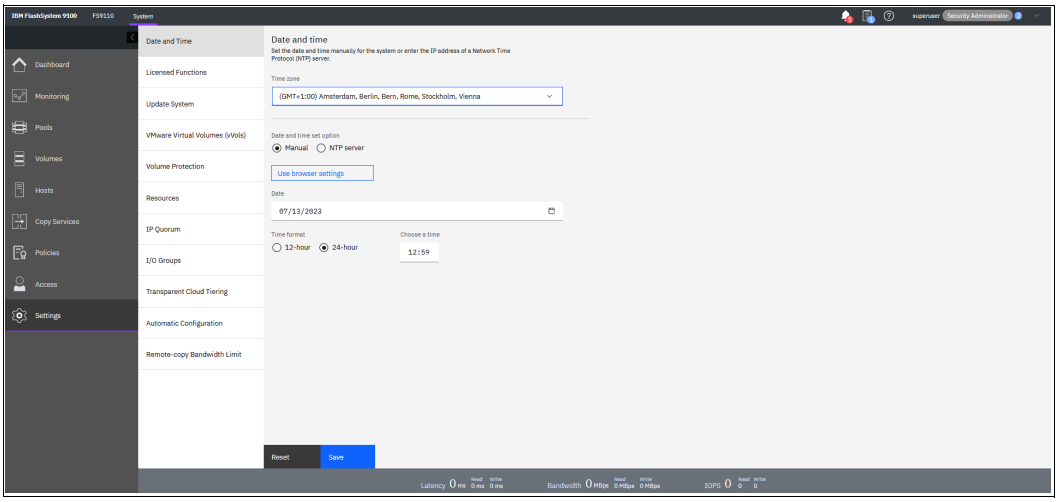


Figure 4-122 Date and Time window

3. From this window, you can modify the following information:
   - Time zone

     Select a time zone for your system by using the drop-down list.
   - Date and time

     The following options are available:
     - If you are not using a Network Time Protocol (NTP) server, manually enter the date and time for your system, as shown in Figure 4-123. You can also click **Use Browser Settings** to automatically adjust the date and time of your system with your local workstation date and time.

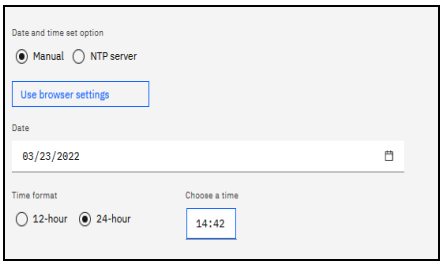


Figure 4-123 Set Date and Time window

- If you use an NTP server, select **Set NTP Server IP Address or Domain** and then enter the IP address of the NTP server, as shown in Figure 4-124.

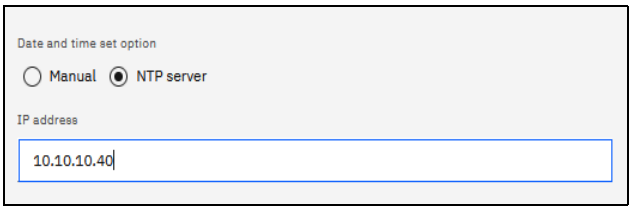


Figure 4-124 Set NTP Server IP Address window

4. Click **Save**.

Licensed functions

The base license that is provided with your system includes the use of its basic functions. However, the extra licenses can be purchased to expand the capabilities of your system. Administrators are responsible for purchasing extra licenses and configuring the systems within the license agreement, which includes configuring the settings of each licensed function on the system.

Depending on the platform, the following license schemes can be used:

- The IBM FlashSystem 5010 and 5015, IBM FlashSystem 5030, 5035 and 5045, IBM FlashSystem 5200 and IBM FlashSystem 7300 system licenses are Licensed Internal Code (LIC). All licenses are controller-based.
- The IBM FlashSystem 5100 system uses enclosure-based licensing, which allows the use of specific licensed functions that are based on the number of enclosures that are indicated in the license.
- IBM FlashSystem 7200, IBM FlashSystem 9100, IBM FlashSystem 9200, and IBM FlashSystem 9500 systems use differential licensing for external virtualization, and capacity-based licensing for other functions.
- The IBM SAN Volume Controller supports differential and capacity-based licensing. For virtualization and compression functions, differential licensing charges different rates for different types of storage, which provides cost-effective management of capacity across multiple tiers of storage. Licensing for these functions is based on the number of Storage Capacity Units (SCUs) that are purchased. With other functions, such as remote mirroring and FlashCopy, the license grants a specific number of terabytes for that function.

Differential licensing is granted per SCU. Each SCU corresponds to a different amount of usable capacity that is based on the type of storage.

Differential licensing charges different rates for different types of virtualized storage, which provides cost-effective management of capacity across multiple tiers of storage. It is based on the number of storage capacity units (SCUs) that are purchased.

Each SCU corresponds to a different amount of usable capacity based on the type of storage.

The different storage types and the associated SCU ratios are listed in Table 4-1.

Table 4-1 SCU ratio per storage type

License	Drive classes	SCU ratio
SCM	Storage-Class Memory (SCM) devices	One SCU equates to 1 TiB of usable Category 1 storage.
Flash	All flash devices, other than SCM drives	One SCU equates to 1.18 TiB of usable Category 1 storage.
Enterprise	10 K or 15 K RPM drives	One SCU equates to 2.00 TiB of usable Category 2 storage.
Nearline (NL)	NL Serial Advanced Technology Attachment (SATA) drives	One SCU equates to 4.00 TiB of usable Category 3 storage.

License settings are initially entered in to a system initialization wizard. They can be changed later. For more information, see Licensed functions.

To view or configure the licensing settings, complete the following steps:

1. From the main Settings window, click **Settings** → **System**.
2. In the left column, click **Licensed Functions**. The example that is shown in Figure 4-125 shows the License Functions window of an IBM SAN Volume Controller system, which uses differential licensing for External Virtualization.

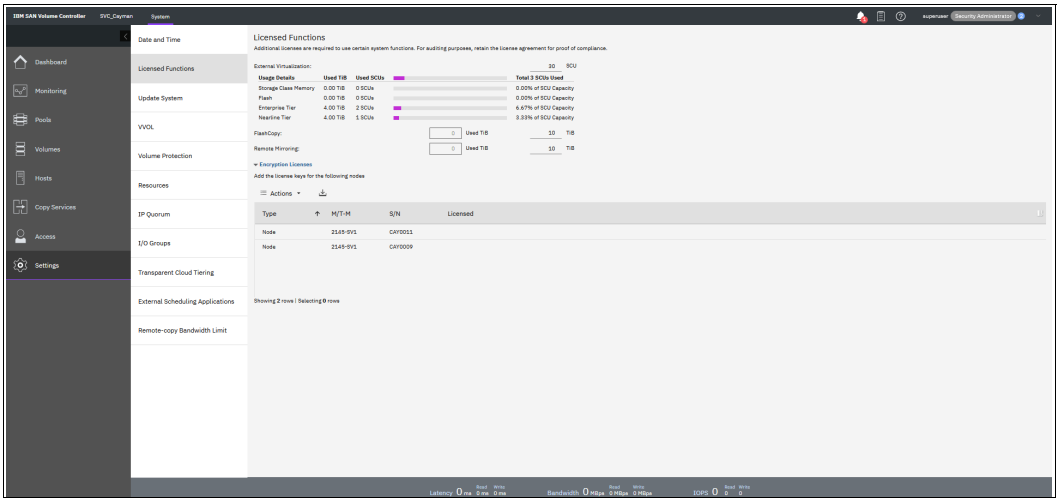


Figure 4-125 Licensing window

3. In the Licensed Functions window, you can view or set the licensing options for the IBM SAN Volume Controller or the IBM FlashSystem storage system for the elements that are described next.

External Virtualization

This value is the total number of licensed SCUs that you are licensed to virtualize across the tiers of storage on your system. The system supports external virtualization licensing for different tiers of storage.

A license can be purchased for a specific quantity of SCUs that can be used to virtualize a designated number of TiB of storage. You can monitor the used capacity for each tier to view the Used SCU value that indicates the amount of SCU that a tier used of the total number of licensed SCUs. The system converts that information into the Used TiB value.

You can monitor how the virtualization license is distributed across the following tiers of storage:

- Storage Class Memory

  Storage Class Memory storage uses persistent memory technologies that improve the endurance and speed of current flash storage device technologies. The Used SCU value indicates the amount of SCU that this tier used of the total number of licensed SCUs. The system converts that information into the Used TiB value.

- Flash

  All flash devices, other than SCM drives. The Used SCU value indicates the amount of SCU that this tier used of the total number of licensed SCUs. The system converts that information into the Used TiB value.

- Enterprise Tier

  The enterprise tier contains Serial Attached SCSI (SAS) drives and Fibre Channel drives in attached external storage. The Used SCU value indicates the amount of SCU that this tier used of the total number of licensed SCUs. The system converts that information into the Used TiB value.

- Nearline Tier

  The nearline tier contains nearline SAS drives in attached external storage. The Used SCU value indicates the amount of SCU that this tier used of the total number of licensed SCUs. The system converts that information into the Used TiB value.

FlashCopy

If required on the platform, the FlashCopy function copies the contents of a source volume to a target volume. It is also used to create cloud snapshots of volumes in systems that include TCT enabled.

FlashCopy can be licensed in terabytes (TB). In this case, the used capacity for FlashCopy mappings is the sum of all of the volumes that are the source volumes of a FlashCopy mapping and volumes with cloud snapshots.

If licensed in enclosures, FlashCopy can be used on a total number of internal enclosures and virtualized (external) enclosures.

Remote mirroring

The remote mirroring function configures a relationship between two volumes. This function mirrors updates that are made to one volume to another volume. The volumes can be in the same system or on two different systems.

If a remote mirroring function is licensed per enclosure, you can use the remote mirroring functions on the total number of enclosures that are licensed. The total number of enclosures must include the enclosures on external storage systems that are licensed for virtualization and the number of control and expansion enclosures that are part of your local system.

If licensed by capacity, the function specifies the amount of data that can be replicated. The used capacity for remote mirroring is the sum of the capacities of all the volumes that are in an MM or GM relationship. Master *and* auxiliary volumes are counted.

The license settings apply only to the system on which you are configuring license settings. For RC partnerships (includes HyperSwap), a license is also required on any remote systems that are in the partnership.

Compression

If required on the platform, data is compressed as it is written to disk with the compression function, which saves extra capacity for the system. A compression license can be purchased for a specific quantity of SCUs, which can be divided among different tiers of storage.

SCU usage for compression is calculated in the same way as for an External Virtualization license. However, *used* capacity (and not *usable* capacity) is calculated for a Compression License.

For example, if your compressed data occupies 4 TB in an NL tier, you need one SCU that is licensed for compression independently of the total compressed virtual disk (VDisk) capacity.

Note: Only IBM Real-time Compression (RtC) for volumes in standard pools is accounted for on a Compression License. Compressed volumes in Data Reduction Pools (DRPs) are *not* accounted for by this license.

Enter the total number of SCUs you are licensed for by using the compression function on the system. With the compression function, data is compressed as it is written to disk, which saves extra capacity for the system.

A compression license can be purchased for a specific quantity of SCU, which can be divided among different tiers of storage. The used capacity for each tier can be monitored to learn how the compression license is distributed across the tiers of storage.

Encryption license

In addition to these enclosure-based licensed functions, the system also supports encryption through a key-based license. Key-based licensing requires an authorization code to activate encryption on the system. Only specific models of the control enclosures support encryption.

During initial setup, you can select to activate the license with the authorization code. The authorization code is sent with the licensed function authorization documents that you receive after purchasing the license. These documents contain the authorization codes that are required to obtain keys for the encryption function that you purchased for your system.

Encryption is activated on a per system basis, and an active license is required for each control enclosure that uses encryption. During system setup, the system detects any SAS attached enclosures and applies the license to these enclosures. If control enclosures are added and require encryption, more encryption licenses must be purchased and activated.

Key-based licensing

For systems that support key-based licensing, an authorization code is used to activate licensed functions on the system. The authorization code is sent with the licensed function authorization documents that you receive after the license is purchased. For each license that you purchase, a separate document with an authorization code is sent to you. An active license is required for each control enclosure that uses the function. If your system supports more than one control enclosure, a license is required for each control enclosure. Several licensed functions have 90-day trial licenses available. You can use the trial licenses to

determine whether the function works as expected, create a business justification for licensed function, or test the benefit in your actual environment. If you use a trial license, the system warns you when the trial is about to expire at regular intervals. If you do not purchase and activate the license on the system before the trial license expires, all configuration that uses the trial licenses is suspended. You can manually remove any configuration that is related to this function or select **Monitoring > Events** in the management GUI. You can run the fix procedure that is related to the event that is generated when the trial license expires. During system setup, the system detects any SAS attached enclosures and applies the license to these enclosures. If additional key-based licenses are acquired after system setup is completed, you can use the Licensed Function page to activate these functions.

The following key-based licensed functions are supported:

- FlashCopy upgrade
- Easy Tier
- Remote Mirroring
- Encryption

Note: To monitor license usage, run the `lslicense` CLI command. For more information, see Licensing and featurization commands.

Updating storage system

Use the Update System panel to install new versions of the software.

Note: Before you start a system update, ensure that there are no problems on the system that might interfere with a successful update of the system.

Use the management GUI to automatically update the system. You can manually update the system. However, the automatic update is the preferred method for updating the system. If you are updating the system from any version to the most recent version, ensure that you follow the guidelines that are available from your support center. System updates can only be completed in a specific order or the update will not be successful.

After you begin the update, select **Dashboard** → **Hardware Components** to display the status of the nodes within the system. The **Hardware Component** tile displays one node as offline as the upgrade cycles through each node in the system. The system remains online during the update process.

If you have transferred the package, click **Check for files** to test and update. If the files for an update are not on the system, use one of the following options to provide a package:

- Use an existing package.
- Use a different package.

Using an existing package

An existing software package is a package that was previously transferred or uploaded to the system. The following actions are available for existing packages:

- Test only
- Test and Update
- Install Patch

To use an existing package, Select **Settings** → **System** → **Update System**.

To test the package without installing it, click **Test only**. On the Run Update Test Utility page, select the **Test utility and Code level**. Click **Test**.

To test and update the package, click **Test and update** and use the following procedure:

1. On the System Update page, select the **Test utility**, **Update package**, and **Code level**:
2. Click **Next**.
3. Select **Automatic update** or **Service Assistant Manual update** based on your requirement.
4. For **Automatic update**, click **Next**. For **Service Assistant Manual update**, click **Finish**.
5. To create intermittent pauses in the update so that you can verify the process, select one of the following options:
   - **Fully automatic**
   - **Pause update after half of the nodes are updated**
   - **Pause update before each node updates**
6. Click **Finish**.

To install the patch file, click **Install patch**. The Install patch action is disabled if a patch file is not uploaded. Select the patch file from the system. Click **Install**. The patch installation process starts, and a task progress dialog displays the progress.

Using a different package

A different software package is a package that needs to be transferred or uploaded to the system.

To obtain the package, use one of the following methods:

- Obtain the package directly.
- Provide the package manually.

To obtain the package directly, enable Call Home with cloud services to connect to support, and directly transfer the latest update available to the storage system. The details are in the following steps:

1. Select **Settings** → **System** → **Update System**.
2. If Transfer is disabled, click Call Home to navigate to the access panel and enable Call Home with Cloud Services. For more information, see Call Home with Cloud Service.
3. On the Update System page, click **Transfer**.
4. On the Transfer Update Package page. Enter your IBMid and password along with either **Version ID** or **Fix ID**. Fix ID can be provided with access key, if one is required. For example, Storage_Disk-2076-8.4.0.6-ifix4:172413436040050112.
5. Click **Transfer**, a progress bar is displayed to provide the status of the transferring package. If you need to cancel the transfer, click **Cancel**.
6. After the transfer is complete, click **Check for files**.
   - If the transfer is successful, then the transferred package appears in Use an existing package.
   - If the transfer is unsuccessful, then **No files found** is displayed.
7. To test and update, go to Use an existing package.

Use this option to upload the update package or patch file to the storage system manually. To provide the package manually, complete these steps:

1. Select **Settings** → **System** → **Update System**.
2. Click **Upload**.
3. On the Upload page, drag the files or click **Select files**. Select the test utility, update package, or patch files that you want to update.
4. Click **Upload**.

5. Monitor the update information in the management GUI to determine when the process completes.

   Figure 4-126 shows you the Test and Test and Upgrade the System options.

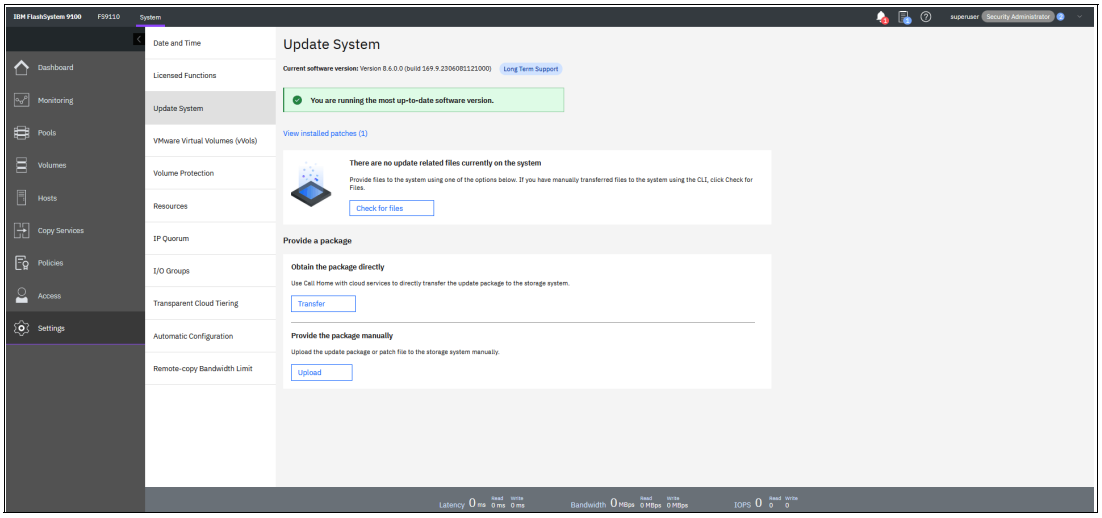


Figure 4-126 Update System

For more information about the update procedure that uses the GUI, see Chapter 11, “Reliability, availability, and serviceability” on page 963.

VMware vSphere virtual volumes

IBM Storage Virtualize can manage VMware vSphere virtual volumes (VVOLs) directly in cooperation with VMware. It enables VMware virtual machines (VMs) to get the assigned disk capacity directly rather than from the Elastic Sky X Integrated (ESXi) data store. This technique enables storage administrators to control the suitable use of storage capacity. It Also enables enhanced features of storage virtualization directly to the VM (such as replication, thin-provisioning, compression, and encryption).

VVOL management is enabled in the System section of the GUI, as shown in Figure 4-127. The NTP server must be configured before enabling VVOL management. As a best practice, use the same NTP server for ESXi and your system.

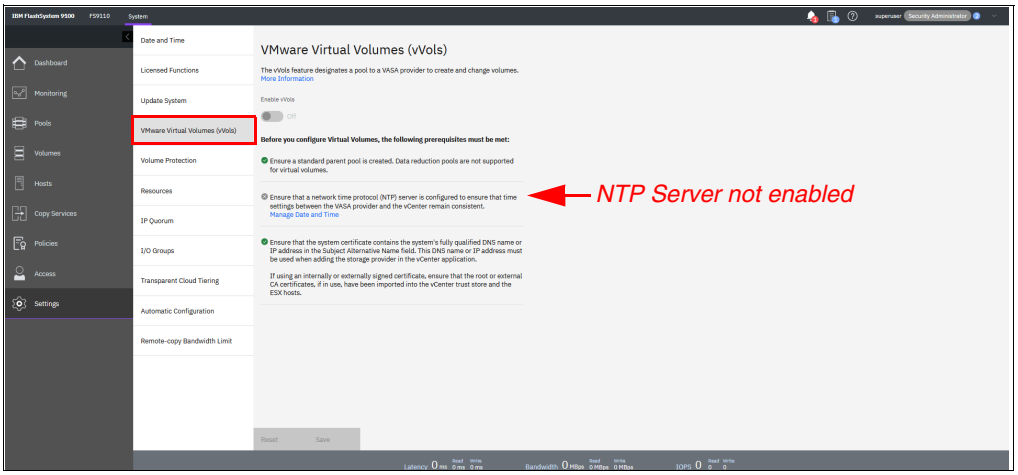


Figure 4-127 Enabling VVOLs management

Restriction: You cannot enable VVOL support until the NTP server is configured in the IBM SAN Volume Controller or the IBM FlashSystem.

For more information about VVOLs, see *IBM Storage Virtualize and VMware: Integrations, Implementation and Best Practices*, SG24-8549.

Volume protection

Volume protection prevents active volumes or host mappings from being deleted inadvertently if the system detects recent I/O activity.

Note: This global setting is enabled by default on new systems. You can set this value to apply to all volumes that are configured on your system, or control whether the system-level volume protection is enabled or disabled on specific pools.

To prevent an active volume from being deleted unintentionally, administrators can use the system-wide setting to enable volume protection. They can also specify a period that the volume must be idle before it can be deleted.

If volume protection is enabled and the period is not expired, the volume deletion fails, even if the **`-force`** parameter is used.

Note: The system-wide volume protection and the pool-level protection both must be enabled for protection to be active on a pool. The pool-level protection depends on the system-level setting to ensure that protection is applied consistently for volumes within that pool.

If system-level protection is enabled but pool-level protection is not enabled, any volumes in the pool can be deleted (even when the setting is configured at the system level).

When you delete a volume, the system verifies whether it is a part of a host mapping, FlashCopy mapping, or an RC relationship. For a volume that contains these dependencies, the volume cannot be deleted unless the **`-force`** parameter is specified on the corresponding remove commands.

However, the **`-force`** parameter does not delete a volume if it includes recent I/O activity and volume protection is enabled. The **`-force`** parameter overrides the volume dependencies, *not* the volume protection setting.

The following actions are affected by this setting:

- Deleting a:
  - Volume
  - Volume copy
  - Host or a host cluster mapping
  - Storage pool
  - Host from an I/O group
  - Host or host cluster
  - Defined host port
- Creating an RC relationship

Figure 4-128 shows the Volume Protection window (Volume Protection is enabled from this window). The actions that are described in “Volume protection” on page 328 are affected by this setting. For example, Volume Protection prevents active volumes or host mappings from being deleted inadvertently if the system detects recent I/O activity.

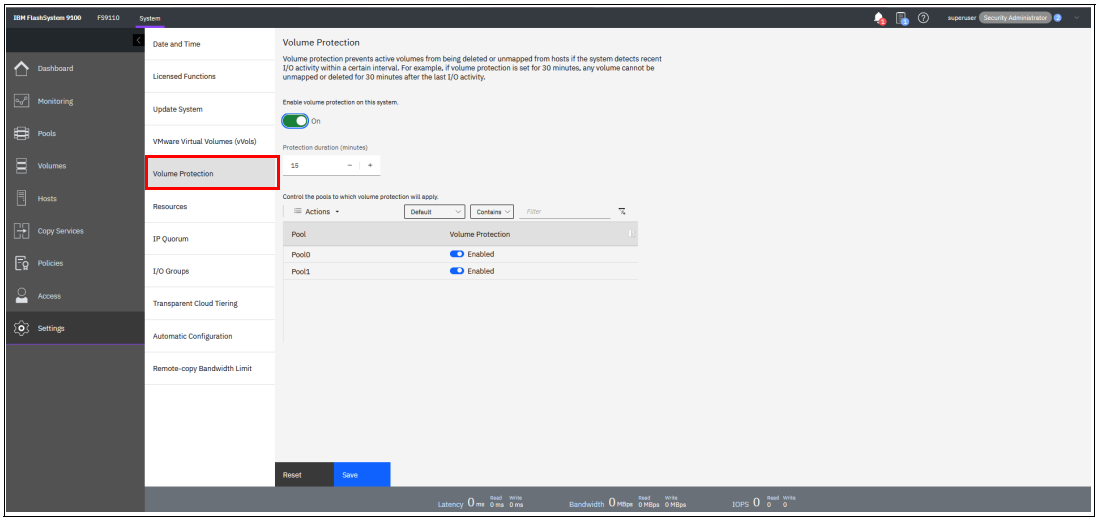


Figure 4-128 Volume Protection window

Resources

Use this option to change the memory limits for the Copy Services and Redundant Array of Independent Disks (RAID) functions for an I/O group.

Copy Services features and RAID require that small amounts of volume cache be converted from cache memory into bitmap memory to enable the functions to operate. If you do not have enough bitmap space that is allocated when you attempt to use one of the functions, you cannot complete the configuration.

The total memory that can be dedicated to these functions is not defined by the physical memory in the system. The memory is constrained by the software functions that use the memory.

In planning the installation for a system, consider the future requirements for the advanced functions.

Before you specify the configuration changes, consider the following factors:

- For FlashCopy mappings, only one I/O group uses bitmap space. By default, the I/O group of the source volume is used.
- For Metro Mirror, Global Mirror, and HyperSwap active-active relationships, two bitmaps exist. For Metro Mirror or Global Mirror relationships, one is used for the master system and one is used for the auxiliary system because the direction of the relationship can be reversed.

  For active-active relationships, which are configured automatically when HyperSwap volumes are created, one bitmap is used for the volume copy on each site because the direction of these relationships can be reversed.
- When you create a reverse mapping (for example, to run a restore operation from a snapshot to its source volume), a bitmap also is created for this reverse mapping.
- When you configure change volumes for use with Global Mirror or Metro Mirror, two internal FlashCopy mappings are created for each change volume.

- The smallest possible bitmap is 4 KiB; therefore, a 512-byte volume requires 4 KiB of bitmap space.

On existing systems, also consider the following factors:

- When you create FlashCopy mappings and mirrored volumes, HyperSwap volumes, or formatted, standard-provisioned volumes, the system attempts to automatically increase the available bitmap space. You do not need to manually increase this space.
- Metro Mirror and Global Mirror relationships do not automatically increase the available bitmap space. You might need to use the **`chiogrp`** command or the management GUI to manually increase the space in one or both of the master and auxiliary systems.

To use the Resource option, select **Settings** → **System** → **Resources** (see Figure 4-129).

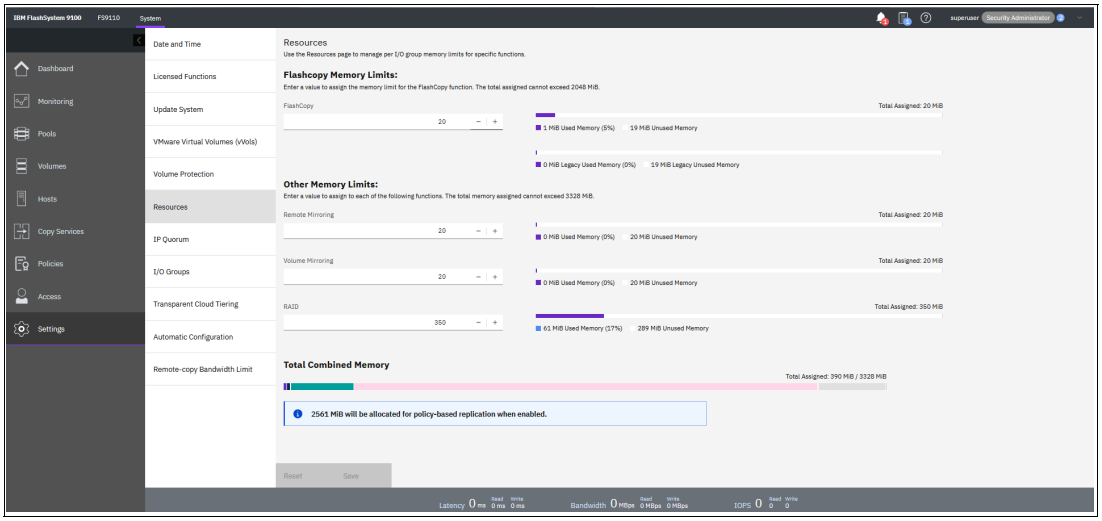


Figure 4-129 Resources allocation

Memory limits can be changed by using the Resources page.

Table 4-2 lists the amount of memory that is required for remote mirroring functions, FlashCopy functions, and volume mirroring.

Table 4-2 Examples of allocation of bitmap memory

Function	Grain size	Provisioned capacity for the specified I/O group that is provided by 1 MIB of memory
Remote Copy	256 KiB	2 TiB of total MM, GM, or HyperSwap provisioned capacity
FlashCopy	256 KiB	2 TiB of total FlashCopy source-provisioned capacity
FlashCopy	64 KiB	512 GiB of total FlashCopy source-provisioned capacity
Incremental FlashCopy	256 KiB	1 TiB of incremental FlashCopy source-provisioned capacity
Incremental FlashCopy	64 KiB	256 GiB of incremental FlashCopy source-provisioned capacity
Volume Mirroring	256 KiB	2 TiB of mirrored provisioned capacity

Notes: Consider the following points:

- For multiple FlashCopy targets, you must consider the number of mappings. For example, for a mapping with a grain size of 256 KiB, 8 KiB of memory allows one mapping between a 16 GiB source volume and a 16 GiB target volume.

  Alternatively, for a mapping with a 256 KiB grain size, 8 KiB of memory allows two mappings between one 8 GiB source volume and two 8 GiB target volumes.
- If you specify an I/O group other than the I/O group of the source volume when a FlashCopy mapping is created, the memory accounting goes toward the specified I/O group, *not* toward the I/O group of the source volume.
- For volume mirroring, the full 512 MiB of memory space enables 1 PiB of total provisioned capacity.
- When creating FlashCopy relationships or mirrored volumes, more bitmap space is automatically allocated by the system, if required.

IP quorum

IBM Storage Virtualize also supports an IP quorum application. By using an IP-based quorum application as the quorum device for the third site, a FC-based system is *not* required. Java applications run on hosts at the third site.

To install the IP quorum device, complete the following steps:

1. If your IBM FlashSystem storage system is configured for IPv4, click **Download IPv4 Application**. If it is configured for IPv6, select **Download IPv6 Application**.

   In our example, IPv4 is the option, as shown in Figure 4-130.

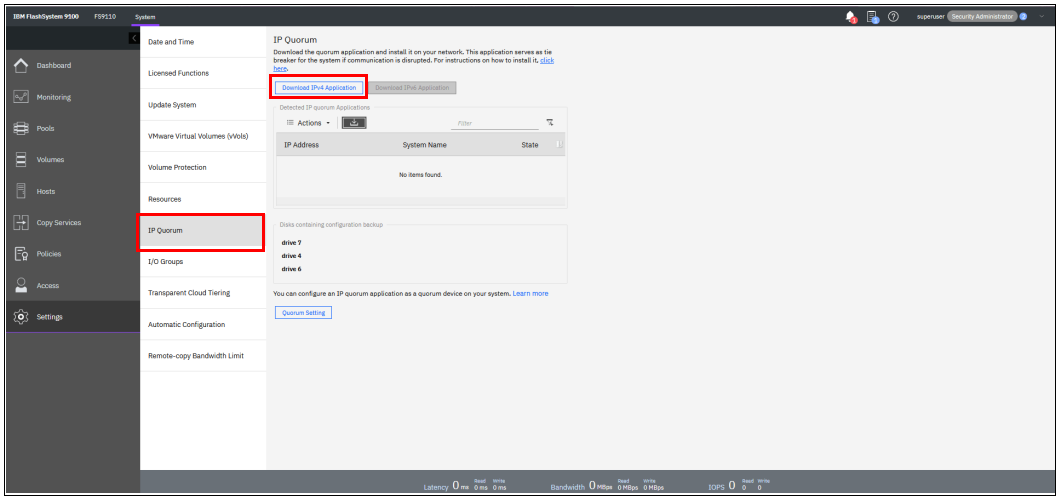


Figure 4-130 IP Quorum settings

2. When you select **Download IPv4 Application**, you are prompted whether you want to download the IP quorum application with or without recovery metadata, as shown in Figure 4-131. IP quorum applications are used to resolve communication problems between nodes and store metadata, which restores system configuration during failure scenarios. If you have a third-site quorum disk that stores recovery metadata, you can download the IP quorum application without the recovery metadata.

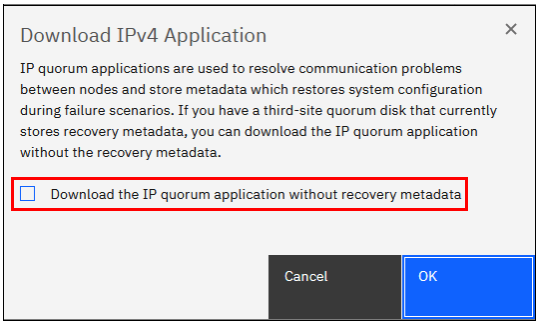


Figure 4-131 IP quorum application metadata

3. After you select your correct IP configuration, IBM Storage Virtualize generates an IP Quorum Java application, as shown in Figure 4-132. The application can be saved and installed in a host that is to run the IP quorum application.

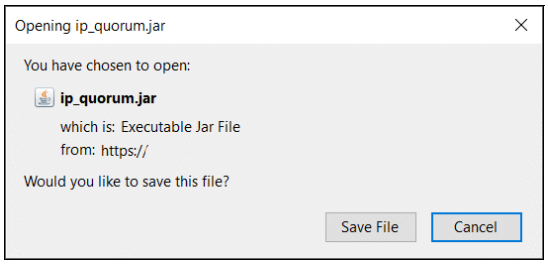


Figure 4-132 IP Quorum Java application

After you download the IP quorum application, save the application on a separate host or server.

If you change the configuration by adding a node, changing a service IP address, or changing Secure Sockets Layer (SSL) certificates, download and install the IP quorum application again.

4. On the host, use the Java command line to initialize the IP quorum application. On the server or host on which you plan to run the IP quorum application, create a separate directory that is dedicated to the IP quorum application.
5. Run the **`ping`** command on the host server to verify that it can establish a connection with the service IP address of each node in the system.
6. Change to the folder where the application is, and run the following command:

```
java -jar ip_quorum.jar
```

Note: The IP quorum application must always be running.

7. To verify that the IP quorum application is installed and active, select **Settings** → **System** → **IP Quorum**. The new IP quorum application is displayed in the table of detected applications. The system automatically selects MDisks for quorum disks.

An IP quorum application can also act as the quorum device for systems that are configured with a single-site or standard topology that does not have any external storage configured.

The IP quorum mode is set to `Standard` when the system is configured for standard topology. The quorum mode of `Preferred` or `Winner` is available only if the system topology is not set to `Standard`.

To change the quorum mode for the IP quorum application, select **Settings** → **System** → **IP Quorum** and set the mode to `Standard`, `Preferred`, or `Winner`, or run the **`chsystem`** command. This configuration provides a system tie-break capability, which automatically resumes I/O processing if half of the system's nodes or enclosures are inaccessible.

Specific quorum settings are shown in Figure 4-133.

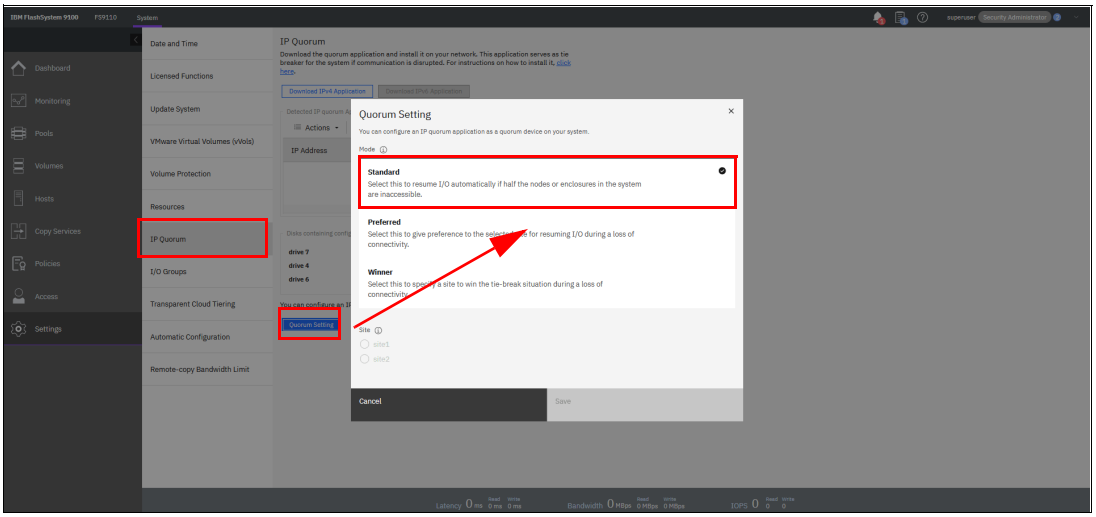


Figure 4-133 Quorum settings

On systems that support multiple-site topologies, you can specify which site resumes I/O after a disruption that is based on the applications that run on the site or other factors, such as whether the environment uses a third site for quorum management.

For example, you can specify whether a selected site is preferred for resuming I/O or if the site automatically "wins" in tie-break scenarios. If only one site runs critical applications, you can configure this site as preferred.

During a disruption, the system delays processing tie-break operations on other sites that are not specified as preferred. The designated preferred site becomes more apt to resume I/O, and critical applications remain online. If the preferred site is the site that is disrupted, the other site continues to win the tie-breaks and continue I/O.

This feature applies only to IP quorum applications. It does *not* apply to FC-based third-site quorum management. In stretched configurations or HyperSwap configurations, an IP quorum application can be used at the third site as an alternative to third-site quorum disks.

No FC connectivity at the third site is required to use an IP quorum application as the quorum device. If you have a third-site quorum disk, you must remove the third site before you use an IP quorum application.

Note: A maximum of five IP quorum applications can be deployed on a single system. Only one instance of the IP quorum application per host or server is supported.

IP quorum applications on multiple hosts or servers can be configured to provide redundancy. If multiple FlashSystem storage systems are in your environment, more than one IP quorum application is allowed per host. However, each IP quorum instance must be dedicated to a single system within the environment.

In addition, the host or server requires available bandwidth to support multiple IP quorum instances. Use the network requirements that are described next to determine bandwidth and latency needs in these types of environments. The recommended configuration remains a single IP quorum application per host or server.

Use the network requirements that are described in “I/O groups” on page 334 to determine bandwidth and latency needs in these types of environments. The recommended configuration remains a single IP quorum application per host or server.

For stable quorum resolutions, an IP network must provide the following requirements:

- Connectivity from the servers that are running an IP quorum application to the service IP addresses of all nodes or node canisters. The network must also deal with possible security implications of exposing the service IP addresses because this connectivity can also be used to access the service assistant interface if the IP network security is configured incorrectly.
- On each server that runs an IP quorum application, ensure that only authorized users can access the directory that contains the IP quorum application. For systems that support metadata for system recovery, metadata is stored in the directory in a readable format; Therefore, ensure access to the IP quorum application and the metadata is restricted to authorized users only.
- Port 1260 is used by the IP quorum application to communicate from the hosts to all nodes or enclosures.
- The maximum round-trip delay must not exceed 80 milliseconds (ms); that is, 40 ms each direction.
- If you are configuring the IP quorum application without a quorum disk for metadata, a minimum bandwidth of 2 MBps must be guaranteed for traffic between the system and the quorum application. However, a minimum bandwidth of 64 MBps between the quorum application and the system is required to support the IP quorum application with quorum disk for metadata.
- If your system supports an IP quorum application with metadata for system recovery, ensure that the directory contains at least 125 MB of available capacity.

I/O groups

For ports within an I/O group, you can enable virtualization of FC ports that are used for host I/O operations. With N_Port ID Virtualization (NPIV), the FC port consists of a physical port and a virtual port.

When port virtualization is enabled, ports do not initialize until they are ready to handle I/O, which improves host behavior. In addition, path failures that occur because of an offline node are masked from hosts.

The following target port mode on the I/O group indicates the current state of port virtualization:

- Enabled: The I/O group contains virtual ports that are available to use. For new systems, the default status for NPIV is set to `Enabled`. However, if you are adding a node to an existing system, you should verify this setting
- Disabled: The I/O group does not contain any virtualized ports.
- Transitional: The I/O group contains physical FC and virtual ports that are being used. You cannot change the target port mode directly from Enabled to Disabled states, or vice versa. The target port mode must be in a transitional state before it can be changed to Disabled or Enabled states.

  The system can be in the transitional state for an indefinite period while the system configuration is changed. However, system performance can be affected because the number of paths from the system to the host doubled. To avoid increasing the number of paths substantially, use zoning or other means to temporarily remove some of the paths until the state of the target port mode is enabled.

The port virtualization settings of I/O groups are available by selecting **Settings** → **System** → **I/O Groups**, as shown in Figure 4-134.

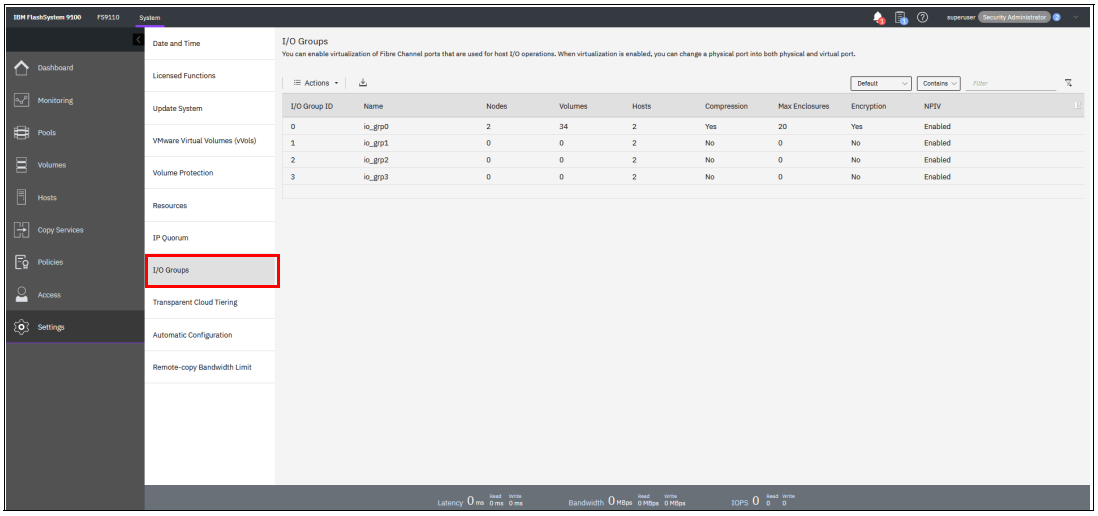


Figure 4-134 I/O group port virtualization

Verifying port virtualization on a new system

New systems have NPIV enabled as default status. However, you must accurately zone your new system for NPIV configuration. For more information on zoning your system, see the **Configuring** → **Configuration details** → **Zoning details** topic in your product's documentation. Verify that the NPIV setting on the new system is set to `Enabled` by completing these steps:

1. In the management GUI, select **Settings** → **System** → **I/O Groups**.
2. On the **I/O Groups** page, verify that the value for the NPIV is set to **`Enabled`**.

Enabling port virtualization on a system

If you are changing the NPIV for an I/O group in a system, complete the following prerequisite steps:

1. Review your SAN fabric layout and zoning rules because NPIV includes stricter requirements. Ensure that equivalent ports are on the same fabric and in the same zone. For more information, see the Zoning Details topic in your product's documentation.

2. Check the path count between your hosts and the system. Make sure that the number of paths is half of the usual supported maximum. For general guidelines about zoning and NPIV, see the Zoning Details topic in your product's documentation.
3. Ensure that Fibre Channel switches permit each physically connected system port the ability to create two other NPIV ports.

After you complete these prerequisite steps, you can enable port virtualization on a system by completing the following steps:

1. In the management GUI, select **Settings** → **System** → **I/O Groups**.
2. Verify that the value for the NPIV is set to `Enabled`. For new systems, the default status for NPIV is set to `Enabled`. However, if you are adding a node to a system, verify this setting. If it is set to **Transitional** or **Disabled**, it must be changed to **Enabled**.
3. If **NPIV** is `Disabled`, complete the following steps to change it to `Enabled`:
   a. Right-click the I/O group and select **Change NPIV Settings**.
   b. On the **Change NPIV Settings** page, select **Transitional** for the new state. Users cannot go directly from `Disabled` to `Enabled` state. The system must be in a transitional state where the I/O group contains both physical and virtual Fibre Channel ports.

      Click **Continue**.
   c. Wait approximately 2 minutes before you verify the changed state for the target ports.
   d. Verify that the new **Transitional** state is displayed on the I/O Groups page.
   e. Right-click the I/O group and select **Change NPIV Settings**.
   f. On the Change NPIV Settings page, select **Enabled** for the new state. Click **Continue**.
   g. After 2 minutes, verify that the new **Enabled** state is displayed on the I/O Groups page.
4. If **NPIV** is already in the **Transitional** state, complete the following steps to change the state to `Enabled`:
   a. Right-click the I/O group and select **Change NPIV Settings**.
   b. On the Change NPIV Settings page, select **Enabled** for the new state. Click **Continue**.
   c. Wait approximately 2 minutes before you verify that the new `Enabled` state is displayed on the I/O Groups page.

You can change the status of the port by selecting an I/O Group ID and right-click the I/O group and selecting **Change NPIV Settings**, as shown in Figure 4-135.

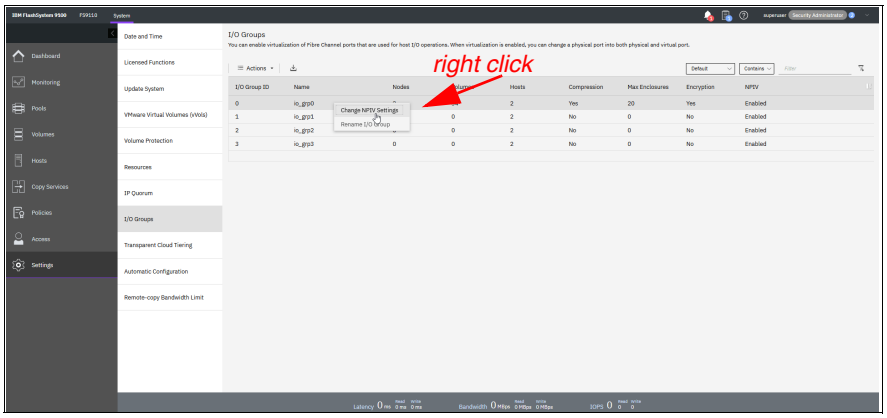


Figure 4-135 Changing NPIV settings

Transparent Cloud Tiering

Transparent Cloud Tiering (TCT) is a licensed function that enables volume data to be copied and transferred to cloud storage. The system supports creating connections to cloud service providers (CSPs) to store copies of volume data in private or public cloud storage.

With TCT, administrators can move older data to cloud storage to free up capacity on the system. PiT snapshots of data can be created on the system and then copied and stored on the cloud storage. An external CSP manages the cloud storage, which reduces storage costs for the system. Before data can be copied to cloud storage, a connection to the CSP must be created from the system.

A cloud account is an object on the system that represents a connection to a CSP by using a specific set of credentials. These credentials differ depending on the type of CSP that is being specified. Most CSPs require the hostname of the CSP and an associated password. Some CSPs also require certificates to authenticate users of the cloud storage.

Public clouds use certificates that are signed by well-known CAs. Private CSPs can use a self-signed certificate or a certificate that is signed by a trusted CA. These credentials are defined on the CSP and passed to the system through the administrators of the CSP. A cloud account defines whether the system can successfully communicate and authenticate with the CSP by using the account credentials.

If the system is authenticated, it can access cloud storage to copy data to the cloud storage or restore data that is copied to cloud storage back to the system. The system supports one cloud account to a single CSP. Migration between providers is *not* supported.

Note: Before enabling TCT, consider the following requirements:

- Ensure that the DNS is configured on your system and accessible.
- Determine whether your company's security policies require enabled encryption. If yes, ensure that the encryption licenses are correctly installed and that encryption is enabled.

Each cloud service provider divides cloud storage into segments for each client that uses the cloud storage. These objects store only data that is specific to that client.

The names of the objects begin with a prefix that you can specify when you create the account for the system. A prefix defines system-specific content that the object stores and supports multiple independent systems to store data to a single cloud account. Each cloud service provider uses different terminology for these storage objects.

Before you create the cloud account, complete the following prerequisites:

- Verify that your hardware model supports this function before proceeding.
- Ensure that a DNS server is configured on the system. During the configuration of the cloud account, the wizard prompts you to create a DNS server if one is not configured.

  On systems that support transparent cloud tiering, at least one DNS server is required if you connect to cloud service providers as part of transparent cloud tiering support, which included establishing a cloud account and connecting to cloud-based storage.

  Before you create a connection to a cloud service provider or connect to cloud storage, ensure that you specify at least one DNS server to manage hostnames. You can have up to two DNS servers that are configured on the system.

  To configure DNS for the system, go to **Settings** → **Network** → **DNS** and enter a valid IP address and name for each server. IPv4 and IPv6 address formats are supported.

- Determine whether encryption is required for your connection to the cloud account. If you are accessing a public cloud solution, encryption protects data during transfers to the external cloud service providers from attack.

  To encrypt data that is sent to the cloud service provider, verify that your system supports encryption and that it is enabled on the system. Some models might require more encryption licenses. Verify these requirements before this function is used.

  For more information about security considerations for encryption and cloud accounts, see this Planning for transparent cloud tiering.

If you configured a cloud account, use the Transparent cloud tiering page to monitor the status and data usage for the account. Use the status values for the cloud account to monitor and troubleshoot connection disruptions between the cloud service provider and the system. You can also display the amount of cloud storage that remains available for cloud snapshots and restore operations.

The system supports connections to various CSPs. Some CSPs require connections over external networks, and others can be created on a private network.

Each CSP requires different configuration options. The system supports the following CSPs:

- IBM Cloud

  The system can connect to IBM Cloud, which is a cloud computing platform that combines platform as a service (PaaS) with infrastructure as a service (IaaS).

- OpenStack Swift

  OpenStack Swift is a standard cloud computing architecture from which administrators can manage storage and networking resources in a single private cloud environment. Standard APIs can be used to build customizable solutions for a private cloud solution.

- Amazon Simple Storage Service (Amazon S3)

  Amazon S3 provides programmers and storage administrators with flexible and secure public cloud storage. Amazon S3 is also based on Object Storage standards and provides a web-based interface to manage, back up, and restore data over the web.

- Microsoft Azure

  Microsoft Azure is an object storage solution that allows large volumes of unstructured data to be stored in the Azure public cloud. It provides an API and web-based interface to help you manage your cloud solution. Microsoft Azure is suitable for archiving, disaster recovery, and backup data storage.

  To view or enable your IBM Storage Virtualize cloud provider settings, from the Settings window, click **Settings** → **System**. Then select **Transparent Cloud Tiering**, as shown in Figure 4-136 on page 339.

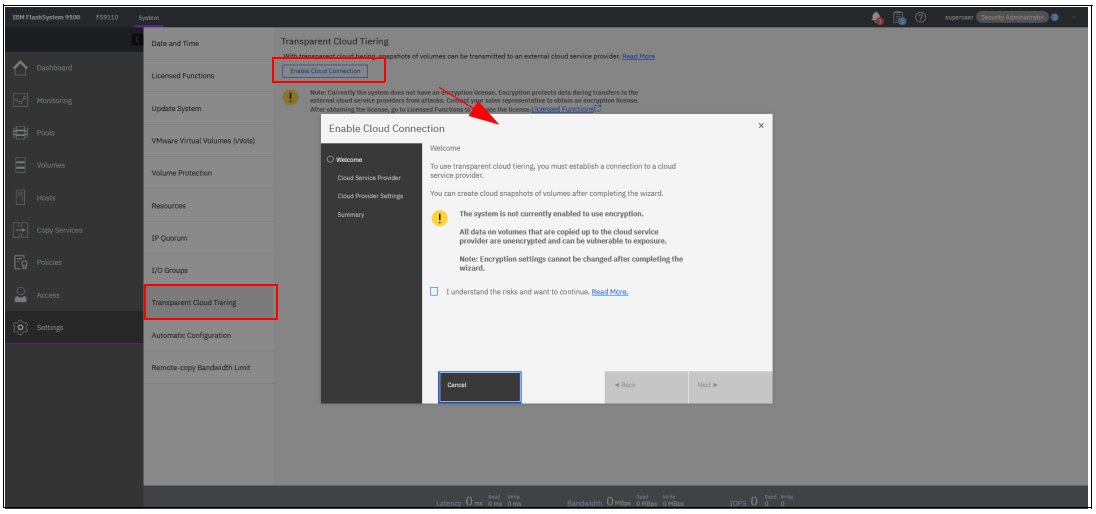


Figure 4-136 Transparent Cloud Tiering settings

By using this view, you can enable and disable features of your TCT and update the system information concerning your CSP. This window allows you to set the following options:

- CSP
- Cloud Object Storage Uniform Resource Locator (URL)
- The tenant or the container information that is associated to your Cloud Object Storage
- Username of the cloud object account
- API key
- The container prefix or location of your object
- Encryption
- Bandwidth

For more information about how to configure and enable TCT, see Chapter 10, “Advanced Copy Services” on page 717.

Automatic configuration

With automatic configuration, you can configure IBM FlashSystem products as external storage to an IBM SAN Volume Controller system. Automatic configuration implements optimal virtualization with an IBM SAN Volume Controller system based on best practices. For these supported systems, the process can be completed in the management GUI during system setup or later as part of storage configuration.

The automatic configuration process is intended for new systems. If you want to virtualize this system by using an IBM SAN Volume Controller system, no other objects (such as volumes or pools) can be configured on the system.

Before the wizard completes automatic configuration, you are prompted to complete the following prerequisite tasks:

- Add enclosures

  If you have any control or expansion enclosures to include as part of the external storage to be virtualized, you can add them. If you do not have more enclosures to add, this part can be skipped.

  If you do have enclosures to add but they are not automatically detected by the management GUI, verify the cabling between the system and the enclosures.

  For more information, see the installation information that came with the system.

- Verify zoning on the IBM SAN Volume Controller system

  In Fibre Channel, zoning is the process of grouping multiple ports to form a virtual, private storage network. Ports that are members of a zone can communicate with each other, but are isolated from ports in other zones. Before automatic configuration can be completed, verify that the IBM SAN Volume Controller is zoned correctly as part of its SAN configuration.

- Define a host cluster to represent the IBM SAN Volume Controller

  As part of the prerequisite steps, you must create a host cluster that represents the IBM SAN Volume Controller system. Creating a host cluster simplifies the port management between the IBM SAN Volume Controller and the systems that are virtualized as part of the automatic configuration.

  In the host cluster that represents the IBM SAN Volume Controller, a host represents a node, and each host port represents a port on a node. The management GUI displays WWPNs that are associated with node ports.

After these prerequisite steps are completed, the automatic configuration process begins. During this process, the following actions are completed automatically:

- Formats drives. Drives remain offline until formatting completes. Do not proceed until all drives are formatted.
- Creates the suitable RAID arrays that are based on the technology type of the drives.
- Creates a pool for each array.
- Provisions all usable capacity in each pool to volumes that are based on best practices.
- Maps all volumes to the IBM SAN Volume Controller system for virtualization as MDisks.

After the process finishes, you are prompted to complete tasks on the IBM SAN Volume Controller system to begin the use of this system as external storage.

Figure 4-137 shows how to enable the **Automatic Configuration for Virtualization** option.

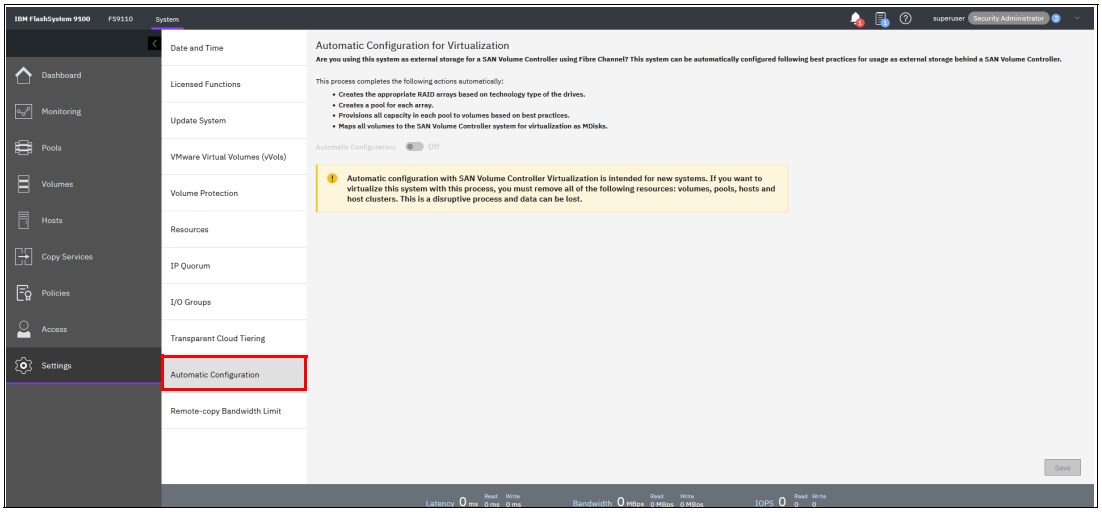


Figure 4-137 Automatic Configuration for Virtualization option

Remote-copy Bandwidth Limit

Use the Remote Copy Bandwidth Limit page to set a system-wide value for the maximum background copy bandwidth that any relationship uses during remote copy operations.

Remote copy bandwidth limit determines the rate at which the background copy for remote copy operation is attempted.

The background copy bandwidth can affect foreground I/O latency in one of the following ways:

- If the background copy bandwidth is set too high for the intersystem link capacity, the following results can occur:
  - The intersystem link cannot process the background copy I/Os fast enough, and the I/Os can back up (accumulate).
  - For synchronous remote-copy operations, a delay occurs in the synchronous secondary write operations of foreground I/Os.
  - For Global Mirror, the work is backlogged, which delays the processing of write operations and causes the relationship to stop. For Global Mirror in multiple-cycling mode, a backlog in the intersystem link can congest the local fabric and cause delays to data transfers.
  - The foreground I/O latency increases as detected by applications.
- If the background copy bandwidth is set too high for the storage at the primary site, background copy read I/Os overload the primary storage and delay foreground I/Os.
- If the background copy bandwidth is set too high for the storage at the secondary site, background copy write operations at the secondary overload the auxiliary storage. As a result, the synchronous secondary write operations of foreground I/Os are again delayed.

  For Global Mirror without cycling mode, the work is backlogged and again the relationship is stopped.

To set the background copy bandwidth optimally, you must consider all three resources (primary storage, intersystem link bandwidth, and auxiliary storage). Provision the most restrictive of these three resources between the background copy bandwidth and the peak foreground I/O workload.

Also consider concurrent host I/O. If other write operations arrive at the primary system for copy to the remote site, these write operations can be delayed by a high level of background copy. As a result, the hosts at the primary site receive poor write-operation response times.

The provisioning for optimal bandwidth for the background copy can also be calculated by determining how much background copy can be allowed before the performance of host I/O becomes unacceptable.

The background copy bandwidth can be decreased slightly to accommodate peaks in workload and provide a safety margin for host I/O. Changing the Bandwidth Limit is shown in Figure 4-138 on page 342.

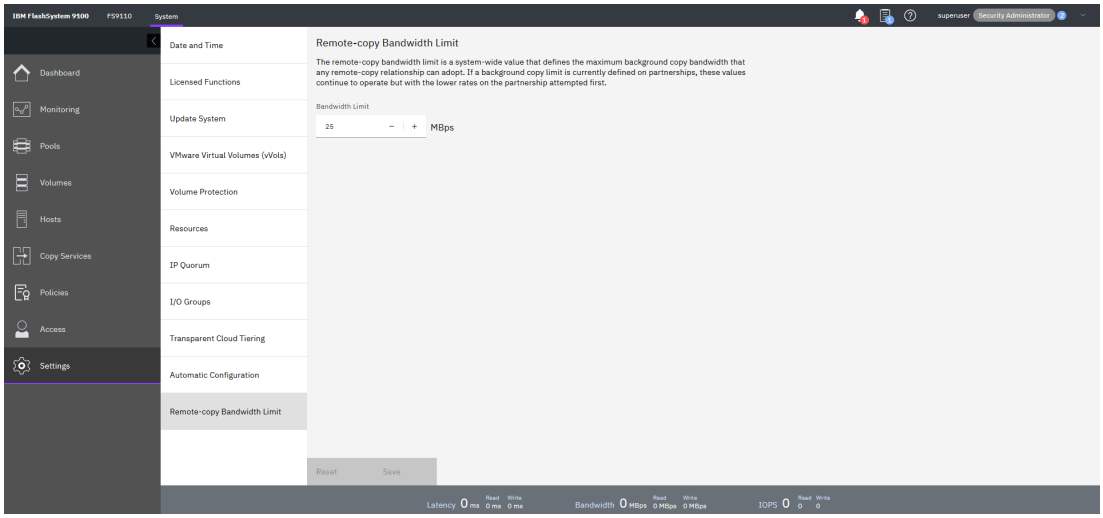


Figure 4-138 Remote-copy Bandwidth Limit

4.7.6 Support menu

Use the Support window to configure and manage connections and upload support packages to the IBM Support Center.

The following options are available:

- Call Home

  The Call Home feature transmits operational and event-related data to you and IBM through a Simple Mail Transfer Protocol (SMTP) server connection in the form of an event notification email. When configured, this function alerts IBM Support personnel about hardware failures and potentially serious configuration or environmental issues.

  This view provides the following useful information about email notification and Call Home information (among others), as shown in Figure 4-139 on page 343:

  - IP of the email server (SMTP server) and port.
  - Call Home email address.
  - Email of one or more users set to receive one or more email notifications.
  - Contact information of the person in the organization that is responsible for the system.

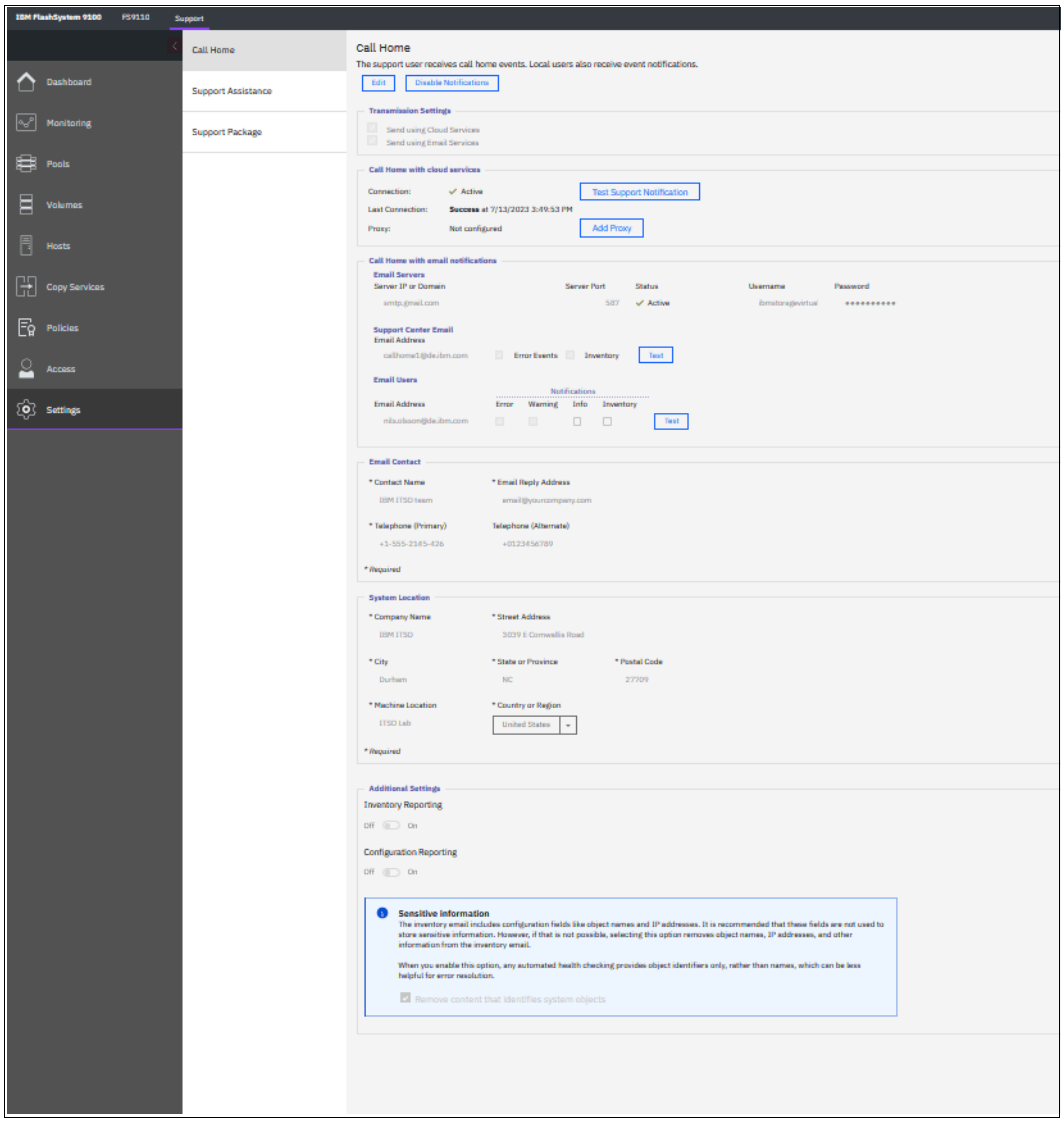


Figure 4-139 Call Home settings

Use the Call Home page to configure or change settings that are used to send notifications to the support center. Call Home connects the system to support personnel who can monitor and respond to system events to ensure that your system remains up and running. Some system models support only email notifications for transmitting information to support.

If a software or hardware error occurs, the Call Home function notifies the support center and then automatically opens a service request. Call Home sends service-related information to inform support personnel so they can quickly act to resolve the problem.

During system setup, you can configure Call Home notifications to improve the response time for issues on the system. Call Home notifications send diagnostic data to support personnel who can quickly determine solutions for these problems that can disrupt operations on the system.

If you want to update current Call Home notification settings, select **Settings** → **Support** → **Call Home**. Call Home is integrated with other support-related services, such as secure remote assistance and IBM Storage Insights.

In addition, inventory information that is related to the overall health of the system and its components can be sent to support at regular intervals. Support can analyze this inventory information to check the system health and recommend actions to improve system health and reliability.

If a severe issue is found, support personnel contact you directly to help you fix the problem. These recommendations are integrated in IBM Storage Insights on the Advisor page.

The system supports Call Home with cloud services and Call Home with email notifications. Call Home with cloud services uses Representational State Transfer (RESTful) APIs, which is the standard for transmitting data through web services.

For new system installations, Call Home with cloud services is configured as the default method to transmit notifications to support. When you update the system software, Call Home with cloud services is also set up automatically. Call Home with cloud services also eliminates email filters that drop notifications to and from support that can delay resolution of problems on the system. The system also supports both these transmission methods that are configured at the same time.

If both cloud services and email notifications are configured, Call Home with cloud services is the primary transmission method to the support center. Call Home with email notifications is used as a backup if the system cannot connect through cloud services to the support center. For sending monitoring information about IBM Storage Virtualize system to the IBM Storage Insights cloud service, Call Home with cloud services is used as the transmission method.

Important: To integrate with IBM Storage Insights, ensure that Call Home with cloud services is configured for the system.

Call Home with cloud services

Call Home with cloud services sends notifications directly to a centralized file repository that contains troubleshooting information that is gathered from customers. Support personnel can access this repository and be assigned issues automatically as problem reports. This method of transmitting notifications from the system to support removes the need for customers to create problem reports manually.

As part of Call Home with cloud services configuration, you can define an internal proxy server within your network to manage connections between the system and the support center. Other connection configurations are supported for Call Home with cloud services, which is described next.

Network considerations for Call Home with cloud services

If you use Call Home with cloud services to send notifications to the support center, you must ensure that the system can connect to the Support center.

Before you configure Call Home with cloud services, ensure that the following prerequisites are configured on the system:

- All of the nodes on the system feature Internet access.
- A valid service IP address is configured on each node on the system.

Call Home with cloud services supports the configurations that include the extra network requirements that are listed in Table 4-3.

Table 4-3 Supported network configurations for Call Home with cloud services

Supported configuration	DNS configuration	Firewall requirements
Call Home with cloud services with an internal proxy server	Required	Configure a firewall to allow outbound traffic on port 443 to esupport.ibm.com
Call Home with cloud services with a DNS server	Defined, but not required	Configure a firewall to allow outbound traffic on port 443 to esupport.ibm.com. Optionally allow outbound traffic on port 443 to the following IP addresses: ► 129.42.56.189 ► 129.42.54.189 ► 129.42.60.189
Call Home with cloud services	None	Configure a firewall to allow outbound traffic on port 443 to the following support IP addresses: ► 129.42.56.189 ► 129.42.54.189 ► 129.42.60.189

Call Home with cloud services with an internal proxy server

This Call Home with cloud services configuration uses your network proxy server, which is the preferred method because it keeps your internal network secure.

To configure an internal proxy server with Call Home with cloud services, select **Settings** → **Support** → **Call Home** and then select **Add Proxy**. If a DNS is not configured, you are prompted to define one. You can also configure an internal proxy server by selecting **Settings** → **Network** → **Internal Proxy Server**.

Call Home with cloud services with DNS server only

When a DNS alone is defined with Call Home with cloud services, you must also update your network firewall settings to allow outbound traffic to `esupport.ibm.com` on port 443. To configure a DNS server, select **Settings** → **Network** → **DNS** and specify valid IP addresses and names for one or more DNS servers.

Call Home with cloud services without a DNS

When a DNS server is not available, you can update firewall settings to enable specific IP addresses and ports to establish a connection to the support center. This option is the least preferred because a DNS certificate is not available during the authentication process.

If you did not configure Call Home during system setup, you can configure this function in the management GUI. You can also change or update current settings on the Call Home page.

To configure or update Call Home with cloud services, complete the following steps:

1. In the management GUI, select **Settings** → **Support** → **Call Home**.
2. On the Call Home page, select **Send data with Call Home cloud services** and then click **Edit**.

3. Verify that the connection status is `Active` and a message displays that indicates the connection was successful. If the connection status displays `Error`, select **Monitoring** → **Events**. If Call Home with cloud services is configured, the following connection statuses can be displayed:
   - `Active`

     Indicates that the connection is active between the system and the support center. A timestamp displays with the last successful connection between the system and the support center.
   - `Error`

     Indicates that the system cannot connect to the support center through Call Home with cloud services. The system attempts connections every 30 minutes and if the connection continually fails for 4 hours, an event error is raised and is displayed.

     A timestamp displays when the failed connection attempt occurred. Select **Monitoring** → **Events** to determine the cause of the problem.

     One common issue that causes connection errors between the system and support center is firewall filters that exclude connections to the support center.

     For more information, see “Network considerations for Call Home with cloud services” on page 344.
   - `Untried`

     Indicates that Call Home with cloud services is enabled, but the system is waiting for the results from the connection test to the support center. After the test completes, the connection status changes to either `Active` or `Error`.
4. To define an internal proxy server to manage connections between the system and support, click **Add Proxy**. A DNS server is required to use an internal proxy server with Call Home with cloud services. The management GUI prompts you to define a DNS server if one is not configured.
5. Under Additional Settings, enter your preferences for inventory intervals and configuration reporting. Inventory reports can be configured with Call Home and provides more information for support personnel.

   An inventory report summarizes the hardware components and configuration of a system. Support personnel can use this information to contact you when relevant updates are available or when an issue that can affect your configuration is discovered. By default, these reports include configuration data that support personnel can use to automatically generate recommendations that are based on your configuration. You can have sensitive data that is redacted from these reports, if necessary.
6. Click **Save**.

Call Home with email notifications

Call Home with email notification sends notifications through a local email server to support and local users or services that monitor activity on the system. With email notifications, you can send notifications to support and designate internal distribution of notifications as well, which alerts internal personnel of potential problems.

Call Home with email notifications requires the configuration of at least one email server and local users. However, external notifications to the support center can be dropped if filters on the email server are active. To eliminate this problem, Call Home with email notifications is not recommended as the only method to transmit notifications to the support center.

Call Home with email notifications can be configured with cloud services for redundancy and internal management of notifications. If you also want notifications that are sent to an internal user or server, you must set up email notifications.

If you did not configure Call Home during system setup, you can configure this function in the management GUI. You can also change or update current settings on the Call Home page.

To configure or update Call Home with email notifications, complete the following steps:

1. In the management GUI, select **Settings** → **Support** → **Call Home**.
2. On the Call Home page, select **Send data with Call Home email notifications** and click then **Edit**.

   Note: Email filters can drop notifications and responses to and from the support center, which can affect resolution times for problems on your system. This transmission method is not recommended as the only way to send notifications to the support center. Use Call Home with email notifications as a backup method when Call Home with cloud services is configured.

3. Verify that the connection status is `Active` and a message displays that indicates the connection was successful. If the connection status displays `Error`, select **Monitoring** → **Events** to determine the cause of the problem.

   If Call Home with email notifications is configured, the following connection statuses can be displayed:

   – `Active`

     Indicates that the system can connect to the email server, which is actively sending notifications to defined email users.

   – `Failed`

     Indicates that the system cannot connect to the configured email server. This message is also displayed when repeated connection attempts fail after a `Failed Temporary` status. If you receive this message, verify that the email server or servers that you use are available. Select **Monitoring** → **Events** to display the error log for the system.

   – `Failed Temporary`

     Indicates that the last connection attempt to the email server failed, but the problem was temporary and other connection attempts are to be repeated.

   – `Untried`

     Indicates that the email server is configured but is not being used to send notifications.

4. Under Email Servers, enter a valid IP address or fully qualified domain name and ports for up to six email servers in your network. If you specify domain names, a DNS server must be configured on your system.

   To configure a DNS server for the system, select **Settings** → **Network** → **DNS**. These email servers send notification to the support center and receive and distribute responses from the support center.

5. Under Call Home, verify the email address for the support center and select the type of notifications that you want to be sent to the support center.
6. Under Email User, enter a valid email address for a business-to-business contact and select the type of notifications that you want them to receive. To comply with privacy regulations, personal contact information for individuals within your organization is *not* recommended.

7. Under Additional Settings, enter your preferences for inventory intervals and configuration reporting. Inventory reports can be configured with Call Home and provides more information to support personnel.

   An inventory report summarizes the hardware components and configuration of a system. Support personnel can use this information to contact you when relevant updates are available or when an issue that can affect your configuration is discovered. By default, these reports include configuration data that Support personnel can use to automatically generate recommendations that are based on your actual configuration. You can have sensitive data that is redacted from these reports, if necessary.

8. Click **Save**.

   If your system also supports both of these transmission methods, you can configure them at the same time. If cloud services *and* email notifications are configured, Call Home with cloud services is the primary transmission method to the support center and email notifications are used as a backup if the system cannot connect through cloud services to the support center.

   For each transmission method, statuses are displayed to indicate the connection health between the system and the support center.

Support assistance

Support assistance enables Support personnel to access the system to complete troubleshooting and maintenance tasks. You can configure local support assistance, where support personnel visit your site to fix problems with the system, or local and remote support assistance. Remote support assistance allows support personnel to access the system remotely from the support center.

Local and remote support assistance uses secure connections to protect data exchange between the support center and system. All actions that are completed with support assistance are recorded for auditing purposes. Local support assistance must be configured before remote support assistance is enabled.

Use local support assistance if you have restrictions that require onsite support only. Unlike other authentication methods, you can audit all actions that support personnel conduct on the system when local support assistance is configured.

Support personnel can log on to your system by using a console or over your Intranet. These users can be authenticated only by using a challenge-response mechanism. Support personnel obtain the challenge-response access through a virtual private network (VPN) or over a telephone call with another support person or the administrator at the support center.

With remote support assistance, support personnel can access the system remotely through a secure connection from the support center. However, before you enable remote support assistance between the system and support, you first must configure local support assistance. You must also ensure that Call Home is configured and a valid email server is specified.

Call Home automatically contacts support when critical errors occur on the system. Call Home sends a return email that communicates information back to the system, such as a Problem Management Report (PMR) number that tracks the problem until it is resolved.

During system initialization, you can optionally set up a service IP address and remote support assistance. If you did not configure a service IP address, click **Settings** → **Network** → **Service IPs** to configure a service IP for each node on the system. Optionally, you must configure a remote proxy server if you use a firewall to protect your internal network.

For more information about proxy configuration requirements, see Setting up Call Home with cloud services.

Figure 4-140 shows the window that is used to enable or reconfigure the Support assistant.

.

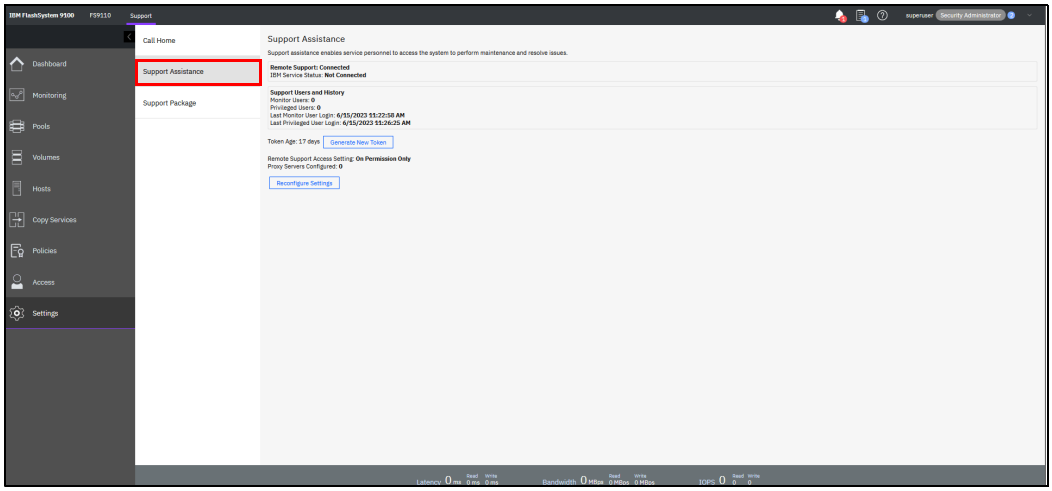


Figure 4-140 Support assistance

When you enable remote support assistance, you can specify an IP address or domain name for support. If you specify domain names, a DNS server must be configured on your system. To configure a DNS server for the system, select **Settings** → **Network** → **DNS**.

In addition, you can define a shared-token that also is generated by the system and sent to the Support center. If the system needs support services, Support personnel can be authenticated onto the system with a challenge-response mechanism.

After Support personnel obtain the response code, it is entered to gain access to the system. Service personnel have three attempts to enter the correct response code. After three failed attempts, the system generates a new random challenge and Support personnel must obtain a new response code.

When you enable local support assistance, you can specify the IP address or domain name for the support connections. When support personnel log on to the systems with local support assistance, they are assigned the Monitor role or the Restricted Administrator role.

The Monitor role can view, collect, and monitor logs and errors to determine the solution to problems on the system. The Restricted Administrator role gives support personnel access to administrator tasks to help solve problems on the system. However, this role restricts these users from deleting volumes or pools, unmapping hosts, or creating, deleting, or changing users.

Roles limit access of the assigned user to specific tasks on the system. Users with the service role can set the time and date on the system, delete dump files, add and delete nodes, apply service, and shut down the system. They can also view objects and system configuration settings, but cannot configure, modify, or manage the system or its resources. They also cannot read user data.

Service personnel might require specific support information to analyze before they can resolve an issue. You can automatically or manually upload new support packages to the support center to help analyze and resolve errors on the system. You can generate support packages and upload them to the support center or select specific logs from a specific time.

Uploading support packages is described next.

Support package

If Support assistance is configured on your systems, you can automatically or manually upload new support packages to the IBM Support Center to help analyze and resolve errors on the system.

The menus are available by selecting **Settings** → **Support** → **Support package**, as shown in Figure 4-141.

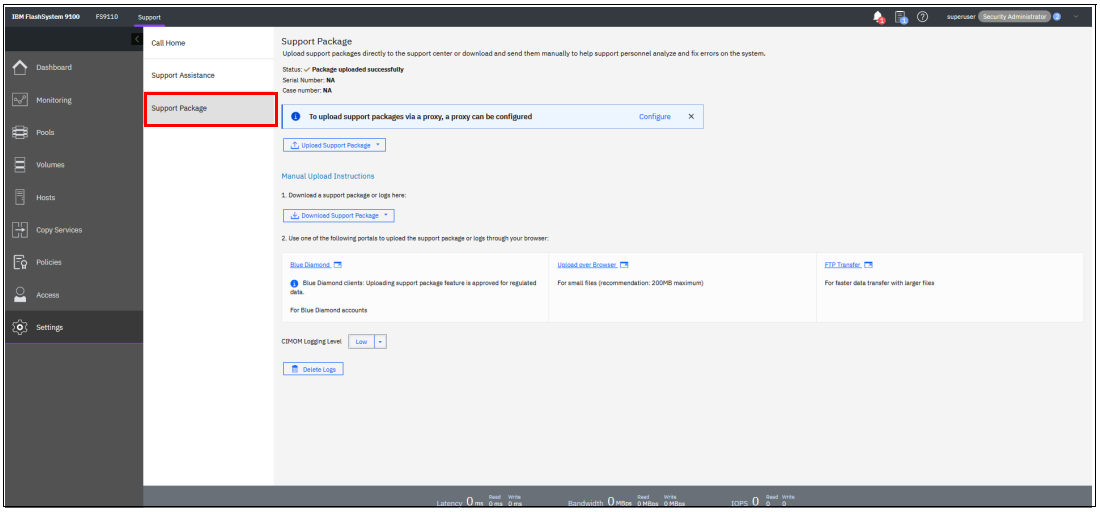


Figure 4-141 Support package menu

For more information about how the Support menu helps with troubleshooting your system or how to back up your systems, see Chapter 11, “Reliability, availability, and serviceability” on page 963.

4.7.7 GUI Preferences menu

The GUI Preferences menu consists of the following options:

- Login
- General
- GUI Features
- Notification Behavior
- Language
- Sidebar Accent Color

Figure 4-142 shows the GUI Preferences selection window.

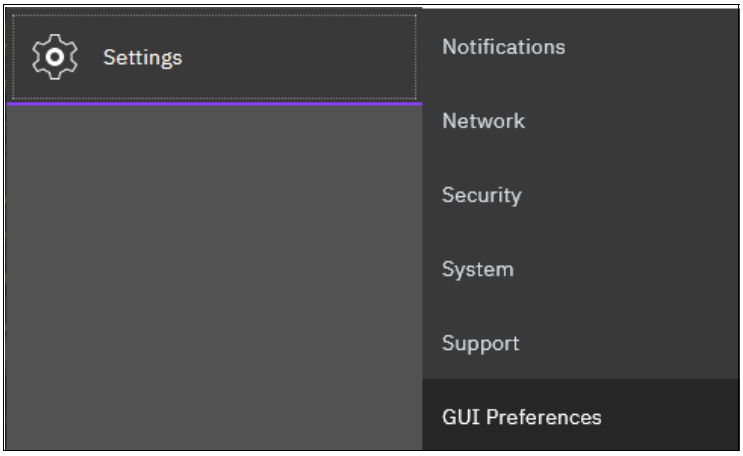


Figure 4-142 GUI Preferences selection window

Login message

IBM Storage Virtualize enables administrators to configure the welcome banner (login message). This message is a text message that appears in the GUI login window or at the CLI login prompt.

The content of the welcome message is helpful when you must notify users about some important information about the system, such as security warnings or a location description. To define and enable the welcome message by using the GUI, edit the text area with the message content and click **Save** (see Figure 4-143).

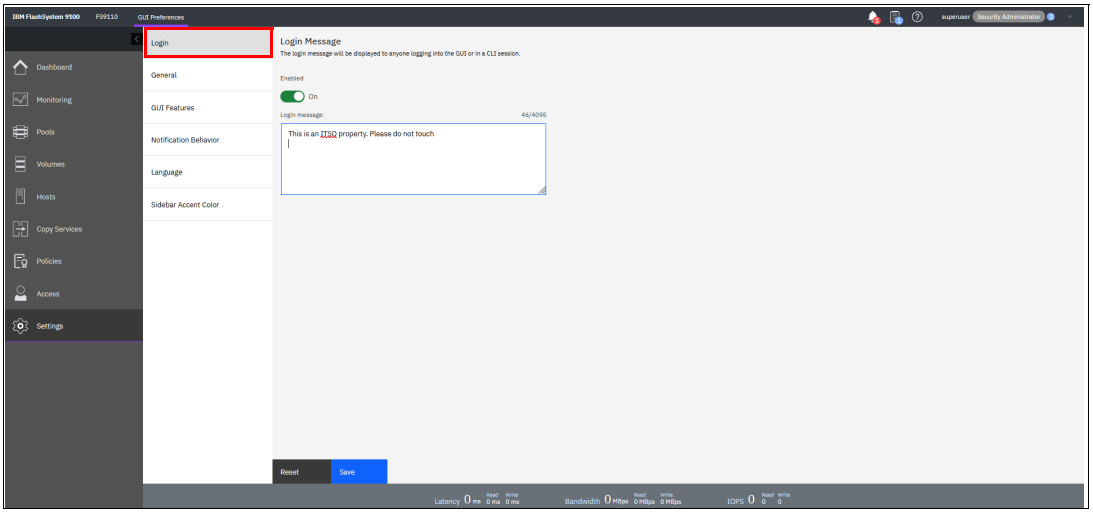


Figure 4-143 Enabling the login message

The resulting login dialog box is shown in Figure 4-144.

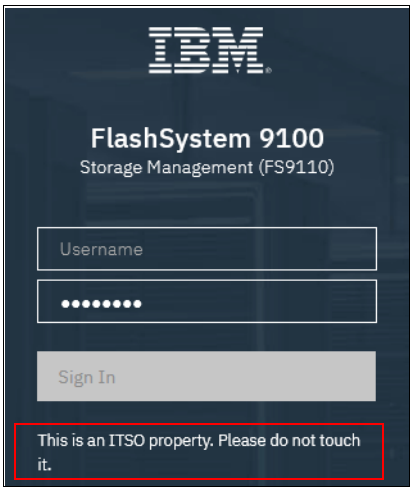


Figure 4-144 Welcome message in the GUI

The banner message also appears in the CLI login prompt window, as shown in Figure 4-145.

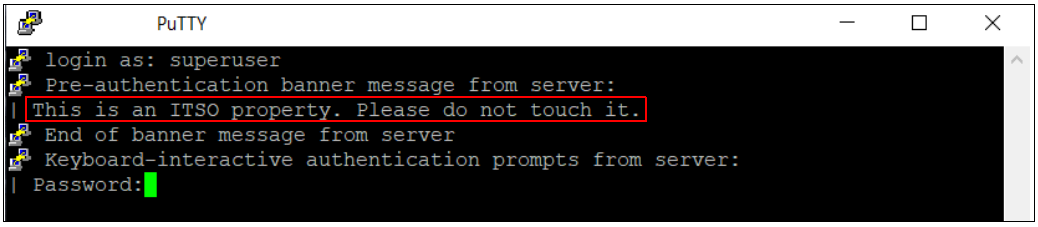


Figure 4-145 Welcome message in CLI

General Settings

With the General Settings menu, you can refresh the GUI cache, set the low graphics mode option, and enable advanced pools settings. To configure general GUI preferences, complete the following steps:

1. From the Settings window, click **Settings** and select **GUI Preferences** → **General** (see Figure 4-146).

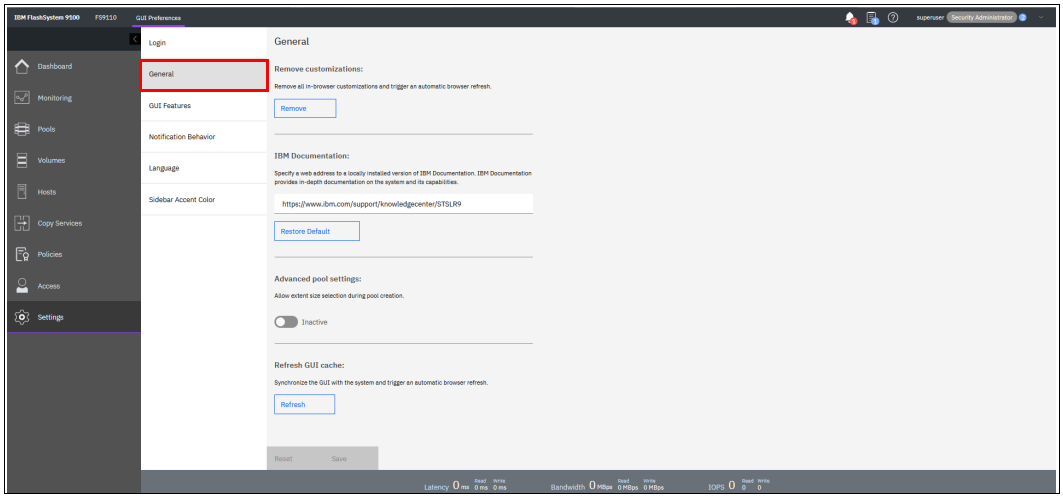


Figure 4-146 General GUI Preferences window

2. You can configure the following elements:
   – Clear customizations

     This option deletes all GUI preferences that are stored in the browser and restores the default preferences.

   – IBM Documentation

     You can change the URL of IBM Documentation for IBM Storage Virtualize.

   – Advanced pool settings

     You can select the extent size during storage pool creation.

   – Refresh GUI cache

     This option causes the GUI to refresh all its views and clears the GUI cache. The GUI looks up every object again. This option is useful if a value or object that is shown in the CLI is not being reflected in the GUI.

GUI Features

GUI features settings are used to update the display of FlashCopy, remote copy, and Safeguarded backup policy in the management GUI.

To update general settings in the management GUI, select **Settings** → **GUI preferences** → **GUI Features**. You can update any of the following settings:

- Display FlashCopy functions under Copy Services

  By default, all FlashCopy functionality under the Copy Services is **Hidden** to display the functionality, set it to **Visible**.

  > **Note:** If FlashCopy mapping exists on a system, this value is automatically set to **Visible**.

  The snapshot functionality under **Volume** → **Volume Groups** replaces FlashCopy functions. The new snapshot function provides enhanced scalability, usability, and an internal scheduler to simplify overall management.

- Display Remote Copy functions in Copy Services

  By default, the Remote Copy function that is related to Remote Copy relationships and Remote Copy consistency groups is hidden. To display this function in the GUI, set this value to **Visible**.

  > **Note:** If remote copy relationships or consistency groups exist on the system, this value is automatically set to **Visible**.

  Policy-based replication replaces asynchronous 2-site replication with Global Mirror function. Policy-based replication provides an automated process to configure and manage replicated data.

- Display Safeguarded Backup policy and External Scheduler Application Settings

  By default, the Safeguarded Backup Policy tile from the Volume Group policy page tab and the External Scheduler Application settings page is not displayed. To display this functionality in the GUI, set this value to **Visible**.

  > **Note:** If Safeguarded backup policy with external scheduling application exists on the system, this value is automatically set to **Visible**.

The system supports Safeguarded snapshots, which eliminate the need for creating a Safeguarded backup copy and selecting an external scheduling application. The system still supports Safeguarded backup policies when the system uses IBM Storage Copy Data Management or an earlier version of Safeguarded Copy function with IBM Copy Services Manager.

A Safeguarded backup policy creates and manages the frequency and retention of Safeguarded copies. The external scheduling applications automate copies and data recovery. The system supports either IBM Storage Copy Data Management or IBM Copy Services Manager as an external scheduling application. Before you can configure a Safeguarded backup policy and an external scheduling application, you must enable these features in the management GUI.

If a Safeguarded Backup Policy is assigned to any volume group, the Display Safeguarded backup policy tile and the External Scheduler Application Settings option is disabled.

See Figure 4-147 for these options.

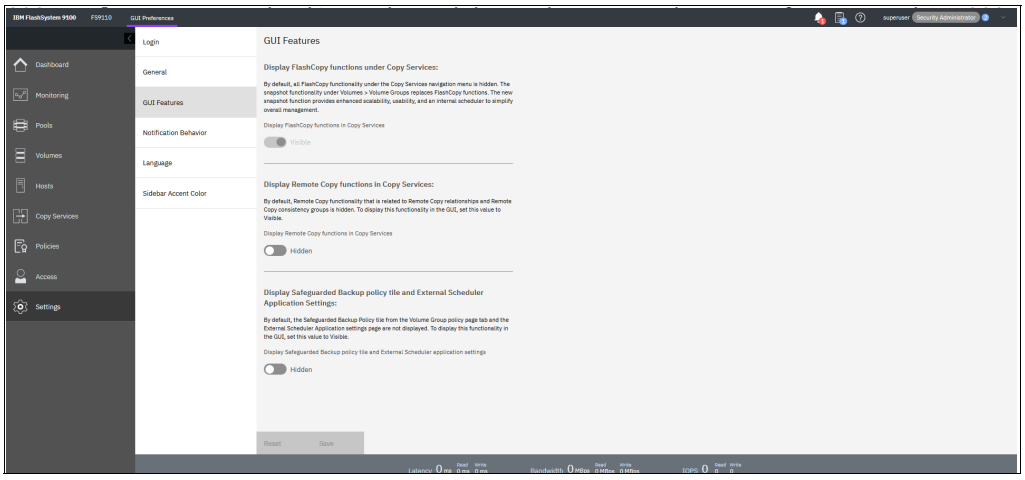


Figure 4-147 GUI Features

Notification Behavior

Figure 4-148 on page 355 shows that you can allow specific notifications to remain on the window until they are manually dismissed. You can select the preferred placement for notifications.

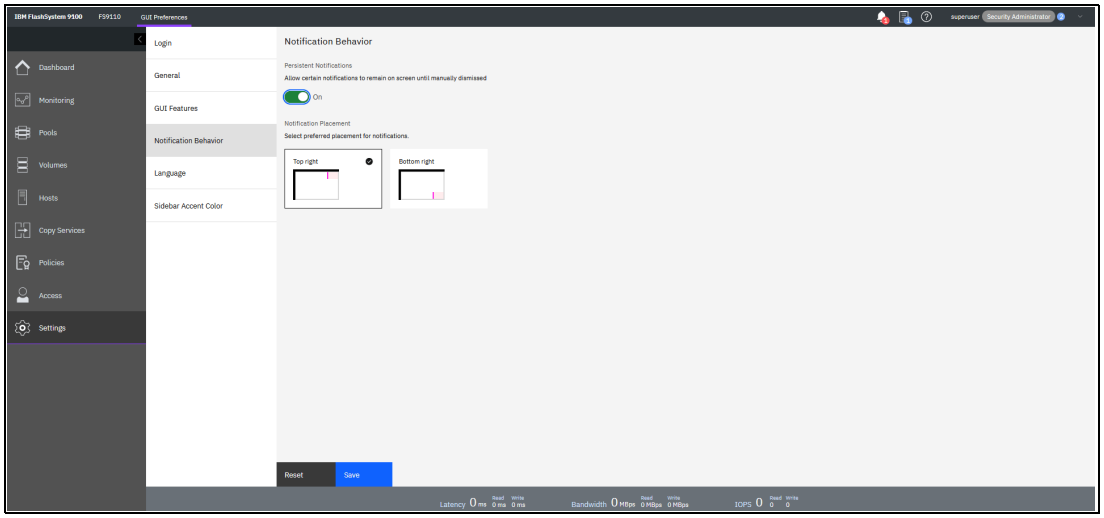


Figure 4-148 Notification Behavior window

Language

You can change the language of the GUI independently of the browser settings (Figure 4-149).

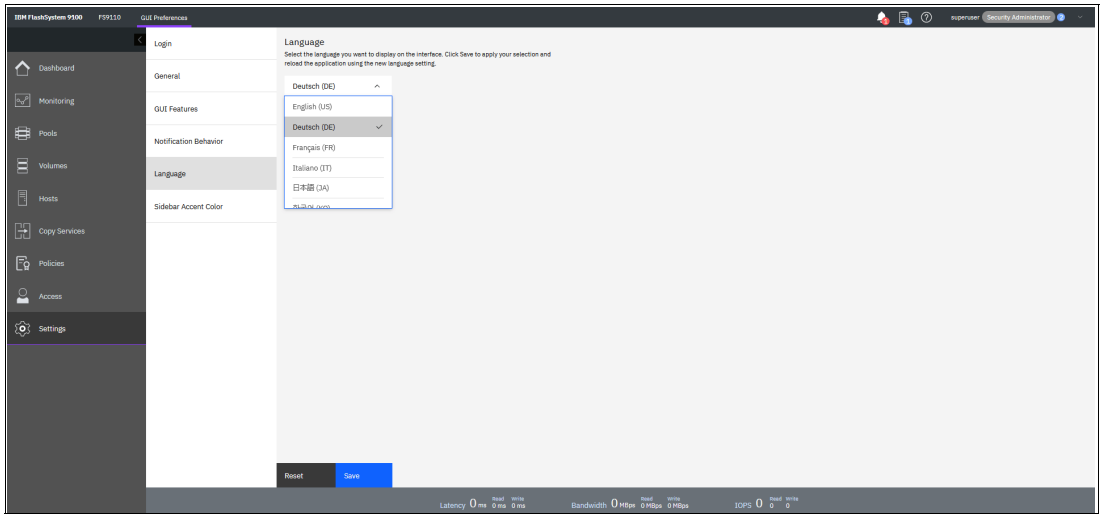


Figure 4-149 Language selection

Sidebar Accent Color

You have the ability to change the color of the sidebars by using the slider on the top to enable the function. This feature helps you to differentiate between different sessions on IBM Storage Virtualize Products (Figure 4-150).

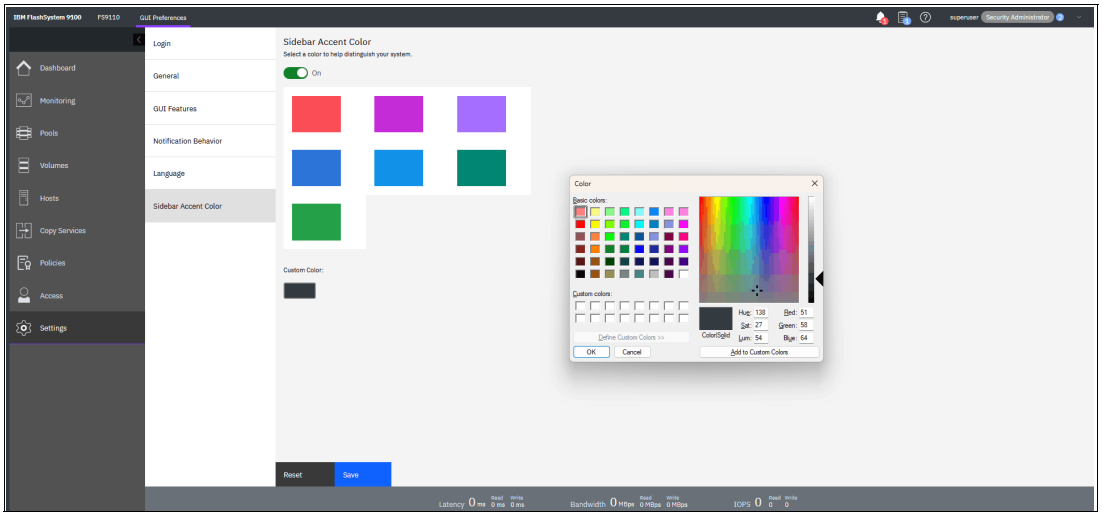


Figure 4-150 Sidebar Accent Color window

Under custom color, you can choose your own color. Figure 4-151 shows a modified sidebar.

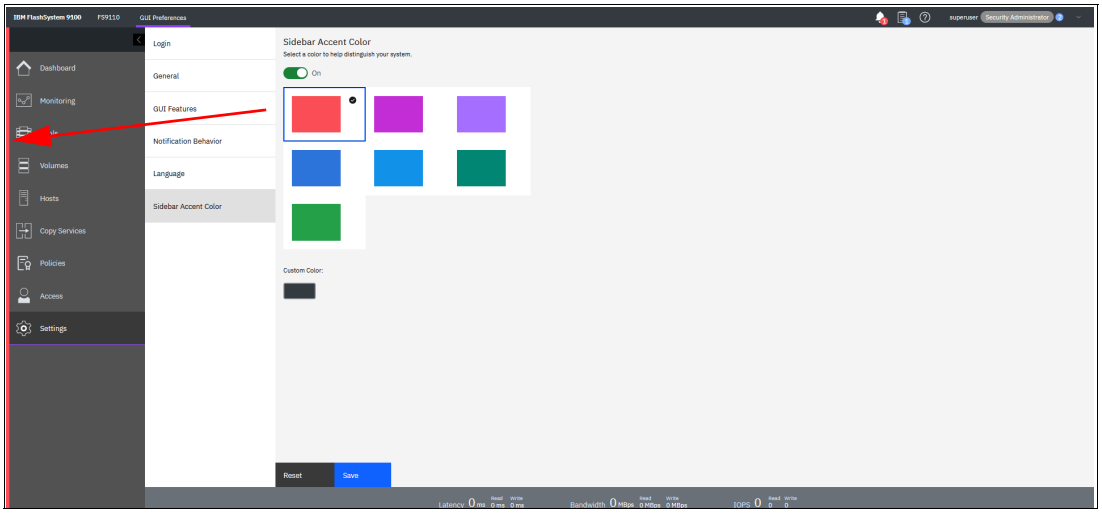


Figure 4-151 Customized sidebar

4.8 Other frequent tasks in the GUI

This section describes other options and tasks that are available in the system GUI that are frequently used by administrators.

4.8.1 Renaming components

These sections provide guidance about how to rename your system and canisters.

Renaming your storage system

All objects in the system feature names that are user-defined or system-generated. Choose a meaningful name when you create an object. If you do not choose a name for the object, the system generates a name for you.

A well-chosen name serves as a label for an object and as a tool for tracking and managing the object. Choosing a meaningful name is important if you decide to use configuration backup and restore.

When you choose a name for an object, apply the following naming rules:

- Names must begin with a letter.

  Important: Do not start names by using an underscore (_) character, even though it is possible. The use of an underscore as the first character of a name is a reserved naming convention that is used by the system configuration restore process.

- The first character cannot be numeric.
- The name can be a maximum of 63 characters, but exceptions exist. The name can be a maximum of 15 characters for RC relationships and groups. The **`lsfabric`** command displays long object names that are truncated to 15 characters for nodes and systems. (**`lsrcrelationshipcandidate`** or **`lsrcrelationship`** commands).
- Valid characters are uppercase letters (A - Z), lowercase letters (a - z), digits (0 - 9), the underscore (_) character, a period (.), a hyphen (-), and a space.
- Names must not begin or end with a space.
- Object names must be unique within the object type. For example, you can have a volume that is called ABC and an MDisk called ABC, but you cannot have two volumes that are called ABC.
- The default object name is valid (an object prefix with an integer).
- Objects can be renamed to their current names.

To rename the system from the System window, complete the following steps:

1. Select **Monitoring** → **System Hardware**, and click **System Actions** in the upper right of the window, as shown in Figure 4-152.

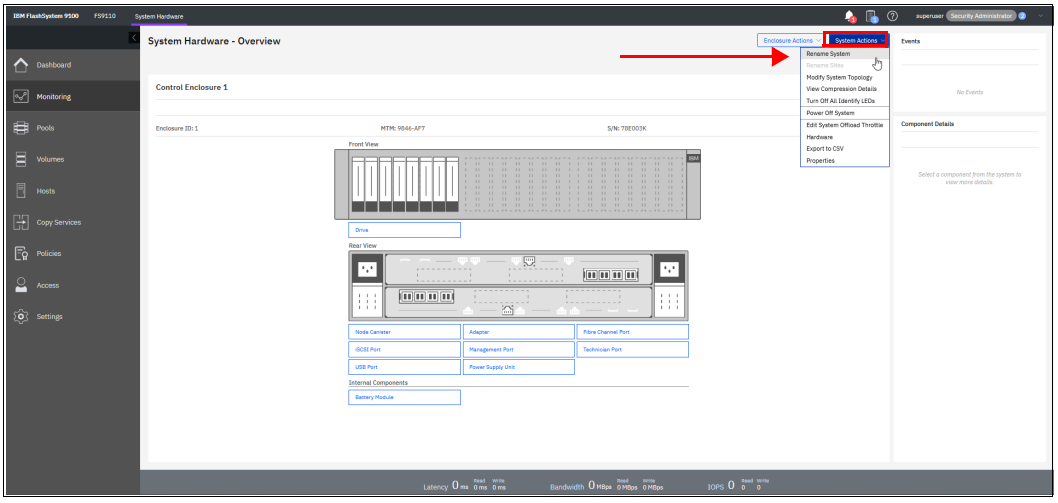


Figure 4-152 Overview window

2. The Rename System window opens (see Figure 4-153). Specify a new name for the system and click **Rename**.

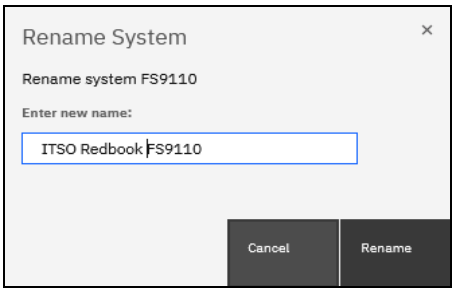


Figure 4-153 Renaming the system

Important: Consider the following points:

- You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The clustered system name can be 1 - 63 characters.
- When you rename your system, the iSCSI name automatically changes because it includes the system name by default. Therefore, this change needs more actions on iSCSI-attached hosts.

Renaming a node canister

To rename a node canister, complete the following steps:

1. Go to the **System Hardware** window and right-click the node that you want to rename, as shown in Figure 4-154. Click **Rename**.

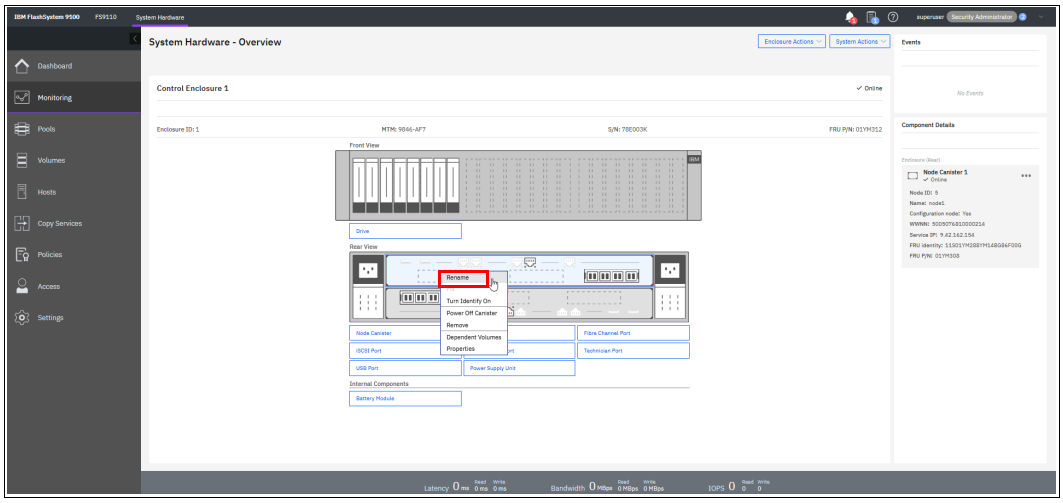


Figure 4-154 Overview window

2. Enter the new name of the node and click **Rename** (see Figure 4-155).

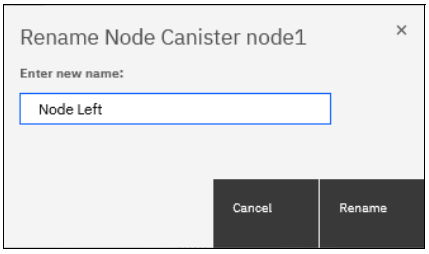


Figure 4-155 Entering the new name of the node

Warning: Changing the node canister name causes an automatic IQN update and requires the reconfiguration of all iSCSI-attached hosts.

4.8.2 Working with enclosures

This section describes how to add or remove expansion enclosures to or from your IBM FlashSystem storage system.

Adding an enclosure

After the expansion enclosure is correctly attached and powered on, complete the following steps to activate it in the system:

1. In the System window that is available from the Monitoring menu, select **SAS Chain View**. Only correctly attached and powered on enclosures appear in the window, as shown in Figure 4-156. The new enclosure is showing as unmanaged, which means it is not part of the system.

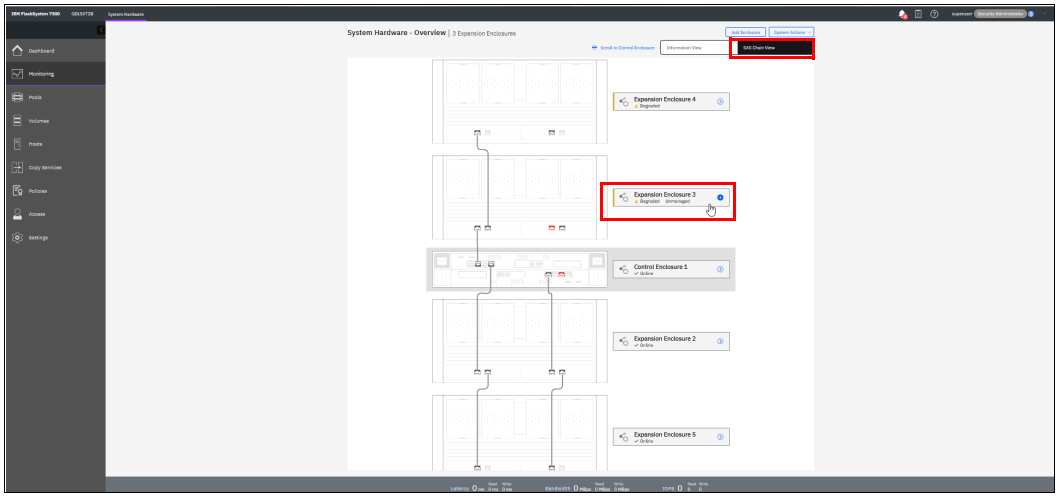


Figure 4-156 Newly detected expansion enclosure

2. Select the **+** next to the enclosure that you want to add or click **Add Enclosure** at the top. These buttons appear only if an unmanaged enclosure exists that is eligible to be added to the system. After they are selected, a window opens, in which you must select the enclosure that you want to add.

   Expansion enclosures that are directly cabled do not need to be selected, as shown in Figure 4-157 on page 360.

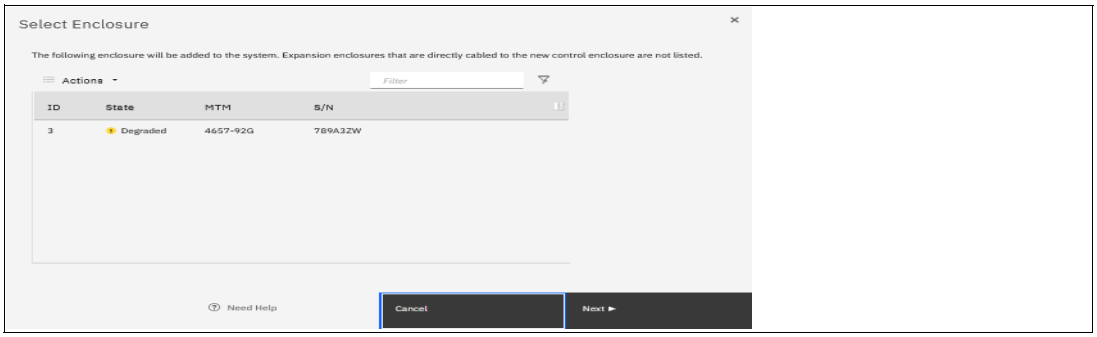


Figure 4-157 Adding an enclosure

3. Select **Next** and then select **Finish** after you are satisfied with your selections. The enclosures are then added to the system and appear as managed. Instead of the **+** button, you see a **>**, which allows you to view details about the enclosure because it is now part of the system, as shown in Figure 4-158.

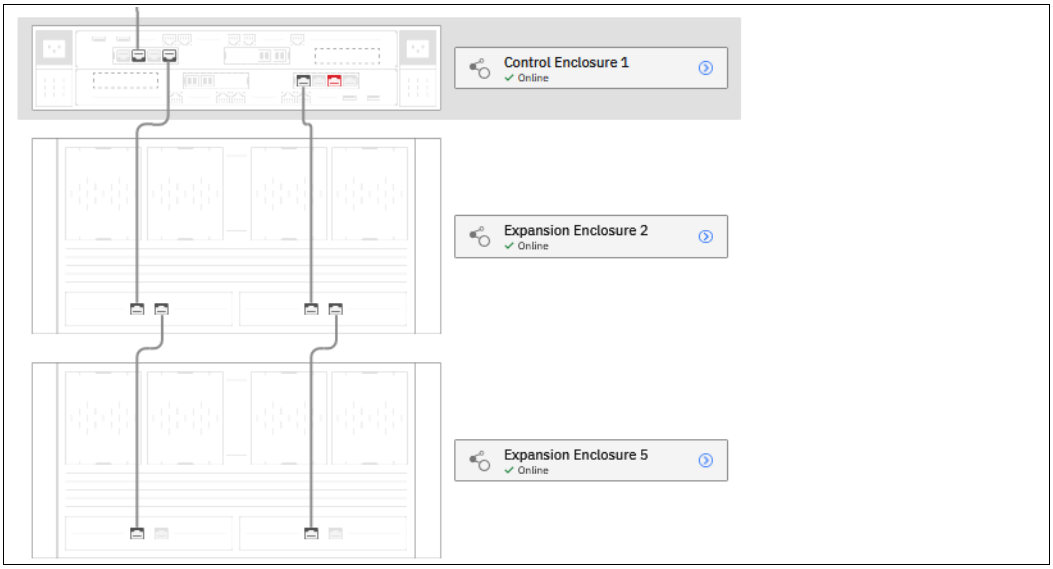


Figure 4-158 Enclosure successfully added

Removing an enclosure

The enclosure removal procedure includes its logical detachment from the system by using a GUI and physically unmounting the systems from the rack. The IBM FlashSystem storage system guides you through this process.

Complete the following steps:

1. In the System window that is available from the **Monitoring** menu, select **>** next to the enclosure that you want to remove. The Enclosure Details window opens. You can then click **Enclosure Actions** and select **Remove**, as shown in Figure 4-159 on page 361.

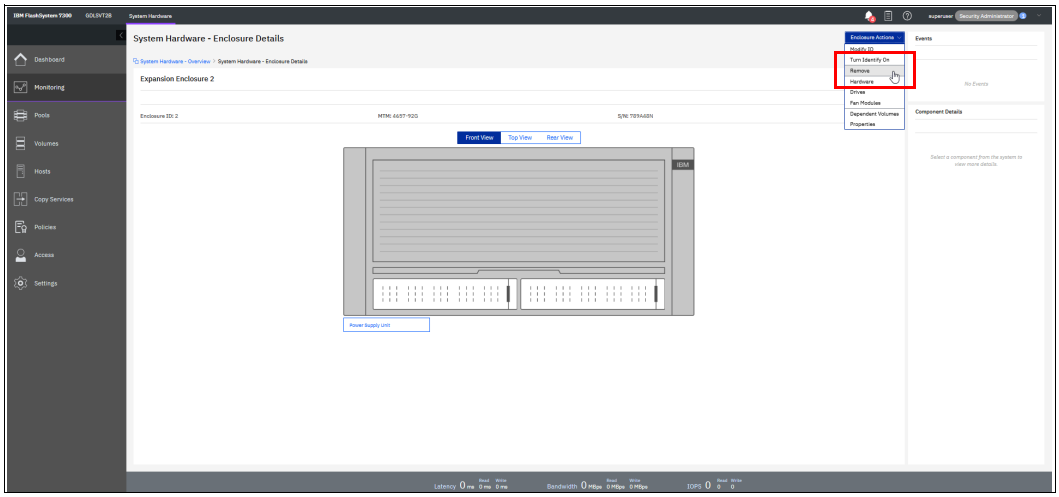


Figure 4-159 Selecting an enclosure for removal

2. The system prompts you to remove the enclosure. All disk drives in the removed enclosure must be in the *Unused* state. Otherwise, the removal process fails (see Figure 4-160).

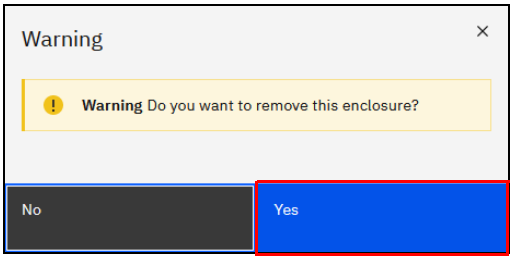


Figure 4-160 Confirming the removal

After the enclosure is logically removed from the system (set to the *Unmanaged* state), the system reminds you about the steps that are necessary for physical removal, such as power off, uncabling, dismantling from the rack, and secure handling (see Figure 4-161).

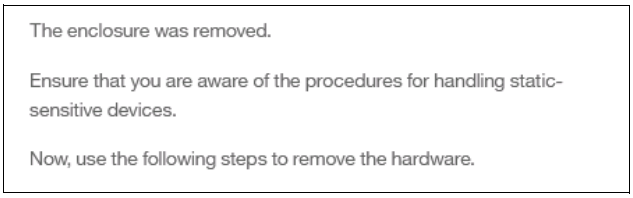


Figure 4-161 Enclosure removed

As part of the enclosure removal process, see your company security policies about how to handle sensitive data on removed storage devices before they leave the secure data center. Most companies require data to be encrypted or logically shredded.

4.8.3 Restarting the GUI service

The service that runs that GUI operates from the configuration node. Occasionally, you might need to restart this service if the GUI is not performing to your expectations (or you cannot connect). To do so, complete the following steps:

1. Log in to the Service Assistant, as shown in Figure 4-162.

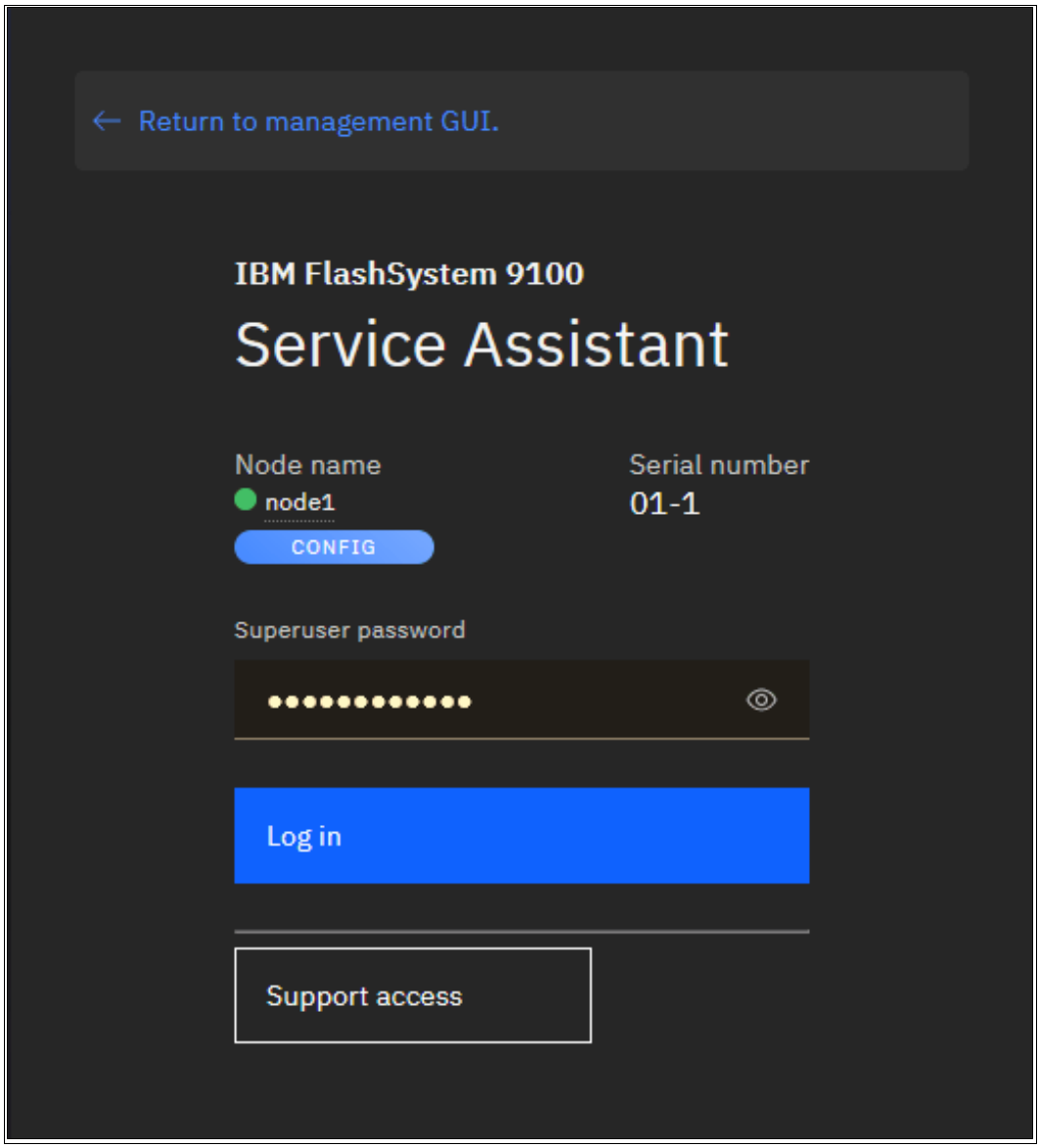


Figure 4-162 Log in Service GUI

2. Identify the configuration node, as shown in Figure 4-163.

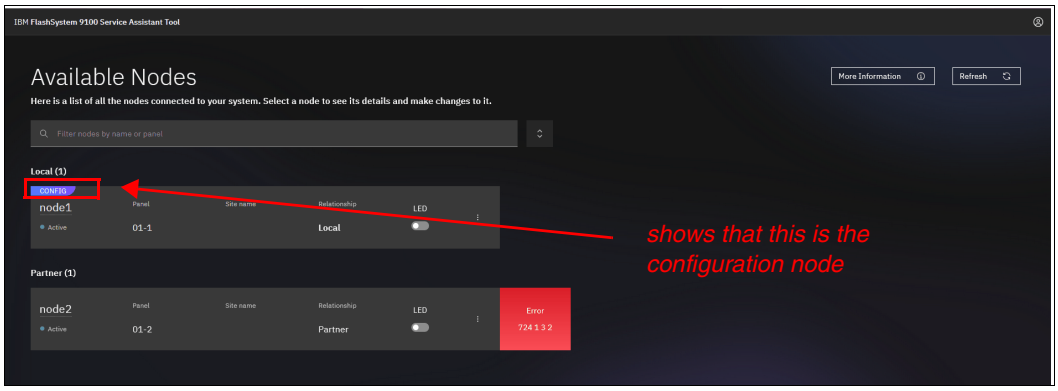


Figure 4-163 Identifying the configuration node on the Service Assistant

3. Select the **config** node to open the Node details window. (Figure 4-164). Select **Restart Service**.

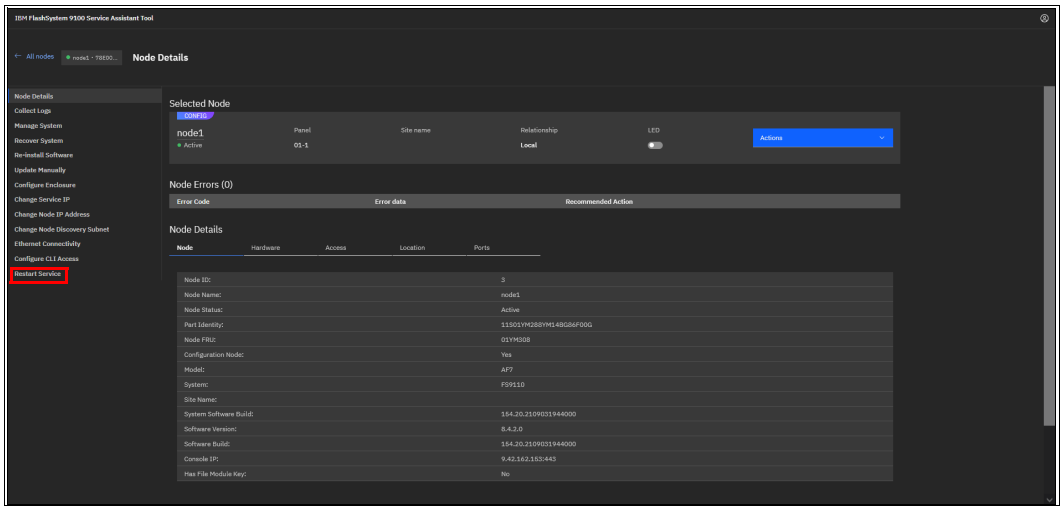


Figure 4-164 Node details window

After the process completes, the Restart option appears, as shown in Figure 4-165.

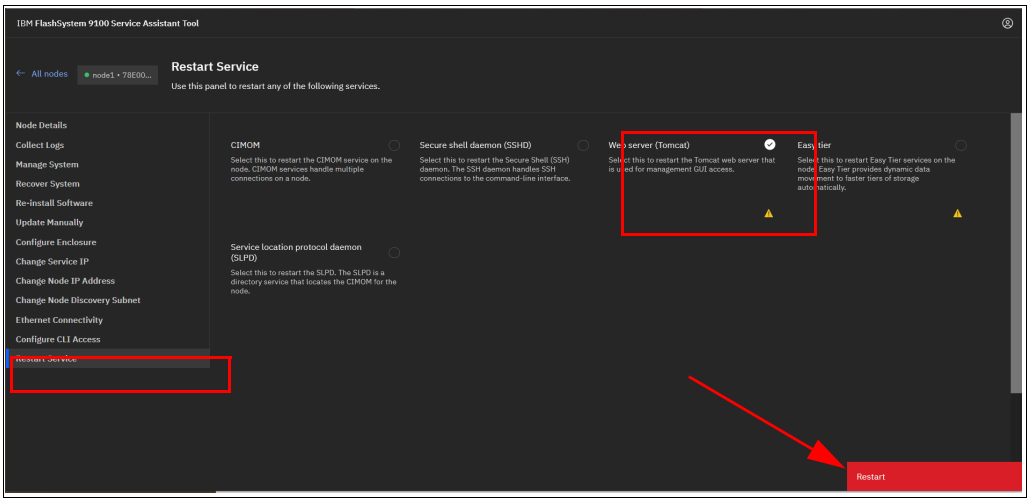


Figure 4-165 Restarting the Tomcat web server

Select **Web Server (Tomcat)**. Click **Restart**, and the web server that runs the GUI restarts. Although this task is a concurrent action, the cluster GUI is unavailable while the server is restarting (the Service Assistant and CLI are not affected). After 5 minutes, check to see whether GUI access was restored.

5

Using storage pools

This chapter describes how the storage system manages physical storage resources. All storage resources under system control are managed by using storage pools, also known as managed disk (MDisk) groups.

Storage pools aggregate internal and external capacity and provide containers in which you can create volumes. Storage pools make it easier to dynamically allocate resources, maximize productivity, and reduce costs.

You can configure storage pools through the management GUI during initial configuration or later. Alternatively, you can configure the storage to your own requirements by using the command line interface (CLI).

This chapter includes the following topics:

- 5.1, “Working with storage pools” on page 366
- 5.2, “Working with provisioning policies” on page 383
- 5.3, “Working with internal drives and arrays” on page 386
- 5.4, “Working with external controllers and MDisks” on page 407

© Copyright IBM Corp. 2024. All rights reserved.

5.1 Working with storage pools

Storage pools act as containers for MDisks, which provide storage capacity to the pool, and volumes that are provisioned from this capacity, which can be mapped to host systems. The system organizes storage in this fashion to ease storage management and make it more efficient.

MDisks can be redundant array of independent disks (RAID) arrays that are created by using internal storage, such as drives and flash modules, or logical units (LUs) that are provided by external storage systems. A single storage pool can contain both types of MDisks, but a single MDisk can be part of only one storage pool. MDisks themselves are not visible to host systems.

Figure 5-1 provides an overview of how storage pools, MDisks, and volumes are related.

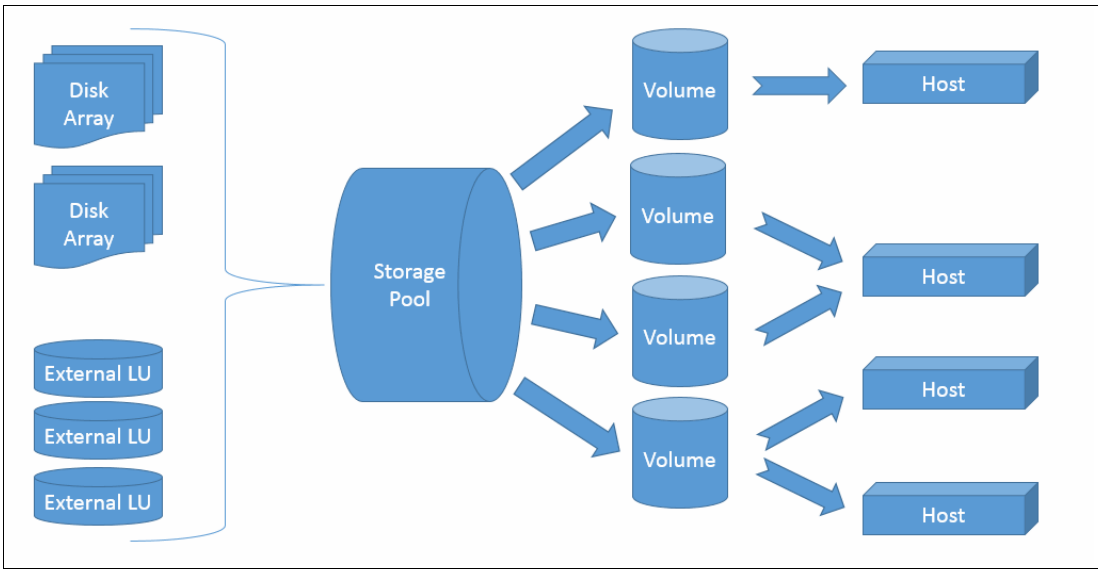


Figure 5-1 Relationship between MDisks, storage pools, and volumes

All MDisks in a pool are split into chunks of the same size, which are called *extents*. Volumes are created from the set of available extents in the pool. The extent size is a property of the storage pool and cannot be changed after the pool is created. The choice of extent size affects the total amount of storage that can be managed by the system.

It is possible to add MDisks to a pool to provide more usable capacity in the form of extents. The system automatically balances volume extents between the MDisks to provide the best performance to the volumes. It is also possible to remove extents from the pool by deleting an MDisk. The system automatically migrates data from extents that are in use by volumes to other MDisks in the same pool.

A storage pool represents a failure domain. If one or more MDisks in a pool become inaccessible, all volumes (except for image mode volumes) in that pool are affected. Volumes in other pools are unaffected.

The system supports standard pools and Data Reduction Pools (DRPs). Both support parent pools and child pools.

Child pools are created from capacity that is assigned to a parent pool instead of created directly from MDisks. When a child pool is created from a standard pool, the capacity for a

child pool is reserved from the parent pool. This capacity is no longer reported as available capacity of the parent pool. In terms of volume creation and management, child pools are similar to parent pools. Child pools created from DRPs are quotaless. Their capacity is not reserved but is shared with a parent pool.

DRPs use a set of techniques, such as compression and deduplication, that can reduce the amount of usable capacity that is required to store data. Data reduction can increase storage efficiency and performance, and reduce storage costs, especially for flash storage. These techniques can be used in addition to compression on Flash Core Modules (FCMs).

In standard pools, there can be no compression on a pool layer, but data is still compressed on the FCM layer, if the pool contains drives with this technology.

DRPs automatically reclaim capacity that is no longer needed by host systems. This reclaimed capacity is returned to the pool as usable capacity and can be reused by other volumes.

For more information about DRP planning and implementation, see Chapter 9, “Advanced features for storage efficiency” on page 669, and *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430.

In general, you manage storage by using the following process:

1. Create storage pools (standard or DRP), depending on your requirements and sizing.
2. Assign storage to these pools by using one or more of the following options:
   – Create array MDisks from internal drives or flash modules.
   – Add MDisks provisioned from external storage systems.
3. If necessary, create child pools, create provisioning policies, and assign policies to child pools or parent pools.
4. Create volumes in these pools and map them to hosts or host clusters.

You can manage storage pools in the Pools panel of the GUI or by using the CLI. To access the Pools panel, select **Pools** → **Pools**, as shown in Figure 5-2.

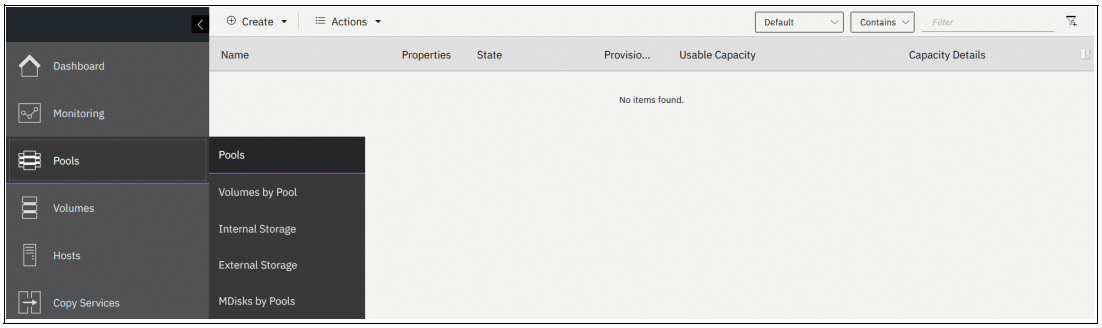


Figure 5-2 Accessing the Pools panel

The panel lists all storage pools and their major parameters. If a storage pool has child pools, they are also shown.

To see a list of configured storage pools by using the CLI, run the **lsmdiskgrp** command without any parameters, as shown in Example 5-1.

Example 5-1 The lsmdiskgrp output (some columns are not shown)

```
IBM_2145:ITSO-SV1:superuser>lsmdiskgrp
id name           status mdisk_count vdisk_count capacity extent_size ...
0  Pool0_DRP      online 1           1           39.99TB  1024        ...
1  Pool1_Std      online 1           4           39.99TB  1024        ...
2  Pool0_Perf     online 0           1           39.99TB  1024        ...
3  Pool0_Capacity online 0           0           39.99TB  1024        ...
```

5.1.1 Creating storage pools

To create a storage pool, complete the following steps:

1. Select **Pools** → **MDisks by Pools** and click **Create Pool** or select **Pools** → **Pools** and click **Create** → **Create Pool**, as shown in Figure 5-3.

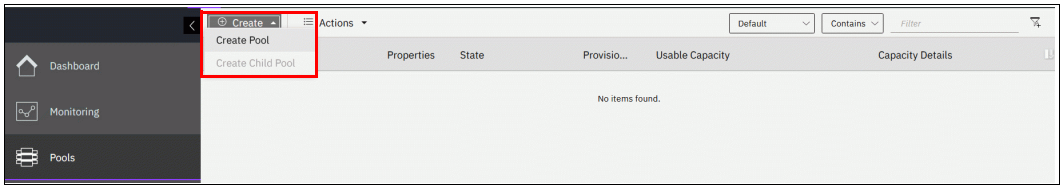


Figure 5-3 Option to create a storage pool in the Pools panel

Both alternatives open the dialog box that is shown in Figure 5-4.

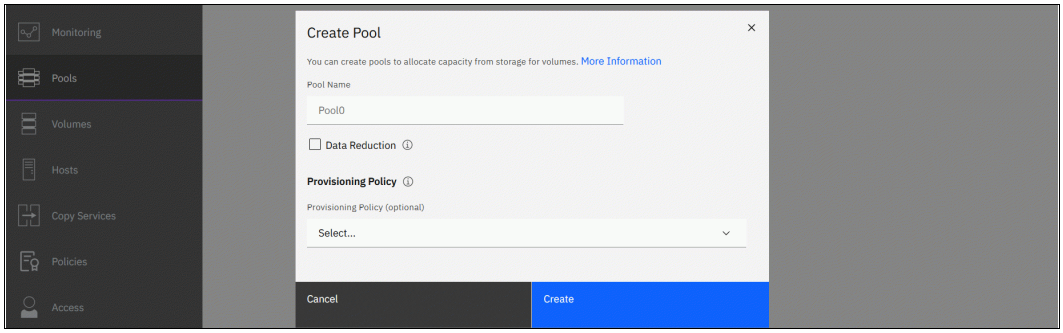


Figure 5-4 Create Pool dialog box

2. Select the **Data reduction** option to create a DRP. Leaving it clear creates a standard storage pool.

> **Note:** Limitations and performance characteristics of DRPs are different from standard pools. Verify with your system architect or IBM representative that your system was sized to be used with DRP and with its reduction features.

For more information about the differences between standard pools and DRPs and for extent size planning, see Chapter 2, “Installation and configuration planning” on page 119.

With the default GUI settings, the dialog will not allow you to select a standard pool extent size, and will create pools with the 1024MB extent size. If your pool is expected to grow beyond 4.0PB at some point during the system lifecycle, enable extent size selection in

Settings → **GUI Preferences** → **General** → **Advanced pool settings**, as shown in Figure 5-5.

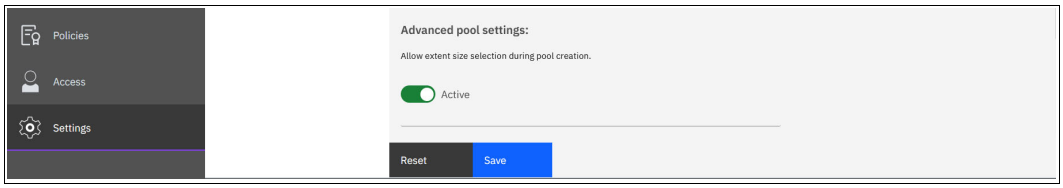


Figure 5-5 Advanced pool settings

For data reduction pools, extent size selection is available even with the default GUI settings.

The size of the extents is selected at creation time and cannot be changed later. The extent size controls the maximum total storage capacity that is manageable per system (across all pools). For DRPs, the extent size also controls the maximum pool stored capacity per IO group. Refer to the Configuration Limits and Restrictions web page for your hardware platform.

Note: Do not create DRPs with small extent sizes. For more information, see this IBM Support alert.

If an encryption license is installed and enabled, you can select whether the storage pool is encrypted, as shown in Figure 5-6. The encryption setting of a storage pool is selected at creation time and cannot be changed later. By default, if encryption is licensed and enabled, the encryption checkbox is selected. For more information about encryption and encrypted storage pools, see 12.7, “Encryption” on page 1098.

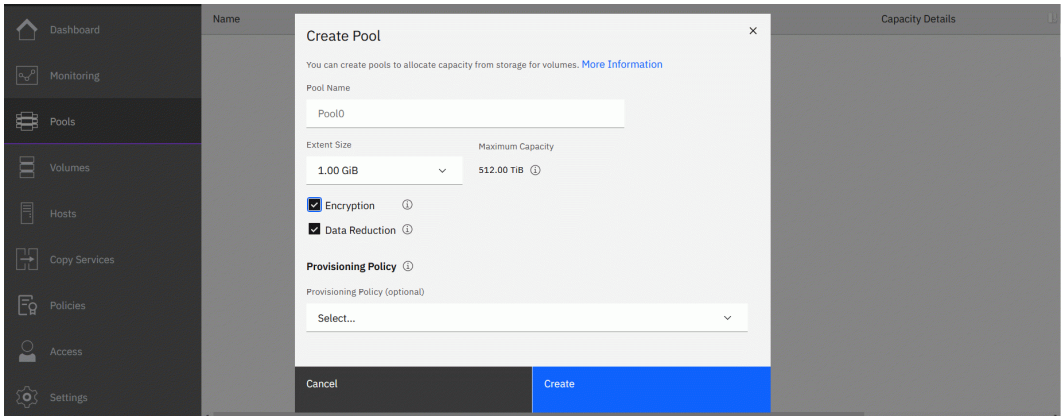


Figure 5-6 Creating a pool with encryption enabled

3. Optionally, assign an existing provisioning policy to the pool. If a policy is assigned, any volumes created from the pool are provisioned based on the capacity savings method defined in the policy. A policy can be created and assigned or unassigned to a pool at any time, but it effects only volumes created while the policy was active.

   For more information about Provisioning policies, refer to 5.2, “Working with provisioning policies” on page 383.

4. Enter the name for the pool and click **Create**.

Naming rules: When you choose a name for a pool, the following rules apply:

- Names must begin with a letter.
- The first character cannot be numerical.
- The name can be a maximum of 63 characters.
- Valid characters are uppercase letters (A - Z), lowercase letters (a - z), digits (0 - 9), underscore (_), period (.), hyphen (-), and space.
- Names must not begin or end with a space.
- Object names must be unique within the object type. For example, you can have a volume that is named `ABC` and a storage pool that is called `ABC`, but not two storage pools that are both called `ABC`.
- The default object name is valid (object prefix with an integer).
- Objects can be renamed at a later stage.

The new pool is created and is included in the list of storage pools, as shown in Figure 5-7. It has no storage in it, so its capacity is zero. Storage in a form of disk arrays or externally-virtualized MDisks must be assigned to the pool before volumes can be created.

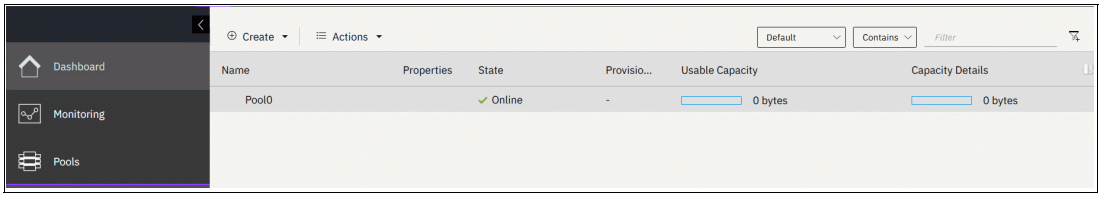


Figure 5-7 Newly created empty pool

To create a pool using the CLI, use the **`mkmdiskgrp`** command. The only required parameter is the extent size, which is specified by the **`-ext`** parameter and must have one of the following values: 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, or 8192 (MB). To create a DRP, specify **`-datareduction yes`**. The minimum extent size of DRPs is 1024, and attempting to use a smaller extent size sets the extent size to 1024.

As shown in Example 5-2, the command creates a DRP that is named `Pool2` with no MDisks in it.

Example 5-2 The mkmdiskgrp command

```
IBM_2145:ITSO-SV1:superuser>mkmdiskgrp -name Pool1 -datareduction yes -ext 8192
MDisk Group, id [1], successfully created
```

5.1.2 Managed disks in a storage pool

A storage pool is created as an empty container with no storage assigned to it. Storage is then added in the form of MDisks. An MDisk can be a RAID array from internal storage (as an array of drives) or an LU from an external storage system. The same storage pool can include both internal and external MDisks.

Arrays are assigned to storage pools at creation time. Arrays cannot exist outside of a storage pool and they cannot be moved between storage pools. It is possible to delete an array by removing it from a pool and re-create it within a new pool.

External MDisks can exist within or outside of a pool. The MDisk object remains on a system if it is visible from external storage, but its access mode changes depending on whether it is assigned to a pool.

MDisks are managed by using the MDisks by Pools panel. To access the MDisks by Pools panel, select **Pools** → **MDisks by Pools**, as shown in Figure 5-8.

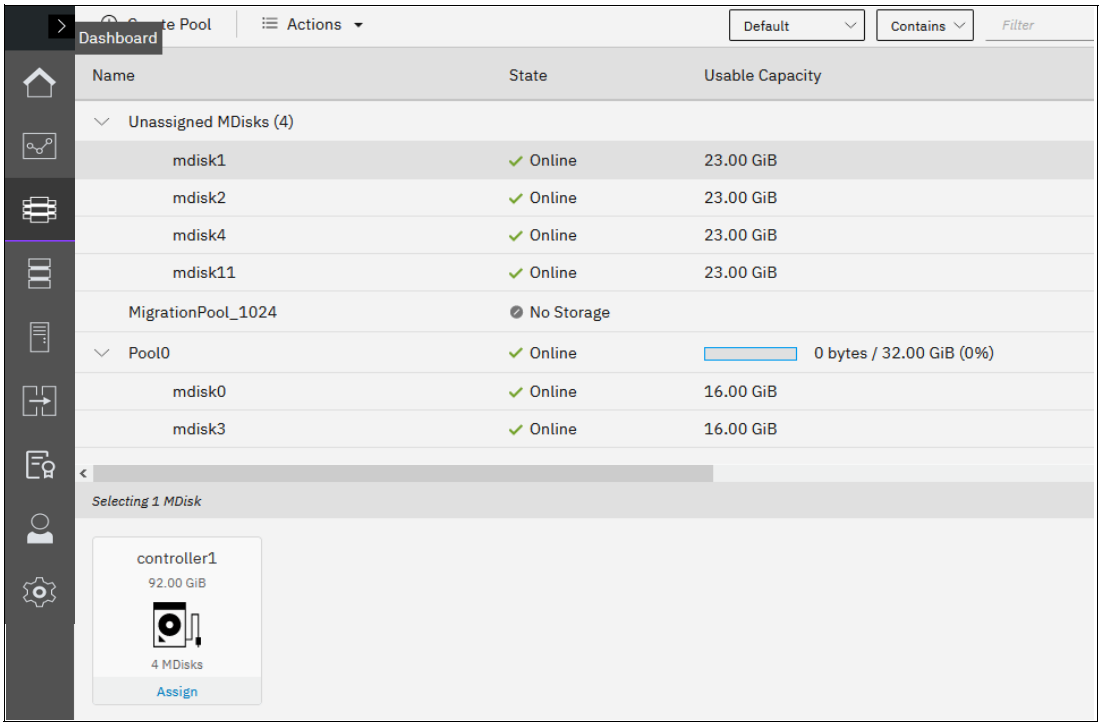


Figure 5-8 MDisks by Pools

The panel lists all the MDisks that are available in the system under the storage pool to which they belong. Unassigned MDisks are listed separately at the top. Both arrays and external MDisks are listed.

For more information about operations with array MDisks, see 5.2, “Working with provisioning policies” on page 383.

For more information about implementing a solution with external MDisks, see 5.2, “Working with provisioning policies” on page 383.

To list all MDisks that are visible by the system by using the CLI, run the **`lsmdisk`** command without any parameters. If required, you can filter output to include only array type MDisks by specifying **`-filtervalue mode=array`**, or to include only external MDisks by using another access modes.

5.1.3 Actions on storage pools

A number of actions can be performed on storage pools. To select an action, select the storage pool that you want to take an action on, and click **Actions**, as shown on Figure 5-9. Alternatively, right-click the storage pool.

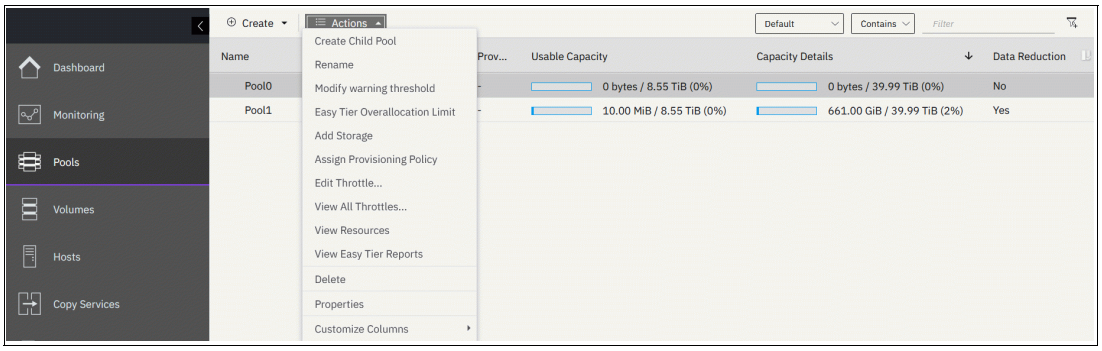


Figure 5-9 Pool actions

A list of available actions on child pools is different from list of actions on parent pools. For example, you cannot add storage to a child pool. Instead, there is a **Modify Capacity** action available for child pools, with the exception of child pools from DRP parent pools. It can be used to adjust the capacity allocated for the child pool.

Create Child Pool window

To create a child storage pool, click **Create Child Pool**. It is not possible to create a child pool from an empty pool.

For more information about child storage pools and this wizard, see 5.1.4, “Child pools” on page 378.

Rename window

To modify the name of a storage pool, click **Rename**. Enter the new name and click **Rename** in the dialog window.

To perform this task by using the CLI, run the **chmdiskgrp** command. Example 5-3 shows how to rename `Pool2` to `StandardStoragePool`. If successful, the command returns no output.

Example 5-3 Using chmdiskgrp to rename a storage pool

```
IBM_2145:ITSO-SV1:superuser>chmdiskgrp -name StandardStoragePool Pool2
IBM_2145:ITSO-SV1:superuser>
```

Modify Threshold window

A warning event is generated when the amount of used capacity in the pool exceeds the warning threshold. When you use thin-provisioned or compressed volumes, monitor the capacity usage and receive warnings before the pool runs out of free extents so that you can add storage before running out of space.

Note: The warning is generated only the first time the threshold is exceeded by the used capacity in the storage pool.

To modify the threshold, select **Modify Threshold** and enter the new value. The default threshold is 80%. To disable warnings set the threshold to 0%.

The threshold is visible in the pool properties and indicated by a red bar, as it can be seen in Figure 5-14 on page 377.

To modify the threshold by using the CLI, run the **chmdiskgrp** command. You can specify the threshold by using a percentage. You can also set an exact value and specify a unit.

Example 5-4 shows the warning threshold set to 750 GB for `Pool0`.

Example 5-4 Changing the warning threshold level by using the CLI

```
IBM_2145:ITSO-SV1:superuser>chmdiskgrp -warning 750 -unit gb Pool0
IBM_2145:ITSO-SV1:superuser>
```

Easy Tier Overallocation Limit window

This action sets the maximum overallocation percentage that controls how much data Easy Tier can migrate onto Flash Core Module arrays when these arrays are used as the top tier in multitier pools. This setting applies to all FlashCore Module arrays in the selected pool, and has no effect if there is no compression-capable back-end storage, such as industry-standard NVMe drives, in the pool.

The default setting is 100%. Overallocation limit can be raised for up to 400% depending on your expected data compression ratio.

This feature is explained in detail in “Overallocation limit” on page 681.

Add Storage to Pool window

Select this action to start the configuration wizard, which assigns storage to the pool, as shown in Figure 5-10.



Figure 5-10 Add Storage to Pool dialog

Consider the following points:

- If Internal Storage is chosen, the system guides you through the process of creating an array MDisk by using internal drives. See 5.2, “Working with provisioning policies” on page 383.
- If External Storage is selected, the system guides you through the selection of external storage MDisks. See 5.2, “Working with provisioning policies” on page 383.
- The GUI aims to create a configuration that is recommended by IBM. If you have unused storage but it is not shown as available in the dialog, it means that adding this storage to this particular pool is against the best practices.

Assign Provisioning Policy window

Select this action to assign an existing provisioning policy to the selected pool. Assigning a policy does not affect existing volumes, but it controls the parameters of newly created volumes.

See 5.2, “Working with provisioning policies” on page 383 for details.

Edit Throttle for Pool window

Select this action to access the window where you set the pool’s throttle configuration.

Throttles can be defined for storage pools to control I/O operations. If a throttle limit is defined, the system processes the I/O for that object, or delays the processing of the I/O. Resources become free for more critical I/O operations.

You can use storage pool throttles to avoid overloading the back-end storage. You can set throttles for both parent and child pools.

Note: To set throttles on child pools, the system must be running code version 8.4.2.0 or later.

You can define a throttle for input/output operations per second (IOPS), bandwidth, or both, as shown in Figure 5-11:

- `IOPS limit` indicates the limit of configured IOPS for both reads and writes combined.
- `Bandwidth limit` indicates the bandwidth limit in megabytes per second (MBps). You can also specify the limit in gigabits per second (Gbps) or terabytes per second (TBps).

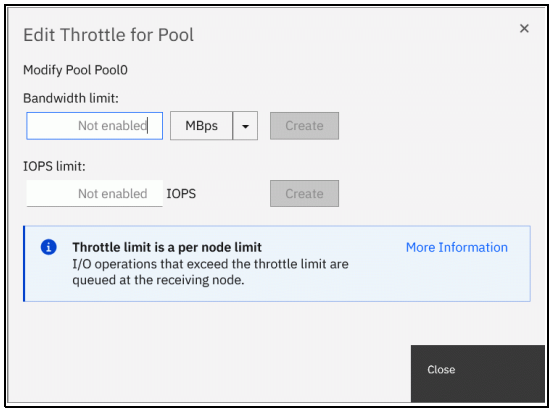


Figure 5-11 Edit throttle for Pool window

If more than one throttle applies to an I/O operation, the lowest and most stringent throttle is used. For example, if a throttle of 100 MBps is defined on a pool and a throttle of 200 MBps is defined on a volume of that pool, the I/O operations are limited to 100 MBps.

The throttle limit is a per node limit. For example, if a throttle limit is set for a pool at 100 IOPS, each node on the system that can access the volumes in the pool allows 100 IOPS. Any I/O operation that exceeds the throttle limit is queued at the receiving nodes. The multipath policies on the host determine how many nodes receive I/O operations and the effective throttle limit.

If a throttle exists for the storage pool, the dialog box that is shown in Figure 5-11 also shows the Remove button that is used to delete the throttle.

To set a storage pool throttle by using the CLI, run the **mkthrottle** command. Example 5-5 shows a storage pool throttle, named `iops_bw_limit` that is set to 3 Mbps and 1000 IOPS on `Pool0`.

Example 5-5 Setting a storage pool throttle by using the mkthrottle command

```
IBM_2145:ITSO-SV1:superuser>mkthrottle -type mdiskgrp -iops 1000 -bandwidth 3
-name iops_bw_limit -mdiskgrp Pool0
Throttle, id [0], successfully created.
```

To remove a throttle by using the CLI, run the **rmthrottle** command. The command uses the throttle ID or throttle name as an argument, as shown in Example 5-6. The command returns no feedback if it runs successfully.

Example 5-6 Removing a pool throttle by running the rmthrottle command

```
IBM_2145:ITSO-SV1:superuser>rmthrottle iops_bw_limit
IBM_2145:ITSO-SV1:superuser>
```

View All Throttles window

You can display the defined throttles by using the Pools panel. Right-click a pool and select **View all Throttles** to display the list of the pool throttles defined for all Pools in the system. If you want to view the throttle of other elements (such as Volumes or Hosts), you can select **All Throttles** in the drop-down list, as shown in Figure 5-12.

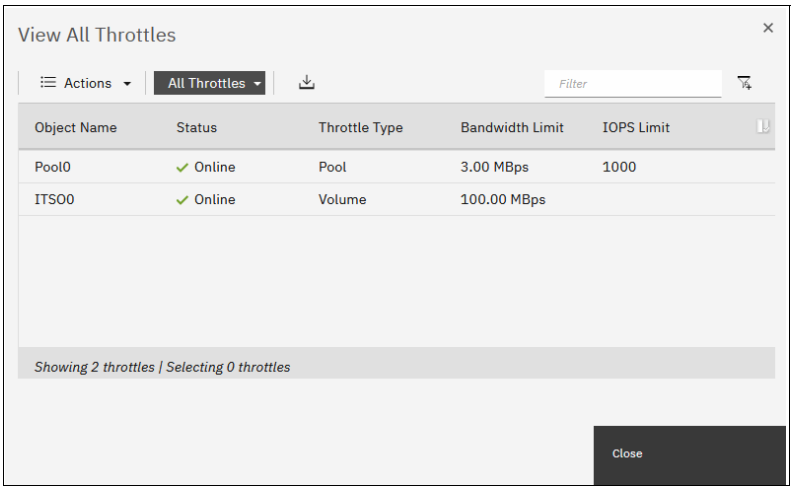


Figure 5-12 Viewing all throttles

To see a list of created throttles by using the CLI, run the **lsthrottle** command. When you run the command without arguments, it displays a list of all throttles on the system.

To list only storage pool throttles, specify the **-filtervalue throttle_type=mdiskgrp** parameter.

View Resources window

Select this action to browse a list of MDisks that are part of the storage pool. Click **View Resources** to open the window that is shown in Figure 5-13.

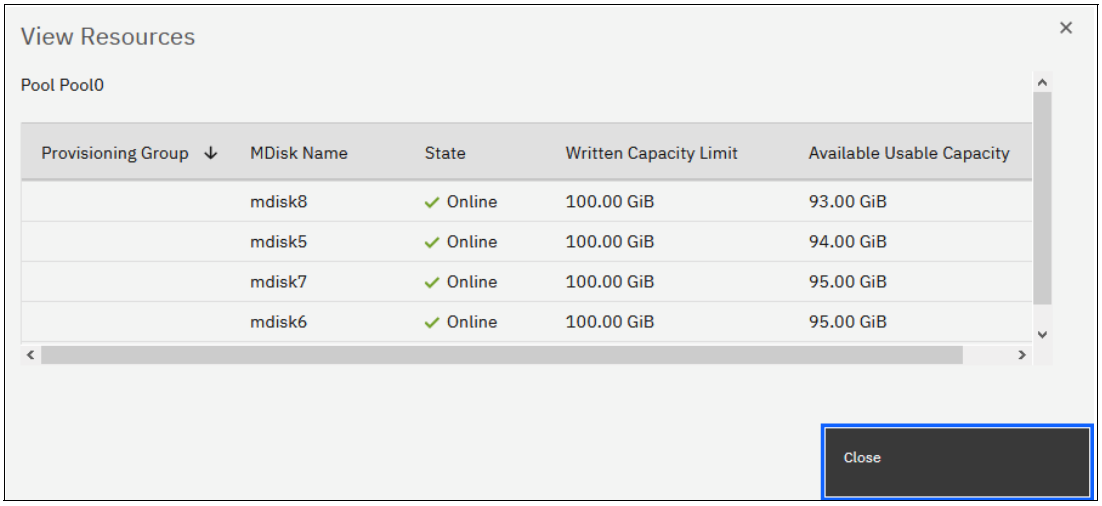


Figure 5-13 List of resources in the storage pool

To list storage pool resources by using the CLI, run the **`lsmdisk`** command. You can filter the output to display MDisk objects that belong only to a single MDisk group (storage pool), as shown in Example 5-7.

Example 5-7 Using lsmdisk (some columns are not shown)

```
IBM_2145:ITSO-SV1:superuser>lsmdisk -filtervalue mdisk_grp_name=Pool0
id name    status mode    mdisk_grp_id mdisk_grp_name    capacity
5  mdisk5  online managed 0            Pool0           100GB
6  mdisk6  online managed 0            Pool0           100GB
7  mdisk7  online managed 0            Pool0           100GB
8  mdisk8  online managed 0            Pool0           100GB
```

View Easy Tier Reports

View the most recent Easy Tier statistics. For more information about Easy Tier Reports, see “Monitoring Easy Tier by using the GUI” on page 683.

Deleting a storage pool

A storage pool can be deleted by using the GUI only if no volumes are associated with it. Select **Delete** to delete the pool immediately without any other confirmation.

If volumes exist in the pool, the Delete option is inactive and cannot be selected. Delete the volumes or migrate them to another storage pool before proceeding. For more information about volume migration and volume mirroring, see Chapter 6, “Volumes” on page 419.

After you delete a pool, the following actions occur:

- All the external MDisks in the pool return to `Unmanaged` mode.
- All the array mode MDisks in the pool are deleted and all member drives return to a Candidate status.

Important: Be careful when you delete a pool with CLI and use **`-force`** parameter. Unlike the GUI, it does not prevent you from deleting a storage pool with volumes. This command deletes all volumes and host mappings on a storage pool, and they **cannot** be recovered.

Properties for Pool window

Select **Properties** to display information about the storage pool, as shown in Figure 5-14. By hovering your cursor over the elements of the window and clicking **[?]**, you see a short description of each property.

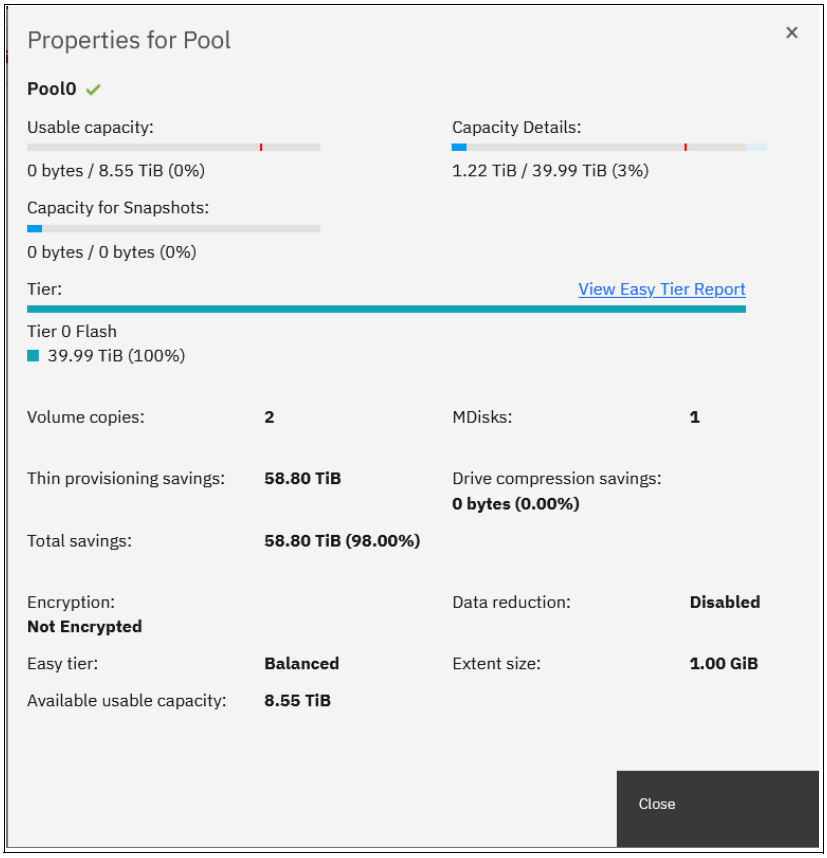


Figure 5-14 Pool properties and details

To display detailed information about the properties by using the CLI, run the **`lsmdiskgrp`** command with a storage pool name or ID as a parameter, as shown in Example 5-8.

Example 5-8 The lsmdiskgrp output (partially shown)

```
IBM_2145:ITSO-SV1:superuser>lsmdiskgrp Pool0
id 2
name Pool0
status online
mdisk_count 4
vdisk_count 2
capacity 397GB
extent_size 1024
free_capacity 377GB
<...>
```

5.1.4 Child pools

A *child pool* is a storage pool that is created within another storage pool. The storage pool in which the child storage pool is created is called the *parent storage pool*. Unlike a parent pool, a child pool does not contain MDisks. Its capacity is provided by the parent pool.

A child pool cannot be created within another child pool. Multiple child pools can be created within a single parent pool.

Child pools created from standard pools and from data reduction pools have a significant difference:

- A child pool with a standard pool as a parent has a type child_thick. Child pools of Standard pools have a fixed capacity, which is taken, or reserved, from the parent pool. Free capacity of a parent pool reduces when a child pool is created. Volumes in a child pool of a standard pool cannot occupy more capacity that is assigned to the child.
- A child pool with DRP as a parent, has type child_quotaless. Quotaless child pools share its free and used capacity with the parent pool and do not have their own capacity limit. Free capacity of a DRP does not change when a new quotaless child pool is created.

The capacity of a child_thick type pool is set at creation time, but can be modified later nondisruptively. The capacity must be a multiple of the parent pool extent size and must be smaller than the free capacity of the parent pool.

Child pools of a child_thick type can be used to implement the following configurations:

- Limit the capacity that is allocated to a specific set of volumes

  It can also be useful when strict control over thin-provisioned volume expansion is needed. For example, you might create a child pool with no volumes in it to act as an emergency set of extents so that if the parent pool uses all its free extents, you can use the ones from the child pool.
- As a container for VMware vSphere virtual volumes (VVOLs). Data Reduction Pools are *not* supported as parent pools for VVOL storage.
- Migrate volumes from non-encrypted parent storage pool to encrypted child pools. When you create a child pool of type child_thick after encryption is enabled, an encryption key is created for the child pool, even when the parent pool is not encrypted. You can then use volume mirroring to migrate the volumes from the non-encrypted parent pool to the encrypted child pool.

  Encrypted child_quotaless type child pools can be created only if the parent pool is encrypted. The data reduction child pool inherits an encryption key from the parent pool.

Both types of child pools, with DRP parent and standard pool as a parent, can be used with the following features:

- Safeguarded Copy Backup. A Safeguarded backup location is a child-pool of a parent-pool. A parent-pool cannot be designated as Safeguarded. A maximum of one child pool per parent pool can be assigned as a Safeguarded backup location.
- Object Based Access Controls (OBAC). Ownership groups can be used to restrict access to a subset of storage resources, designated to a child pool, to a specific set of users, as described in 12.5, “Ownership groups principles of operations” on page 1083.
- Specifying different provisioning policies. If you want to have multiple provisioning policies within a single storage pool, you can create multiple child pools inside it and assign a different provisioning policy to each one.

- limiting performance of a subset of volumes. It can be achieved by creating child pool throttles.

Child pools inherit most properties from their parent pools, and these properties cannot be changed. The inherited properties include:

- Extent size
- Easy Tier setting

Creating a child storage pool

To create a child pool, complete the following steps:

1. Select **Pools** → **Pools**. Right-click the parent pool that you want to create a child pool from and select **Create Child Pool**, as shown in Figure 5-15.

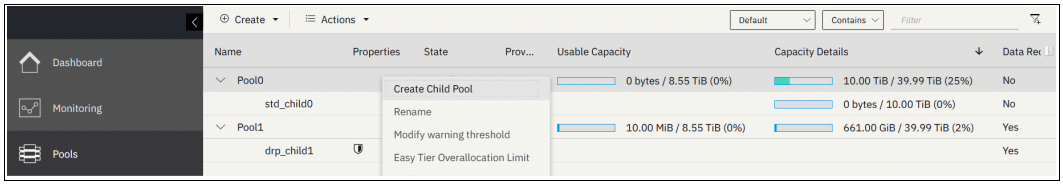


Figure 5-15 Creating a child pool

2. When the dialog box opens, enter the name of the child pool and click Create. Figure 5-16 on page 380 shows a dialog for pool type child_thick, created from a parent Pool0, without any provisioning policy assigned and with the capacity of 10TB.

   You can enable the child pool to store backups of volumes that are protected by the Safeguarded Copy function by selecting the **Safeguard** option. For more information about Safeguarded copy function, see *IBM FlashSystem Safeguarded Copy Implementation Guide*, REDP-5654. If ownership groups are configured, you will also be suggested to assign a group to the child pool being created.

   After a child pool is created, it is listed in the Pools panel under its parent pool. You can toggle the sign to the left of the parent storage pool icon to show or hide the list of child pools.

To create a child pool by using the CLI, run the **mkmdiskgrp** command. You must specify the parent pool for your new child pool and its size for pool type child_thick, as shown in Example 5-9. The size is in megabytes by default (unless the **-unit** parameter is used) and must be a multiple of the parent pool's extent size. In this case, it is 100 * 1024 MB = 100 GB.

Example 5-9 The mkmdiskgrp command to create child pools

```
IBM_2145:ITSO-SV1:superuser>mkmdiskgrp -parentmdiskgrp Pool0 -size 102400 -name
Pool0_child0
MDisk Group, id [4], successfully created
```

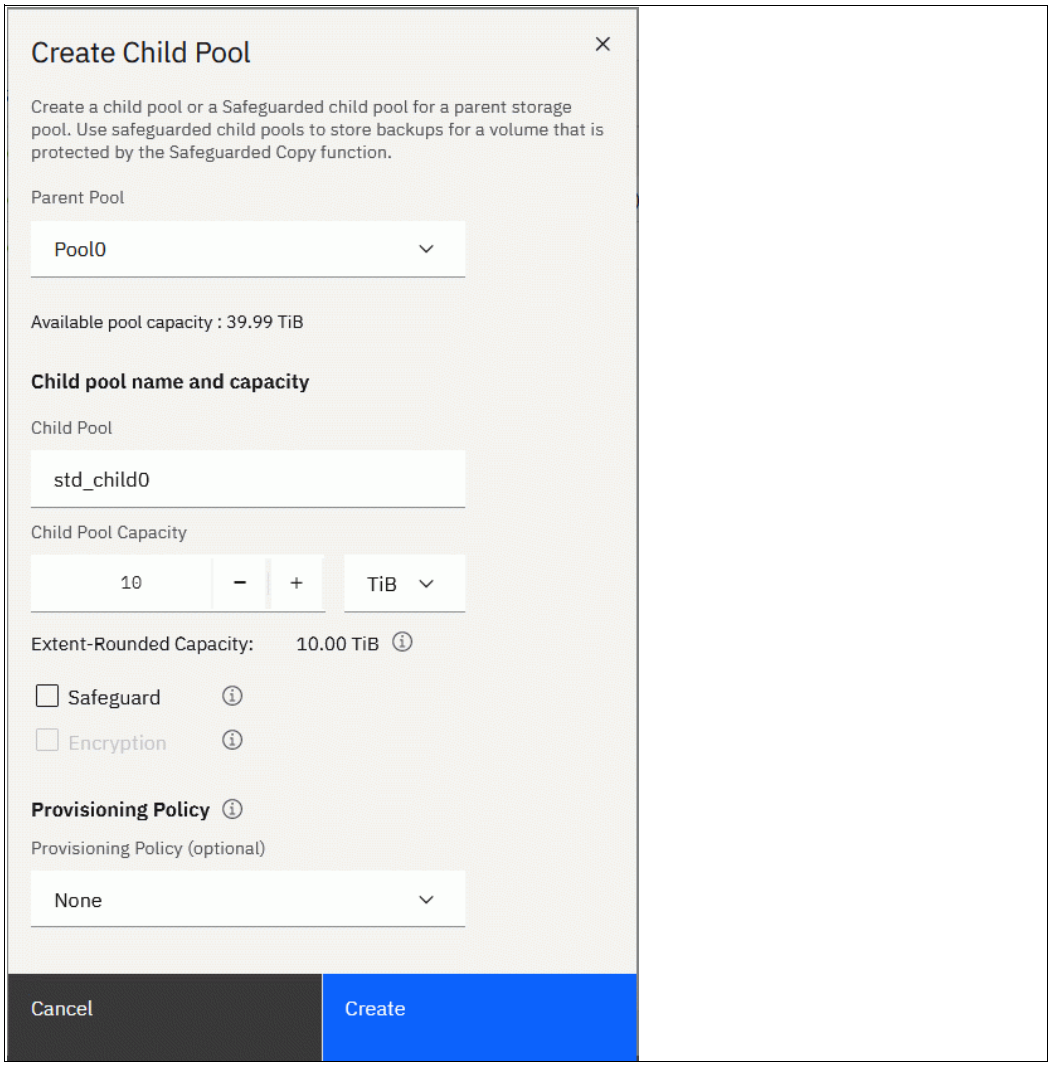


Figure 5-16 Child pool creation window

Child storage pool actions

A list of actions available on child pools has most of the entries available for a parent pool, shown on Figure 5-9 on page 372. You can rename a child pool, assign provisioning policy to it, modify its capacity only for child_thick type, set and modify pool throttle, and delete a child pool. Also, its warning threshold can be modified and it can be assigned it to an ownership group.

To resize the child pool, complete the following steps:

1. Right-click the child storage pool. Alternatively, select the storage pool and then click **Actions**.
2. Select Resize to increase or decrease the capacity of the child storage pool type child_thick, as shown in Figure 5-17 on page 381. Enter the new pool capacity and click **Resize**.

Note: You cannot shrink a child pool below its real capacity. Therefore, the new size of a child pool must be larger than the capacity that is used by its volumes.

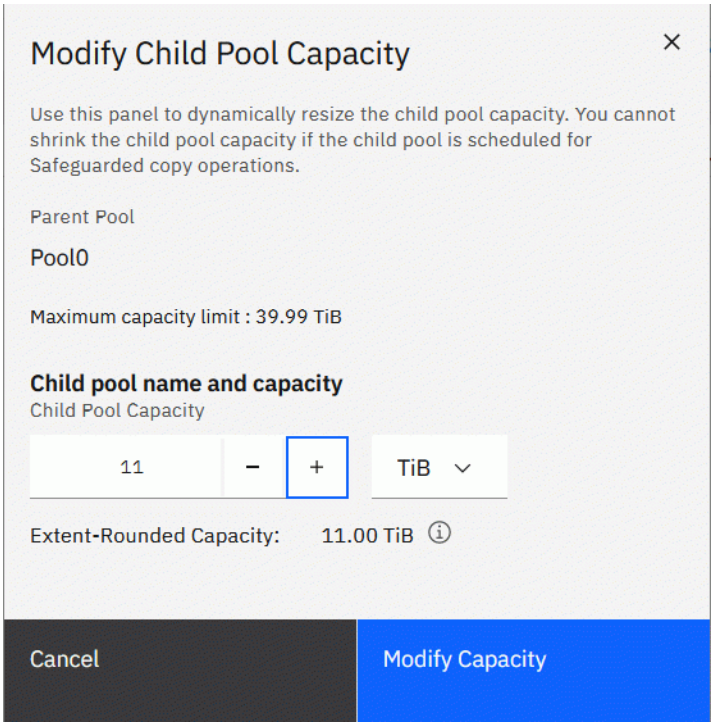


Figure 5-17 Resizing a child pool

When the child pool is reduced, the system resets the warning threshold and issues a warning if the threshold is reached.

To rename and resize child pool by using the CLI, run the **chmdiskgrp** command. Example 5-10 renames the child pool `std_child0` to `child_new` and reduces its size to 44 GB. If successful, the command returns no feedback.

Example 5-10 Running the chmdiskgrp command to rename a child pool

```
IBM_2145:ITSO-SV1:superuser>chmdiskgrp -name child_new -size 61440 std_child0
IBM_2145:ITSO-SV1:superuser>
```

Deleting a child pool is a task that is like deleting a parent pool. As with a parent pool, the Delete action is disabled if the child pool contains volumes, as shown in Figure 5-18.

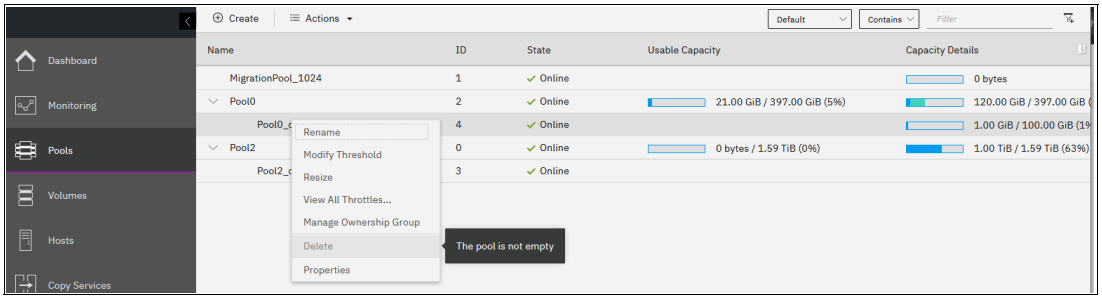


Figure 5-18 Deleting a child pool

After deleting a child pool type child_thick, the extents that it occupied return to the parent pool as free capacity.

To delete a child pool by using the CLI, run the **rmmdiskgrp** command.

To assign an ownership group to a child pool, click **Manage Ownership Group**, as shown in Figure 5-19. All volumes created in the child pool inherit the ownership group of the child pool. For more information, see 12.5, “Ownership groups principles of operations” on page 1083.

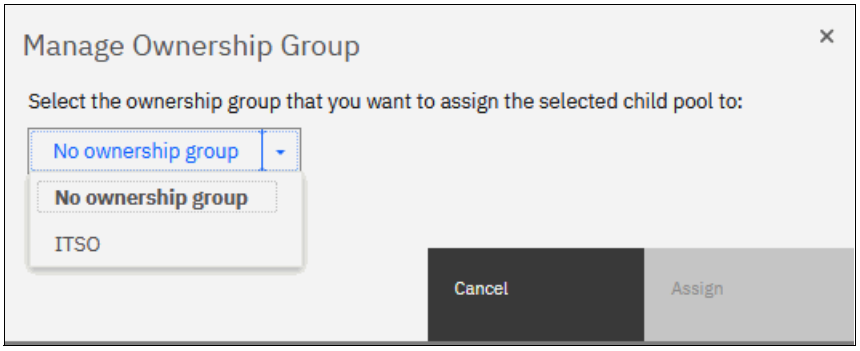


Figure 5-19 Managing the ownership group of a child pool

Migrating volumes to and from child pools

To move a volume to another pool, you can use migration or volume mirroring in the same way that you use them for parent pools. For more information about volume migration and volume mirroring, see Chapter 6, “Volumes” on page 419.

The system supports migrating volumes between child pools within the same parent pool or migration of a volume between a child pool and its parent pool. Migrations between a source and target child pool with different parent pools are not supported. However, you can migrate the volume from the source child pool to its parent pool. Then, the volume can be migrated from the parent pool to the parent pool of the target child pool. Finally, the volume can be migrated from the target parent pool to the target child pool.

The migration of a volume to a safeguarded child pool is not supported. For more information about Safeguarded copy function, see *IBM FlashSystem Safeguarded Copy Implementation Guide*, REDP-5654.

During a volume migration within a parent pool, such as between a child pool and its parent or between child pools with the same parent, no data is moved, but extent reassignments occur.

Volume migration between a child storage pool and its parent storage pool can be performed by going to the Volumes by Pool page and clicking **Volumes**. Right-click a volume and select it to migrate it into a suitable pool.

In the example that is shown in Figure 5-20, the volume `vdisk0` was created in the child pool `Pool0_child_new`. The child pools appear the same as the parent pools in the Volumes by Pool panel.

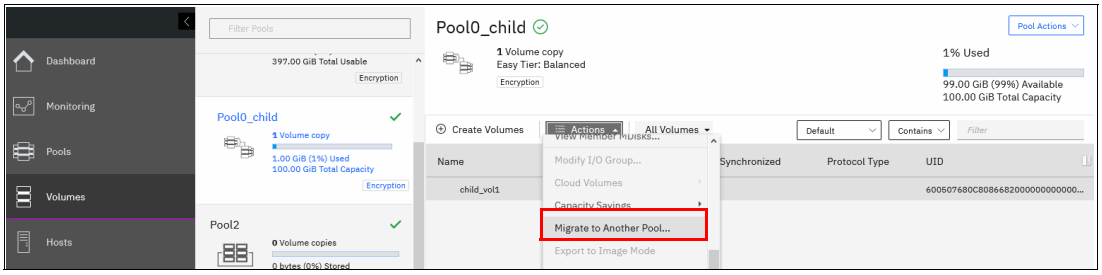


Figure 5-20 Actions menu in Volumes by Pool

For more information about CLI commands for migrating volumes to and from child pools, see Chapter 6, “Volumes” on page 419.

5.1.5 Encrypted storage pools

The system supports optional encryption for Data-at-Rest. After you have an encryption license installed and encryption enabled, all new pools by default will have encryption enabled.

All existing unencrypted pools remain unencrypted. To encrypt data on such pools, migrate volumes into the new encrypted pool or a child pool.

Working with encrypted pools is explained in detail in 12.7, “Encryption” on page 1098.

5.2 Working with provisioning policies

Provisioning policy is a set of rules that define capacity savings attributes of a volume. Only provisioning policies are described in this section. For replication policies and snapshot policies, refer to Chapter 10, “Advanced Copy Services” on page 717.

5.2.1 Introduction to provisioning policies

Each pool or a child pool can be associated with a provisioning policy. When a policy is assigned, all volumes that are created in the pool adopt capacity savings parameters from the defined policy. Provisioning policies allow a simpler and consistent provisioning process, and they are also useful with automation software.

When policy-based replication is implemented, provisioning policy is used to define capacity savings of an automatically-created replication secondary volumes.

Only one provisioning policy can be assigned to a pool or to a child pool. A single policy can be assigned to multiple pools.

Note: Provisioning policy will not change any parameters of volumes that already exist in the pool when a policy is assigned. If you already have volumes in the pool, after assigning a provisioning policy you might need to change volumes capacity savings settings manually.

5.2.2 Creating and deleting provisioning policies

To create or delete a provisioning policy using the GUI, navigate to **Policies** → **Provisioning Policies,** as show in Figure 5-21 on page 384. Figure 5-21 on page 384 also shows default policies that are created when a new system is deployed.

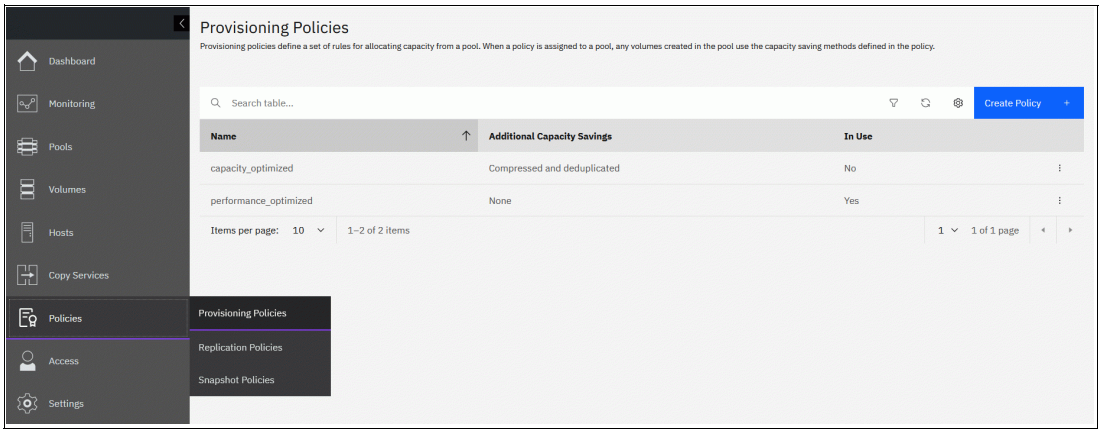


Figure 5-21 Provisioning Policies list

To create a new provisioning policy, click **Create Policy** button. A dialog is shown in Figure 5-22. A new policy can have one of five capacity savings types:

1. None: with this setting, volumes are created as Fully Allocated, without any capacity savings. The default policy *performance_optimized* is defined with this type.
2. Thin-provisioned
3. Thin-provisioned and deduplicated
4. Compressed
5. Compressed and deduplicated. Default policy *capacity_optimized* uses this setting.

After specifying the policy name and capacity savings type, click **Create** to complete the provisioning policy creation process.

To delete or rename a provisioning policy, click **Rename** at the end of the line of that policy to open the actions menu, as shown in Figure 5-23 on page 385.

You cannot delete a policy if it is in use, assigned to any of the pools or child pools. If you attempt to delete it, a message is shown and the action is not performed.

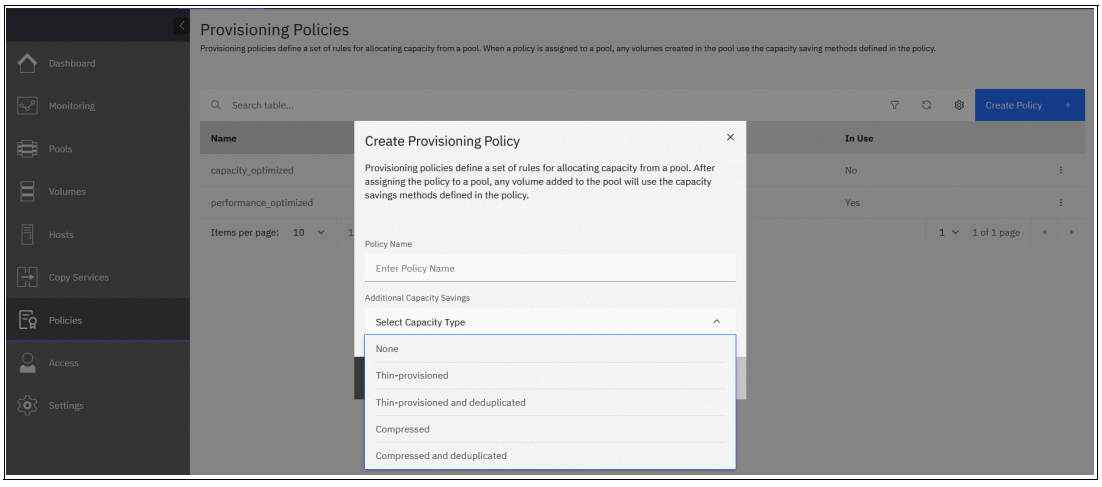


Figure 5-22 Create provisioning policy

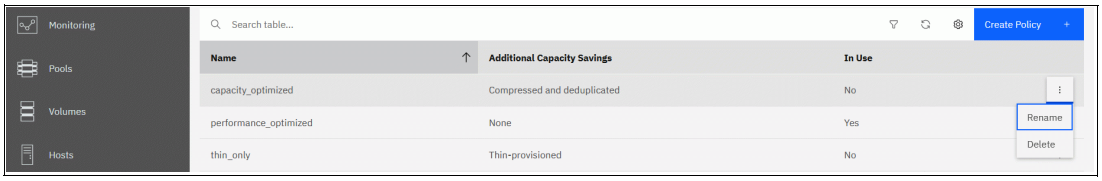


Figure 5-23 Delete or rename provisioning policy

5.2.3 Assigning and unassigning provisioning policies

You can assign a provisioning policy to a storage pool or a child pool during creation time, or anytime later. Note that if you assign a new policy, or change an assigned policy to a pool, all the volumes that already exist in the pool will not change their capacity savings type. Only volumes created after the moment when a policy is assigned or changed adhere to the provisioning policy.

To assign provisioning policy to an existing pool, navigate to **Pools** → **Pools**, right-click the existing pool and select **Assign Provisioning Policy**. The action item changes to **Manage Provisioning Policy**, if there is a policy defined for the pool.

A dialog that opens suggests only policies that match to the best practices and supported capacity savings method for the pool and the platform. For example, as shown in Figure 5-24, policies that define a use of compression and deduplication are not suggested for the standard pools. Pool-level compression, which is different than FCM-level compression, and deduplication is supported only in Data Reduction Pools.

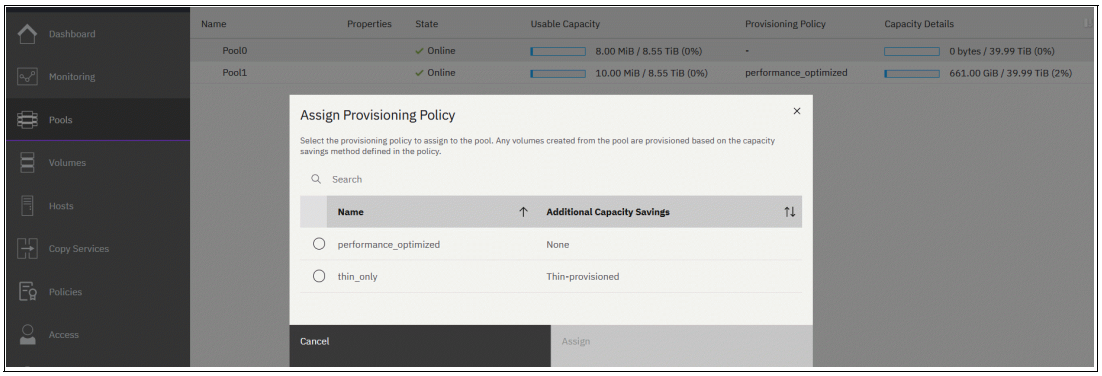


Figure 5-24 Assigning policies to the standard pool

For the Data Reduction Pools with FCM arrays, it is not recommended to use thin-provisioned volumes without pool-level compression because it causes complications in free capacity calculations. Because of this, a provisioning policy that defines the capacity saving type Thin-provisioned is not listed. For DRP with FCMs, recommended volume capacity savings are None (Fully allocated), and Compressed and Deduplicated, so only two possible policies are included in the list, as shown in Figure 5-25 on page 386.

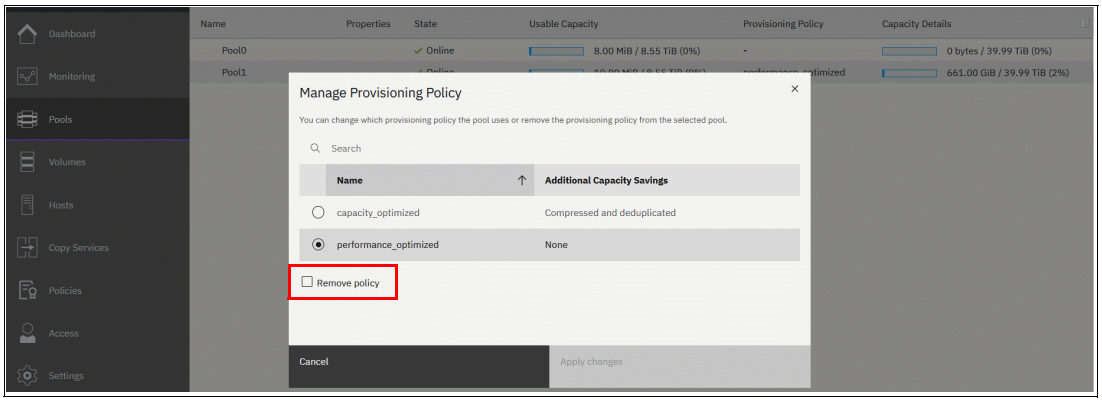


Figure 5-25 Assigning a policy to a DRP and removing the assigned policy

To remove an assigned policy from a pool, select the **Remove policy** checkbox as shown in Figure 5-25, and click **Apply changes**.

5.3 Working with internal drives and arrays

An array is a type of MDisk that consists of disk drives or flash modules. These drives are members of the array. To create high availability (HA) and high-performance groupings of drives, RAID (Redundant Array of Independent Drives) technologies are used.

5.3.1 Working with drives

This section describes how to manage internal storage disk drives and configure them to be used in arrays.

Listing disk drives

The system provides an Internal Storage window for managing all internal drives. To access the Internal Storage window, click **Pools** → **Internal Storage**, as shown in Figure 5-26.

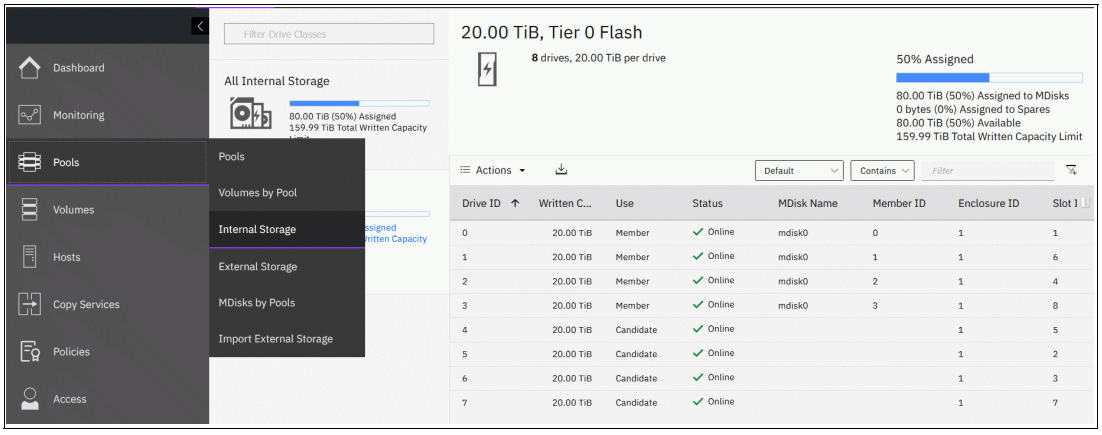


Figure 5-26 Internal storage panel

The window gives an overview of the internal drives in the system. Select **All Internal Storage** in the drive class filter to display all drives that are managed in the system, including all I/O groups and expansion enclosures.

Alternatively, you can filter the drives by their type or class. For example, you can choose to show only Enterprise disks, Nearline (NL) disks, or Flash. Select the class on the left side of the window to filter the list and display only the drives of the selected class.

You can find information about the capacity allocation of each drive class in the upper right corner:

Assigned to MDisks Shows the storage capacity of the selected drive class that is assigned to MDisks.

Assigned to Spares Shows the storage capacity of the selected drive class that is used for spare drives.

Available Shows the storage capacity of the selected drive class that is not yet assigned to MDisks or Spares.

Total Written Capacity Limit Shows the total amount of storage capacity of the drives in the selected class.

If **All Internal Storage** is selected under the Drive Class filter, the values that are shown refer to the entire internal storage.

The percentage bar indicates how much of the total written capacity limit is assigned to MDisks and Spares. MDisk capacity is represented by the solid portion and spare capacity by the shaded portion of the bar.

To list all internal drives available in the system, run the **`lsdrive`** CLI command. If needed, you can filter the output to list only drives that belong to a specific enclosure, that have a specific capacity, or by another attribute (see Example 5-11).

Example 5-11 lsdrive output (some lines and columns are not shown)

```
IBM_2145:ITSO-SV1:superuser>lsdrive
id status error_sequence_number use       tech_type       capacity mdisk_id
0  online                       spare     tier_enterprise 1.1TB
1  online                       spare     tier_enterprise 1.1TB
2  online                       member    tier_nearline   931.0GB  16
3  online                       candidate tier_enterprise 1.1TB
4  online                       candidate tier_enterprise 1.1TB
5  online                       member    tier_nearline   931.0GB  16
<...>
```

The drive list shows the status of each drive. A drive can be `Online`, which means that the drive is fully accessible by both nodes in the I/O group. A `Degraded` drive is accessible only by one of the two nodes. An `Offline` status indicates that the drive is not accessible by any of the nodes; for example, because it was physically removed from the enclosure or because it is unresponsive or failing.

The drive `Use` attribute describes the role that it plays in the system. The values and meanings are available:

Unused

The system can access the drive but was not told to take ownership of it. Most actions on the drive are not permitted. This state is a safe state for newly added hardware.

Candidate

The drive is owned by the system, and is not part of the RAID configuration. It is available to be used in an array MDisk.

Spare

The drive is a hot spare that protects nondistributed, traditionalRAID arrays. If any member of the array fails, a spare drive is taken and becomes a `Member` for rebuilding the array.

Member

The drive is part of a RAID array.

Failed

The drive is owned by the system and was diagnosed as faulty. It is waiting for a service action.

The drive use can transition between different values, but not all transitions are valid, as shown in Figure 5-27.

		to				
		unused	candidate	failed	member	spare
from	unused	-	yes	no	no	no
	candidate	yes	-	yes	yes	yes
	failed	no	yes	yes	yes	yes
	member	no	yes	yes	-	no
	spare	no	yes	yes	yes	-

Figure 5-27 Drive use transitions

The system automatically sets the drive use to `Member` when creating a RAID array. The drive can be changed from `Member` to `Failed` only if the array does not depend on the drive and more confirmation is required when taking a drive offline when no spare is available. Transitioning a `Candidate` drive to `Failed` is possible only by using the CLI.

Note: To start configuring arrays in a new system, all `Unused` drives must be configured as `Candidates`. The Initial setup GUI wizard or Assign storage GUI wizard can complete this process automatically.

Several actions can be performed on internal drives. To perform any action, select one or more drives and right-click the selection, as shown in Figure 5-28, or click **Actions**.

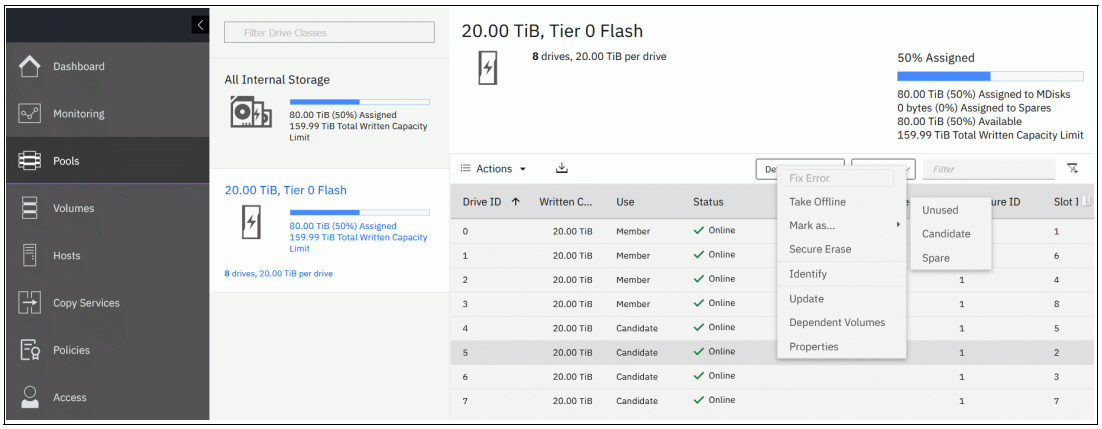


Figure 5-28 Actions on internal storage

The actions that are available in the drop-down menu depend on the status and use of the drives that are selected. Some actions can be performed only on drives in a specific state, and some are possible only when a single drive is selected.

Action: Fix error

This action is available only if the selected drive has an error event that is associated with it. Select **Fix Error** to start the directed maintenance procedure (DMP) for the selected drive. For more information about DMPs, see Chapter 11, “Reliability, availability, and serviceability” on page 963.

Action: Take Offline

If a problem is identified with a specific drive, you can select **Take Offline** to take the drive offline. You must confirm the action by clicking **Take Offline**. If a drive is an array member, taking it offline will decrease array redundancy, so a warning is displayed, as shown in Figure 5-29.

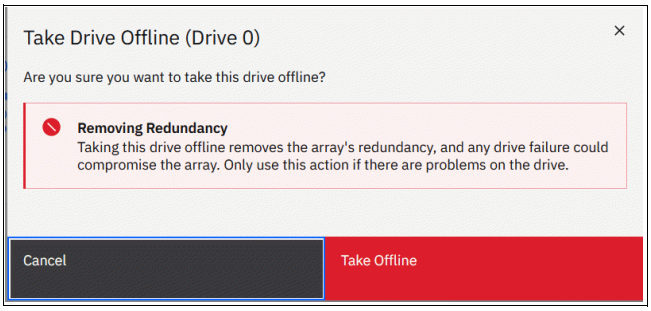


Figure 5-29 Taking a drive offline if a spare or rebuild area is available

If a spare drive or rebuild area is available and the drive is taken offline, the associated MDisk remains `Online` and the RAID array starts a rebuild. If no spare is available, the status of the associated MDisk becomes `Degraded`. The status of the storage pool to which the MDisk belongs becomes `Degraded` as well.

A drive that is taken offline is considered `Failed`, as shown in Figure 5-30.

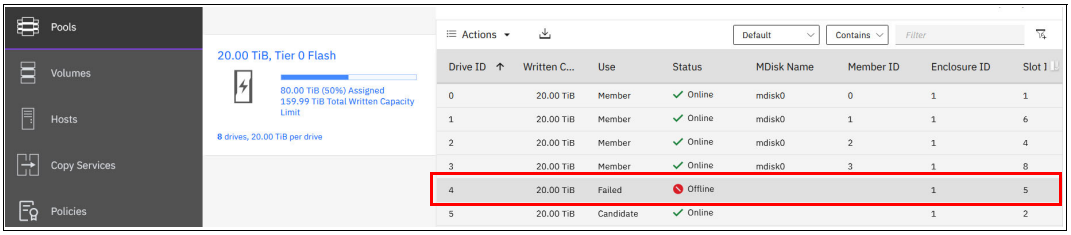


Figure 5-30 An offline drive is marked as failed

To take a drive offline with the CLI, run the **`chdrive`** command (Example 5-12). This command returns no feedback. Use the **`-allowdegraded`** parameter to set a member drive offline, even if no suitable spare is available.

Example 5-12 Setting drive offline with CLI

```
IBM_2145:ITSO-SV1:superuser>chdrive -use failed 3
IBM_2145:ITSO-SV1:superuser>
```

The system prevents you from taking a drive offline if the RAID array depends on that drive and doing so results in a loss of access to data, as shown in Figure 5-31 on page 390.

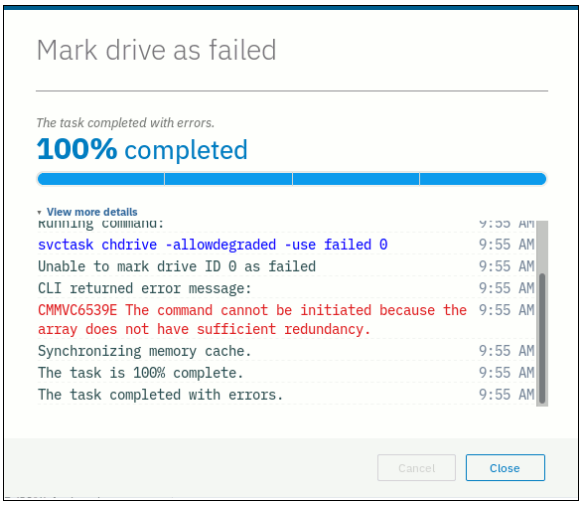


Figure 5-31 Taking a drive offline fails if it might result in a loss of access to data

Action: Mark as

Select **Mark as** to change the designated use of the drive, as shown in Figure 5-28 on page 388. The list of available options depends on the current drive use and state. See the allowed state transitions that are shown in Figure 5-27 on page 388.

To change the drive role with the CLI, run the **chdrive** command (Example 5-13). It shows the drive that was set offline with a previous command is set to spare. Notice that it cannot go from `Failed` to `Spare` use in one step. It must be assigned as a `Candidate` before it be assigned as a `Spare`. Distributed Arrays (DRAIDs) do not use spares. Marking a drive as a spare fails for drive types that support DRAIDs only.

Example 5-13 Changing drive role with CLI

```
IBM_2145:ITSO-SV1:superuser>lsdrive -filtervalue status=offline
id status  error_sequence_number use    tech_type       capacity mdisk_id
3  offline                      failed tier_enterprise 558.4GB
IBM_2145:ITSO-SV1:superuser>chdrive -use spare 3
CMMVC6537E The command cannot be initiated because the drive that you have
specified has a Use property that is not supported for the task.
IBM_2145:ITSO-SV1:superuser>chdrive -use candidate 3
IBM_2145:ITSO-SV1:superuser>chdrive -use spare 3
IBM_2145:ITSO-SV1:superuser>
```

Action: Secure Erase

Select **Secure Erase** to securely delete all the data from the drive before it is retired or decommissioned. The drive status must be `Candidate`, `Unused`, or `Spare`. A drive with a status of `Member` can't be erased. You need to take the drive out of the array by member replacement) or delete the array first.

The exact method which is used for erasing depends on the drive type and the drive's supported commands. For more information, see Secure data deletion.

Action: Identify

Select **Identify** to turn on the light-emitting diode (LED) light of the enclosure slot of the selected drive. This LED helps you to easily locate a drive that must be replaced or that you

want to troubleshoot. A dialog box opens, which confirms that the LED was turned on, as shown in Figure 5-32.

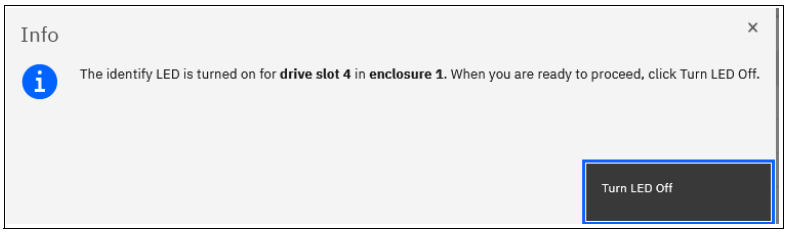


Figure 5-32 Identifying an internal drive

This action makes the amber LED that is associated with the drive that had this action performed flash continuously.

Click **Turn LED Off** when you are finished. The LED is returned to its initial state.

On the CLI, run the **chenclosureslot** command to turn on the LED. See Example 5-14 for commands to locate the enclosure and slot for drive 21 and then to turn the identification LED of slot 3 in enclosure 1 on and off again.

Example 5-14 Changing slot LED to identification mode with CLI

```
IBM_2145:ITSO-SV1:superuser>lsdrive 21
id 21
<...>
enclosure_id 1
slot_id 4
<...>
IBM_2145:ITSO-SV1:superuser>chenclosureslot -identify yes -slot 4 1
IBM_2145:ITSO-SV1:superuser>lsenclosureslot -slot 4 1
enclosure_id 1
slot_id 4
fault_LED slow_flashing
powered yes
drive_present yes
drive_id 21
IBM_2145:ITSO-SV1:superuser>chenclosureslot -identify no -slot 4 1
```

Action: Update and Update All

Select **Update** to install a new version of a drive firmware, as shown in Figure 5-33 on page 392. You can choose to update an individual drive, or select multiple drives. Also, if you want to update all drives, ensure that no drives are selected and then select **Actions** → **Update All**.

To apply drive firmware, run the Update Test Utility. The Test Utility is the same as the one used for system code update. Even if **Test Only** is selected, a drive update package must be uploaded because the system verifies the package validity.

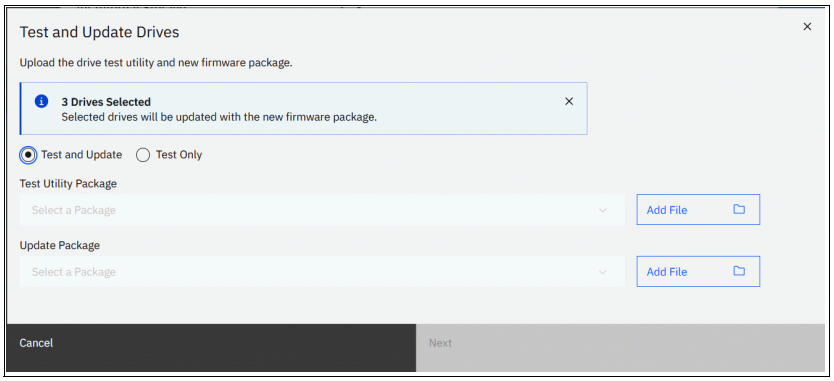


Figure 5-33 Upgrading a drive or a set of drives

When multiple drives are selected for update they are all put in the queue of processes and updated one by one. You can monitor progress with **`lsdriveupgradeprogress`** CLI or GUI **Monitoring** → **Background Processes** page.

For more information about updating drive firmware, see Chapter 11, “Reliability, availability, and serviceability” on page 963.

Action: Dependent volumes

Select **Dependent Volumes** to list the volumes that depend on the selected drives. A volume depends on a drive or a set of drives when removal or failure of that drive or set of drives results in a loss of access or a loss of data for that volume. Use this option before you perform maintenance operations to confirm which volumes (if any) are affected.

Figure 5-34 shows the list of volumes that depend on a set of three drives that belong to the same MDisk. All listed volumes go offline if *all* selected drives go offline at the same time. This does *not* mean that volumes go offline if a single drive or two of the three drives were to go offline.

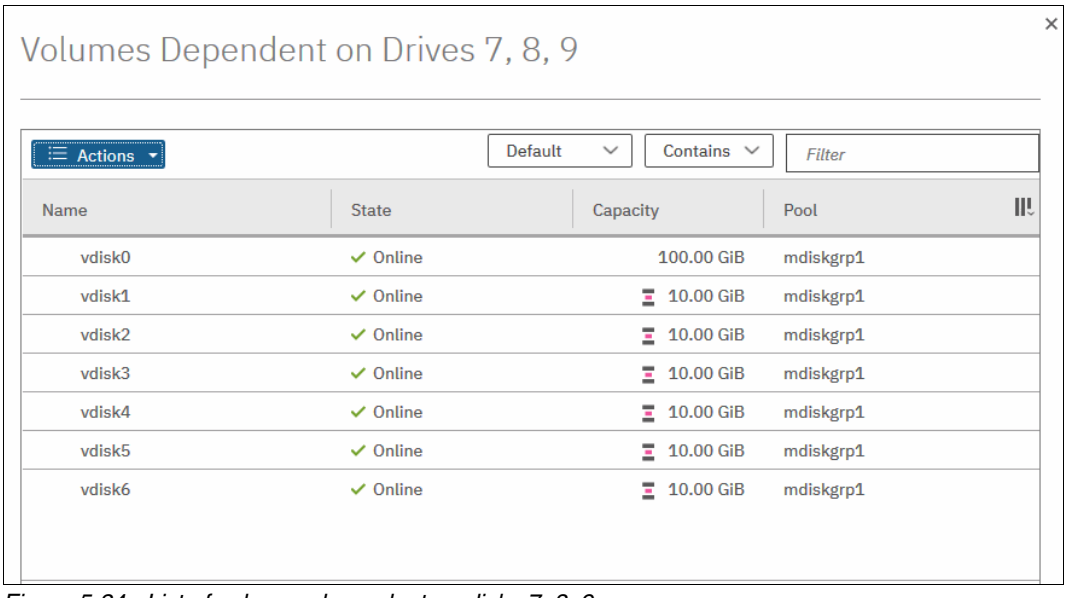


Figure 5-34 List of volumes dependent on disks 7, 8, 9

Whether dependent volumes exist depends on the redundancy of the RAID array at a specific moment. It is based on the RAID level, state of the array, and state of the other member drives in the array. For example, it takes three or more drives going offline at the same time in a healthy RAID 6 array for dependent volumes to exist.

Note: A lack of dependent volumes does not imply that no volumes exist that use the drive. Volume dependency shows the list of volumes that become unavailable if the drive or the set of selected drives becomes unavailable.

You can get the same information by running the CLI command **lsdependentvdisks**. Use the parameter **-drive** with a list of drive IDs that you are checking, separated with a colon (**:**), as shown in Example 5-15.

Example 5-15 Listing volumes dependent on drives with the CLI

```
IBM_2145:ITSO-SV1:superuser>lsdependentvdisks -drive 7:8:9
vdisk_id vdisk_name
0        vdisk0
1        vdisk1
2        vdisk2
3        vdisk3
4        vdisk4
5        vdisk5
6        vdisk6
```

Action: Properties

Select **Properties** to view more information about the drive, as shown in Figure 5-35.

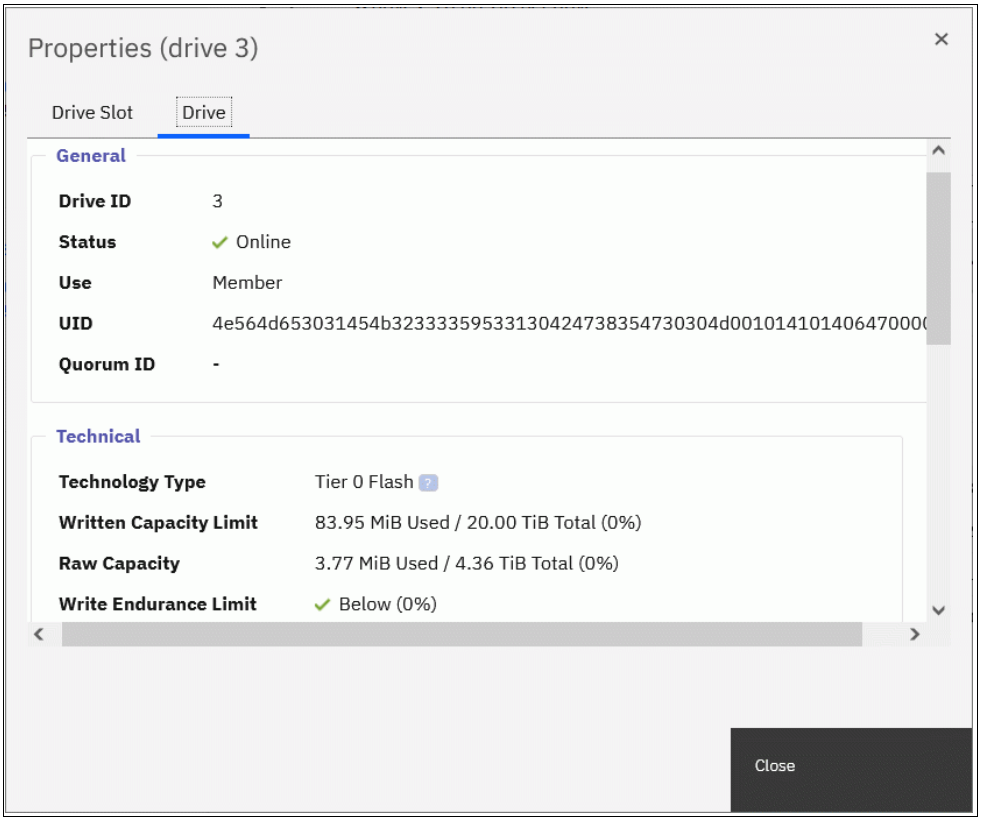


Figure 5-35 Drive properties

You can find a short description of each drive property by hovering over it and then clicking **[?]**. You can also display drive slot details by clicking the **Drive Slot** tab.

To view all available information about the specific drive, run the **`lsdrive`** CLI command with drive ID as the parameter. To view slot information, run the **`lsenclosureslot`** command.

5.3.2 RAID and distributed RAID

To use internal disks in storage pools, they must be joined into RAID arrays to form array mode MDisks.

RAID provides the following key design goals:

- Increased data reliability
- Increased input/output (I/O) performance

RAID technology can provide better performance for data access, HA for the data, or a combination of both. RAID levels define a trade-off between HA, performance, and cost.

In a traditional RAID Storage Virtualize approach (whether it is RAID 10, RAID 5, or RAID 6), data is spread among up to 16 drives in an array. Separate spare drives exist that do not belong to an array and might protect multiple arrays. When one of the drives within the array fails, the system rebuilds the array by using a spare drive. Consider the following points:

- For RAID 10, all data is read from the mirrored copy.
- For RAID 5 or RAID 6, data is calculated from remaining data stripes and parity.

This data is then written to a spare drive. The spare becomes a member of the array when the rebuild starts. After the rebuild is complete and the failed drive is replaced, a member exchange is performed to add the replacement drive to the array and to restore the spare to its original state, so it can act as a hot spare again for another drive failure in the future.

During a rebuild of a RAID array, writes are submitted to a single spare drive that can become a bottleneck and might affect I/O performance. With increasing drive capacity, the rebuild time increases significantly. The probability of a second failure during the rebuild process also becomes more likely. Outside of any rebuild activity, the spare drives are idle and do not process I/O requests for the system.

Distributed Redundant Array of Independent Disks (DRAID) addresses these shortcomings.

Note: Traditional RAID (TRAID) is not supported on the latest platforms. It is recommended to use only Distributed RAID on all Storage Virtualize systems.

Distributed RAID

In DRAID, spare capacity is spread across all member drives to form one or more *rebuild areas*. During a rebuild, the write workload is distributed across all drives, which removes the single drive bottleneck of traditional arrays.

By using this approach, DRAID reduces the rebuild time, the effect on I/O performance during rebuild, and the probability of a second failure during rebuild. As with TRAID, a DRAID 6 array can tolerate two drive failures and survive. If another drive fails in the same array before the array is rebuilt, the MDisk and the storage pool go offline. That is, DRAID has the same redundancy characteristics as a traditional RAID of the same level.

A rebuild after a drive failure reconstructs the data on the failed drive and distributes it across all drives in the array by using a rebuild area. After the failed drive is replaced, a copyback

process copies the data to the replacement drive to free up the rebuild area so that it can be used for another drive failure in the future.

Figure 5-36 shows an example of a DRAID 6 with 10 disks. The capacity on the drives is divided into many packs. The reserved spare capacity (marked in yellow) is equivalent to two spare drives, but the capacity is distributed across all of the drives (depending on the pack number) to form two rebuild areas. The data is striped similar to a TRAID array, but the number of drives in the array can be larger than the stripe width.

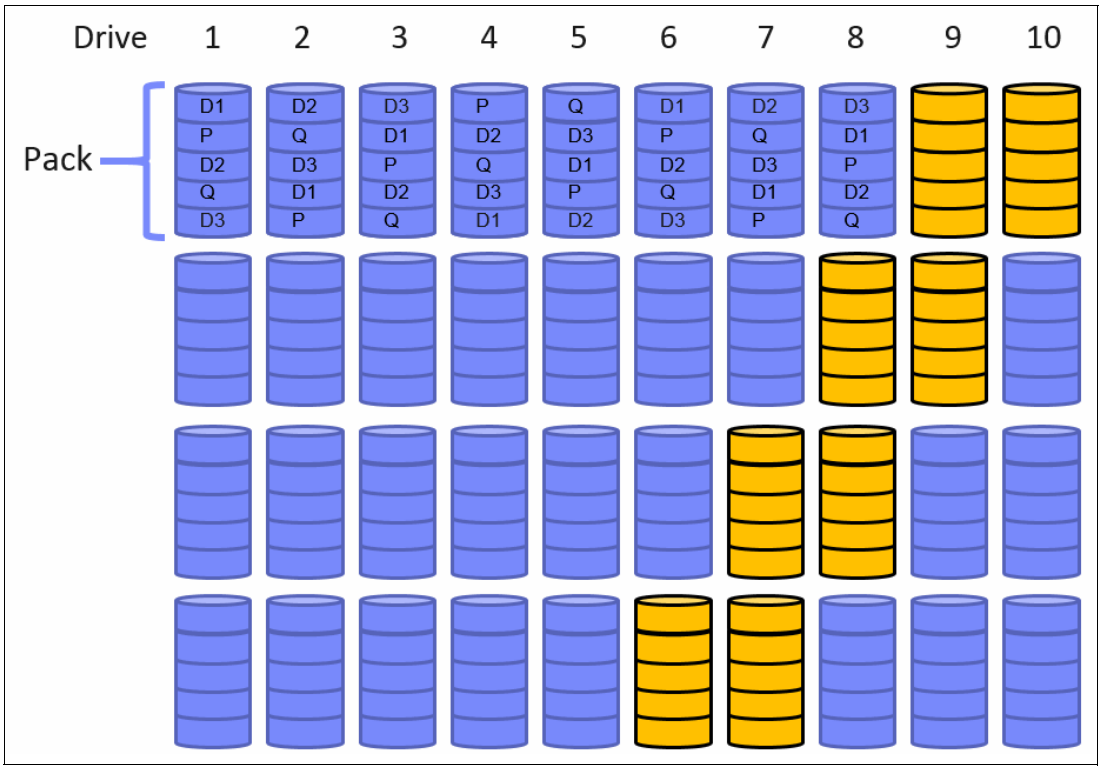


Figure 5-36 Distributed RAID 6 (for simplification, not all packs are shown)

Figure 5-37 on page 396 shows what happens after a single drive failure in this DRAID 6 array. Drive 3 failed and the array uses half of the spare capacity in each pack (marked in green) to rebuild the data of the failed drive.

All drives are involved in the rebuild process, which significantly reduces the rebuild time. One of the two distributed rebuild areas is in use, but the second rebuild area can be used to rebuild the array after another failure.

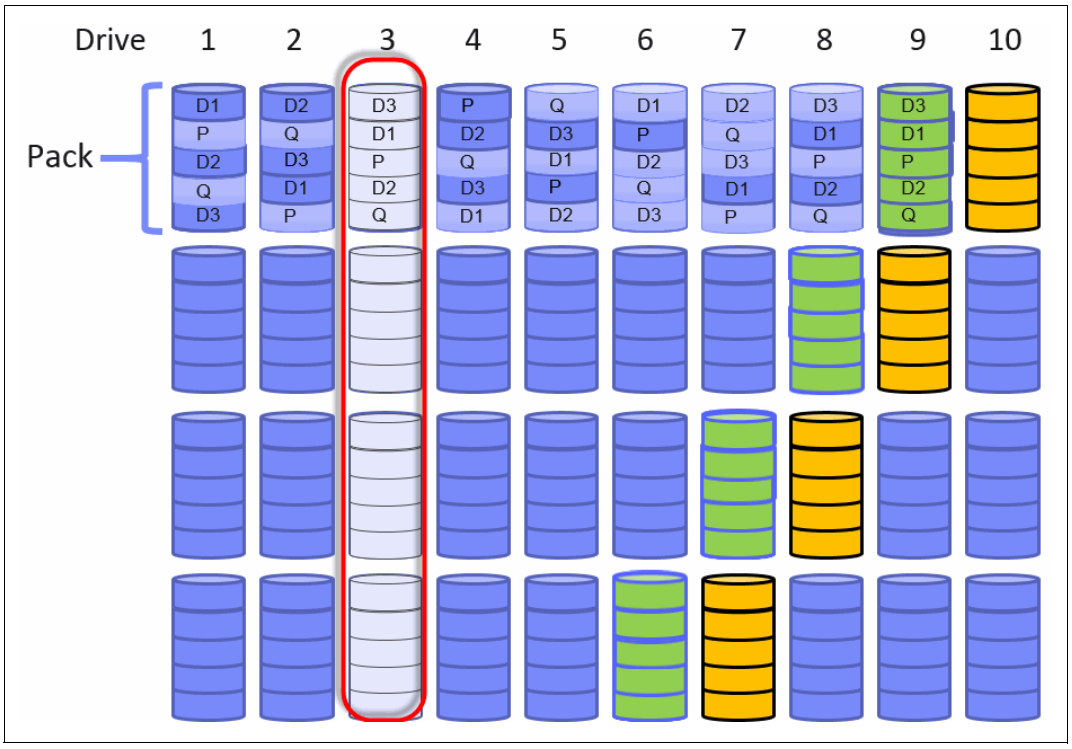


Figure 5-37 Single drive failure with DRAID 6 (for simplification, not all packs are shown)

After the rebuild completes, the array can sustain two more drive failures, even before drive 3 is replaced. If no rebuild area is available to perform a rebuild after another drive failure, the array becomes `Degraded` until a rebuild area is available again and the rebuild can start.

After drive 3 is replaced, a copyback process copies the data from the occupied rebuild area to the replacement drive to empty the rebuild area and make sure it can be used again for a new rebuild.

DRAID addresses the main disadvantages of TRAID while providing the same redundancy characteristics:

- In a drive failure, data is read from many drives and written to many drives. This process minimizes the effect on performance during the rebuild process. Also, it significantly reduces rebuild time. Depending on the distributed array configuration and drive sizes, the rebuild process can be up to 10 times faster.
- Spare space is distributed throughout the array, which means more drives are processing I/O, and no dedicated spare drives are idling.

The DRAID implementation has the following other advantages:

- Arrays can be much larger than before and can span many more drives, which improves the performance of the array. A DRAID can contain a maximum of 128 drives.
- Distributed arrays can be expanded by adding one or more drives. Traditional arrays cannot be expanded.
- Distributed arrays use all node CPU cores to improve performance, especially in configurations with a few arrays.

5.3.3 Creating arrays

Only RAID arrays (array mode MDisks) can be added to a storage pool. It is not possible to add a Just A Bunch Of Disks (JBOD) or a single drive. It is also not possible to create a RAID array without assigning it to a storage pool.

Note: DRAID 6 is recommended for most use cases. DRAID technology dramatically reduces rebuild times, decreases the exposure volumes have to the extra load of recovering redundancy, and improves performance. For more information, see *Performance and Best Practices Guide for IBM Storage FlashSystem and IBM SAN Volume Controller: Updated for IBM Storage Virtualize Version 8.6*, SG24-8543.

To create a RAID array from internal storage, select **Pools** → **Pools**, **Actions** and then, select **Add Storage** or right-click the storage pool to which you want to add arrays and select **Add Storage**, as shown in Figure 5-38.

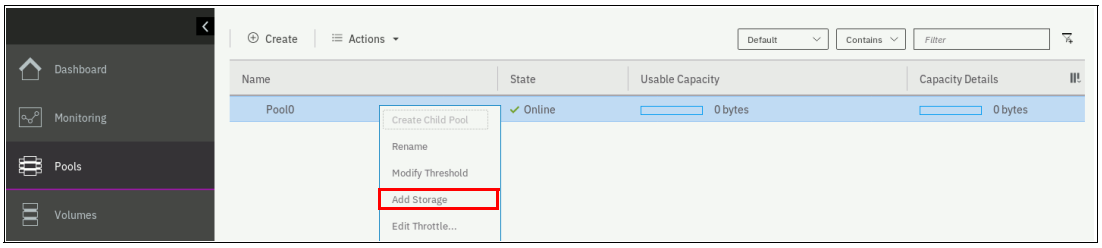


Figure 5-38 Adding storage to a pool

Alternatively, select **Pool** → **Mdisk by pools** and click **Assign**, as shown in Figure 5-39.

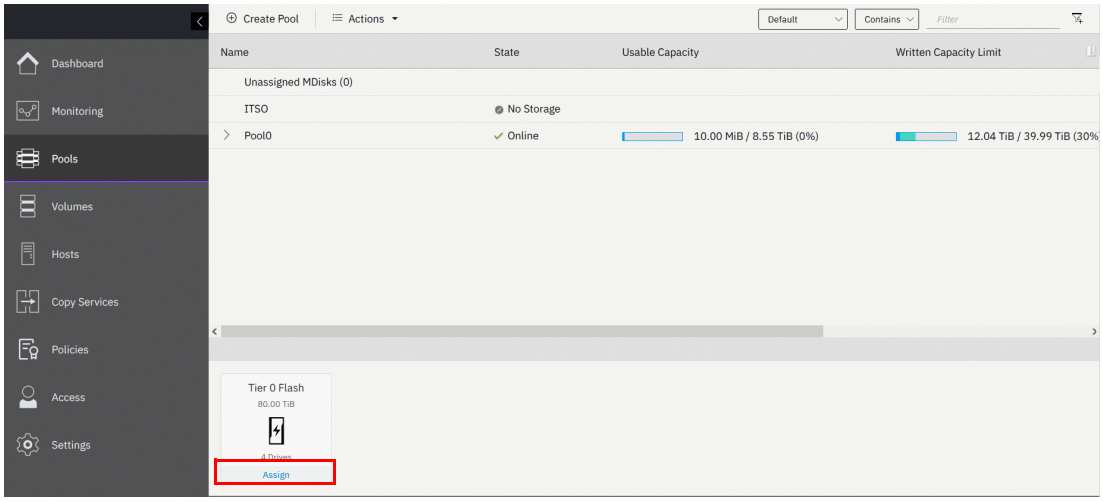


Figure 5-39 Assign drives

This action starts the configuration dialog box that is shown in Figure 5-40. You must click **Define Array** if you use the first method of opening the dialog. If any of the drives are found with an `Unused` role, reconfigure them as `Candidates` to be included into the configuration.

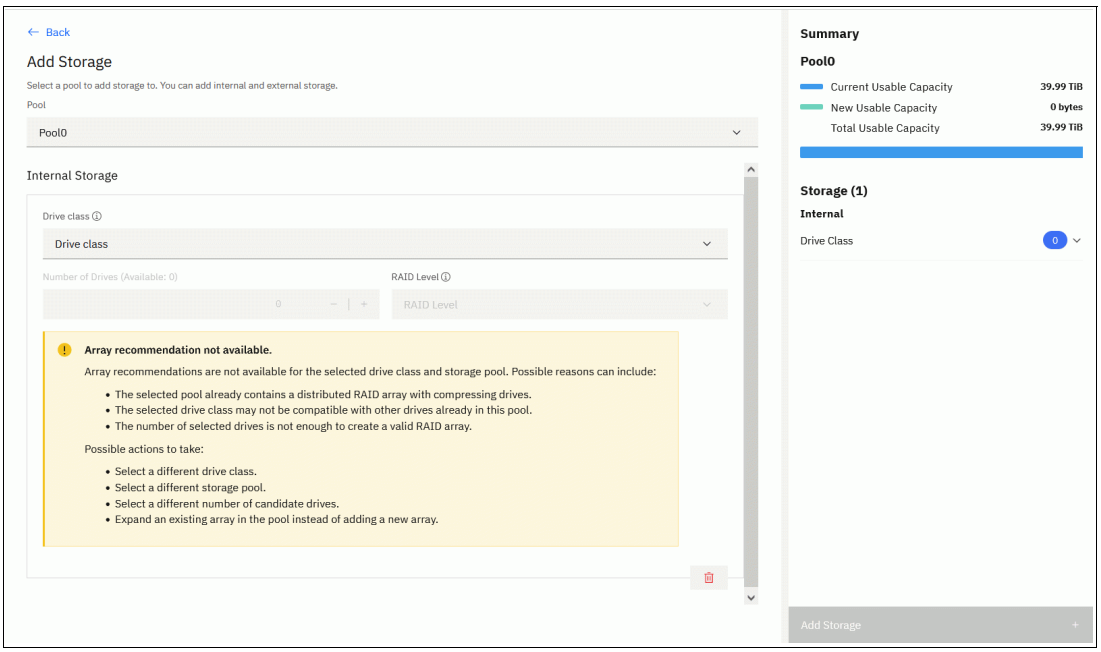


Figure 5-40 Assigning storage to a pool

If Internal Storage is chosen, the system guides you through the array MDisk creation process. If External Storage is selected, the system guides you through the selection of external storage.

Select the pool from the drop-down menu if no pool is selected. The summary view in the right panel shows the Current Usable Capacity of the selected pool. After you defined Internal or External storage or both, click **Add storage**.

Defining an Array

Select **Define Array,** and choose the drive class for the array from the drop-down menu. Only drive classes for which candidate drives exist are displayed. The system automatically recommends a raid type and level that is based on the available candidate drives.

Note: If you are adding storage to a pool that already has storage assigned, the system adheres to best practices when suggesting the new array settings. The goal is to achieve a balanced configuration. Therefore, some properties are inherited from existing arrays in the pool for a specific drive class. It is not possible to add RAID arrays that are different from arrays in a pool by using the GUI. Also it is not possible to add a second array with compressing drives to the same pool, only one is allowed.

Select **Advanced**, as shown in Figure 5-41 on page 399, to adjust the number of spares or rebuild areas, the stripe width and array width before the array is created. Depending on the adjustments that are made, the system might select a different RAID type and level. The summary view on the right panel can be expanded to preview the details of the arrays that are going to be created.

Note: It is not possible to change the RAID level or stripe width of an existing array. You also cannot decrease the drive count of an array after it is created. To change these properties you must delete the array MDisk and re-create it with the required settings.

The summary window reflects the usable capacity based on the selected settings.

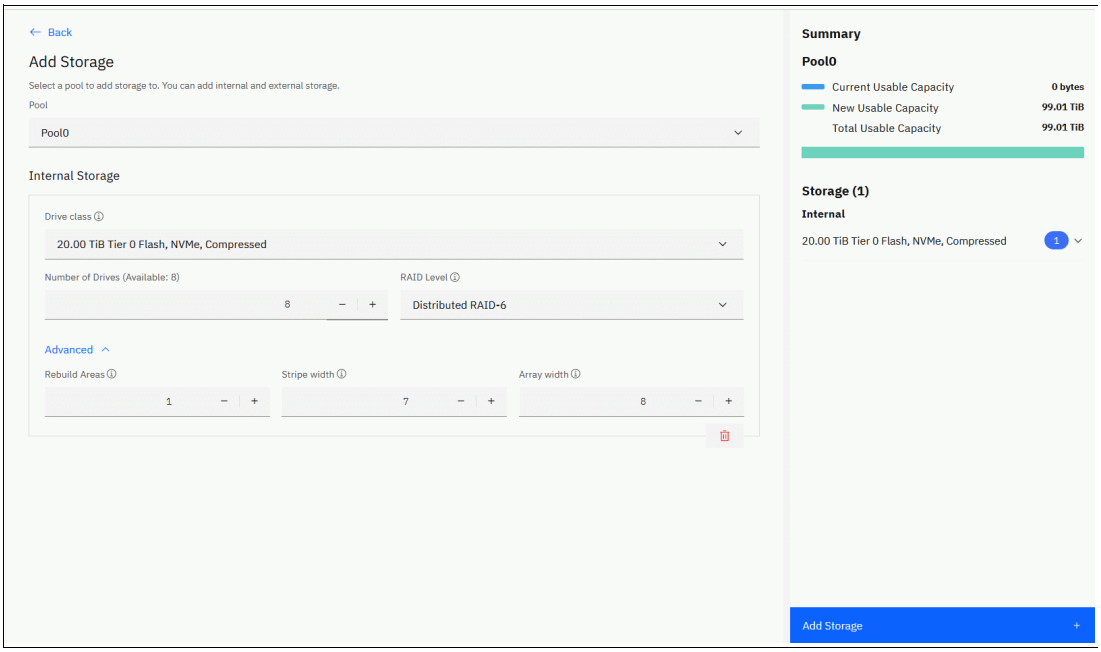


Figure 5-41 Assign Storage to Pool

The stripe width indicates the number of strips of data that can be written at one time when data is rebuilt after a drive fails. This value is also referred to as the *redundancy unit width*.

A stripe, which can also be referred to as a redundancy unit, is the smallest amount of data that can be addressed. Distributed array strip size is 256 KB.

If the system has multiple drive classes, for example, flash and enterprise drives, use the plus symbol (**+**) to create more arrays from other drive classes to take advantage of Easy Tier. The plus symbol is displayed only if multiple drive classes exist on the system. For more information about Easy Tier, see Chapter 9, “Advanced features for storage efficiency” on page 669.

If the pool has a DRAID 6 array that is made of 16 drives, you cannot add a two-drive RAID 1 array to the same pool from the same drive class because this configuration imbalances the storage pool. You can still add any array of any configuration to a pool by using the CLI if the platform is supporting the raid level.

When you are satisfied with the configuration that is presented, click **Add Storage**. The RAID arrays are then created, added as array mode MDisks to the pool, and initialized in the background.

You can monitor the progress of the initialization by selecting the corresponding task under **Running Tasks** in the upper-right corner of the GUI. The array is available for I/O during this process and you do not need to wait for it to complete.

The time that it takes to initialize an array depends on the type of drives of which it consists. For example, an array of Flash drives is much quicker to initialize than NL-serial-attached SCSI (SAS) drives.

Configuring arrays with the CLI

When you use the CLI, run the **mkdistributedarray** command to create DRAID. For this process, it is required to retrieve a list of drives that are ready to become array members.

For more information about how to list all available drives, and read and change their use modes, see 5.3.1, “Working with drives” on page 386.

To view the recommended array configuration on the CLI, run the **lsdriveclass** command to list the available drive classes and the **lsarrayrecommendation** to see what RAID types can be created out of the existing set of drives, and what will be the capacity of a new array, as shown in Example 5-16. The recommendations are listed in the order of preference.

Example 5-16 Listing array recommendations by using the CLI

```
IBM_Storwize:ITSOV7K:superuser>lsarrayrecommendation -driveclass 0 -drivecount 10
Pool2
raid_level distributed stripe_width rebuild_areas drive_count array_count capacity
raid6      yes         9            1             10          1           7.6TB
raid6      no          10           0             10          1           8.7TB
raid5      yes         9            1             10          1           8.7TB
raid5      no          9            0             9           1           8.7TB
raid10     no          8            0             8           1           4.4TB
raid1      no          2            0             2           5           5.5TB
```

To create the recommended DRAID 6 array, specify the RAID level, drive class, number of drives, stripe width, number of rebuild areas, and the storage pool. The system automatically chooses drives for the array from the available drives in the class. As shown in Example 5-17, a DRAID 6 array is created out of 10 drives of class 0 by using a stripe width of 9 and a single rebuild area and adds it to `Pool2`.

Example 5-17 Creating DRAID with mkdistributedarray

```
IBM_2145:ITSO-SV1:superuser>mkdistributedarray -level raid6 -driveclass 0
-drivecount 10 -stripewidth 9 -rebuildareas 1 Pool2
MDisk, id [0], successfully created
```

Default values are available for the stripe width and the number of rebuild areas, depending on the RAID level and the drive count. In this example, it was required to specify the stripe width because for DRAID 6 it is 12 by default. The drive count value must equal or be greater than the sum of the stripe width and the number of rebuild areas.

The storage pool must already exist at the moment when an array is being created. See 5.1.1, “Creating storage pools” on page 368.

To check array initialization progress with the CLI, run the **lsarrayinitprogress** command.

5.3.4 Actions on arrays

MDisks that are created from internal storage support specific actions that are not supported on external MDisks.

To choose an action, open **Pools** → **MDisks by Pools,** select the array (MDisk) and click **Actions**. Alternatively, right-click the array, as shown in Figure 5-42.

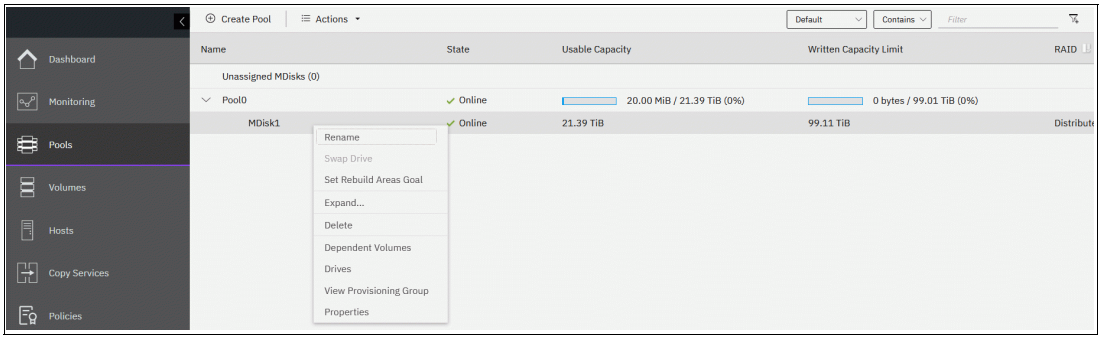


Figure 5-42 Actions on arrays

Rename

Select this option to change the name of an MDisk.

The CLI command for this operation is **`charray`** (see Example 5-18). No feedback is returned.

Example 5-18 Renaming array MDisk with charray

```
IBM_2145:ITSO-SV1:superuser>charray -name Distributed_array mdisk1
IBM_2145:ITSO-SV1:superuser>
```

Swap drive

Select **Swap Drive** to replace a drive in the array with another drive. This action item is active only if there are available candidate or spare drives.

The other drive must have use of Candidate or Spare. Use this action to perform proactive drive replacement to replace a drive that is not failed but is expected to fail soon, for example, if an error message is posted in the Event log.

Figure 5-43 shows the dialog box that opens. Select the member drive to be replaced and the replacement drive, and click **Swap**.

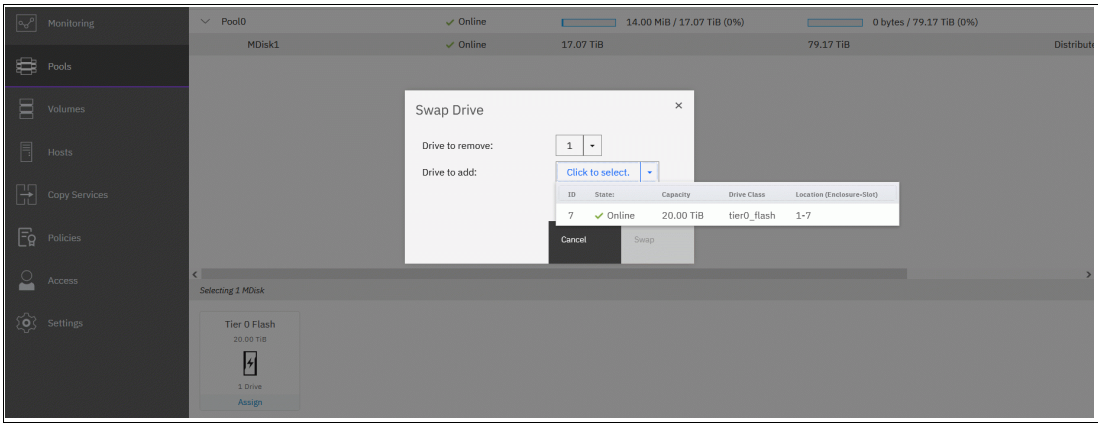


Figure 5-43 Swapping array member with another candidate or spare drive

The exchange of the drives starts running in the background. The volumes on the affected MDisk remain accessible during the process.

In a distributed array, the system immediately removes the old member from the array and performs a rebuild to an available rebuild area. After the rebuild is complete, a copyback is started to copy the data to the new member drive. This process is nondisruptive, but reduces the redundancy of the array during the rebuild process.

The **charraymember** CLI command is run to perform this task. Example 5-19 shows the replacement of array member ID 7 that was assigned to drive ID 12, with drive ID 17. Notice that the **-immediate** parameter is required for distributed arrays to acknowledge that a rebuild starts.

Example 5-19 Replacing array member with CLI (some columns are not shown)

```
IBM_2145:ITSO-SV1:superuser>lsarraymember 16
mdisk_id mdisk_name        member_id drive_id new_drive_id spare_protection
16       Distributed_array 6         18                    1
16       Distributed_array 7         12                    1
16       Distributed_array 8         15                    1
<...>
IBM_2145:ITSO-SV1:superuser>lsdrive
id status error_sequence_number use       tech_type       capacity
16 online                       member    tier_enterprise 558.4GB 16
17 online                       spare     tier_enterprise 558.4GB
18 online                       member    tier_enterprise 558.4GB 16
<...>
IBM_2145:ITSO-SV1:superuser>charraymember -immediate -member 7 -newdrive 17
Distributed_array
IBM_2145:ITSO-SV1:superuser>
```

Set Rebuild Areas Goal

Select this action to set the number of rebuild areas that are expected to be available to protect the array from drive failures. If the number of available rebuild areas is below the configured goal, an error is logged in the Event log. This error can be fixed by replacing failed drives in the DRAID array.

Note: This option does not change the actual number of rebuild areas. It specifies only at which point a warning event is generated. Setting the goal to 0 does not prevent the array from rebuilding.

Expand

Select **Expand** to expand the array by adding one or more drives to it to increase the available capacity of the array or to create rebuild areas. Only distributed arrays can be expanded, and the option is not available for traditional arrays.

Note: The array cannot be shrunk. It is not possible to reduce the number of drives in an array.

Candidate drives of a drive class that is compatible to the drive class of the array must be available in the system. Otherwise, an error message is shown and the array cannot be expanded. A drive class is compatible to another if its characteristics, such as capacity and performance, are an exact match or are superior. In most cases, drives of the same class should be used to expand an array.

The dialog box that is shown in Figure 5-44 on page 403 is displayed to give the user an overview of the current size of the array, the number of available candidate drives in the selected drive class, and the new array capacity after the expansion. The drive class and the

number of drives to add can be modified as required, and the projected new array capacity is updated as needed. To add rebuild areas to the array, click **Advanced Settings** and modify the number of extra spares.

Note: You can't have more than one rebuild area in an array of FCMs.

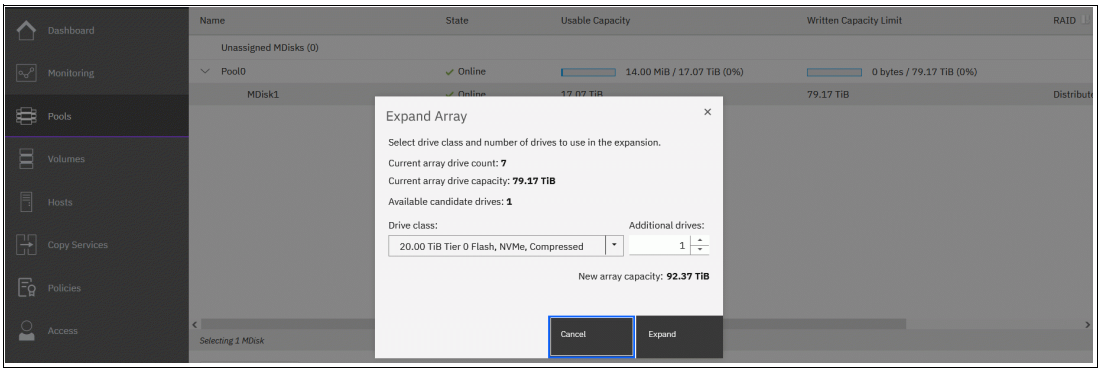


Figure 5-44 Expand a distributed array

Clicking **Expand** starts a background process that adds the selected number of drives to the array. As part of the expansion, the system automatically migrates data for optimal performance for the new expanded configuration.

You can monitor the progress of the expansion by clicking the **Running Tasks** icon in the upper-right corner of the GUI, or by selecting **Monitoring** → **Background tasks**, as shown in Figure 5-45.

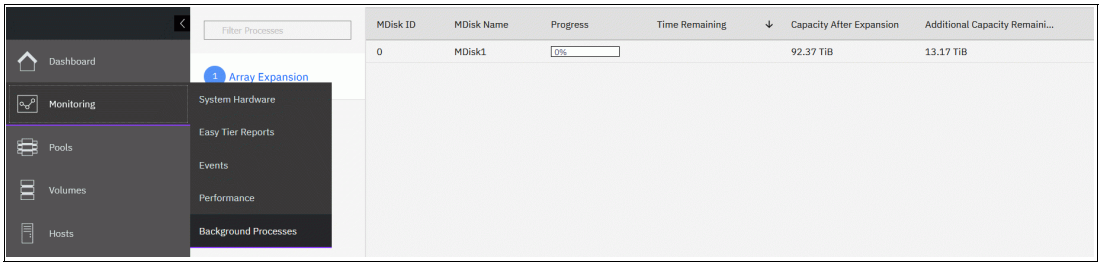


Figure 5-45 Array expansion progress

On the CLI, this task is performed by running the **`expandarray`** command. For a list of compatible drive classes, run the **`lscompatibledriveclasses`** command. See Example 5-20.

Example 5-20 Expanding an array by using the CLI

```
IBM_2145:ITSO-SV1:superuser>lsarray 0
<..>
capacity 3.2TB
<..>
drive_class_id 0
drive_count 6
<..>
rebuild_areas_total 1
IBM_2145:ITSO-SV1:superuser>lscompatibledriveclasses 0
id
0
```

```
IBM_2145:ITSO-SV1:superuser>expandarray -driveclass 0 -totaldrivecount 10
-totalrebuildareas 2 0
IBM_2145:ITSO-SV1:superuser>lsarrayexpansionprogress
progress estimated_completion_time  target_capacity additional_capacity_remaining
29       191018233758               5.17TB          1.38TB
```

Note: The **expandarray** command on the CLI expects the total drive count *after* the expansion as a parameter, which is the sum of the number of new drives and the number of drives in the array before the expansion. The same is true for the number of rebuild areas.

Delete

Select **Delete** to remove the array from the storage pool and delete it. An array MDisk does not exist outside of a storage pool. Therefore, an array cannot be removed from the pool without being deleted. All drives that belong to the deleted array return to **Candidate**.

If no volumes use extents from this array, the command runs immediately without more confirmation. If volumes use extents from this array, you are prompted to confirm the action, as shown in Figure 5-46.

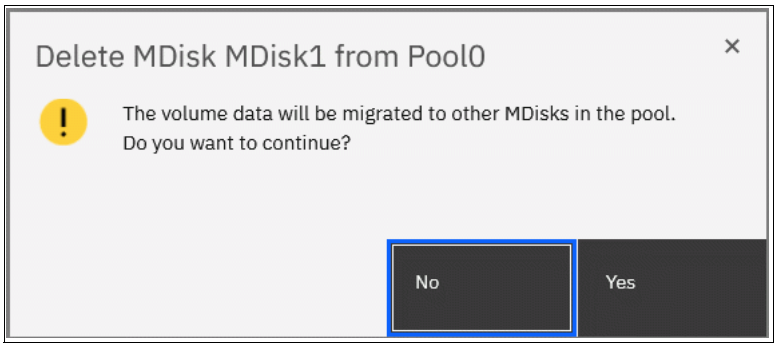


Figure 5-46 Deleting an array from a non-empty storage pool

Confirming the deletion starts a background process that migrates used extents on the MDisk to other MDisks in the same storage pool. After completion of that process, the array is removed from the storage pool and deleted.

Note: The command fails if not enough available capacity remains in the storage pool to allocate the capacity that is being migrated from the removed array.

To delete the array with the CLI, run the **rmarray** command. The **-force** parameter is required if volume extents must be migrated to other MDisks in a storage pool.

Use the **Running Tasks** section in the GUI or run the **lsmigrate** command on the CLI to monitor the progress of the migration. The MDisk continues to exist until the migration completes.

Dependent volumes

A volume depends on an MDisk if the MDisk that is becoming unavailable results in a loss of access or a loss of data for that volume. Use this option before you conduct maintenance operations to confirm which volumes (if any) are affected.

If an MDisk in a storage pool goes offline, the entire storage pool goes offline. That is, all volumes in a storage pool usually depend on each MDisk in the same pool, even if the MDisk does not have extents for each of the volumes. Clicking the **Dependent Volumes Action** menu of an MDisk lists the volumes that depend on that MDisk, as shown in Figure 5-47.

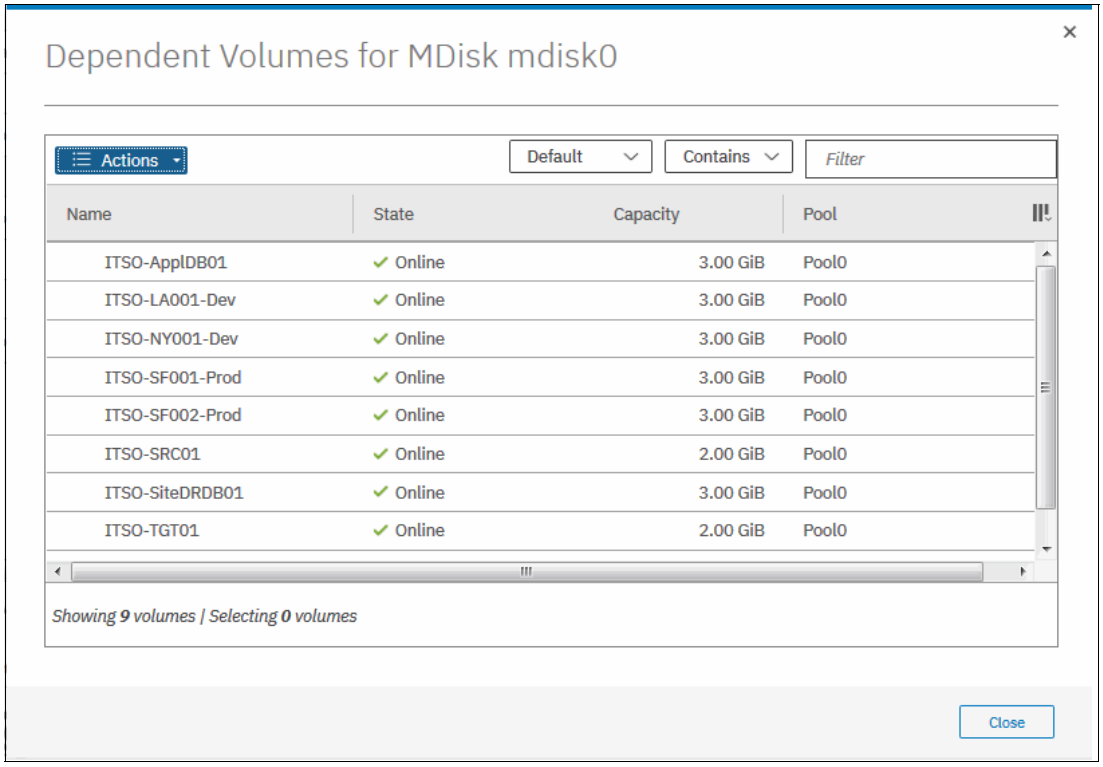


Figure 5-47 Dependent volumes for MDisk mdisk0

You can get the same information by running the CLI command **`lsdependentvdisks`** (see Example 5-21).

Example 5-21 Listing VDisks that depend on MDisk with CLI

```
IBM_2145:ITSO-SV1:superuser>lsdependentvdisks -mdisk mdisk0
vdisk_id vdisk_name
0        ITSO-SRC01
1        ITSO-TGT01
2        ITSO-ApplDB01
<...>
```

Drives

Select **Drives** to see information about the member drives that are included in the array, as shown in Figure 5-48.

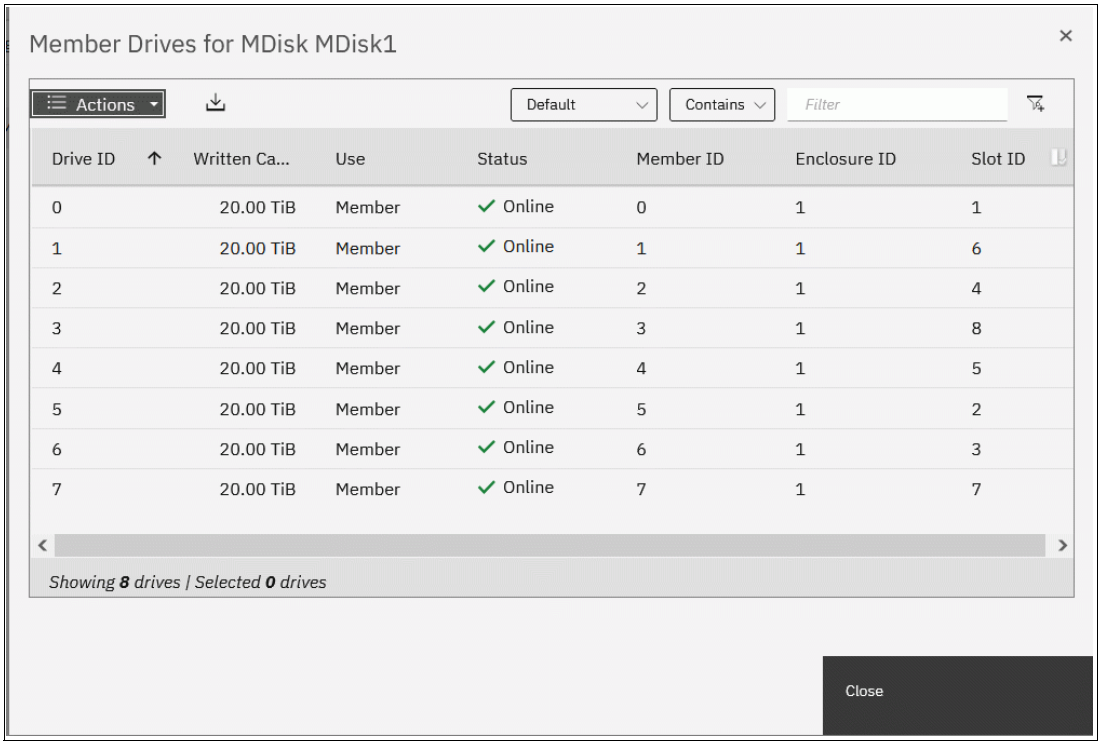


Figure 5-48 List of drives in an array

You can run the CLI command **`lsarraymember`** to get the same information with the CLI. Provide an array name or ID as the parameter to filter output by the array. If run without arguments, the command lists all members of all configured arrays.

Properties

This section shows all available array MDisk parameters: its state, capacity, and storage pool. For more information, click **View more details**. See Figure 5-49.

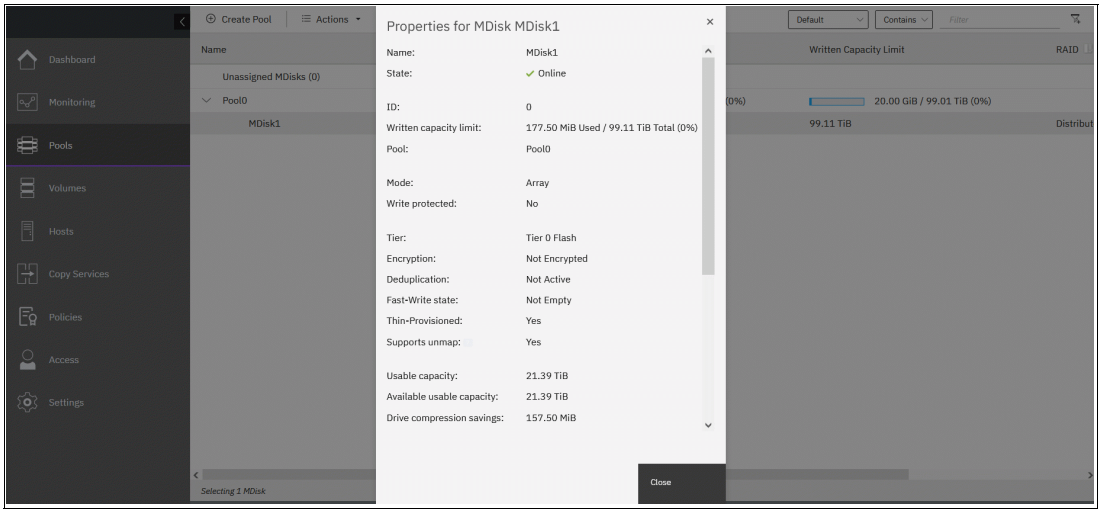


Figure 5-49 Array properties with expanded list of details

Run the CLI command **lsarray** to get a list of all configured arrays and **lsarray** with array name or ID as the parameter to get extended information about the selected array, as shown in Example 5-22.

Example 5-22 lsarray output (truncated)

```
IBM_2145:ITSO-SV1:superuser>lsarray
mdisk_id mdisk_name        status mdisk_grp_id mdisk_grp_name capacity
0        mdisk0            online 0            mdiskgrp0      1.3TB
16       Distributed_array online 1            mdiskgrp1      2.2TB
IBM_2145:ITSO-SV1:superuser>lsarray 16
mdisk_id 16
mdisk_name Distributed_array
status online
mode array
mdisk_grp_id 1
mdisk_grp_name mdiskgrp1
capacity 2.2TB
<...>
```

5.4 Working with external controllers and MDisks

Controllers are external storage systems that provide storage resources that are used as MDisks. The system supports external storage controllers that are attached through internet Small Computer Systems Interface (iSCSI) and through Fibre Channel (FC).

A key feature of the system is its ability to consolidate disk controllers from various vendors into storage pools. The storage administrator can manage and provision storage to applications from a single user interface and use a common set of advanced functions across all of the storage systems under the control of the system.

This concept is called External Virtualization, which makes your storage environment more flexible, cost-effective and easy to manage.

Storage Virtualize hardware platforms can support External Virtualization for migration only, or for both migration and permanent use. Most follow a differential licensing scheme that is based on Storage Capacity Units (SCU).

For more information about how to configure external storage systems, see Chapter 2, “Installation and configuration planning” on page 119.

5.4.1 External storage controllers

External storage controllers can be attached by using FC and iSCSI. The following sections describe how to attach external storage controllers to the system and how to manage them by using the GUI.

System layers

A system layer affects how the system interacts with other external IBM Storwize or IBM FlashSystem family systems. IBM SAN Volume Controller is always in the replication layer. A Storwize or FlashSystem family system is in the default storage layer or the replication layer.

With these default settings, the IBM SAN Volume Controller can virtualize other IBM Storwize or IBM FlashSystem family systems. However, if the IBM Storwize or IBM FlashSystem family system was moved to the replication layer, it must be configured back to the storage layer for the IBM SAN Volume Controller to use it as external storage.

The IBM SAN Volume Controller system layer cannot be changed. The changes must be made on the external IBM Storwize or IBM FlashSystem family system instead.

Note: Before you change the system layer, the following conditions must be met:

- No host object can be configured with worldwide port names (WWPNs) from an IBM Storwize or IBM FlashSystem family system.
- No system partnerships can be defined.
- No IBM Storwize or IBM FlashSystem family system can be visible on the storage area network (SAN) fabric.

Attachment by using Fibre Channel

A controller that is connected through FC is detected automatically by the system if the cabling, zoning, and system layer are configured correctly.

Any supported controller can be temporarily direct-attached for migrating the data from that controller onto the system. For permanent virtualization, external storage controllers are connected through SAN switches.

For more information about how to attach and zone back-end storage controllers to the system, see 2.7, “Fibre Channel SAN configuration planning” on page 130.

Note: For more information about the supported permanently directly attached storage systems, see this Direct Attachment of Storwize and SAN Volume Controller Systems.

If the external controller is not detected, ensure that the system is cabled and zoned into the same SAN as the external storage system. Check that layers are set correctly on both virtualizing and virtualized systems if they belong to the IBM Storage Virtualize family.

Attachment by using iSCSI

You must manually configure iSCSI connections between the storage system controller and the external storage controller. Until then, the controller is not listed in the External Storage panel. For more information about how to attach back-end storage controllers to the system, see Chapter 2, “Installation and configuration planning” on page 119.

To start virtualizing an iSCSI back-end controller, see SAN Volume Controller to perform configuration steps that are specific to your back-end storage controller. You can see the steps by selecting **Configuring** → **Configuring and servicing storage systems** → **External storage system configuration with iSCSI connections**.

For more information about configuring Storage systems to virtualize back-end storage controller with iSCSI, see *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

Managing external storage controllers

You can manage FC and iSCSI storage controllers through the External Storage panel. To access the External Storage panel, select **Pools** → **External Storage**, as shown in Figure 5-50 on page 409.

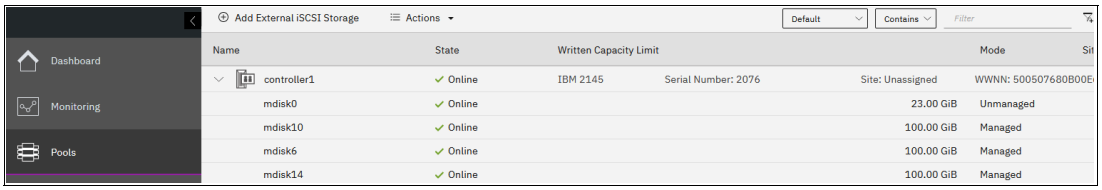


Figure 5-50 External Storage panel

Note: A controller that is connected through FC is detected automatically by the system. The cabling, zoning, and system layers must be configured correctly. A controller that is connected through iSCSI must be added to the system manually.

Depending on the type of back-end system, it might be detected as one or more controller objects.

The External Storage panel lists the external controllers that are connected to the system and all the external MDisks that are detected by the system. The MDisks are organized by the external storage system that presents them. Toggle the sign to the left of the controller icon to show or hide the MDisks that are associated with the controller.

If you configured logical unit names on your external storage systems, the system cannot determine these names because they are local to the external storage system. However, you can use the LU unique identifiers (UIDs), or external storage system worldwide node names (WWNNs) and LU number to identify each device.

To list all visible external storage controllers with CLI, run the **`lscontroller`** command, as shown in Example 5-23.

Example 5-23 Listing controllers by using the CLI (some columns are not shown)

```
IBM_2145:ITSO-SV1:superuser>lscontroller
id controller_name ctrl_s/n             vendor_id            product_id_low
0  controller1      2076                 IBM                   2145
1  controller0     2076                  IBM                   2145
```

5.4.2 Actions on external storage controllers

You can perform many actions on external storage controllers. Some actions are available for external iSCSI controllers only.

To select any action, select **Pools** → **External Storage** and right-click the controller, as shown in Figure 5-51. Alternatively, select the controller and click **Actions**.

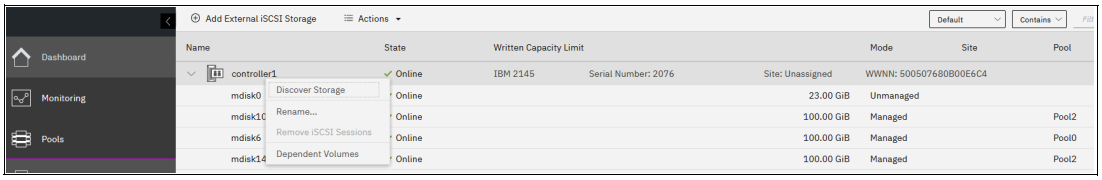


Figure 5-51 Actions for external storage

Discover Storage

When you create or remove LUs on an external storage system, the change might not be detected immediately. In this case, click **Discover Storage** so that the system can rescan the FC or iSCSI network. In general, the system automatically detects disks when they appear on the network. However, some FC controllers do not send the required SCSI primitives that are necessary to automatically discover the new disks.

The rescan process runs the **detectmdisk** command and discovers any new MDisks that were added to the system and rebalances MDisk access across the available ports. It also detects any loss of availability of the controller ports.

Rename

To modify the name of an external controller to help simplify administration tasks, click **Rename**. The naming rules are the same as for storage pools. For more information, see 5.1.1, “Creating storage pools” on page 368.

To rename a storage controller by using the CLI, run the **chcontroller** command. A use case is shown in Example 5-24.

Removing iSCSI sessions

This action is available only for external controllers that are attached with iSCSI. To remove the iSCSI session that is established between the source and target port, right-click the session and select **Remove**.

For more information about the CLI commands and detailed instructions, see *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

Modifying a site

This action is available only for systems that are configured as an Enhanced Stretched Cluster (ESC) or HyperSwap topology. To change the site with which the external controller is associated, select **Modify Site**, as shown in Figure 5-52.

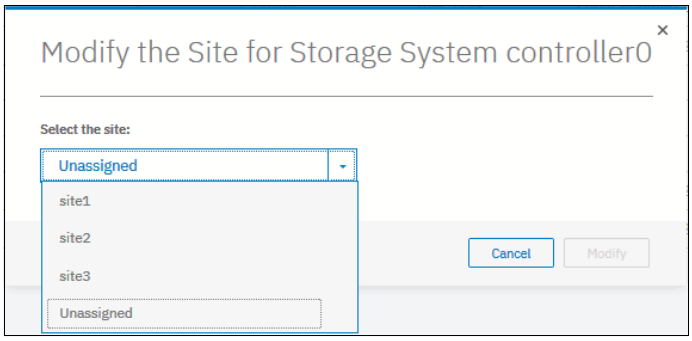


Figure 5-52 Modifying the site of an external controller

To change the controller site assignment by using the CLI, run the **chcontroller** command. Example 5-24 shows that `controller0` was renamed and reassigned to a different site.

Example 5-24 Changing a controller’s name and site

```
IBM_2145:ITSO-SV1:superuser>chcontroller -name site3_controller -site site3
controller0
IBM_2145:ITSO-SV1:superuser>
```

5.4.3 Working with external MDisks

After an external back-end storage controller is configured, attached to the system, and detected as a controller, you can work with LUs that are provisioned from it. Each LU is represented by an MDisk object.

External MDisks can have one of the following modes:

- Unmanaged

  External MDisks are initially discovered by the system as unmanaged MDisks. An unmanaged MDisk is not a member of any storage pool. It is not associated with any volumes, and has no metadata that is stored on it.

  The system does not write to an MDisk that is in unmanaged mode except when it attempts to change the mode of the MDisk to one of the other modes. Removing an external MDisk from a pool returns it to unmanaged mode.

- Managed

  When unmanaged MDisks are added to storage pools, they become managed. Managed mode MDisks are always members of a storage pool, and their extents contribute to the storage pool. This mode is the most common and normal mode for an MDisk.

- Image

  Image mode provides a direct block-for-block conversion from the MDisk to a volume. This mode is provided to satisfy the following major usage scenarios:

  – Presenting data that is on an MDisk through the system to an attached host
  – Importing data that is on an MDisk into the system
  – Exporting data that is on a volume by performing a migration to an image mode MDisk

Listing external MDisks

You can manage external MDisks by using the External Storage panel, which is accessed by selecting **Pools** → **External Storage**, similarly to listing connected controllers, as shown in Figure 5-50 on page 409.

To list all MDisks that are visible by the system by using the CLI, run the **`lsmdisk`** command without any parameters. If required, you can filter output to include only external or only array type MDisks by adding **`-filtervalue mode=array`** or **`-filtervalue mode=managed`** argument.

Assigning MDisks to pools

You can add unmanaged MDisks to a pool or create a pool to include them. If no storage pool exists yet, follow the procedure that is described in 5.1.1, “Creating storage pools” on page 368.

Figure 5-53 shows how to add a selected MDisk to a storage pool. Click **Assign** from the **Actions** menu or right-click the MDisk and select **Assign**.

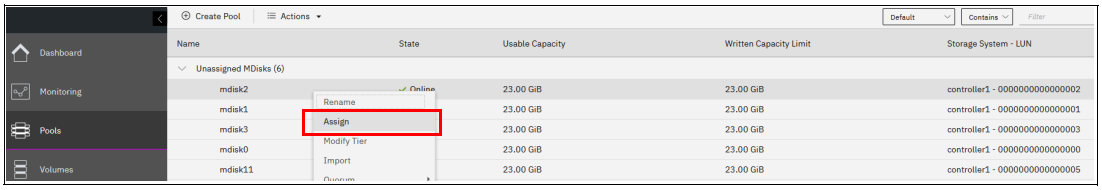


Figure 5-53 Assigning an unmanaged MDisk

After you click **Assign**, a dialog box opens, as shown in Figure 5-54. Select the target pool, MDisk storage tier, and external encryption setting.

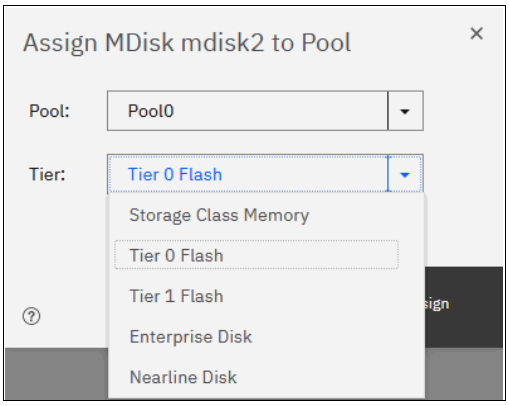


Figure 5-54 Assign MDisk dialog box

When you add MDisks to pools, you must assign them to the correct storage tiers. It is important to set the tiers correctly if you plan to use the Easy Tier feature. The use of an incorrect tier can mean that the Easy Tier algorithm might make wrong decisions and thus affect system performance.

For more information about storage tiers, see Chapter 9, “Advanced features for storage efficiency” on page 669.

The storage tier setting can also be changed after the MDisk is assigned to the pool.

Select the **Externally encrypted** option if your back-end storage performs data encryption. For more information about encryption and encrypted storage pools, see 12.7, “Encryption” on page 1098.

After the task completes, click **Close.**

Note: If the external storage LUs that are presented to the storage system contain data that must be retained, do not use the Assign option to add the MDisks to a pool. This option destroys the data on the LU. Instead, use the Import option to create an image mode MDisk. For more information, see Chapter 8, “Storage migration” on page 635.

To see the external MDisks that are assigned to a pool within the system, select **Pools** → **MDisks by Pools**.

When an MDisk is added to a pool that contains MDisks and volumes, the Easy Tier feature automatically balances volume extents between the MDisks in the pool as a background process. The goal of this process is to distribute extents in a way that provides the best performance to the volumes. It does *not* attempt to balance the amount of data evenly between all MDisks.

The data migration decisions that Easy Tier makes between tiers of storage (inter-tier) or within a single tier (intra-tier) are based on the I/O activity that is measured. Therefore, when you add an MDisk to a pool, extent migrations are not necessarily performed immediately. Extents are not migrated until sufficient I/O activity exists to trigger it.

If Easy Tier is turned off, no extent migration is performed. Only newly allocated extents are written to a new MDisk.

For more information about the Easy Tier feature, see Chapter 9, “Advanced features for storage efficiency” on page 669.

To assign an external MDisk to a storage pool by using the CLI, run the **addmdisk** command. You must specify the MDisk name or ID, MDisk tier, and target storage pool, as shown in Example 5-25. The command returns no feedback.

Example 5-25 The addmdisk command

```
IBM_2145:ITSO-SV1:superuser>addmdisk -mdisk mdisk2 -tier enterprise Pool0
IBM_2145:ITSO-SV1:superuser>
```

5.4.4 Actions for external MDisks

External MDisks support specific actions that are not supported on RAID arrays that are made from internal storage. Some actions are supported only on unmanaged external MDisks, and some are supported only on managed external MDisks.

To choose an action, select **Pools** → **External Storage** or **Pools** → **MDisks by Pools**. Select the external MDisk and then click **Actions** as shown in Figure 5-55. Alternatively, right-click the external MDisk.

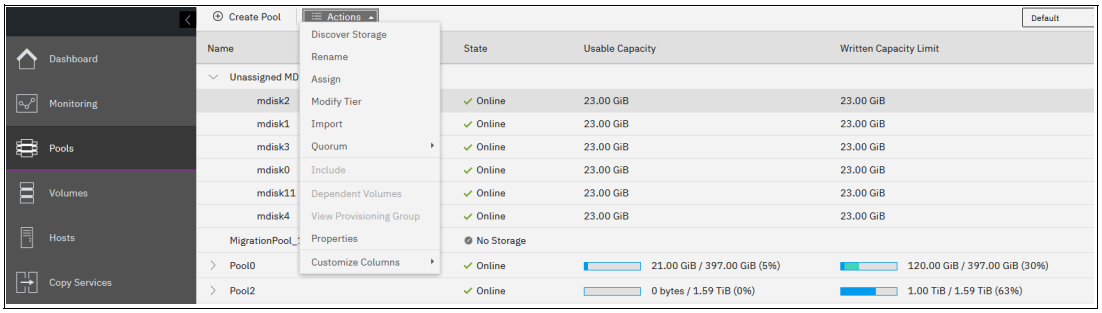


Figure 5-55 Actions for MDisks

Discover Storage

This option is available even if no MDisks are selected. By running it, you cause the system to rescan the iSCSI and FC network for the following purposes:

- Find any new MDisks that might be added
- Rebalance MDisk access across all available controller device ports

This action runs the **detectmdisk** command.

Assign

This action is available for unmanaged MDisks only. Select **Assign** to open the dialog box, as described in “Assigning MDisks to pools” on page 411.

Modify Tier

To modify the tier to which the external MDisk is assigned, select **Modify Tier,** as shown in Figure 5-56 on page 414. This setting is adjustable because the system cannot always detect the tiers that are associated with external storage automatically, unlike with internal arrays.

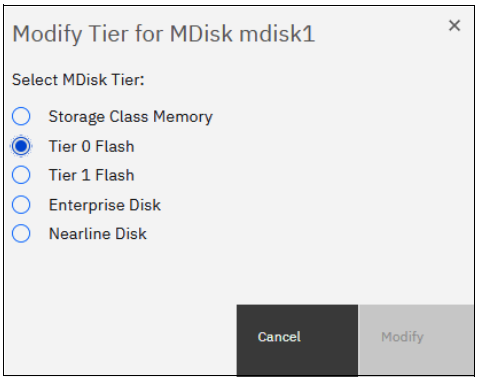


Figure 5-56 Modifying an external MDisk tier

For more information about storage tiers and their importance, see 9.1, “IBM Easy Tier” on page 670.

To change the external MDisk storage tier, run the **chmdisk** command. Example 5-26 shows setting the new tier to `mdisk2`. No feedback is returned.

Example 5-26 Changing the tier setting by using the CLI

```
IBM_2145:ITSO-SV1:superuser>chmdisk -tier tier1_flash mdisk2
IBM_2145:ITSO-SV1:superuser>
```

Modify Encryption option

To modify the encryption setting for the MDisk, select **Modify Encryption**. This option is available only when encryption is enabled.

If the external MDisk is encrypted by the external storage system, change the encryption state of the MDisk to **Externally encrypted**. This setting stops the system from encrypting the MDisk again if the MDisk is part of an encrypted storage pool.

For more information about encryption, encrypted storage pools, and self-encrypting MDisks, see 12.7, “Encryption” on page 1098.

To perform this task by using the CLI, run the **chmdisk** command, as shown in Example 5-27.

Example 5-27 Using chmdisk to modify encryption

```
IBM_2145:ITSO-SV1:superuser>chmdisk -encrypt yes mdisk5
IBM_2145:ITSO-SV1:superuser>
```

Importing and migrating external MDisks to another pool can be done by selecting **Pools** → **System Migration** to start the system migration wizard. For more information, see Chapter 8, “Storage migration” on page 635.

Quorum

This menu option enables you to introduce a new set of quorum disks. When three online managed MDisks are selected, the **Quorum** → **Modify Quorum Disks** menu becomes available, as shown on Figure 5-57 on page 415. For HyperSwap and Enhanced Stretched Cluster configurations, selected MDisks must belong to storage controllers assigned to three different sites.

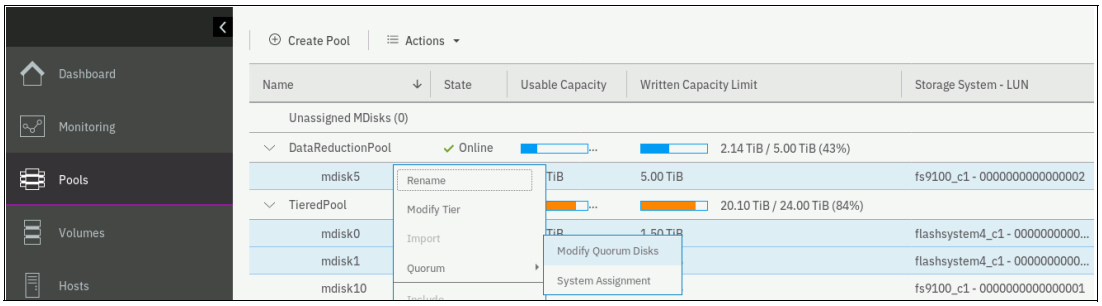


Figure 5-57 Selecting new quorum disks

To list and change the quorum configuration, run the **`lsquorum`** and **`chquorum`** commands.

Include

The system can exclude an MDisk from its storage pool if it has multiple I/O failures or has persistent connection errors. Exclusion ensures that no excessive error recovery exists that might affect other parts of the system. If an MDisk is automatically excluded, run the DMP to resolve any connection and I/O failure errors.

If no error event is associated with the MDisk in the log and the external problem is corrected, click **Include** to add the excluded MDisk back to the storage pool.

The **`includemdisk`** command performs the same task. The command needs the MDisk name or ID to be provided as a parameter, as shown in Example 5-28.

Example 5-28 Including a degraded MDisk by using the CLI

```
IBM_2145:ITSO-SV1:superuser>includemdisk mdisk3
IBM_2145:ITSO-SV1:superuser>
```

Remove

In some cases, you might want to remove external MDisks from their storage pool. To remove the MDisk from the storage pool, click **Remove**.

After the MDisk is removed, it returns to the `Unmanaged` state. If no volumes exist in the storage pool to which this MDisk is allocated, the command runs immediately without more confirmation. If volumes are in the pool, you are prompted to confirm the action, as shown in Figure 5-58.

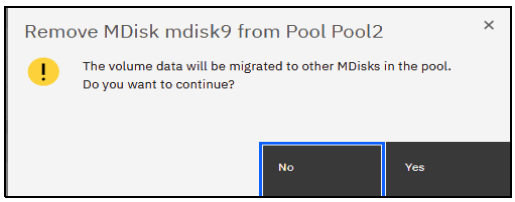


Figure 5-58 Removing an MDisk from a pool

Confirming the action starts a migration of volumes to extents on that MDisk to other MDisks in the pool. During this background process, the MDisk remains a part of the storage pool. Only when the migration completes is the MDisk removed from the storage pool and returned to `Unmanaged` mode.

Ensure that you have enough available capacity remaining in the storage pool to allocate the data that is being migrated from the removed MDisk, or this command fails.

Important: The MDisk that you are removing must remain accessible to the system while all data is copied to other MDisks in the same storage pool. If the MDisk is unmapped before the migration finishes, all volumes in the storage pool go offline and remain in this state until the removed MDisk is connected again.

To remove an MDisk from a storage pool by using the CLI, run the **`rmmdisk`** command. You must use the **`-force`** parameter if you must migrate volume extents to other MDisks in a storage pool.

The command fails if you do not have enough available capacity remaining in the storage pool to allocate the data that you are migrating from the removed array.

Dependent volumes

A volume depends on an MDisk if the MDisk becoming unavailable results in a loss of access or loss of data for that volume. Use this option before you complete maintenance operations to confirm which volumes (if any) are affected. Selecting an MDisk and clicking **Dependent Volumes** lists the volumes that depend on that MDisk. An example is shown in Figure 5-59.

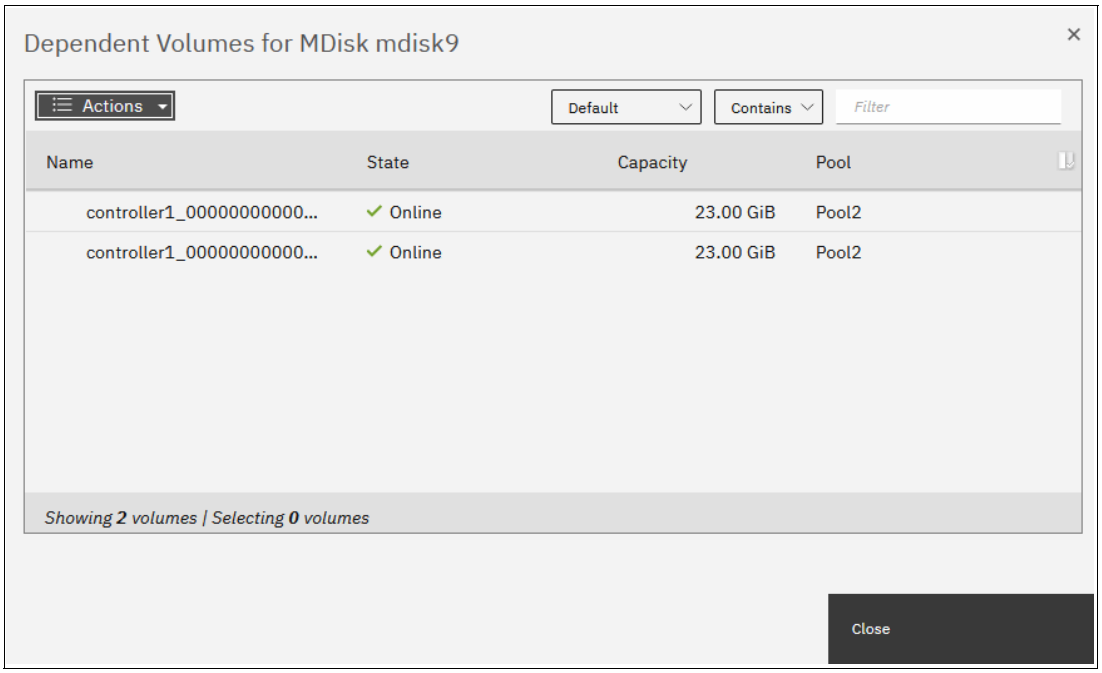


Figure 5-59 Dependent volumes for an MDisk

You can get the same information by running the **`lsdependentvdisks`** command (Example 5-29).

Example 5-29 Listing VDisks that depend on an MDisk by using the CLI

```
IBM_2145:ITSO-SV1:superuser>lsdependentvdisks -mdisk mdisk9
vdisk_id  vdisk_name
17        controller1_0000000000000000
16        controller1_0000000000000001
<...>
```

View provisioning groups

Provisioning groups are used for capacity reporting and monitoring of over-provisioned external storage controllers. Each over-provisioned MDisk is part of a provisioning group that defines the physical storage resources that are available to a set of MDisks.

Storage controllers report the usable capacity of an over-provisioned MDisk that is based on its provisioning group. If multiple MDisks are part of the same provisioning group, these MDisks share the physical storage resources and report the same usable capacity. However, this usable capacity is not available to each MDisk individually because it is shared among all these MDisks.

To know the usable capacity that is available to the system or to a pool when over-provisioned storage is used, you must account for the usable capacity of each provisioning group.

To show a summary of over-provisioned external storage, including controllers, MDisks, and provisioning groups, click **View Provisioning Groups**, as shown in Figure 5-60.

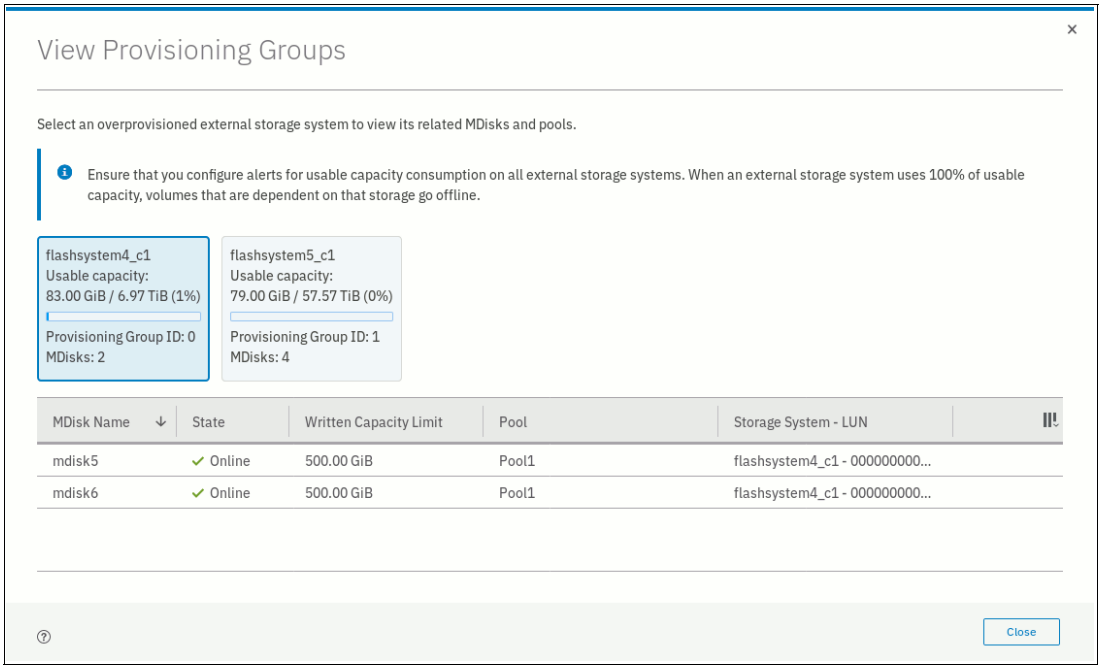


Figure 5-60 View Provisioning Groups

6

Volumes

In IBM Storage Virtualize, a *volume* is storage space that is provisioned out of a storage pool and presented to a host as a Small Computer System Interface (SCSI) logical unit (LU), which is also called a logical disk.

This chapter describes how to create and provision volumes on IBM Storage Virtualize systems. The first part of this chapter provides a brief overview of IBM Storage Virtualize volumes, the classes of volumes that are available, and the available volume customization options.

The second part of this chapter describes how to create, modify, and map volumes by using the GUI.

The third part of this chapter provides an introduction to volume manipulation from the command line interface (CLI).

This chapter includes the following topics:

- 6.1, “Introduction to volumes” on page 420
- 6.2, “Volume characteristics” on page 420
- 6.3, “Volume groups” on page 438
- 6.4, “Virtual volumes” on page 444
- 6.5, “Volumes in multi-site topologies” on page 444
- 6.6, “Operations on volumes” on page 447
- 6.7, “Volume operations by using the CLI” on page 488

© Copyright IBM Corp. 2024. All rights reserved.

6.1 Introduction to volumes

For an IBM Storage Virtualize system cluster, the volume that is presented to a host is internally represented as a virtual disk (VDisk). A VDisk is an area of usable storage that was allocated out of a pool of storage that is managed by an IBM Storage Virtualize cluster. The term *virtual* is used because the volume that is presented does not necessarily exist on a single physical entity.

Note: Volumes are composed of extents that are allocated from a storage pool. Storage pools group the managed disks (MDisks). MDisks are redundant arrays of independent disks (RAIDs) that are configured by using internal storage, or LUs from external storage that are presented to and virtualized by an IBM Storage Virtualize system. Each MDisk is divided into sequentially numbered extents (zero-based indexing). The extent size is a property of a storage pool, and is used for all MDisks that make up the storage pool.

MDisks are internal objects that are used for storage management. They are not directly visible to or used by host systems.

Every volume is presented to hosts by an I/O group. One of the nodes within that group is defined as a preferred node; that is, a node that by default serves I/O requests to that volume. When a host requests an I/O operation to a volume, the multipath driver on the host identifies the preferred node for the volume and by default uses only paths to this node for I/O requests.

6.2 Volume characteristics

The following parameters characterize each volume. They must be set correctly to match the requirements of the storage user, which can be an application running on a host:

- Size
- Performance:
  - I/O operations per second (IOPS)
  - response time
  - bandwidth)
- Resiliency
- Storage efficiency
- Security (data-at-rest encryption)
- Extent allocation policy
- Management mode

Volumes can also be configured as VMware vSphere virtual volumes (vVols).

vVols change the approach to VMware virtual machines (VMs) disk configuration. If vVols are not used, then the VM disk is a file on a VMware virtual machine file system (VMFS) volume. If vVols are used, then the VM disk is a one-to-one mapping between the VM disk and a storage volume. vVols can be managed by the VMware infrastructure so that the storage system administrator can delegate VM disk management to VMware infrastructure specialists, which can simplify storage allocation for a virtual infrastructure. It can also reduce the need for the storage management team to support the VMware infrastructure administrators.

The downside of the use of vVols is that the number of volumes that are presented by a storage system are increased because typically multiple VM disks are configured on every VMFS volume. Excessive proliferation of volumes that is presented to ESXi clusters can

negatively affect performance. Therefore, it is a best practice to carefully plan a storage system configuration before production deployment and include in the assessment the projected system growth.

Note: If too many logical unit numbers (LUNs) are presented to a sufficiently large ESXi cluster, I/O requests that are simultaneously generated by ESXi hosts might exceed the storage system command queue. Such overflow leads to I/O request retries, which reduce storage system performance as perceived by the connected hosts.

To provide storage users adequate service, all parameters must be correctly set. Importantly, the various parameters might be *interdependent*; that is, setting one of them might affect other properties of the volume.

The volume parameters and their interdependencies are described next.

6.2.1 Volume type

The type attribute of a volume defines the method of allocation of extents that make up the volume copy:

- A striped volume contains a volume copy that has extents that are allocated from multiple MDisks from the storage pool (see Figure 6-1). By default, extents are allocated from all MDisks in the storage pool that uses a round-robin algorithm. However, it is possible to supply a list of MDisks to use for volume creation.

Attention: By default, striped volume copies are striped across all MDisks in the storage pool. If some of the MDisks are smaller than others, the extents on the smaller MDisks are used up before the larger MDisks run out of extents. Manually specifying the stripe set in this case might result in the volume copy not being created.

If you are unsure whether sufficient free space is available to create a striped volume copy, use one of the following approaches:

- Check the free space on each MDisk in the storage pool by running the **`lsfreeextents`** command. Also, ensure that each MDisk that is included in the manually specified stripe set has enough free extents.
- Allow the system to automatically create the volume copy by not supplying a specific stripe set.

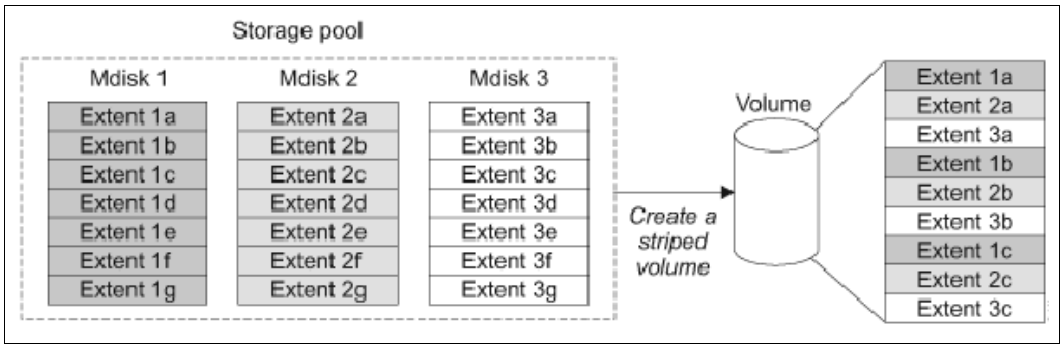


Figure 6-1 Striped extent allocation

- A sequential volume contains a volume copy with extents that are allocated sequentially on one MDisk.
- An image mode volume is a special type of volume that has a direct one-to-one mapping to one image mode MDisk.

For striped volumes, by default, the extents are allocated from all the MDisks in the storage pool. Consider the following points:

- An MDisk is picked by using a pseudo-random algorithm and an extent is allocated from this MDisk. This approach minimizes the probability of triggering the *striping effect*, which might lead to poor performance for workloads that generate many metadata I/Os, or that create multiple sequential streams.
- All subsequent extents, if required, are allocated from the MDisk set by using a round-robin algorithm.
- If an MDisk has no free extents when its turn arrives, the algorithm moves to the next MDisk in the set that has a free extent.

Note: The striping effect can occur when multiple logical volumes on a set of physical storage devices (MDisks) store their metadata or file system transaction log on the same MDisk. This can impact performance if there's frequent access to that specific MDisk

Because of the way the file systems work, system metadata disk regions are typically busy. For example, in a journaling file system, a write to a file might require two or more writes to the file system journal: At minimum, one to make a note of the intended file system update, and one marking the successful completion of the file write.

If multiple volumes, each with their own file system, are defined on the same set of MDisks, and all or most of them store their metadata on the same MDisk, a disproportionately large I/O load is generated on this MDisk, which can result in suboptimal performance of the storage system. Pseudo-randomly allocating the first MDisk for new volume extent allocation minimizes the probability that multiple file systems that are created on these volumes place their metadata regions on the same physical MDisk.

Some file systems allow specifying different logical disks for data and metadata storage. When taking advantage of this file system feature, you can allocate differently configured volumes that are dedicated to data and metadata storage.

6.2.2 Managed mode and image mode

Volumes are configured within IBM Storage Virtualize by allocating a set of extents off one or more managed mode MDisks in the storage pool. Extents are the smallest allocation unit at the time of volume creation. Therefore, each MDisk extent maps to exactly one volume extent.

Note: An MDisk extent maps to only one volume extent. For volumes with two copies, one volume extent maps to two MDisk extents, one for each volume copy.

Figure 6-2 on page 423 shows this mapping. It also shows a volume that consists of several extents that are shown as V0 - V7. Each of these extents is mapped to an extent on one of the MDisks: A, B, or C. The mapping table stores the details of this indirection.

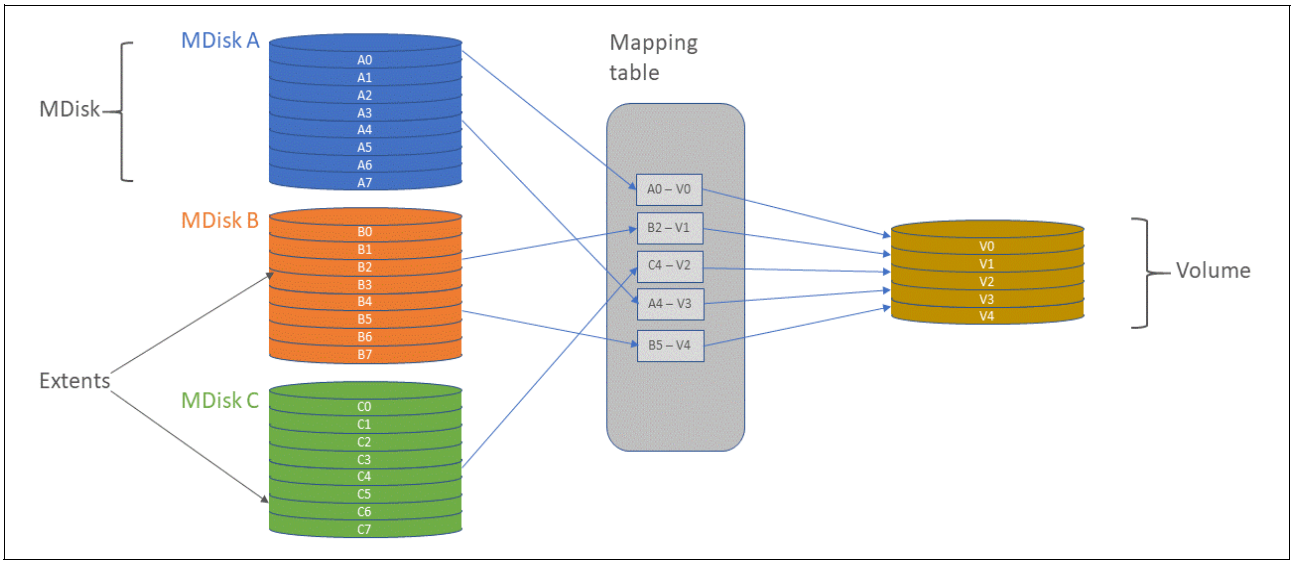


Figure 6-2 Simple view of block virtualization

Several of the MDisk extents are unused. That is, no volume extent maps to them. These unused extents are available for volume creation, migration, and expansion.

The default and most common type of volumes in IBM Storage Virtualize are managed mode volumes. Managed mode volumes are allocated from a set of MDisks that belong to a storage pool, and they can be subjected to the full set of virtualization functions. In particular, they offer full flexibility in mapping between logical volume representation (a continuous set of logical blocks) and the physical storage that is used to store these blocks. This function requires that the physical storage (MDisks) is fully managed by IBM Storage Virtualize, which means that the LUs that are presented to IBM Storage Virtualize by the back-end storage systems do not contain any data when they are added to the storage pool.

Image mode volumes enable IBM Storage Virtualize to work with LUs that were directly mapped to hosts, which are often required when IBM Storage Virtualize is introduced into a storage environment. In such scenarios, image mode volumes are used to enable seamless migration of data and a smooth transition to virtualized storage.

Image mode volumes establish a one-to-one mapping of logical block addresses (LBAs) between the volume and a single MDisk. They have a minimum size of one block (512 bytes) and always occupy the entire capacity of the MDisk as a single extent.

An image mode MDisk cannot be used as a quorum disk and no IBM Storage Virtualize system metadata extents are allocated from it. All the IBM Storage Virtualize copy services functions can be applied to image mode disks.

The difference between a managed mode volume with striped extent allocation and an image mode volume is shown in Figure 6-3.

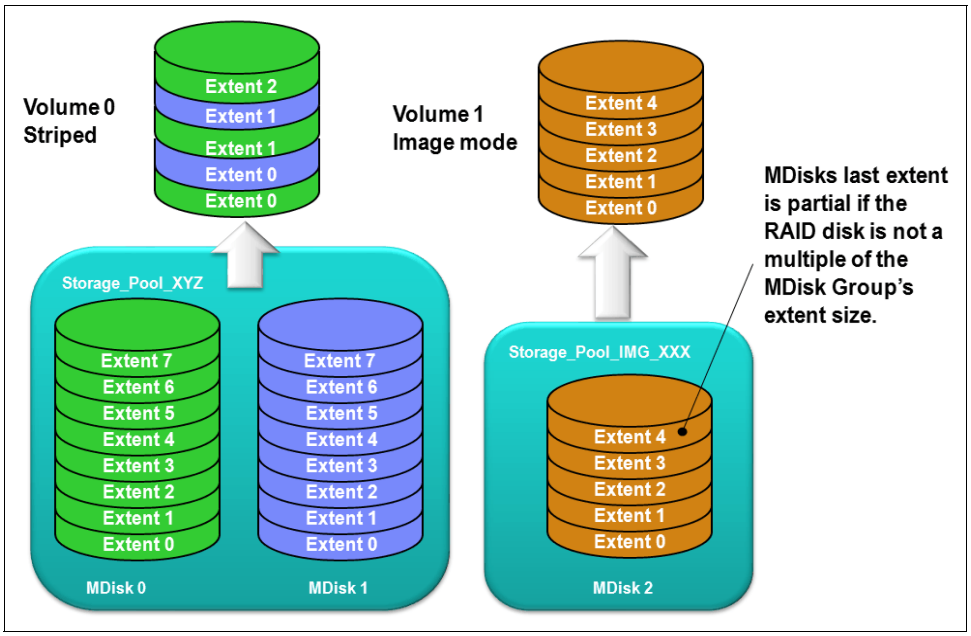


Figure 6-3 An image mode volume versus a striped volume

An image mode volume is mapped to only one image mode MDisk, and it is mapped to the entirety of this MDisk. Therefore, the image mode volume capacity is equal to the size of the corresponding image mode MDisk. If the size of the image mode MDisk is not a multiple of the MDisk group's extent size, the last extent is marked as partial or not filled.

When you create an image mode volume, you map it to an MDisk that must be in unmanaged mode and must not be a member of a storage pool. As the image mode volume is configured, the MDisk is made a member of the specified storage pool. It is a best practice to use a dedicated pool for image mode MDisks with a name indicating its role, such as `Storage Pool_IMG_`*`xxx`*.

An image mode volume can be migrated to a managed mode volume. This standard procedure is used to perform nondisruptive migration of the organization's SAN to an environment that is managed by or based on IBM Storage Virtualize systems.

After the data is migrated off the managed image volume, the space it used on the source storage system can be reclaimed. After all data is migrated off the storage system, it can be decommissioned or used as a back-end storage system that is managed by the IBM Storage Virtualize system. For more information, see 2.10, "Back-end storage configuration" on page 146.

IBM Storage Virtualize also supports the reverse process in which a managed mode volume can be migrated to an image mode volume. During the migration, the volume is identified in the system as being in managed mode. It changes to image mode only after the process completes.

6.2.3 VSize

Each volume has the following associated values that describe its size:

- The real capacity, also called the physical capacity, is the size of storage space that is allocated to the volume from the storage pool. It determines how many MDisk extents are allocated to form the volume. The real capacity is used to store the user data, and in the case of thin-provisioned volumes, the metadata of the volume.
- The virtual capacity is capacity that is reported to the host, but also to any other IBM Storage Virtualize components or functions, such as IBM FlashCopy, cache, and Remote Copy (RC)) that operate based on a volume size.

In a standard-provisioned volume, the real and virtual capacities are the same. In a thin-provisioned volume, the real capacity can be as little as a few percent of virtual capacity. The volume size can be specified in units down to 512-byte blocks (see Figure 6-4). The real capacity can be specified as an absolute value or as a percentage of the virtual capacity.

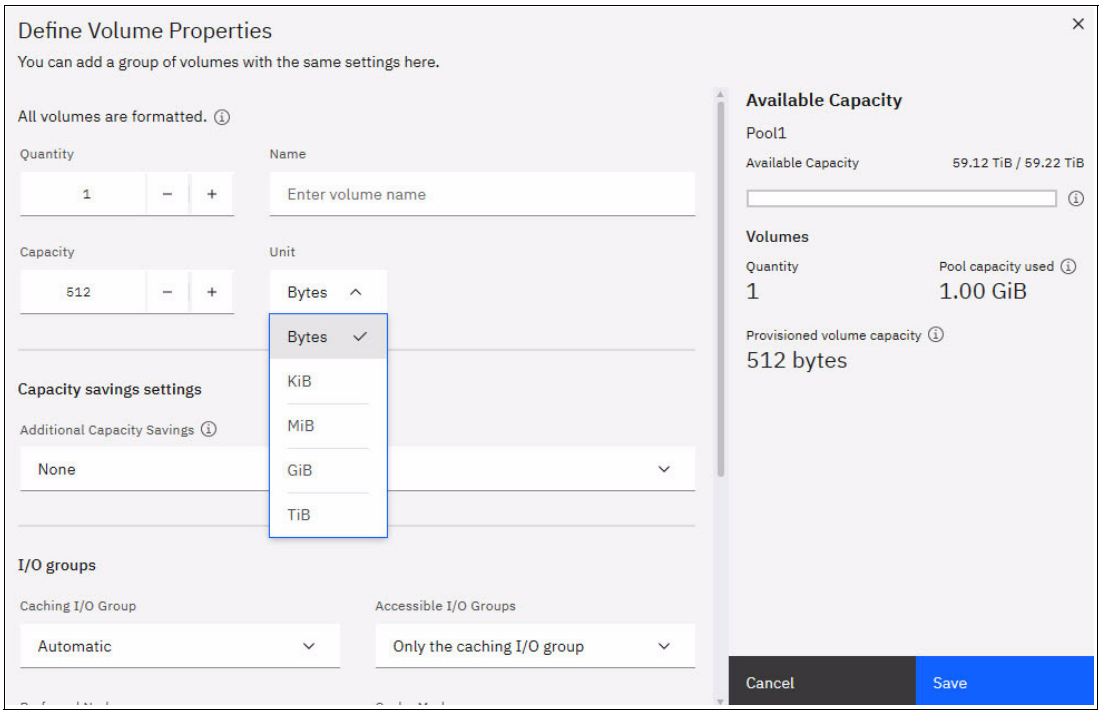


Figure 6-4 Smallest possible volume size

A volume is composed of storage pool extents; therefore, it is not possible to allocate less than one extent to create a volume. Effectively, the internal unit of volume size is the extent size of the pool (or pools) in which the volume is created.

For example, a basic volume of 512 bytes that is created in a pool with the default extent size (1024 mebibytes [MiB]) uses 1024 MiB of the pool space because an entire extent must be allocated to provide the space for the volume.

In practice, this rounding up of volume size to the whole number of extents has little effect on storage use efficiency unless the storage system serves many small volumes. For more information about storage pools and extents, see Chapter 5, “Using storage pools” on page 365.

6.2.4 Performance

The basic metrics of volume performance are the number of IOPS the volume can provide, the time to service an I/O request (average, median, and first percentile), and the bandwidth of the data that is served to a host.

Volume performance is defined by the pool or pools that are used to create the volume. The pool determines the media bus (Non-Volatile Memory Express [NVMe] or serial-attached SCSI [SAS]); media type (IBM FlashCore Module [FCM] drives, solid-state drives [SSDs], or hard disk drives [HDDs]); redundant array of independent disks (RAID) level and number of drives per RAID array; and the possibility for the Easy Tier function to optimize the performance of a volume.

However, volumes that are configured in the same storage pool or pools might still have different performance characteristics, depending on the storage resiliency, efficiency, security, and allocation policy configuration settings of a volume.

Another factor that influences the performance of volumes is the back-end storage that is used to serve the volume. Back-end storage cache size controller performance and the load level on the specific backend storage system are also factors that affect the performance of a volume.

6.2.5 Volume copies

A volume can have one or two physical copies. Each copy of the volume has the same virtual capacity, but the two copies can have different characteristics, including different real capacity. However, each volume copy is not a separate object and can be manipulated only in the context of the volume.

A mirrored volume behaves in the same way as any other volume, such as the following examples:

- All its copies are expanded or shrunk when the volume is resized.
- It can participate in FlashCopy and RC relationships.
- It is serviced by an I/O group.
- It has a preferred node.

Volume copies are identified in the GUI by a copy ID, which can have value 0 or 1. Copies of the volume can be split, which provides a point-in-time (PiT) copy of a volume. An overview of volume mirroring is shown in Figure 6-5 on page 427.

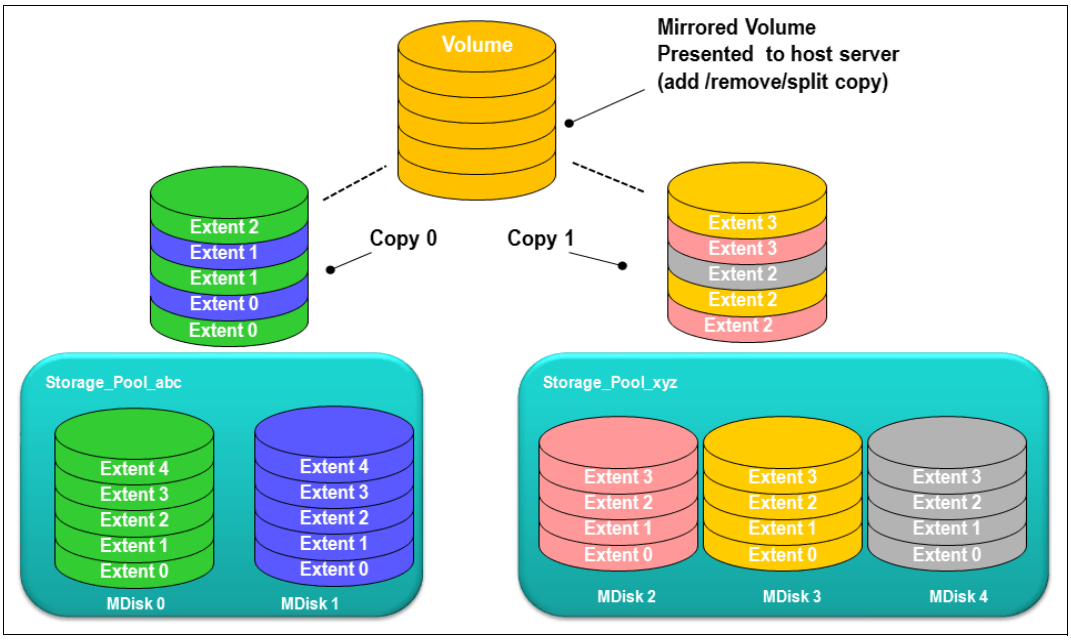


Figure 6-5 Volume mirroring overview

A copy can be added to a volume with a single copy or removed from a volume with two copies. Internal safety mechanisms prevent accidental removal of the only remaining copy of a volume.

A newly created, unformatted volume with two copies initially has the two copies in an out-of-synchronization state. The primary copy is defined as “fresh” and the secondary copy is defined as “stale”, and the volume is immediately available for use.

The synchronization process updates the secondary copy until it is fully synchronized; that is, data that is stored on the secondary copy matches the data that is on the primary copy. This update is done at the *synchronization rate* that is defined when the volume is created, but can be modified after volume creation. The synchronization status for mirrored volumes is recorded on the storage system quorum disk.

If a mirrored volume is created by using the **`format`** parameter, both copies are formatted in parallel. The volume comes online when both operations are complete with the copies in sync.

If it is known that MDisk space (which is used for creating volume copies) is formatted or if the user does not require read stability, a **`no synchronization`** option can be used that declares the copies as `synchronized` even when they are not.

Creating a volume with more than one copy is beneficial in multiple scenarios, including the following examples:

- Improving volume resilience by protecting it from a single back-end storage system failure (requires each volume copy to be configured on a different back-end storage system).
- Providing concurrent maintenance of a storage system that does not natively support concurrent maintenance (for volumes on external virtualized storage).
- Providing an alternative method of data migration with improved availability characteristics. While a volume is being migrated by using the data migration feature, it is vulnerable to failures on the source and target storage pool.

Volume mirroring provides an alternative migration method that is not affected by the destination volume pool availability.

For more information about this volume migration method, see “Volume migration by adding a volume copy” on page 485.

Note: When migrating volumes to a data reduction pool (DRP), volume mirroring is the only migration method because DRPs do not support **`migrate`** commands.

- Converting between standard-provisioned volumes and thin-provisioned volumes (in either direction).

Typically, each volume copy is allocated from a different storage pool. Although not required, the use of different pools that are backed by different back-end storage for each volume copy is the typical configuration because it markedly increases volume resiliency.

If one of the mirrored volume copies becomes temporarily unavailable (for example, because the storage system that provides its pool is unavailable), the volume remains accessible to hosts. The storage system remembers which areas of the volume were modified after a loss of access to a volume copy and resynchronizes only these areas when both copies are available.

Note: Volume mirroring is not a disaster recovery (DR) solution because both copies are accessed by the same node pair and addressable by only a single cluster. However, if correctly planned, it can improve availability.

The storage system tracks the synchronization status of volume copies by dividing the volume into 256 kibibyte (KiB) grains and maintaining a bitmap of stale grains (on the quorum disk), mapping 1 bit to one grain of the volume space. If the mirrored volume needs resynchronization, the system copies to the out-of-sync volume copy only these grains that were written to (changed) since the synchronization was lost. This approach is known as an *incremental synchronization*, and it minimizes the time that is required to synchronize the volume copies.

Important: Mirrored volumes can be taken offline if no quorum disk is available. This behavior occurs because the synchronization status of mirrored volumes is recorded on the quorum disk.

A volume with more than one copy can be checked to see whether all of the copies are identical or consistent. If a medium error is encountered while it is reading from one copy, a check is repaired by using data from the other copy. This consistency check is performed asynchronously with host I/O.

Because mirrored volumes use bitmap space at a rate of 1 bit per 256 KiB grain, 1 MiB of bitmap space supports up to 2 TiB of mirrored volumes. The default size of the bitmap space is 20 MiB, which allows a configuration of up to 40 TiB of mirrored volumes. If all 512 MiB of variable bitmap space is allocated to mirrored volumes, 1 PiB of mirrored volumes can be supported.

For bitmap space configuration options specific to your system’s hardware see:

- IBM SAN Volume Controller
- IBM FlashSystem 9500, 9200 and 9100
- IBM FlashSystem 7200 and 7300
- IBM FlashSystem 5000 and 5200

I/O operations data flow

Although a mirrored volume looks to its users the same as a volume with a single copy, some differences exist in how I/O operations are performed internally for volumes with single or two copies.

Read I/O operations data flow

If the volume is mirrored (that is, two copies of the volume exist), one copy is known as the *primary copy*. If the primary copy is available and synchronized, host read requests are directed to that copy. The choice of the primary copy is part of the initial configuration of a mirrored volume (this setting can be changed at any time).

In the management GUI, an asterisk indicates the primary copy of the mirrored volume. Placing the primary copy on a high-performance controller maximizes the read performance of the volume.

For nonmirrored volumes, only one volume copy exists; therefore, no choice exists for the read source, and all reads are directed to the single volume copy.

Write I/O operations data flow

The host sends all write I/O operation requests to any volume to the preferred node for this volume. The preferred node is responsible for destaging the data from cache to persistent storage. Figure 6-6 shows the data flow for this scenario.

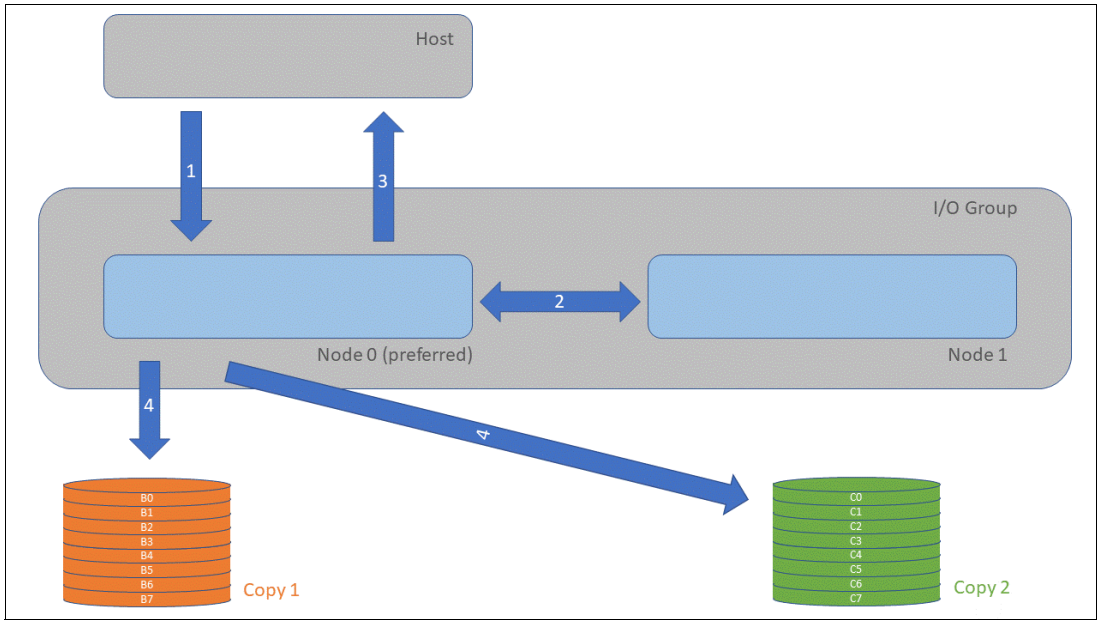


Figure 6-6 Data flow for write I/O processing in a mirrored volume

As shown in Figure 6-6, the writes are sent by the host to the preferred node for the volume (1). Then, the data is mirrored to the cache of the partner node in the I/O group (2), and acknowledgment of the write operation is sent to the host (3). The preferred node then destages the written data to all volume copies (4).

The example that is shown in Figure 6-7 on page 430 shows an example with destaging to a mirrored volume; that is, one with two physical data copies.

With Version 7.3, the cache architecture changed from an upper-cache design to a two-layer cache design. With this change, the data is written once, and then it is directly destaged from the controller to the disk system.

Figure 6-7 shows the data flow in a stretched environment.

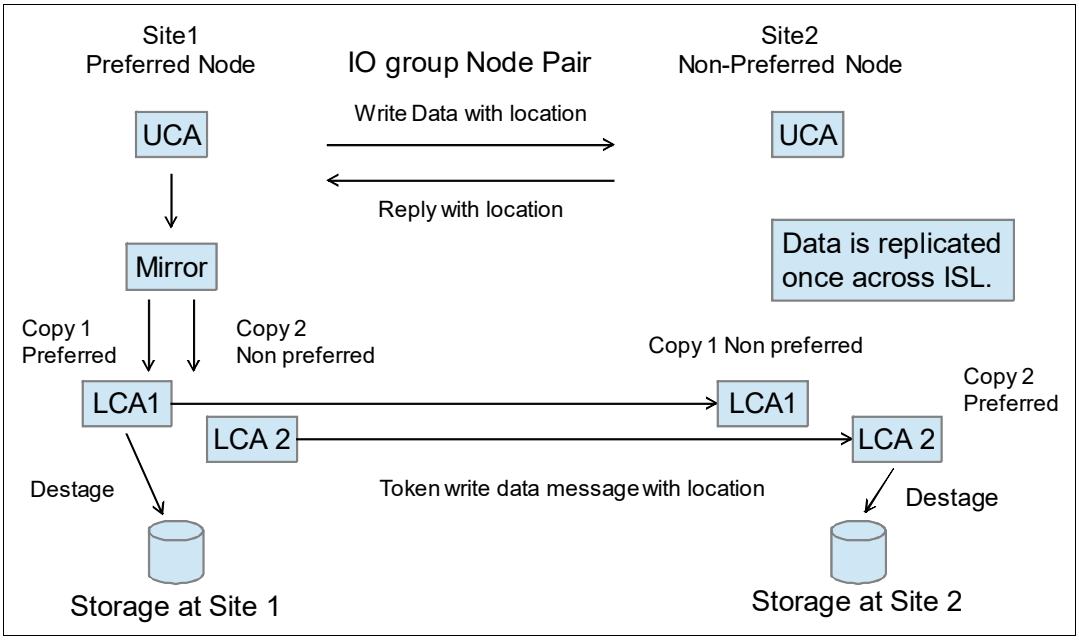


Figure 6-7 Design of an enhanced stretched cluster

6.2.6 Storage efficiency

When aiming for the maximum efficiency of data storage with IBM Storage Virtualize, you can configure DRPs that provide several technologies that increase the efficiency of physical storage use:

- Thin provisioning
- Deduplication (block-level and pattern-matching)
- Compression
- SCSI UNMAP support

Note: Storage efficiency options might require extra licenses and hardware components, depending on the model and configuration of your storage system.

Implementation of DRPs requires careful planning and sizing. Before configuring the first space-efficient volume on a storage system, see the relevant sections in Chapter 2, “Installation and configuration planning” on page 119, and Chapter 10, “Advanced Copy Services” on page 717.

DRPs use multithreading and hardware acceleration (where available) to provide storage efficiency functions on IBM Storage Virtualize storage systems. When considering the use of storage efficiency options, consider that they increase the number of I/O operations that the storage system must realize compared to accessing a basic volume. Space-efficient volumes require the storage system to write the data that is sent by the host and the metadata that is required to maintain a space-efficient volume.

Note: Because FCM drives include compression hardware, it provides data set size reduction with no performance penalty.

For more information about the storage efficiency functions of IBM Storage Virtualize, see Chapter 5, “Using storage pools” on page 365, and *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430.

It is possible to benefit from compression and data-at-rest encryption because encryption is done after compression. However, the size of data that is encrypted at the host level is unlikely to be reduced by compression or deduplication at the storage system.

Standard and thin-provisioned volumes

A standard-provisioned volume directly maps logical blocks on the virtual volume to physical blocks on storage media. Therefore, its virtual and physical capacities are identical.

A thin-provisioned volume has a virtual capacity that is larger than its physical capacity. Thin provisioning is the base technology for all space-efficient volumes. When a thin-provisioned volume is created, a small amount of the real capacity is used for initial metadata. This metadata holds a mapping of a set of continuous LBAs in the volume to a *grain* on a physically allocated extent.

Note: If you use thin-provisioned volumes, it is recommended to closely monitor the available space in the pool that contains these volumes. If a thin-provisioned volume does not have enough real capacity for a write operation, the volume is taken offline, and an error is logged. Limited ability exists to recover with UNMAP. Also, consider creating a fully allocated, empty, emergency space volume that can be deleted to temporarily create more space.

The grain size is defined when the volume is created and cannot be changed afterward. The grain size can be 32 KiB, 64 KiB, 128 KiB, or 256 KiB. The default grain size is 256 KiB, which is the preferred option. However, the following factors must be considered when deciding the grain size that is used:

- A smaller grain size helps to save space. If a 16 KiB write I/O requires a new physical grain to be allocated, the used space is 50% of a 32 KiB grain, but just over 6% of a 256 KiB grain. If no subsequent writes to other blocks of the grain occur, the volume provisioning is less efficient for volumes with larger grain.
- A smaller grain size requires more metadata I/O to be performed, which increases the load on the physical back-end storage systems.
- When a thin-provisioned volume is a FlashCopy source or target volume, specify the same grain size for FlashCopy and the thin-provisioned volume configuration. Use 256 KiB grain to maximize performance.
- The grain size affects the maximum size of the thin-provisioned volume. For 32 KiB size, the volume size cannot exceed 260 TiB.

Figure 6-8 on page 432 shows the thin-provisioning concept.

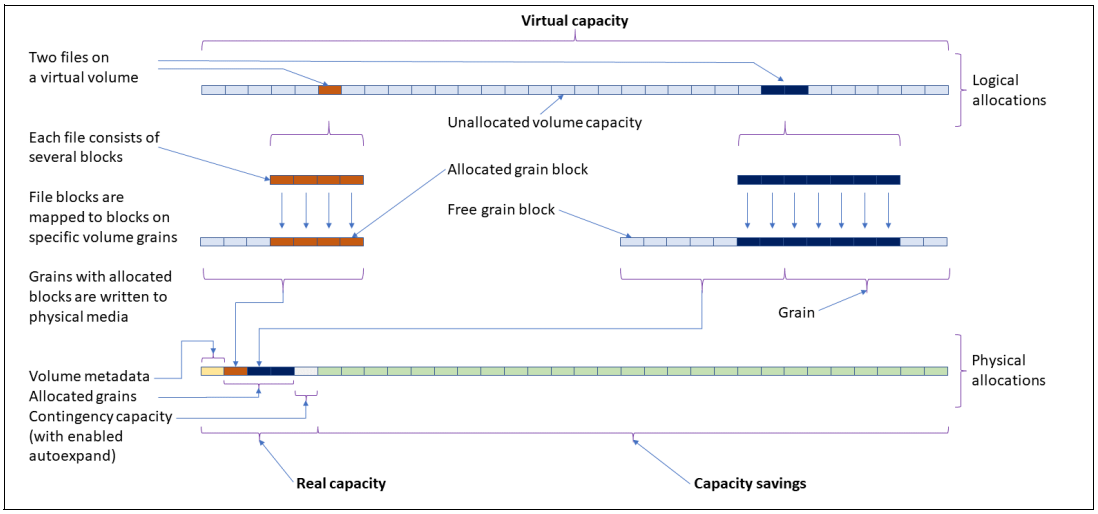


Figure 6-8 Conceptual diagram of a thin-provisioned volume

Thin-provisioned volumes use metadata to enable capacity savings, and each grain of user data requires metadata to be stored. Therefore, the I/O rates that are obtained from thin-provisioned volumes are lower than the I/O rates that are obtained from standard-provisioned volumes.

When a write request comes from a host, the block address for which the write is requested is checked against the mapping table. If the write is directed to a block that maps to a grain with physical storage that is allocated by a previous write, physical storage was allocated for this LBA and can be used to service the request. Otherwise, a new physical grain is allocated to store the data, and the mapping table is updated to record that allocation.

The metadata storage that is used is never greater than 0.1% of the user data. The resource usage is independent of the virtual capacity of the volume.

Thin-provisioned volume format: Thin-provisioned volumes do not need formatting. A read I/O, which requests data from deallocated data space, returns zeros. When a write I/O causes space to be allocated, the grain is “zeroed” before use. Also, when a full-grain write consists of “all zeros”, no space is physically allocated on disk.

The real capacity of a thin-provisioned volume can be changed if the volume is not in image mode. Thin-provisioned volumes use the grains of real capacity that is provided in ascending order as new data is written to the volume. If the user initially assigns too much real capacity to the volume, the real capacity can be reduced to free storage for other uses.

A thin-provisioned volume can be configured to autoexpand. This feature causes IBM Storage Virtualize to automatically add a fixed amount of extra real capacity to the thin-provisioned volume as required. Autoexpand does not cause the real capacity to grow much beyond the used capacity. Instead, it attempts to maintain a fixed amount of unused real capacity for the volume, which is known as the contingency capacity.

The contingency capacity is initially set to the real capacity that is assigned when the volume is created. If the user modifies the real capacity, the contingency capacity is reset to be the difference between the used capacity and real capacity.

A volume that is created without the autoexpand feature and has zero contingency capacity goes offline when the real capacity is used. It also receives a write request that requires real capacity allocation.

To facilitate management of the auto expansion of thin-provisioned volumes, a capacity warning must be set for the storage pools from which they are allocated. When the used capacity of the pool exceeds the warning capacity, a warning event is logged. For example, if a warning of 80% is specified, an event is logged when 20% of the pool capacity remains free.

A thin-provisioned volume can be converted nondisruptively to a standard-provisioned volume (or vice versa) by using the volume mirroring function. You can create a thin-provisioned copy to a standard-provisioned primary volume and then remove the standard-provisioned copy from the volume after they are synchronized.

The standard-provisioned to thin-provisioned migration procedure uses a zero-detection algorithm so that grains that contain all zeros do not use up any real capacity.

Thin-provisioned volumes can be used as volumes that are assigned to the host by FlashCopy to implement thin-provisioned FlashCopy targets. When creating a mirrored volume, a thin-provisioned volume can be created as a second volume copy, whether the primary copy is a standard or thin-provisioned volume.

Deduplicated volumes

Deduplication is a specialized data set reduction technique. However, in contrast to the standard file-compression tools that work on single files or sets of files, deduplication is a technique that is applied on a block level to larger scale data sets, such as a file system or volume. In IBM Storage Virtualize, deduplication can be enabled for thin-provisioned and compressed volumes that are created in DRPs.

IBM Storage Virtualize uses two techniques to detect duplicate data:

- Pattern matching
- Data signature (hash)

Deduplication works by identifying repeating chunks in the data that is written to the storage system. Pattern matching looks for a known data patterns (for example, “all ones”), and the data signature-based algorithm calculates a signature for each data chunk (by using a hash function) and checks whether the calculated signature is present in the deduplication database.

If a known pattern or a signature match is found, the data chunk is replaced by a reference to a stored chunk, which reduces the storage space that is required for storing the data. Conversely, if no match is found, the data chunk is stored without modification, and its signature is added to the deduplication database.

To maximize the space that is available for the deduplication database, the system distributes it between all nodes in the I/O groups that contain deduplicated volumes. Each node holds a distinct portion of the records that are stored in the database. If nodes are removed or added to the system, the database is redistributed between the anodes to ensure optimal use of available resources.

Depending on the data type that is stored on the volume, the capacity savings can be significant. Examples of use cases that typically benefit from deduplication are virtual environments with multiple VMs running the same operating system, and backup servers. In both cases, it is expected that multiple copies of identical files exist, such as components of the standard operating system or applications that are used in the organization.

Note: If data is encrypted by the host, expect no benefit from deduplication because the same cleartext (for example, a standard operating system library file) that is encrypted with different keys results in different output, which makes deduplication impossible.

Although deduplication (and other features of IBM Storage Virtualize) is transparent to users and applications, it must be planned for and understood before implementation because it might reduce the redundancy of a solution. For example, if an application stores two copies of a file to reduce the chances of data corruption because of a random event, the copies are deduplicated and the intended redundancy is removed from the system if these copies are on the same volume.

When planning the use of deduplicated volumes, be aware of update and performance considerations and the following software and hardware requirements:

- Nodes must have at least 32 GB to support deduplication. Nodes that have more than 64 GB can use a bigger deduplication fingerprint database, which might lead to better deduplication.

You must run supported hardware. For more information about the valid hardware and features combinations, see the **Planning the configuration** section within the IBM Documentation web page for your system's hardware:

  - IBM SAN Volume Controller
  - IBM FlashSystem 9500, 9200 and 9100
  - IBM FlashSystem 7200 and 7300
  - IBM FlashSystem 5000 and 5200

Compressed volumes

A volume that is created in a DRP can be compressed. Data that is written to the volume is compressed before committing it to back-end storage, which reduces the physical capacity that is required to store the data. Because enabling compression does not incur an extra metadata handling penalty, it is a best practice in most cases to enable compression on thin-provisioned volumes.

Notes: Consider the following points:

- When a volume is backed by FCM drives that compress data at line speed, the volume must be configured with compression that is turned on. IBM Storage Virtualize is tightly integrated with the storage controller and uses knowledge of the logical and physical space.
- You can use the management GUI or the CLI to run the built-in compression estimation tool. This tool can be used to determine the capacity savings that are possible for data on the system by using compression.
- Another benefit of data compression for volumes that are backed by flash-based storage is the reduction of write amplification, which has a beneficial effect on media longevity.

Capacity reclamation

File deletion in modern file systems is realized by updating file system metadata and marking the physical storage space that is used by the removed file as unused. The data of the removed file is not overwritten, which improves file system performance by reducing the number of I/O operations on physical storage that is required to perform file deletion.

However, this approach affects the management of the real capacity of volumes with enabled capacity savings. File system deletion frees space at the file system level, but physical data blocks that are allocated by the storage for the file still take up the real capacity of a volume.

To address this issue, file systems added support for the SCSI UNMAP command, which can be run after file deletion. It informs the storage system that physical blocks that are used by the removed file must be marked as no longer in use so that they can be freed. Modern operating systems run SCSI UNMAP commands to only storage that advertises support for this feature.

Version 8.1.0 and later releases support the SCSI UNMAP command on IBM Storage Virtualize systems, which enables hosts to notify the storage controller of capacity that is no longer required and can be reused or deallocated, which might improve capacity savings.

Note: For volumes that are outside DRPs, the complete stack from the operating system down to the back-end storage controller must support UNMAP to enable the capacity reclamation. SCSI UNMAP is passed only to specific back-end storage controllers.

Consider the following points:

- Version 8.1.2 can also reclaim capacity in DRPs when a host runs SCSI UNMAP commands.
- By default, Version 8.2.1 does not advertise support for SCSI UNMAP to hosts.
- In Version 8.3.1, support for the host SCSI UNMAP command is enabled by default.

Before enabling SCSI UNMAP, see SCSI Unmap support in IBM Storage Virtualize systems. It provides more information on SCSI Unmap commands, which software versions support the commands and other topics.

Analyze your storage stack to optimally balance the advantages and costs of data reclamation.

Data reduction at two levels

It is possible to design and implement a solution where data reduction technologies are applied at the IBM SAN Volume Controller and back-end storage. Such solutions realize small extra savings from compressing metadata

However, such cases require meticulous planning. Incorrect implementation can lead, for example, to an out-of-space event being triggered by DRP garbage collection process or Easy Tier hot data migrations.

Note: For more information about correctly planning advanced data reduction architectures, see the following resources:

- *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430
- Planning for data reduction pools and deduplicated volumes

6.2.7 Encryption

IBM Storage Virtualize systems can be configured to enable data-at-rest encryption. This function is realized in hardware (self-encrypting drives or in SAS controller for drives that do not support self-encryption and are connected through the SAS bus) or in software (external virtualized storage).

Note: Data-at-rest encryption can also be realized by back-end storage. When planning data encryption, consider the performance and operational aspects of all data encryption options. Incorrect encryption configuration that relies on back-end storage capabilities can result in inadvertent creation of volumes that are reported as encrypted but in reality, they store data in clear text.

For more information about creating and managing encrypted volumes, see 12.7, “Encryption” on page 1098.

6.2.8 Cache mode

Another volume parameter is its cache characteristics. Under normal conditions, a volume’s read and write data is held in the cache of its preferred node with a mirrored copy of write data that is held in the partner node of the same I/O group. However, it is possible to create a volume with different cache characteristics if this configuration is required.

The cache setting of a volume can have the following values:

`readwrite` All read and write I/O operations that are performed by the volume are stored in the cache. This mode is the default cache mode for all volumes.

`readonly` Read I/O operations that are performed on the volume are stored in the cache. Writes to the volume are not cached.

`disabled` No I/O operations on the volume are stored in the cache. I/Os are passed directly to the back-end storage controller rather than being held in the node’s cache.

Having cache-disabled volumes makes it possible to use the native copy services in the underlying RAID array controller for MDisks; that is, logical unit numbers (LUNs) that are used as IBM Storage Virtualize image mode volumes. However, the use of IBM Storage Virtualize Copy Services rather than the underlying disk controller copy services provides better results.

Note: Disabling the volume cache is a prerequisite for the use of native copy services on image mode volumes that are defined on storage systems that are virtualized by IBM Storage Virtualize. Contact IBM Support before turning off the cache for volumes in your production environment to avoid performance degradation.

6.2.9 I/O throttling

You can set a limit on the rate of I/O operations that are realized by a volume. This limitation is called I/O throttling or governing.

The limit can be set in terms of the number of IOPS or bandwidth measured in megabytes per second (MBps), gigabytes per second (GBps), or terabytes per second (TBps). By default, I/O throttling is disabled, but each volume can have up to two throttles that are defined: one for bandwidth and one for IOPS.

When deciding between the use of IOPS or bandwidth as the I/O governing throttle, consider the disk access profile of the application that is the primary volume user. Database applications generally issue large amounts of I/O operations, but transfer a relatively small amount of data. In this case, setting an I/O governing throttle that is based on bandwidth might not achieve much. A throttle that is based on IOPS is better suited for this use case.

Conversely, a video streaming or editing application issues a small amount of I/O, but transfers large amounts of data. Therefore, it is better to use a bandwidth throttle for the volume in this case.

An I/O governing rate of **0** does *not* mean that zero IOPS or bandwidth can be achieved for this volume. It means that no throttle is set for this volume.

Note: Consider the following points:

- I/O governing does not affect FlashCopy and data migration I/O rates.
- I/O governing on MM or GM secondary volumes does not affect the rate of data copy from the primary volume.

6.2.10 Volume protection

Volume protection prevents volumes or host mappings from being deleted if the system detects recent I/O activity. This global setting is enabled by default on new systems. You can set this value to apply to all volumes that are configured on your system or control whether the system-level volume protection is enabled or disabled on specific pools.

The volume protection must be enabled at two levels to be effective: system level and pool level. Both levels must be enabled for protection to be active on a pool. The pool-level protection depends on the system-level setting to ensure that protection is applied consistently for volumes within that pool. If system-level protection is enabled, but pool-level protection is not enabled, any volumes in the pool can be deleted.

When you enable volume protection at the system level, you specify a period in minutes that the volume must be idle before it can be deleted. If volume protection is enabled and the period is not expired, the volume deletion fails, even if the **-force** parameter is used. The following CLI commands and the corresponding GUI activities are affected by the volume protection setting:

- **rmvdisk**
- **rmvdiskcopy**
- **rmvvolume**
- **rmvdiskhostmap**
- **rmvolumehostclustermap**
- **rmmdiskgrp**
- **rmhostiogrp**
- **rmhost**
- **rmhostcluster**
- **rmhostport**
- **mkrcrelationship**

Volume protection can be set from the GUI (See 6.6.3, “Volume protection” on page 463) and CLI (See 6.7.9, “Volume protection” on page 502).

6.2.11 Secure data deletion

The system provides methods to securely erase data from a drive or boot drive when a control enclosure is decommissioned.

Secure data deletion effectively erases or overwrites all traces of data from a data storage device. The original data on that device becomes inaccessible and cannot be reconstructed. You can securely delete data on individual drives and on a boot drive of a control enclosure. The methods and commands that are used to securely delete data enable the system to be used in compliance with European Regulation EU2019/424.

For more information, see Secure data deletion.

6.3 Volume groups

A volume group is a container for managing a set of related volumes as a single object. The volume group provides consistency across all volumes in the group.

Volume groups can be used with the following functions:

- Safeguarded Copy function
- Policy-based replication
- Snapshot function

Safeguard Copy function

One implementation of volume groups is to group volumes to be configured as Safeguarded. Safeguarded copy function is a cyber-resiliency feature that creates immutable copies of data that cannot be changed or manipulated.

A Safeguarded volume group describes a set of source volumes that can span different pools and are backed up collectively with the Safeguarded Copy function. Safeguarded snapshots are supported on the system through an internal scheduler that is defined in the snapshot policy or can be configured with an external snapshot scheduling application such as IBM Copy Services Manager.

Policy-based replication

You can use volume groups for policy-based replication. Policy-based replication is configured on all volumes in a volume group by assigning a replication policy to that volume group. The system automatically replicates the data and configuration for volumes in the group based on the values and settings in the replication policy. As part of policy-based replication, a recovery volume group is created automatically on the recovery system. Recovery volume groups cannot be created, changed, or deleted. A single replication policy can be assigned to multiple volume groups to simplify replication management. When additional volumes are added to the group, replication is automatically configured for these new volumes. Policy-based replication supports configuration changes while the partnership is disconnected. After the partnership is reconnected, the system automatically reconfigures the recovery system.

Snapshot function

Snapshots are the read-only point-in-time copies of a volume group that cannot be directly accessible from the hosts. To access the snapshot contents, you can create a clone or thin clone of a volume group snapshot. You can use the command line interface or management GUI to configure volume groups to use snapshot policies for multiple

volumes for consistent management. Safeguarded snapshot with internal scheduler can be created by using snapshot function.

The volumes in a volume group are supposed to be mutually consistent. This means that volume group makes sense only as a group. When a clone or thin-clone of a volume group is populated, the snapshot function ensures that the images are mutually consistent. When volumes are added or removed from a group, the host applications ensure that the volume groups are mutually consistent.

6.3.1 Creating volume groups

The section focuses on the means of defining volume groups in the management GUI. As with other management GUI functions, there are equivalent CLI commands.

One means of creating a volume group is by selecting **Volumes** → **Volume Groups** (see Figure 6-9)

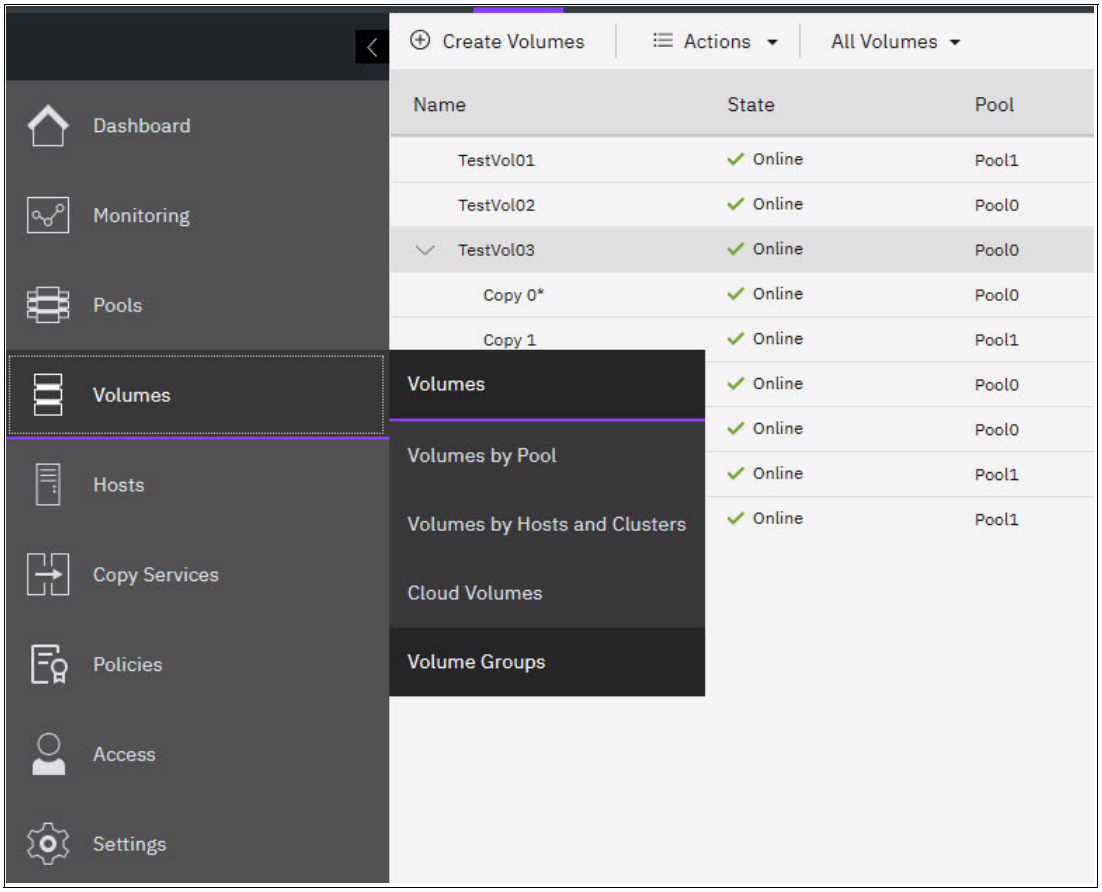


Figure 6-9 Creating a volume group

The main Volume Groups view is presented (see Figure 6-10).

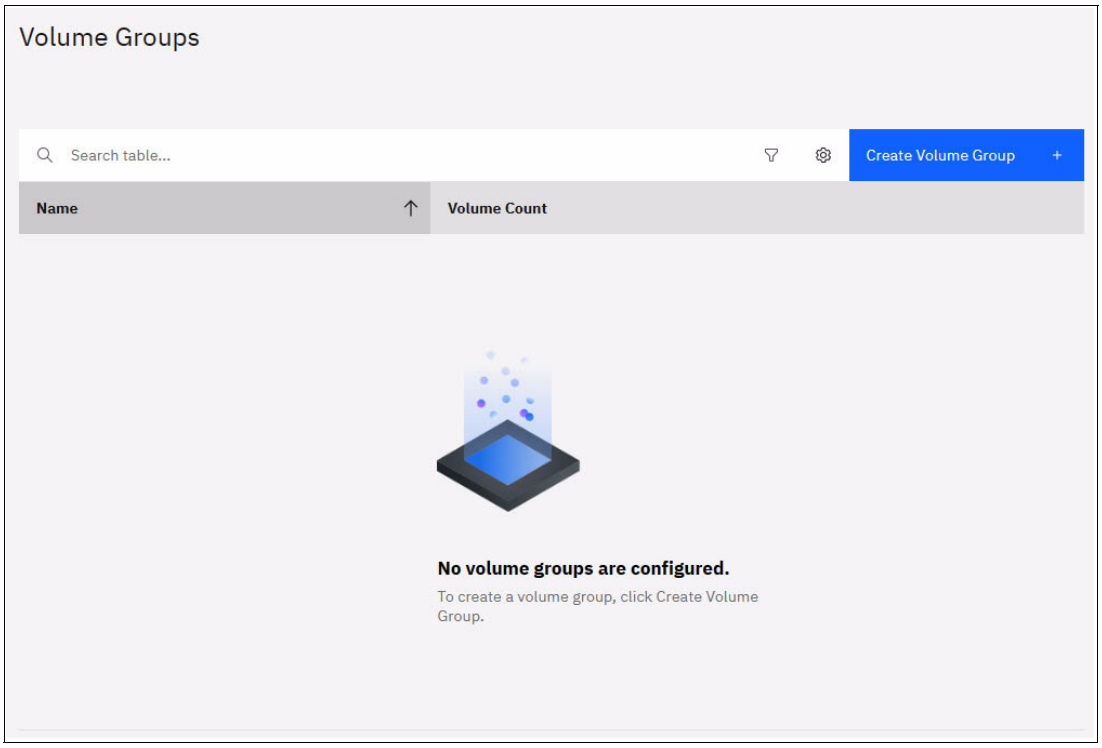


Figure 6-10 Volume Groups view

This view allows you to manage all volume groups. Initially, the view is empty. Select **Create Volume Group** to define a new group.

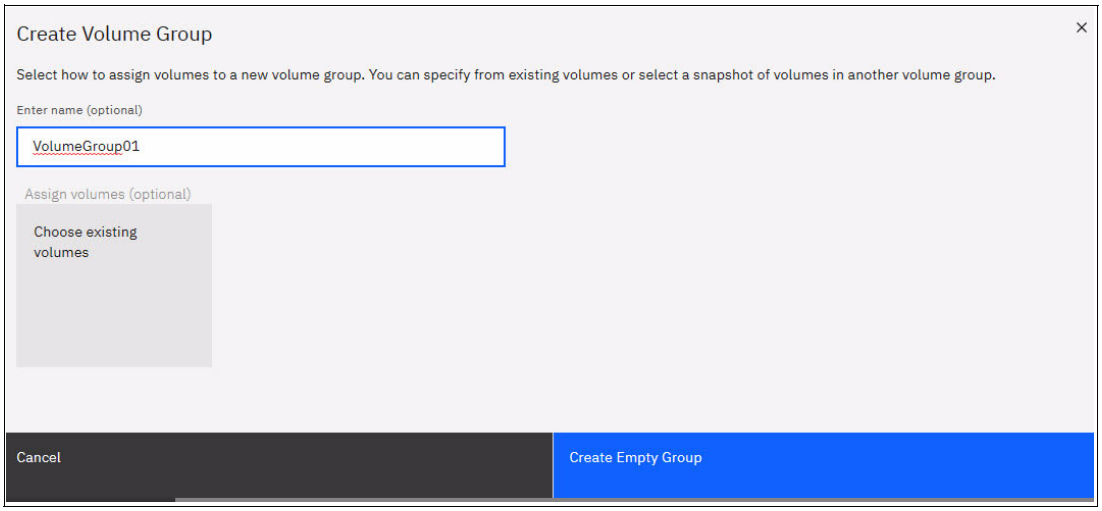


Figure 6-11 Create Volume Group dialog

You have the option to create an empty group, or you can add existing volumes to the group. In this example, an empty group was defined. After the process completes, it is listed in the group view. See Figure 6-12 on page 441.

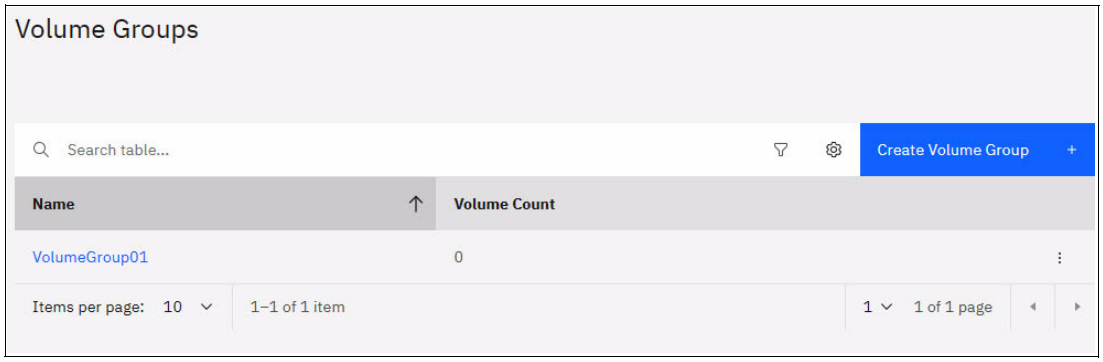


Figure 6-12 Updated Volume Group view

To add existing volumes, select **Choose existing volumes**. See Figure 6-13.

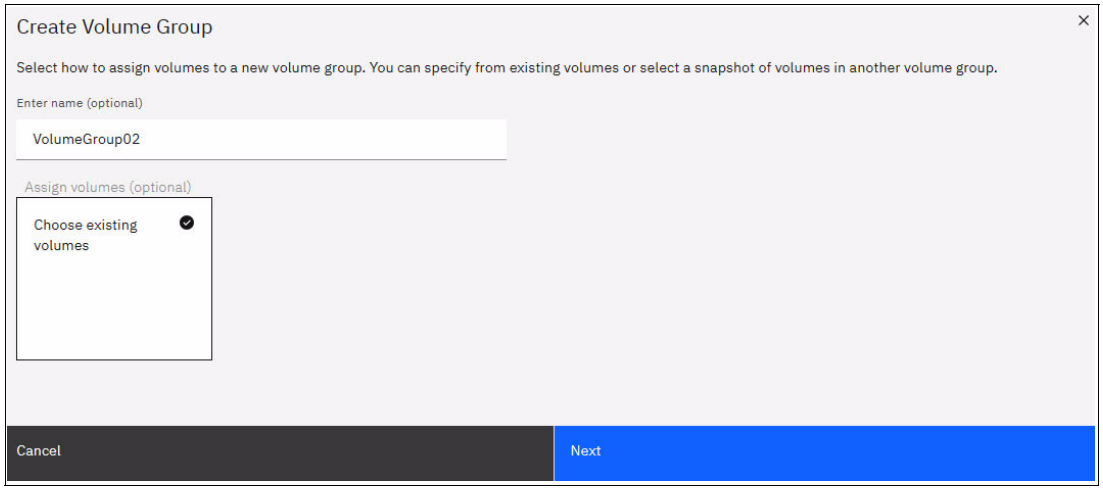


Figure 6-13 Selecting existing volumes for a volume group

Selecting **Next** opens the list of available volumes.

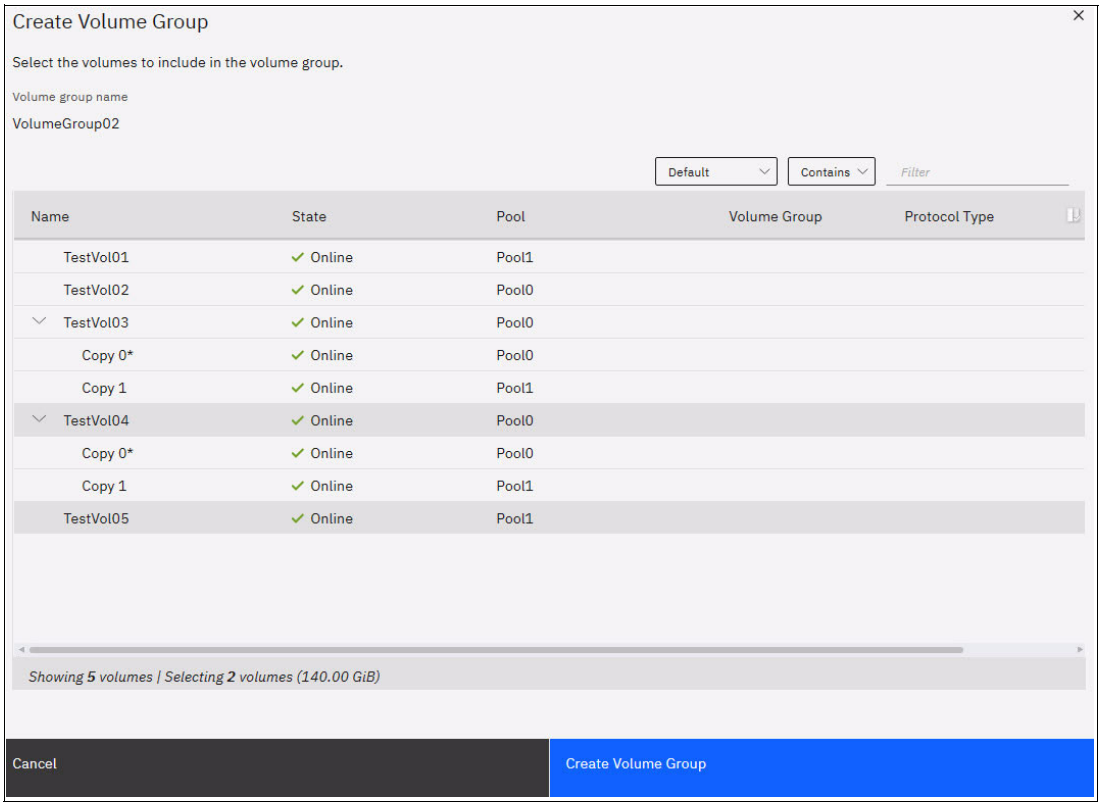


Figure 6-14 Volume selection

Clicking **Create Volume Group** completes the process and adds the newly defined group to the list of groups. See Figure 6-15.

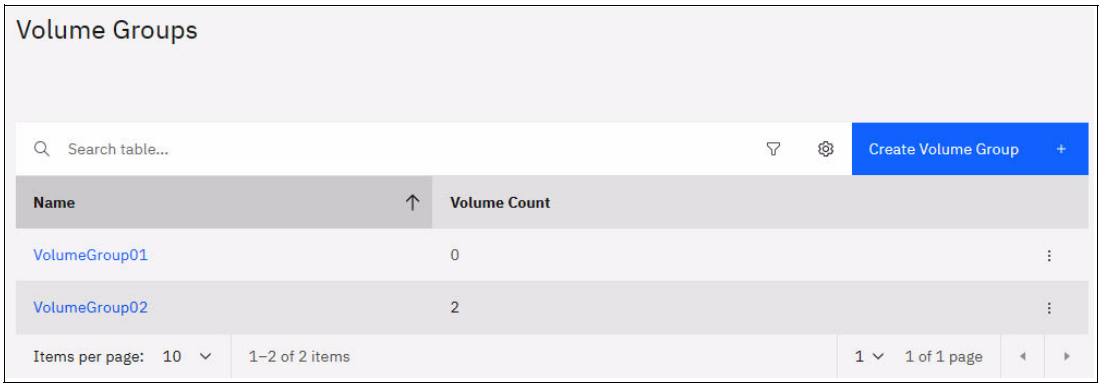


Figure 6-15 Updated Volume Group list

You can also manage volume groups via **Actions** in the main Volumes view.

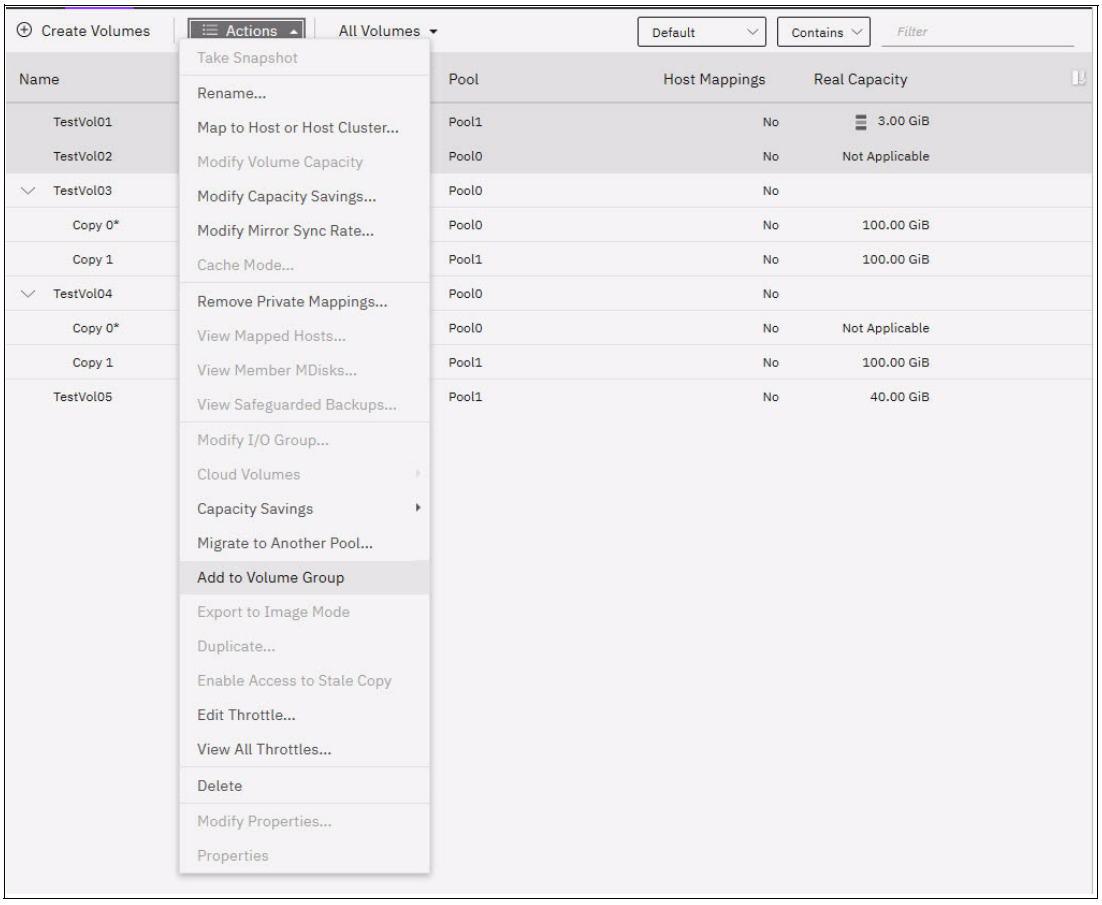


Figure 6-16 Managing volume groups by using the Volumes view

You cannot create a volume group from the list of volumes. Volumes can be added only to existing groups. In this case TestVol01 and TestVol02 were added to VolumeGroup01. See Figure 6-17.

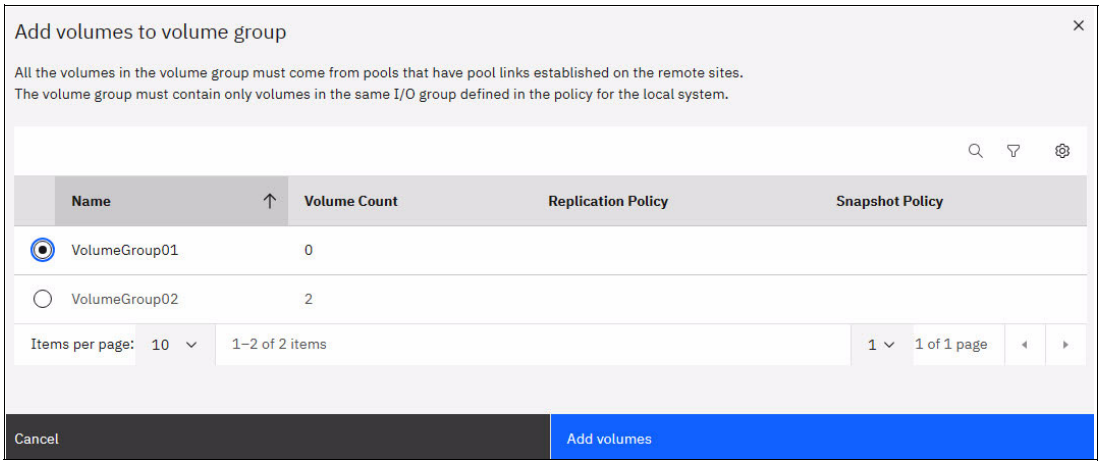


Figure 6-17 Add volumes to volume group

Clicking **Add volumes** initiates the addition.

6.4 Virtual volumes

IBM Storage Virtualize V7.6 introduced support for virtual volumes. Virtual volumes enable support for vVols, which allow VMware vCenter to manage system objects, such as volumes and pools. The IBM Storage Virtualize system administrators can create volume objects of this class, and assign ownership to VMware administrators to simplify management.

For more information about configuring vVols with IBM Storage Virtualize, see *Configuring VMware Virtual Volumes for Systems Powered by IBM Spectrum Virtualize*, SG24-8328.

6.5 Volumes in multi-site topologies

IBM Storage Virtualize can be set up in a multi-site configuration, which makes the system aware of which system components (I/O groups, nodes, and back-end storage) are at which site. For the storage system topology description, a site is defined as an independent failure domain, which means that if one site fails, the other site can continue to operate without disruption.

Depending on the type and scale of the failure that the solution must survive, the sites can be two places in the same data center room (one end of the IBM Storage Virtualize system) or buildings in different cities on different tectonic plates and powered from independent grids (the other end of the IBM Storage Virtualize system).

The following storage system topologies are available:

- Standard topology, which is intended for single-site configurations that do not allow site definition and assume that all components of the solution are at a single site. You can use GM or MM to maintain a copy of a volume on a different system at a remote site, which can be used for DR.
- IBM HyperSwap topology, which is a three-site high availability (HA) configuration in which each I/O group is at a different site. A volume can be active on two I/O groups so that if one site is not available, it can immediately be accessed through the other site.
- Stretched topology (Enhanced Stretched Cluster). When set up in the stretched topology, each node of an I/O group of the storage system is at a different site and volumes have a copy at each site. Access to a volume can continue when one site is not available but with reduced performance. This topology is also known as an *enhanced stretched system*.

  The reason an enhanced stretched system is used rather than the HyperSwap topology might be because use of Global Mirror (GM) or Metro Mirror (MM) to a third site, or because the system was configured as an enhanced stretched system before HyperSwap being released.

The stretched topology uses fewer system resources compared to HyperSwap, which allows more highly available volumes to be configured. However, during a disaster that makes one site unavailable, the system cache on the nodes of the surviving site are disabled.

The HyperSwap topology uses extra system resources to support a full independent cache on each site, which allows full performance, even if one site is lost. In some environments, a HyperSwap topology provides better performance than a stretched topology.

If the objective of your solution design is HA, it is better to use an IBM HyperSwap topology. However, if the objectives include DR, complex copy services, or highest scalability, see the *Planning for high availability* section of the IBM Documentation site for your system:

  - IBM SAN Volume Controller

- IBM FlashSystem 9500, 9200 and 9100
- IBM FlashSystem 7200 and 7300
- IBM FlashSystem 5000 and 5200

Note: Multi-site topologies of IBM Storage Virtualize use two sites as locations for the nodes and back-end storage. The third site is used as a location for a tie-breaker component that prevents split-brain scenarios if the storage system components lose communication with each other.

The Create Volumes menu provides the following options, depending on the configured system topology:

- Standard topology which includes Basic, Mirrored, and Custom
- HyperSwap topology which includes Basic, HyperSwap, and Custom
- Stretched topology which includes Basic, Stretched, and Custom

The HyperSwap function provides HA volumes that are accessible through two sites up to 300 km (186.4 miles) apart. A fully independent copy of the data is maintained at each site.

Note: The determining factor for HyperSwap configuration validity is the time that it takes to send the data between the sites. Therefore, while estimating the distance, consider the fact that the distance between the sites that is measured along the data path is longer than the geographic distance. Also, each device on the data path that adds latency increases the effective distance between the sites.

6.5.1 HyperSwap topology

In the HyperSwap topology, both nodes of an I/O group are in the same site. Therefore, to get a volume resiliently stored on both sites, the storage system must be composed of least two I/O groups.

When data is written by hosts at either site, both copies are synchronously updated before the write operation completion is reported to the host. The HyperSwap function automatically optimizes to minimize data that is transmitted between sites and to minimize host read and write latency.

If the nodes or storage at either site goes offline, the HyperSwap function automatically fails over access to the other copy. The HyperSwap function also automatically resynchronizes the two copies when possible.

The HyperSwap function is built on a foundation of earlier technologies:

- The Nondisruptive Volume Move (NDVM) function that was introduced in IBM Storage Virtualize V6.4
- RC features that include MM, GM, and GMCV.

The HyperSwap volume configuration is possible only after the IBM Storage Virtualize system is configured in the HyperSwap topology. After this topology change, the GUI presents an option to create a HyperSwap volume and creates them by running the **mkvolume** command instead of the **mkvdisk** command. The GUI continues to use the **mkvdisk** command when all other classes of volumes are created.

Note: For more information, see *IBM Storwize V7000, Storage Virtualize, HyperSwap, and VMware Implementation*, SG24-8317.

For more information, see HyperSwap system configuration details.

From the perspective of a host or a storage administrator, a HyperSwap volume is a single entity, but it is realized by using four volumes, a set of FlashCopy maps, and an RC relationship (see Figure 6-18).

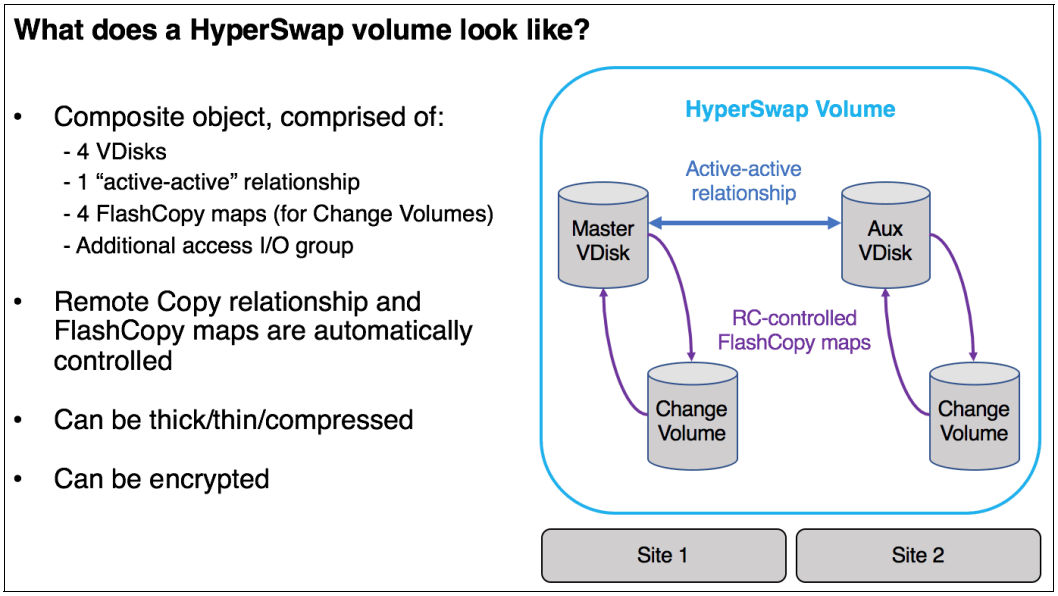


Figure 6-18 What makes up a HyperSwap volume

The GUI simplifies the HyperSwap volume creation process by asking about required volume parameters only and automatically configuring all the underlying volumes, FlashCopy maps, and volume replication relationships.

6.5.2 Stretched topology

In the stretched topology each I/O group in the system has one node on one site, and one node on the other site. The topology works with any number of I/O groups from 1 to 4.

The Enhanced Stretched Cluster function should be used if:

- You need synchronous or asynchronous replication by using MM or GM to a third site. It is possible to extend the configuration to four sites if the remote system is also configured with a stretched system topology.
- You need the absolute maximum number of highly available volumes in the system.

It is important to note that stretched topology is only available for SAN Volume Controller systems.

6.6 Operations on volumes

This section describes how to perform operations on volumes by using the GUI. The following operations can be performed on a volume:

- Volumes can be created and deleted.
- Volumes can have their characteristics modified, including:
  - Size (expanding or shrinking)
  - Number of copies (adding or removing a copy)
  - I/O throttling
  - Protection
- Volumes can be migrated at run time to another MDisk or storage pool.
- A PiT volume snapshot can be created by using FlashCopy. Multiple snapshots and quick restore from snapshots (reverse FlashCopy) are supported.
- Volumes can be mapped to (and unmapped from) hosts.

Note: It is possible to prevent accidental deletion of volumes if they recently performed any I/O operations. This feature is called volume protection, and it prevents active volumes or host mappings from being deleted inadvertently. This process is done by using a global system setting.

For more information, see 6.7.9, “Volume protection” on page 502 and Changing volume protection settings.

6.6.1 Creating volumes

This section focuses on the use of the Create Volumes menu which has been redesigned.

Individual volumes

Perform the following steps to create a volume.

1. To create a volume select **Volumes** → **Volumes** in the IBM Storage Virtualize GUI, as shown in Figure 6-19 on page 448.

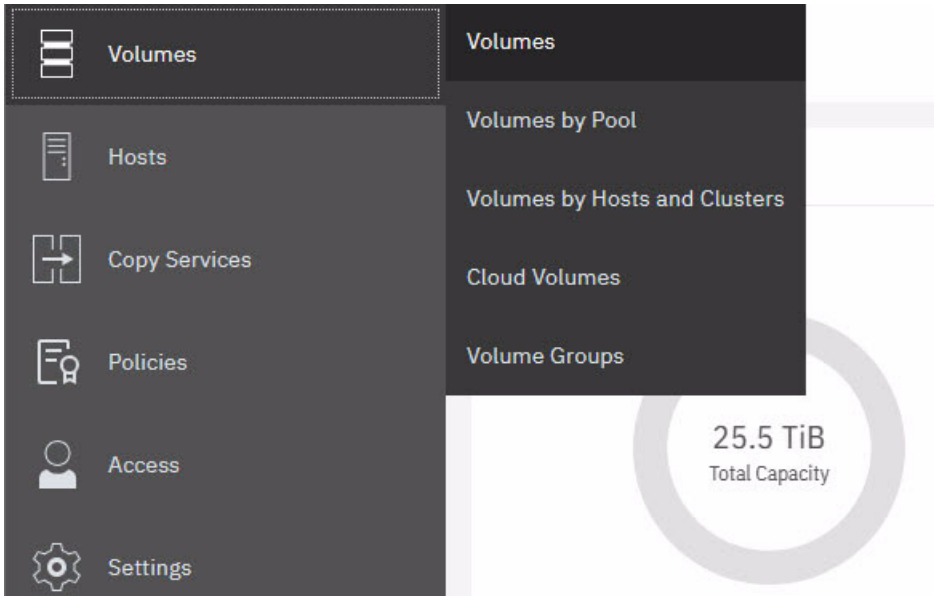


Figure 6-19 Volumes menu

2. A list of volumes, their state, capacity, and associated storage pools are displayed. Clicking **Create Volumes**, as shown in Figure 6-20, opens the Create Volumes view, see Figure 6-21 on page 449.

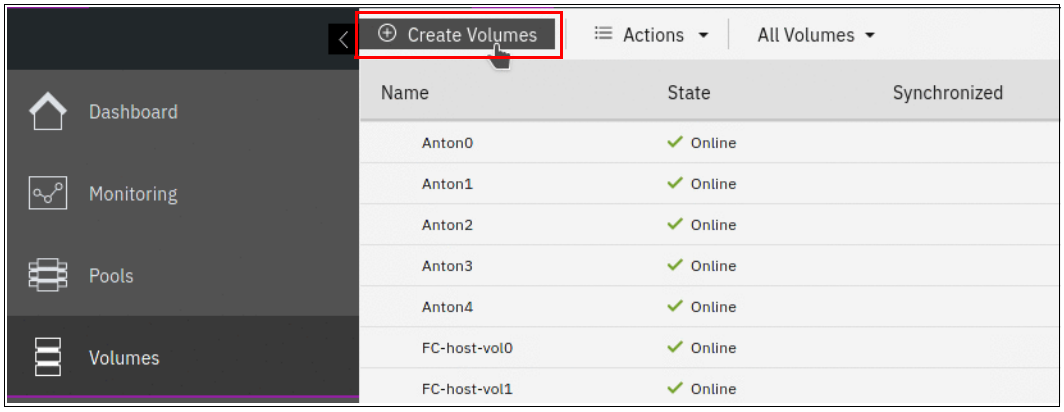


Figure 6-20 Create Volumes

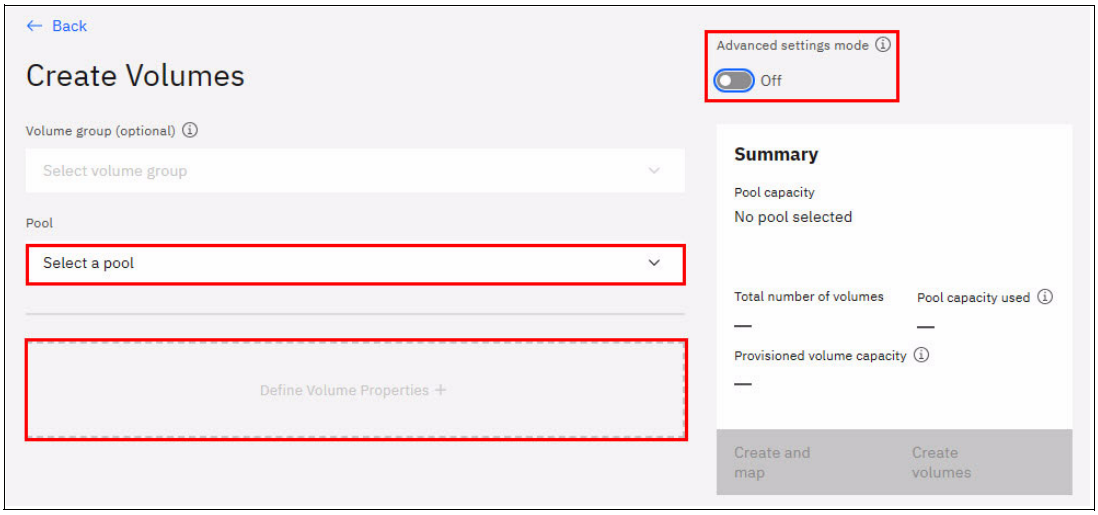


Figure 6-21 Creating volumes, advance settings mode disable

3. When **Advanced setting mode** is disabled there is one required selection, Pool, and one option selection, Volume group. After selecting the pool, the **Define Volume Properties** link is enabled. See Figure 6-22.

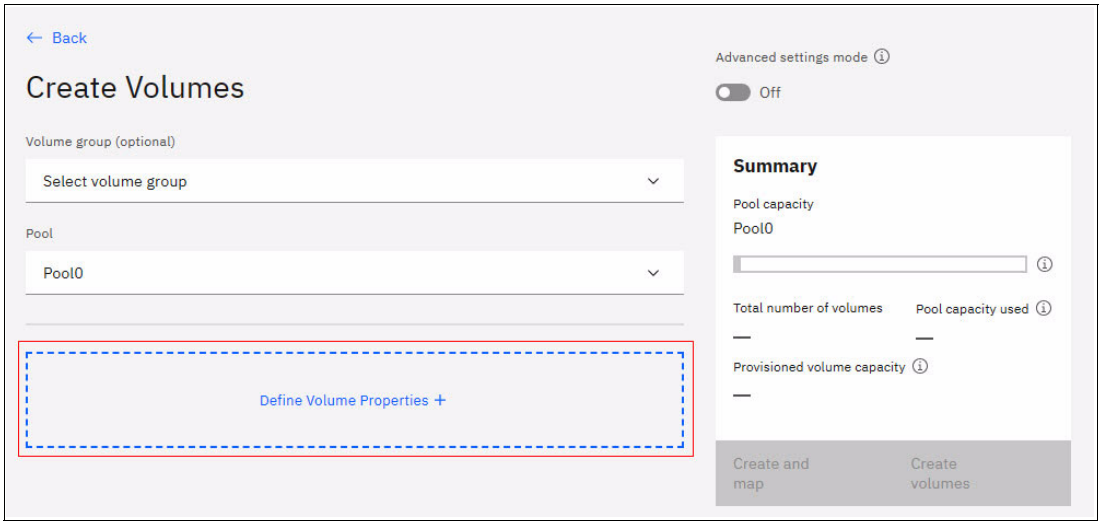


Figure 6-22 Define volume properties

4. The content of the **Define Volume Properties** dialog is dependent on the pool type and topology of the cluster. In Figure 6-22 a data reduction pool was selected. For DRP you can implement capacity savings by defining the volume as compressed and deduplicated. See Figure 6-23 on page 450.

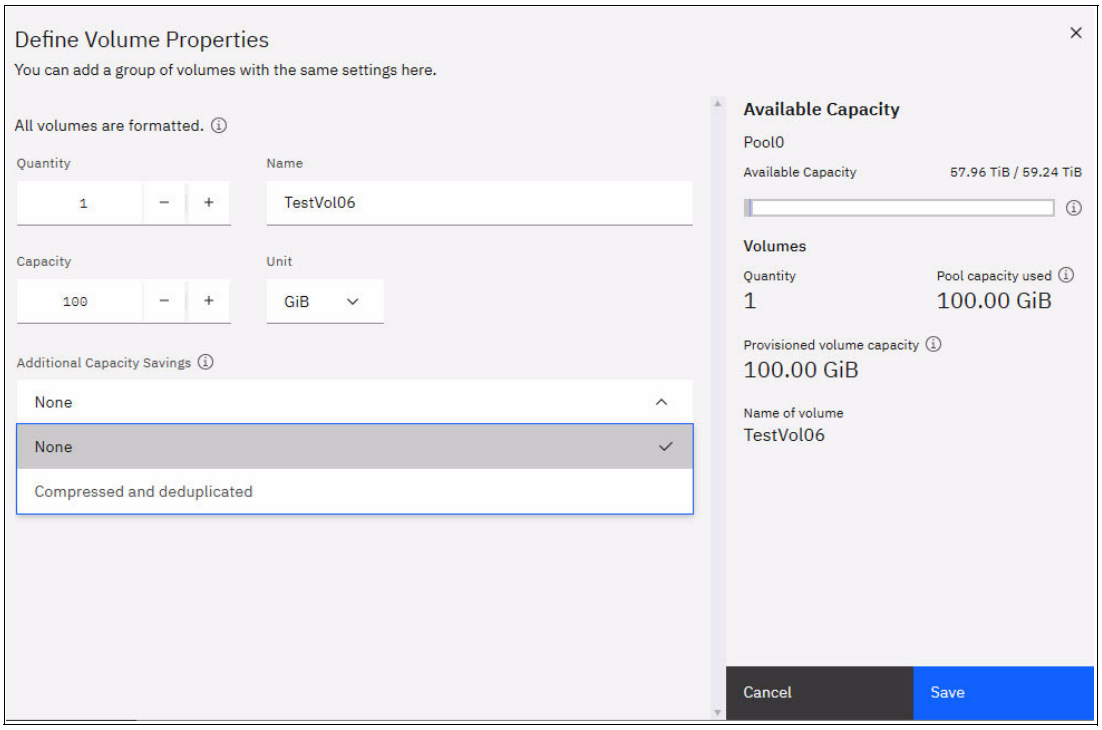


Figure 6-23 Create volume, DRP (compressed and deduplicated)

5. Click **Save** to return focus to the **Create Volumes** view. You can select **Create and map** the volume or **Create volumes**. See Figure 6-24.

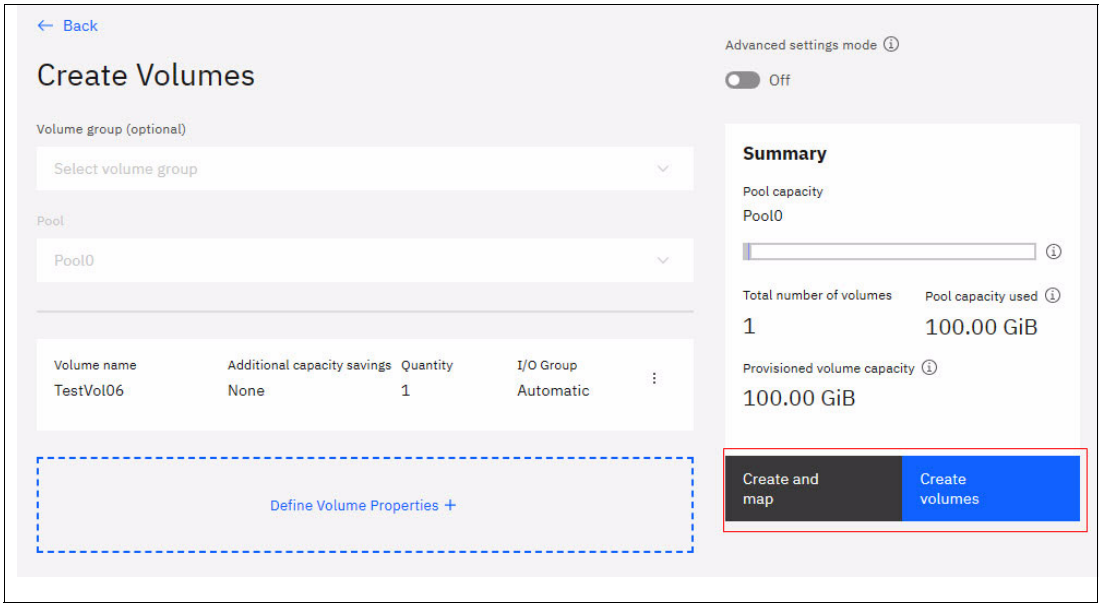


Figure 6-24 Create volume, DRP (compressed and deduplicated)

Advanced settings mode

More control over volume definitions is available when you toggle the **Advanced settings mode** switch to the **On** position. see Figure 6-25 on page 451.

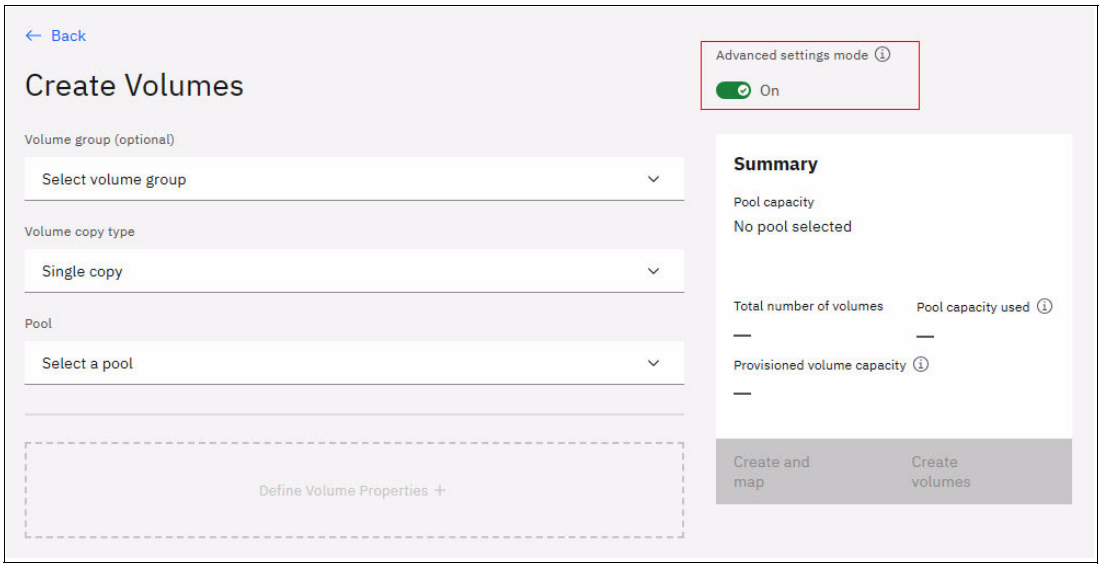


Figure 6-25 Advanced settings mode enabled

Enabling advanced settings allows you to amend the volume copy type, capacity requirements and I/O settings. When you define a mirrored volume, select the pool for Copy 0 and Copy 1. See Figure 6-26.

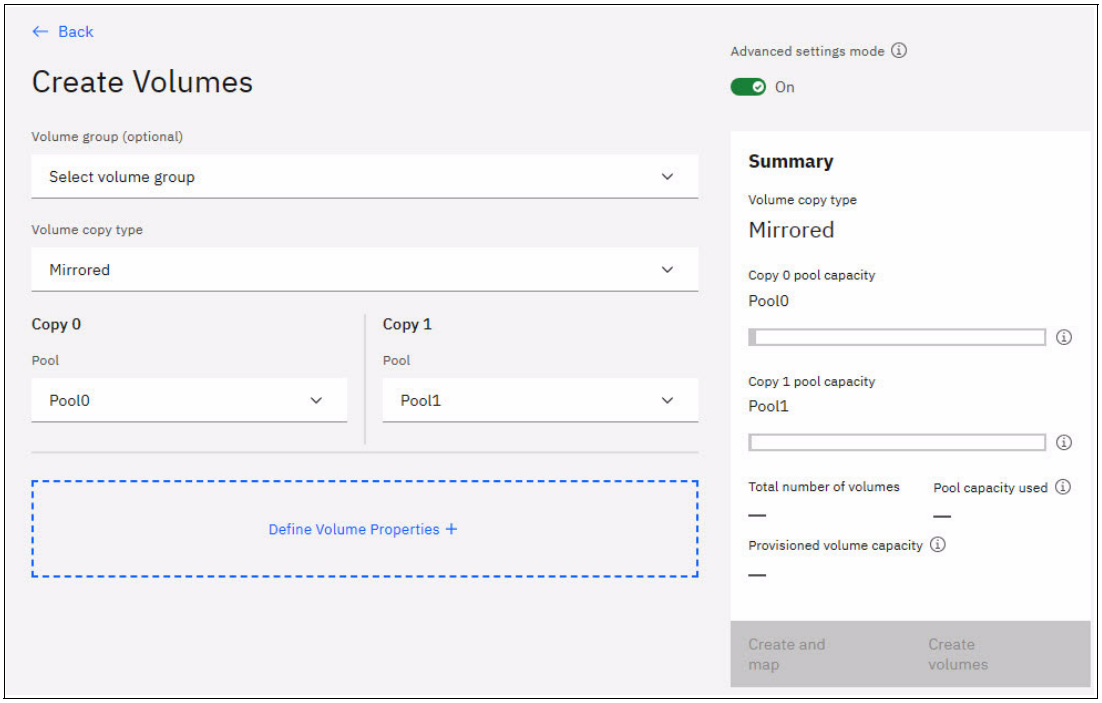


Figure 6-26 Mirrored volume

In the **Defined Volume Properties** dialog, in addition to specifying the name, capacity, and capacity savings, you can also configure the following options (Figure 6-27):

- Caching I/O Group
- Accessible I/O Group
- Preferred Node
- Cache Mode
- Mirror Sync Rate

After you click **Save**, you can select **Create and map** to map the volume to a host or select **Create volumes** to create the volume without mapping it.

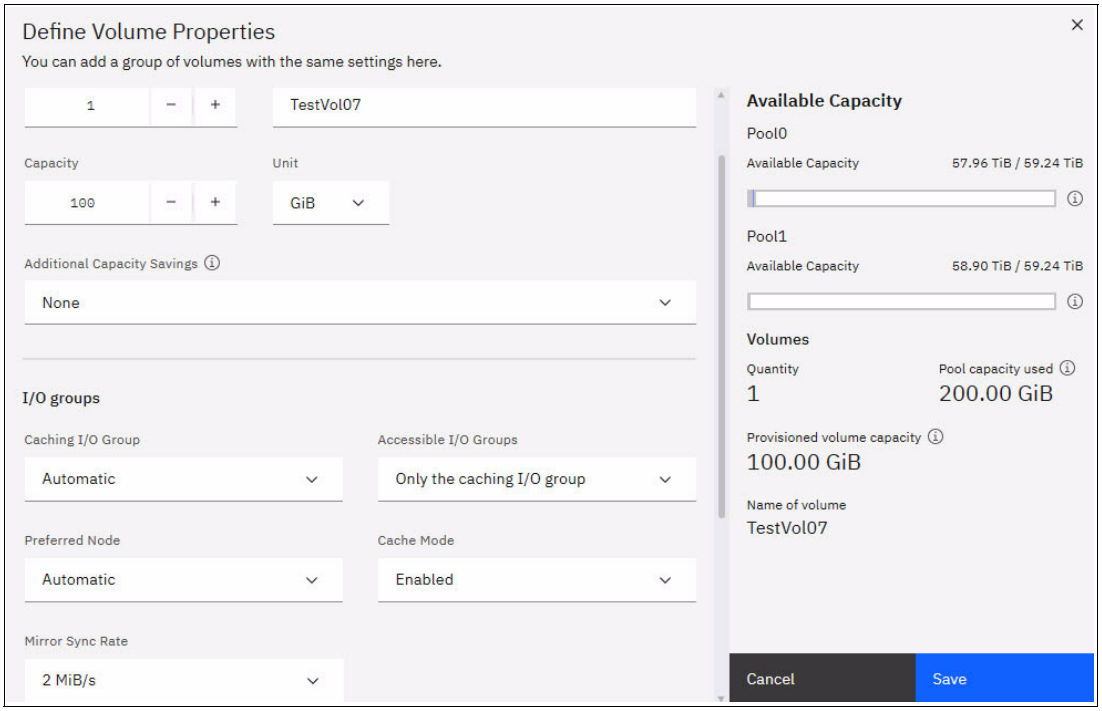


Figure 6-27 Mirrored volume advanced settings

Multiple volumes

You can define multiple volumes. Increment the **Quantify** field to define more than one volume, and note that a **Suffix Range** field is shown. (Figure 6-28).

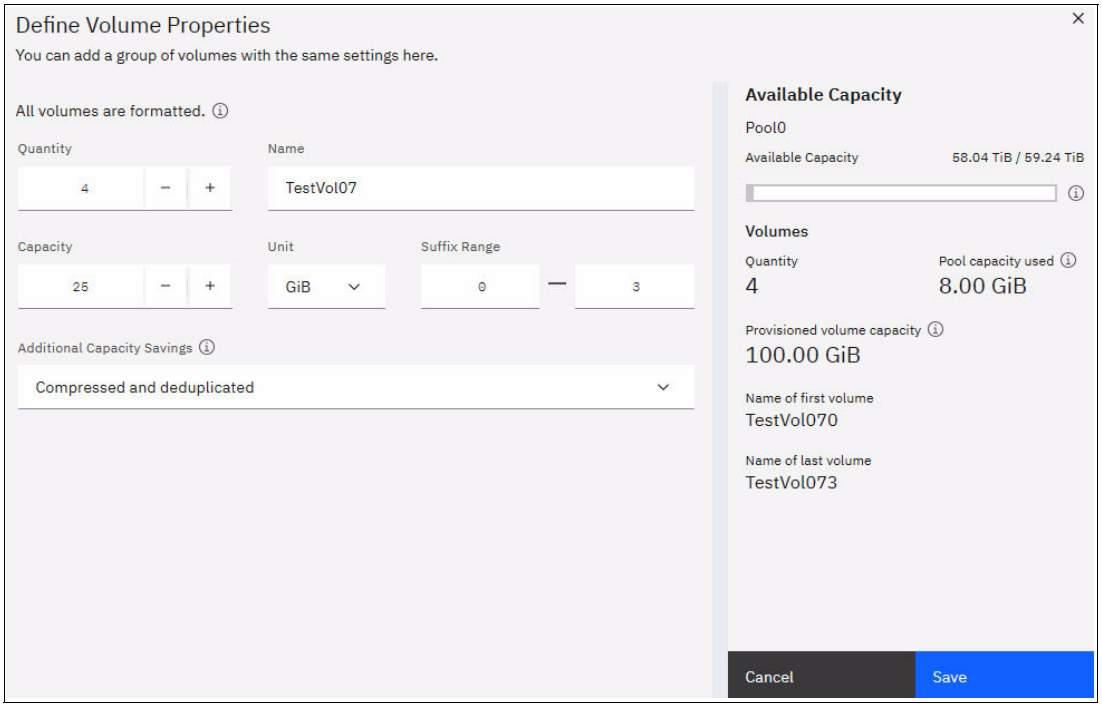


Figure 6-28 Multiple volumes advanced settings disabled

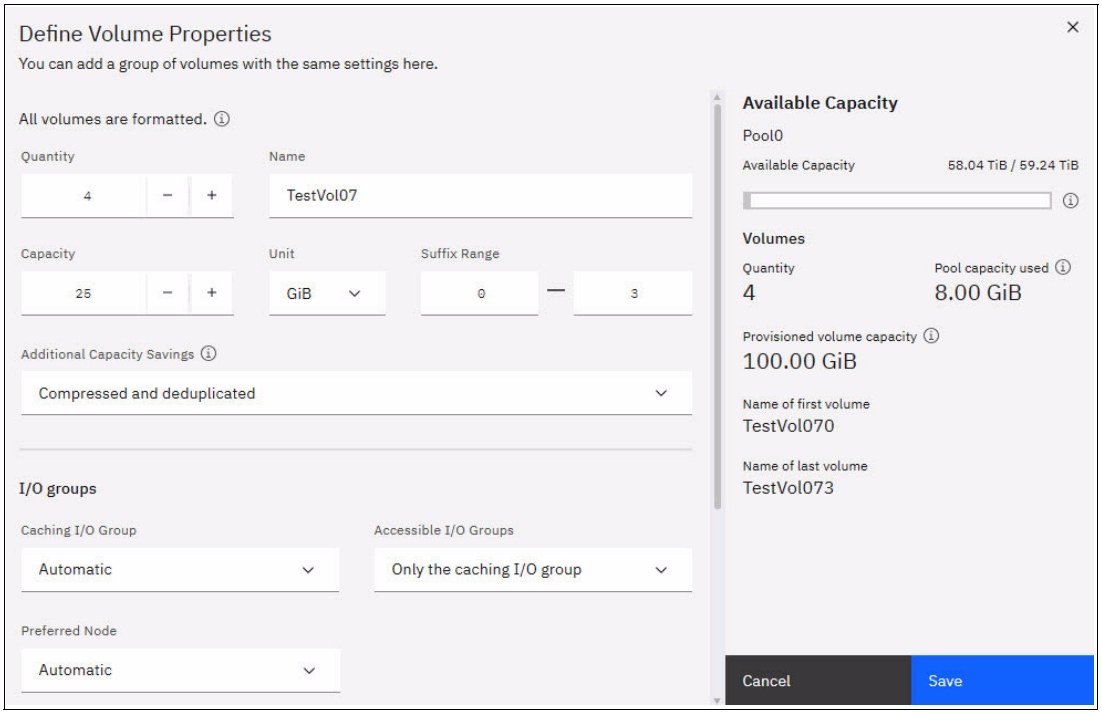


Figure 6-29 Multiple volumes advanced settings enabled

The suffix values will be appended to the name that is provided. In this case, the specified name is **TestVol07**, and the names of the 4 volumes are **TestVol070**, **TestVol071**, **TestVol072**, and **TestVol073** are created. See Figure 6-30.

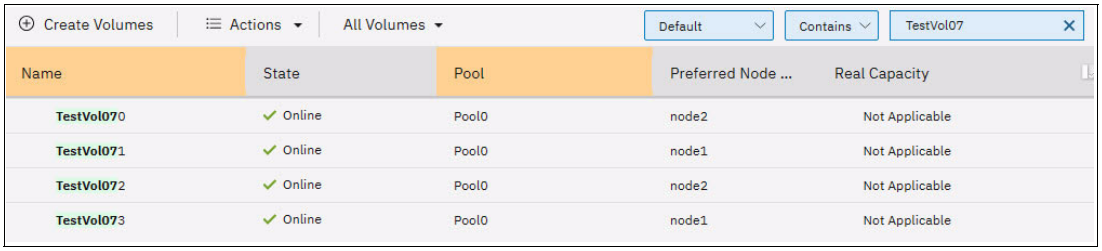


Figure 6-30 Multiple volumes created

HyperSwap volumes

The GUI to create HyperSwap volumes includes pool selection at each site. The same general steps are required. You select a pool at each site and select a volume group for the volumes being created. See Figure 6-31.

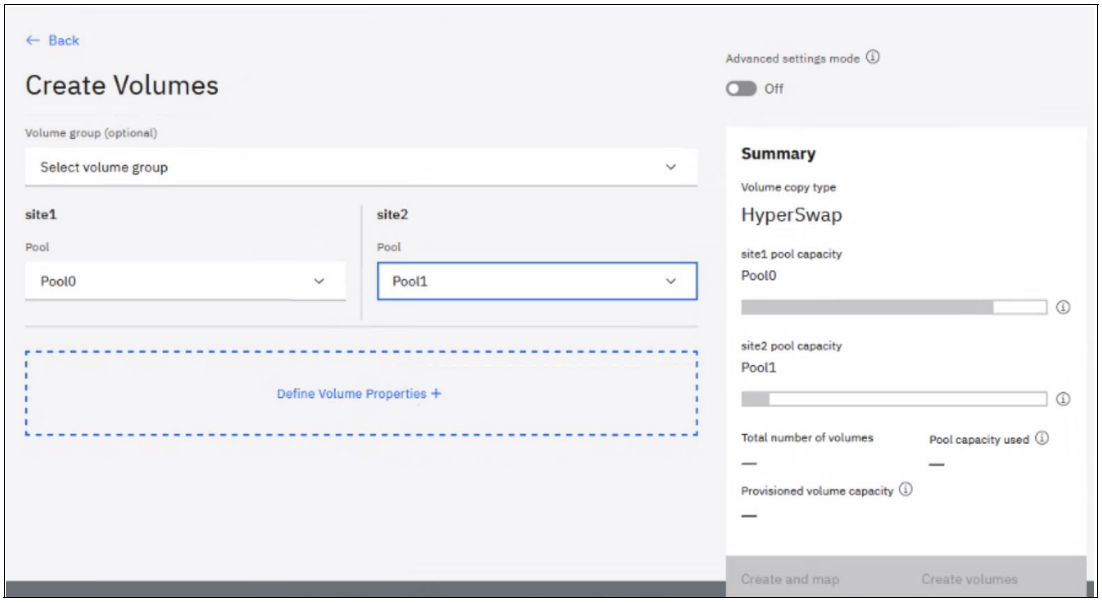


Figure 6-31 Create HyperSwap volume, advanced settings mode disable

After you select the Pools, select **Define Volume Properties**. You can define the following properties (Figure 6-32 on page 455):

- Quantity
- Name
- Capacity
- Additional capacity savings

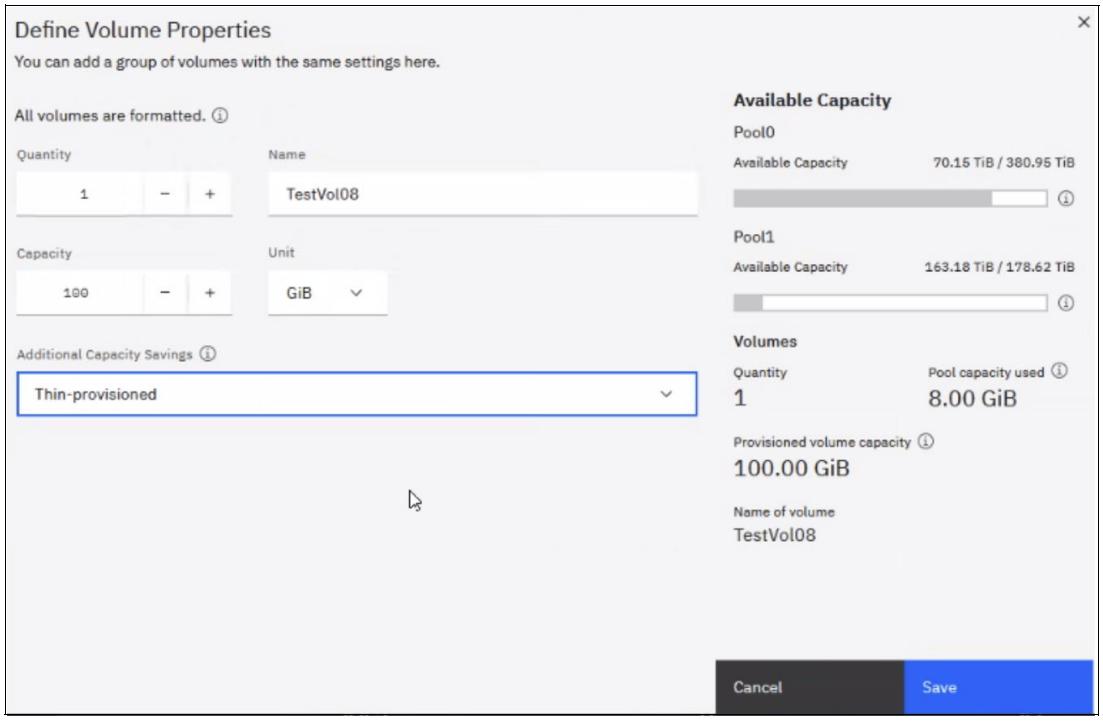


Figure 6-32 Define HyperSwap volume properties, advanced settings disabled

After you click **Save**, you can select **Create and map** to create the volumes and map the new volumes to a host or select **Create volumes** to create the volumes without mapping them.

Advanced settings mode

If you toggle the **Advanced settings mode** switch to **On**, then you can specify the preferred site for the created volume. See Figure 6-33.

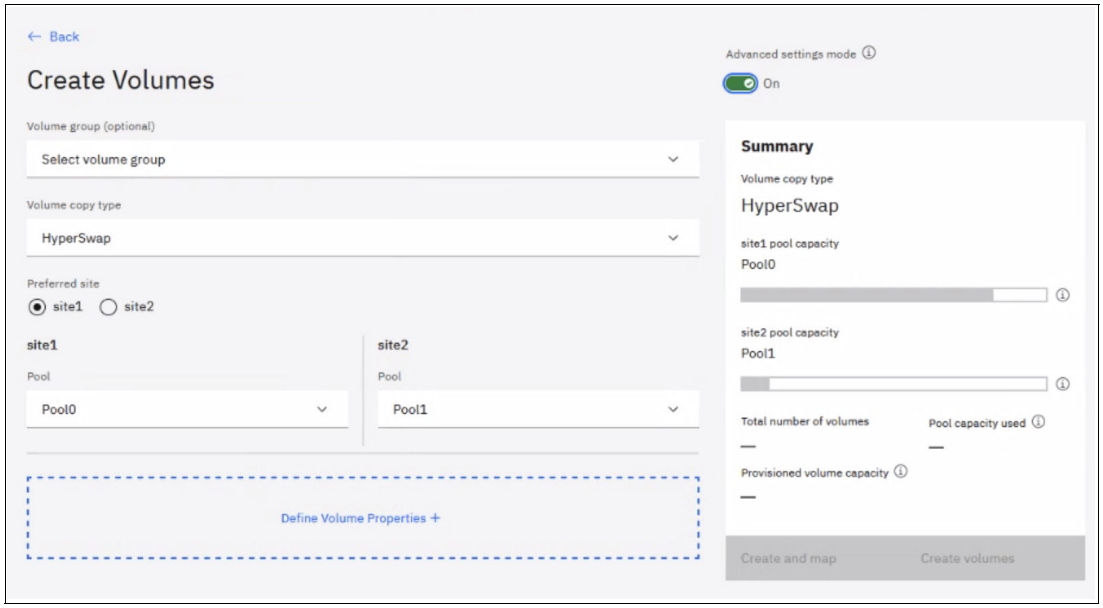


Figure 6-33 Create HyperSwap volume, advanced settings mode enable

Additionally, in the **Defined Volume Properties** window (Figure 6-34), you can configure the following properties:

- Name
- Capacity
- Capacity savings
- Caching I/O Group for both sites
- Accessible I/O Group
- Cache Mode
- Mirror Sync Rate

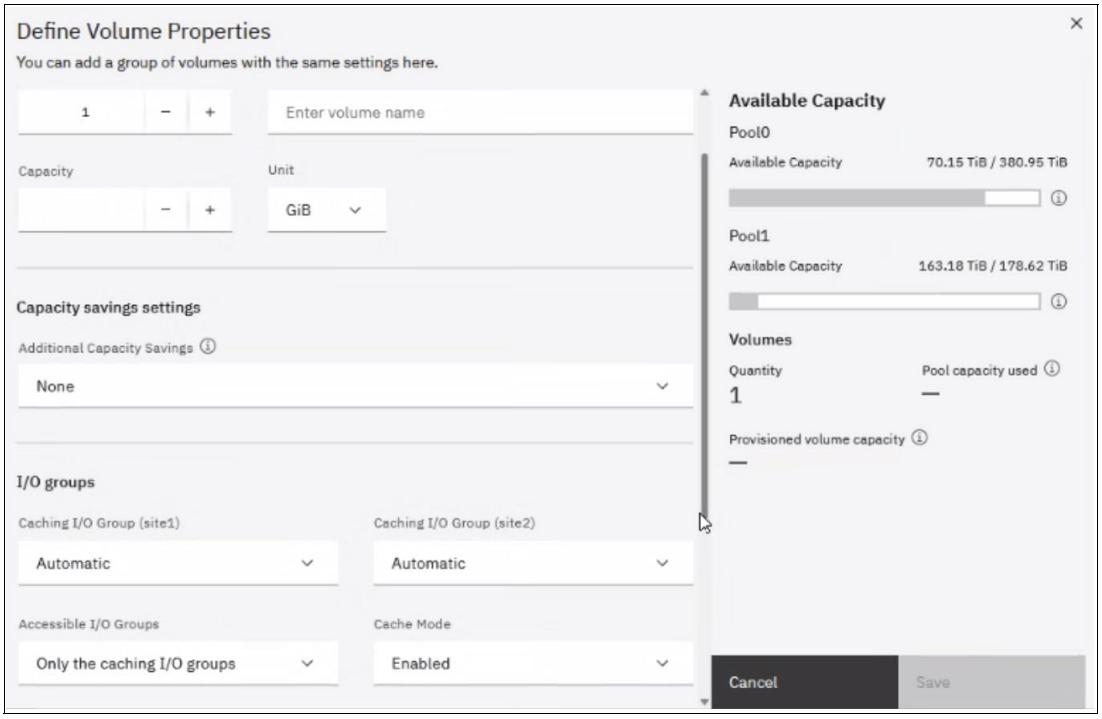


Figure 6-34 Define HyperSwap volume properties, advanced settings enabled

After you click **Save**, you can select **Create and map** to create the volumes and map the new volumes to a host or select **Create volumes** to create the volumes without mapping them.

6.6.2 I/O throttling

This section describes how to use I/O throttling on a volume.

Defining a volume throttle

To set a volume throttle, complete the following steps:

1. Select **Volumes** → **Volumes** and then the volume requires a throttle. Next, select **Actions** → **Edit Throttle**, as shown in Figure 6-35 on page 457.

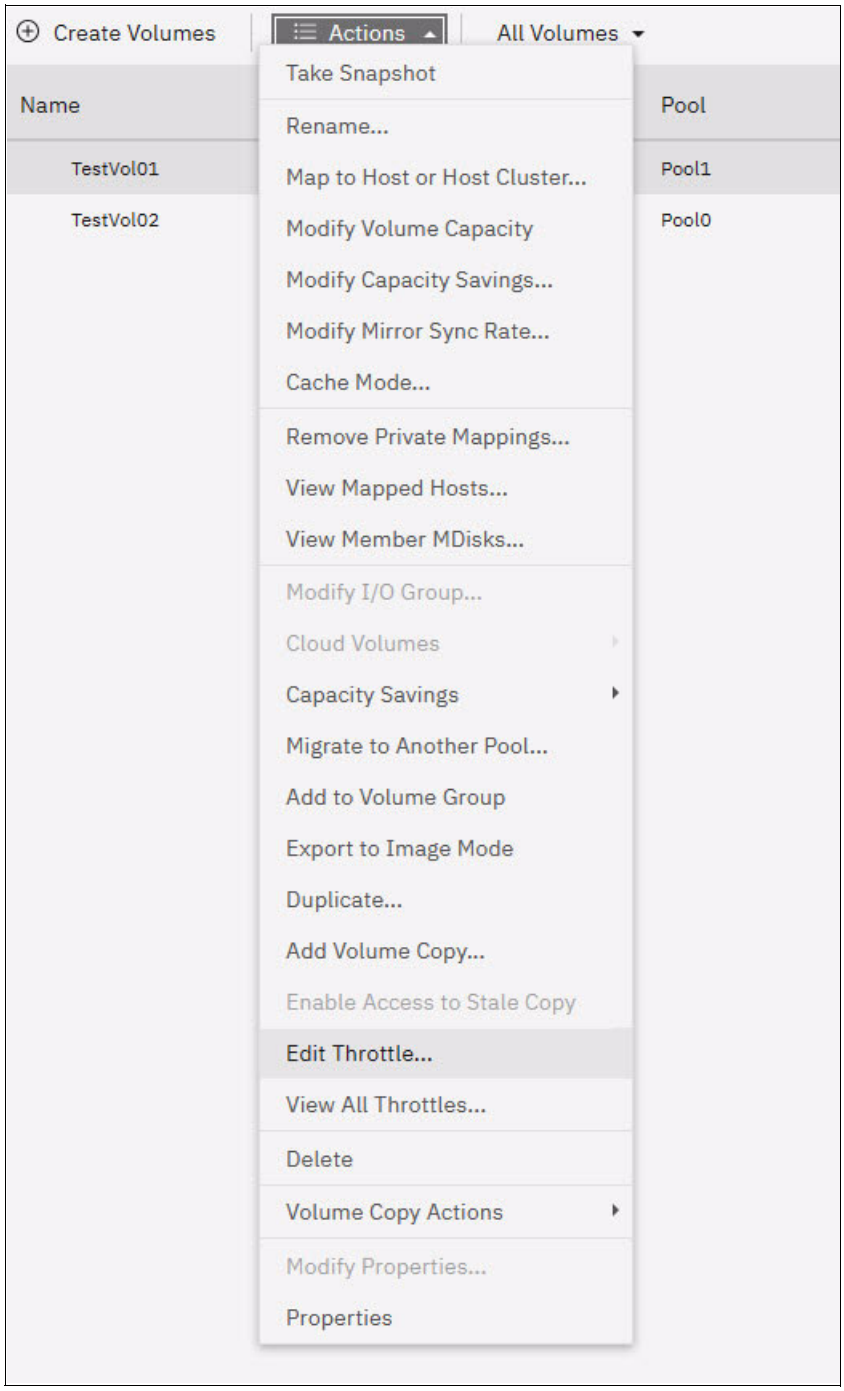


Figure 6-35 Edit Throttle menu item

2. In the Edit Throttle window (Figure 6-36), define the throttle in terms of the number of IOPS or bandwidth. After defining the throttle, click **Create**.

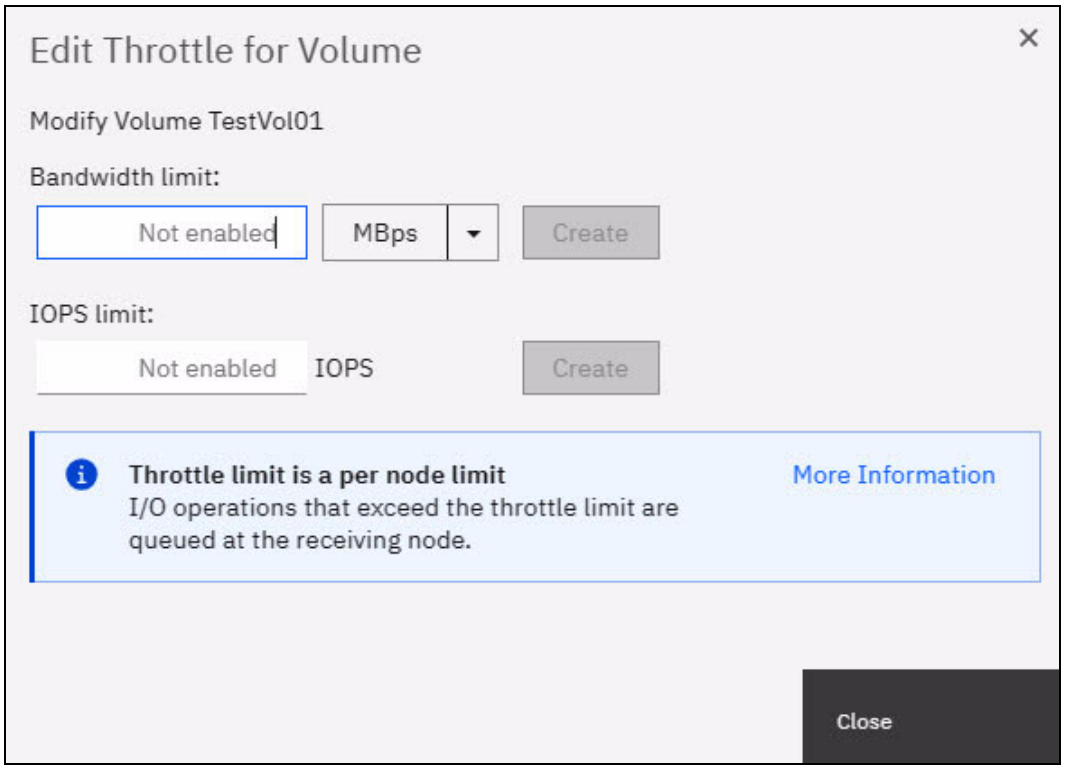


Figure 6-36 IOPS throttle on a volume

3. After the Edit Throttle task completes successfully, the Edit Throttle window opens again. You can now set the throttle based on the different metrics, modify the throttle, or close the window without performing further actions by clicking **Close**.

Listing volume throttles

To view volume throttles, select **Volumes** → **Volumes**, and then select **Actions** → **View All Throttles**, as shown in Figure 6-37.

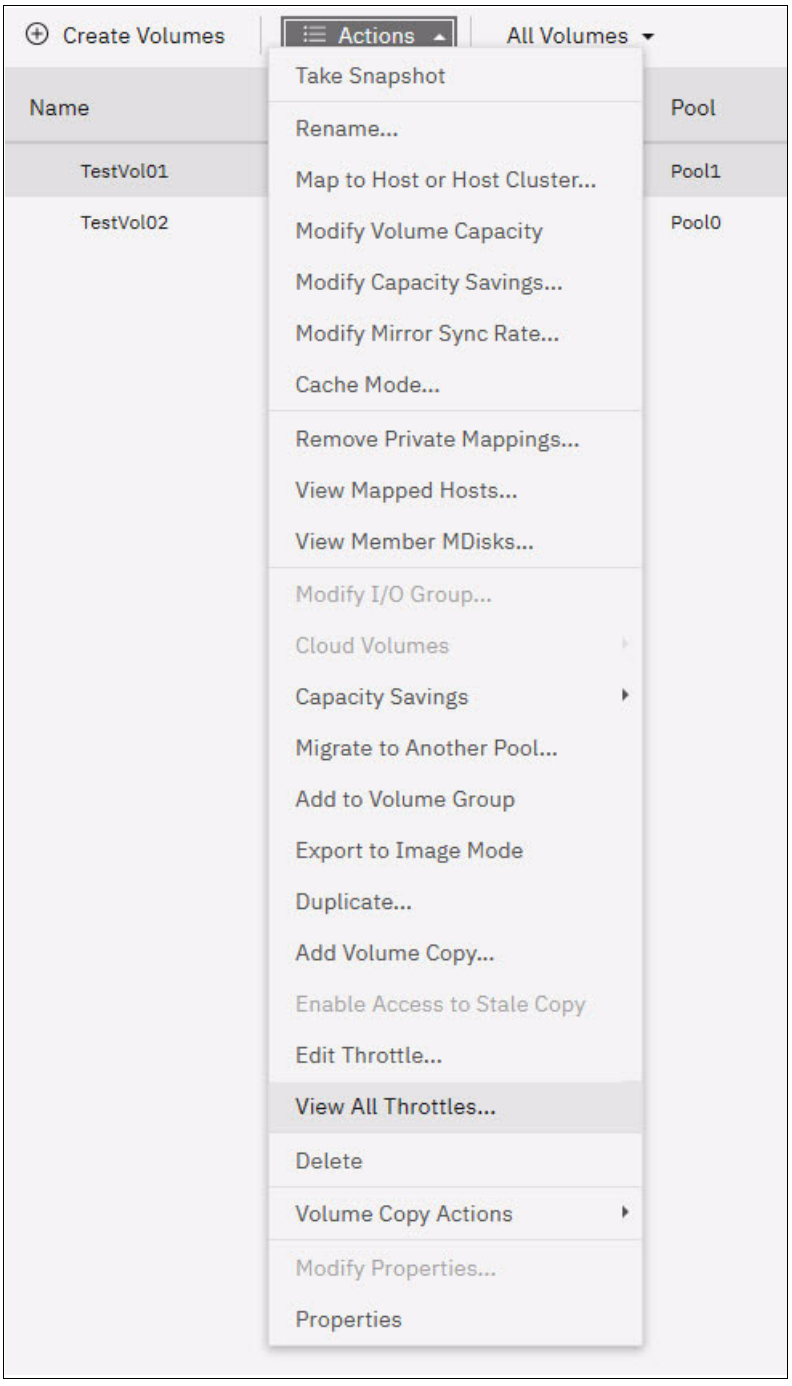


Figure 6-37 View All Throttles

The View All Throttles window shows all volume throttles that are defined in the system, as shown in Figure 6-38.

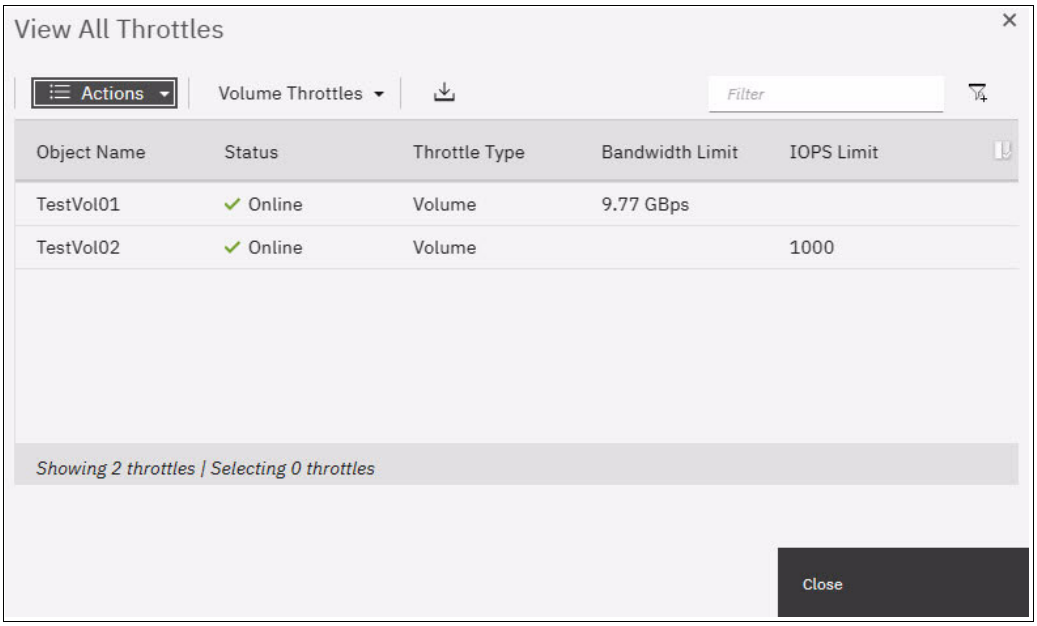


Figure 6-38 View All Throttles window

From this window you can view other throttles by selecting **Volume Throttles**, as shown in Figure 6-39.

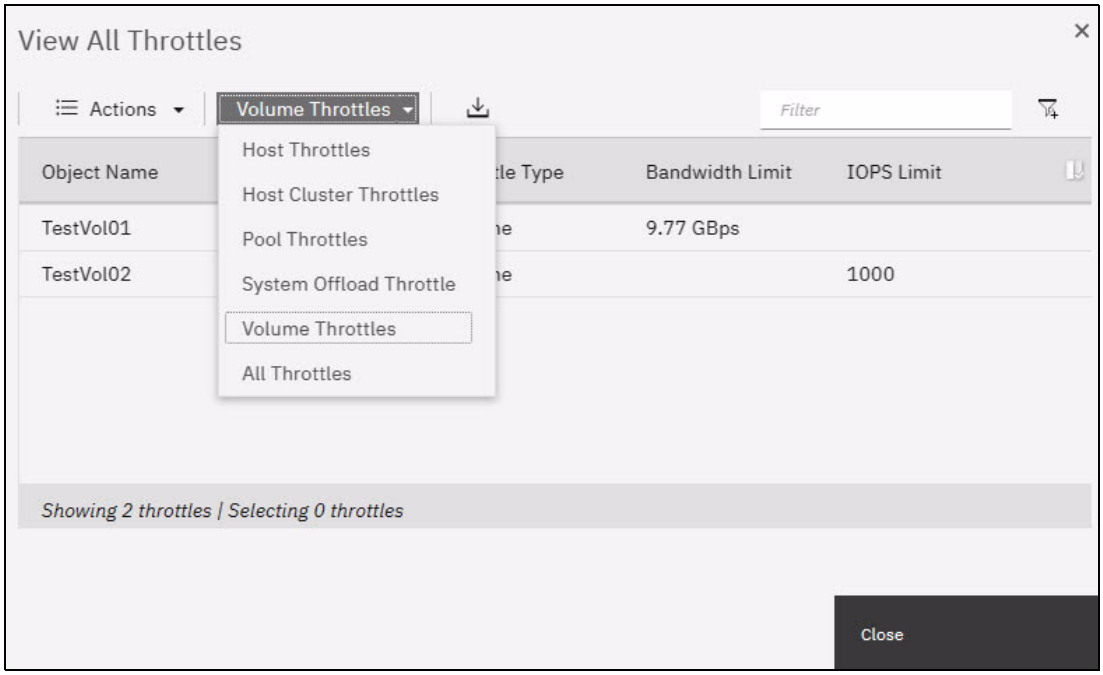


Figure 6-39 Filtering the throttle type

Modifying or removing a volume throttle

To modify or remove a volume throttle, complete the following steps:

1. From the **Volumes** menu, select the volume that is attached to the throttle that you want to change or delete. Select **Actions** → **Edit Throttle**, as shown in Figure 6-40.

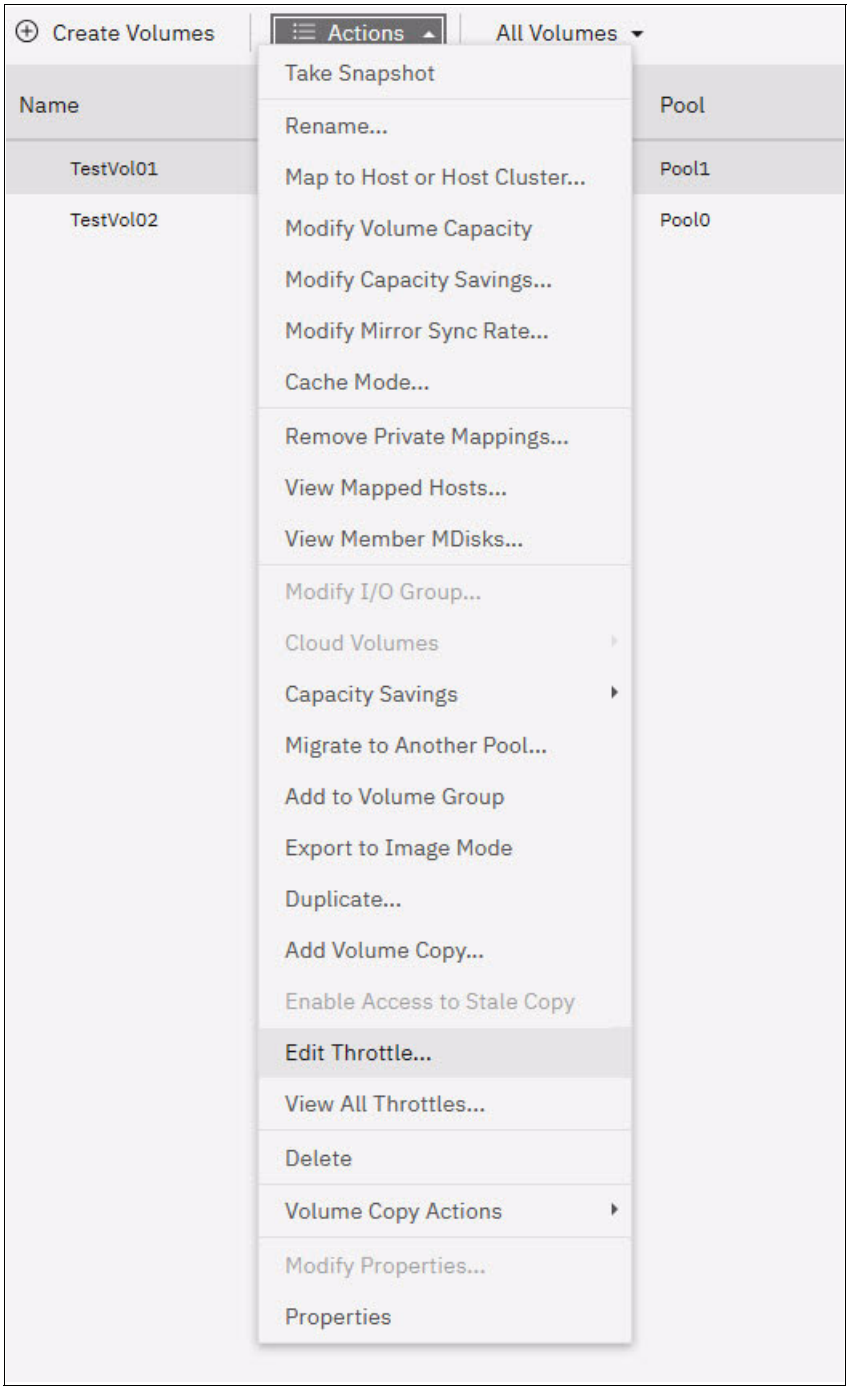


Figure 6-40 Edit Throttle

2. To modify the volume throttle, enter new throttling parameters and click **Save**, as shown in Figure 6-41. In this example, the I/O throttling limit is increased to 15,000 IOPS.

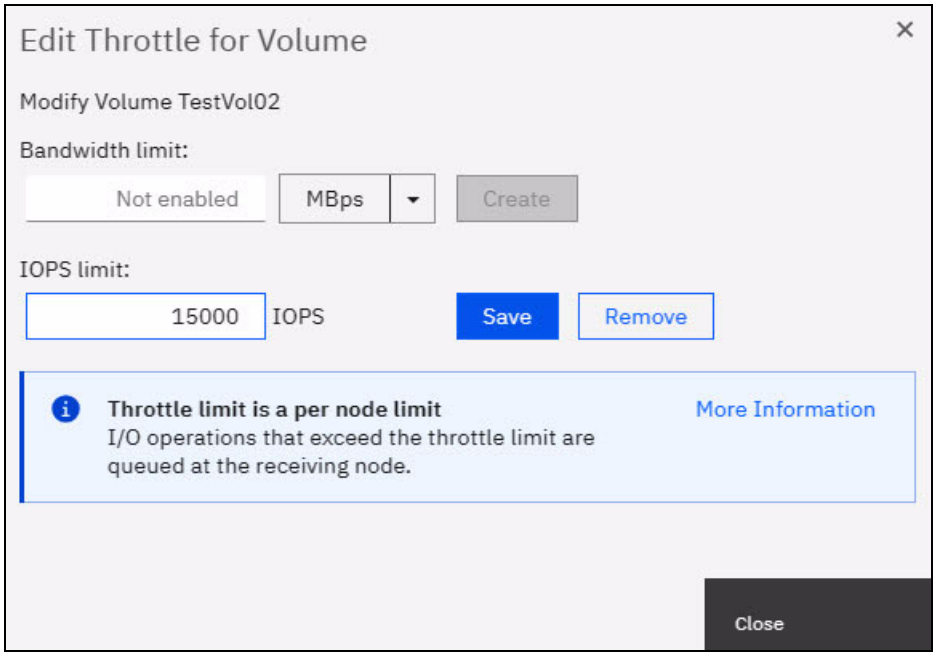


Figure 6-41 Modifying a volume throttle

3. To remove the throttle completely, click **Remove** for the throttle that you want to remove, as shown in Figure 6-42.

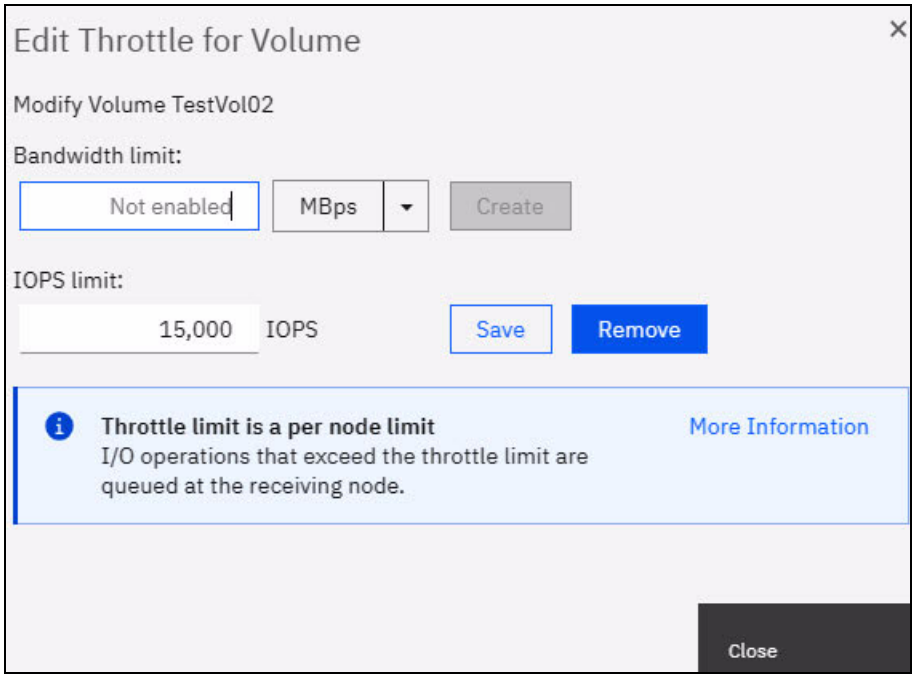


Figure 6-42 Removing a volume throttle

After the Edit Throttle task completes successfully, the Edit Throttle window opens again. You can modify the throttle, or you can click **Close**.

6.6.3 Volume protection

To configure volume protection, select **Settings** → **System** → **Volume Protection**, as shown in Figure 6-43.

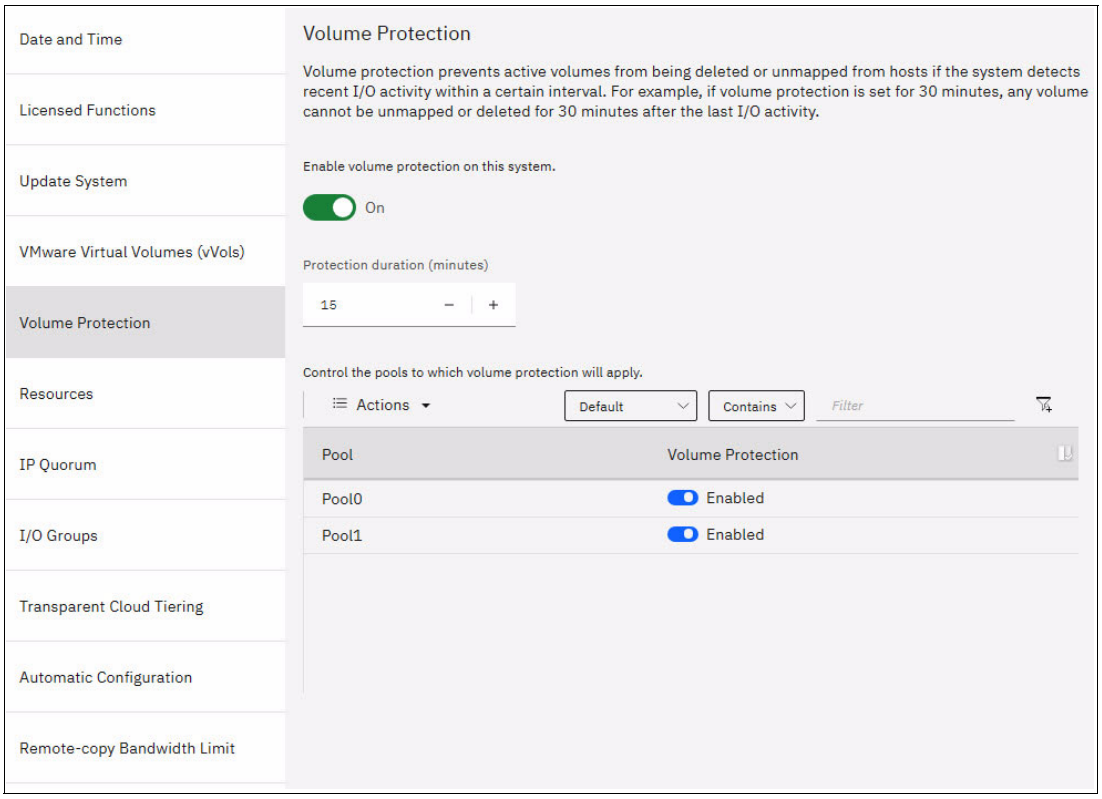


Figure 6-43 Volume Protection configuration

In this view, you can configure system-wide volume protection (enabled by default), set the minimum inactivity period that is required to allow volume deletion (protection duration), and configure volume protection for each configured pool (enabled by default). In the example, volume protection is enabled with the 15-minute minimum inactivity period and is turned on for all configured pools.

6.6.4 Modifying a volume

After a volume is created, it is possible to modify many of its characteristics. The following sections show how to perform those configuration changes.

Shrinking

To shrink a volume, complete the following steps:

1. Ensure that you have a current and verified backup of any in-use data that is stored on the volume that you intend to shrink.

2. From the **Volumes** menu, select the volume that you want to shrink. Select **Actions** → **Shrink...**, as shown in Figure 6-44.

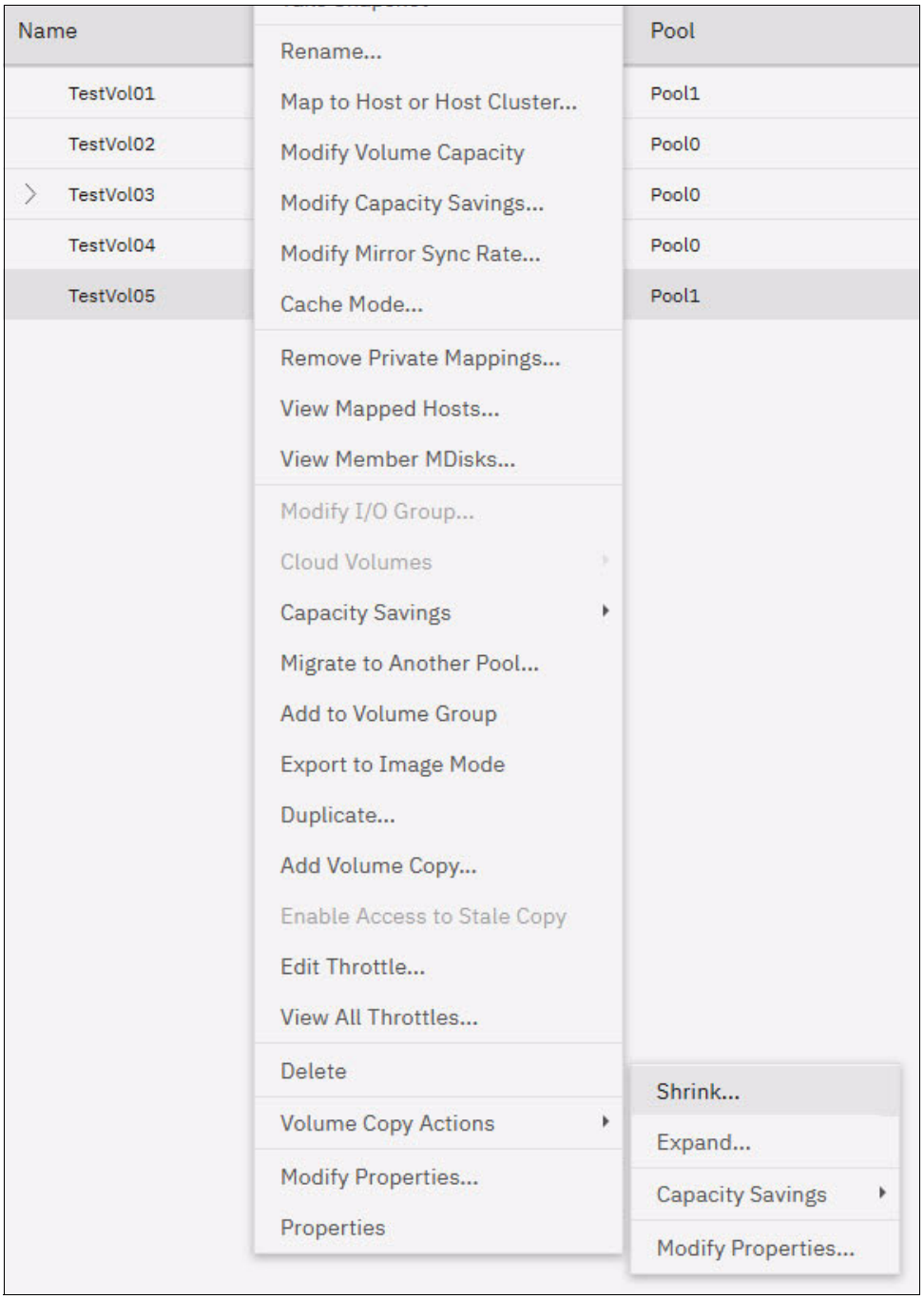


Figure 6-44 Volume Shrink menu item

3. Specify **Shrink by** or **Final size.** The other choice is calculated automatically. See Figure 6-45.

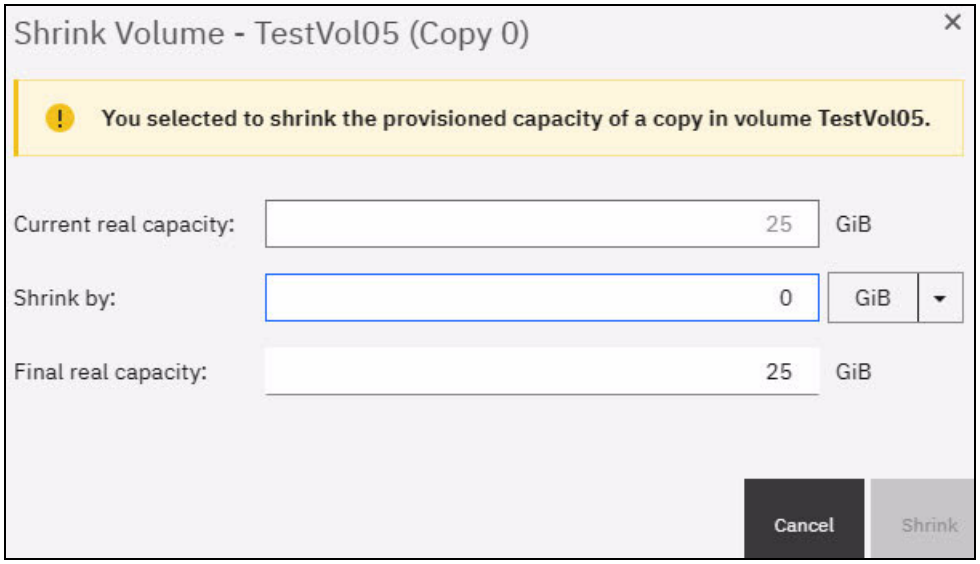


Figure 6-45 Specifying the size of the shrunk volume

Note: The storage system reduces the volume capacity by removing one or more arbitrarily selected extents. Do not shrink a volume that contains data that is being used unless you have a current and verified backup of the data.

4. Click **Yes** to confirm the action, as shown in Figure 6-46.

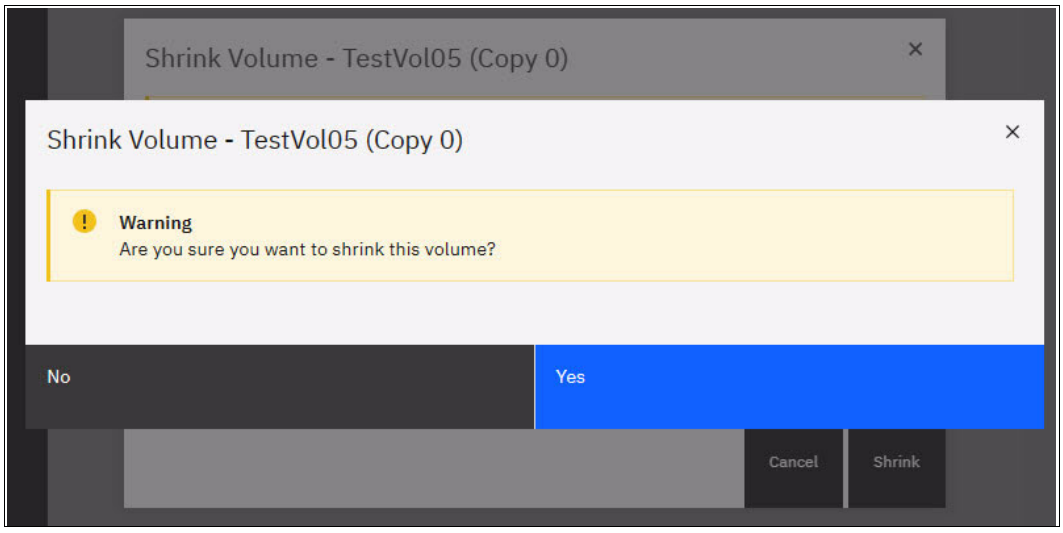


Figure 6-46 Confirming the volume shrinking operation

5. After the operation finishes, you can see the volume with the new size by selecting **Volumes** → **Volumes**, as shown in Figure 6-47 on page 466.

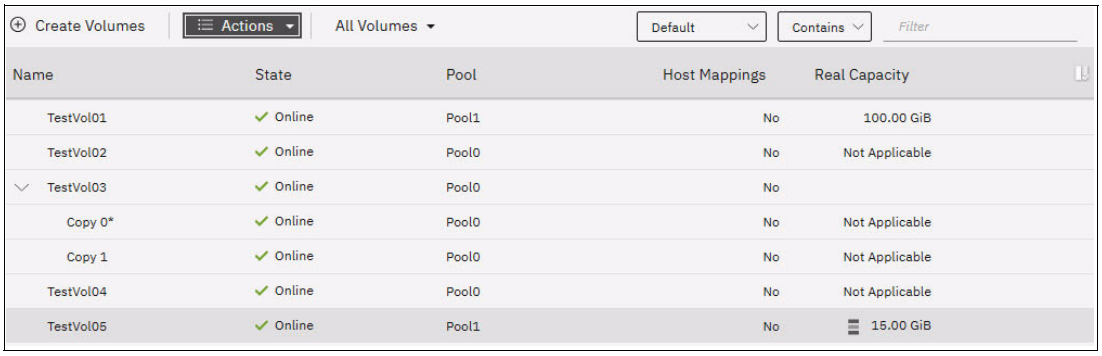


Figure 6-47 Shrunk volume size

Expanding

To expand a volume, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to expand. Select **Actions** → **Volume Copy Actions** → **Expand**, as shown in Figure 6-48 on page 467.
2. Enter a value in the **Expand by** or **Final size** field. The other value is calculated automatically. Click **Expand**, as shown in Figure 6-49 on page 468.
3. After the operation finishes and formats the extra space, you can see the volume with the new size by selecting **Volumes** → **Volumes**, as shown in Figure 6-50 on page 468.

Notes: Expanding a volume is not sufficient to increase the available space that is visible to the host. The host must become aware of the changed volume size at the operating system level through, for example, a bus rescan. More operations at the logical volume manager (LVM) or file system levels might be needed before more space is visible to applications running on the host.

Storage Virtualize Version 8.4 introduced the ability to expand a volume while it is formatting.

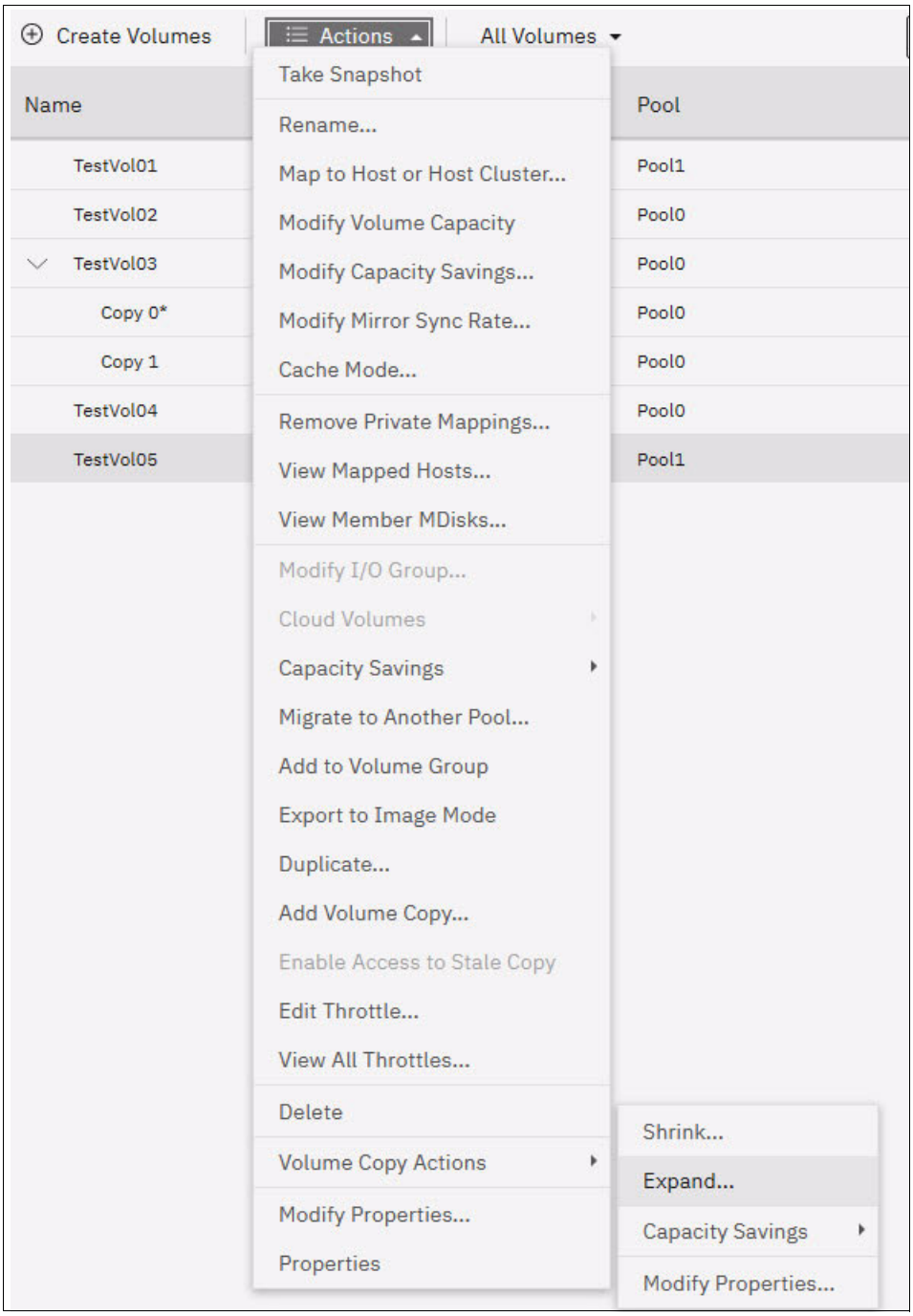


Figure 6-48 Volume Expand menu item

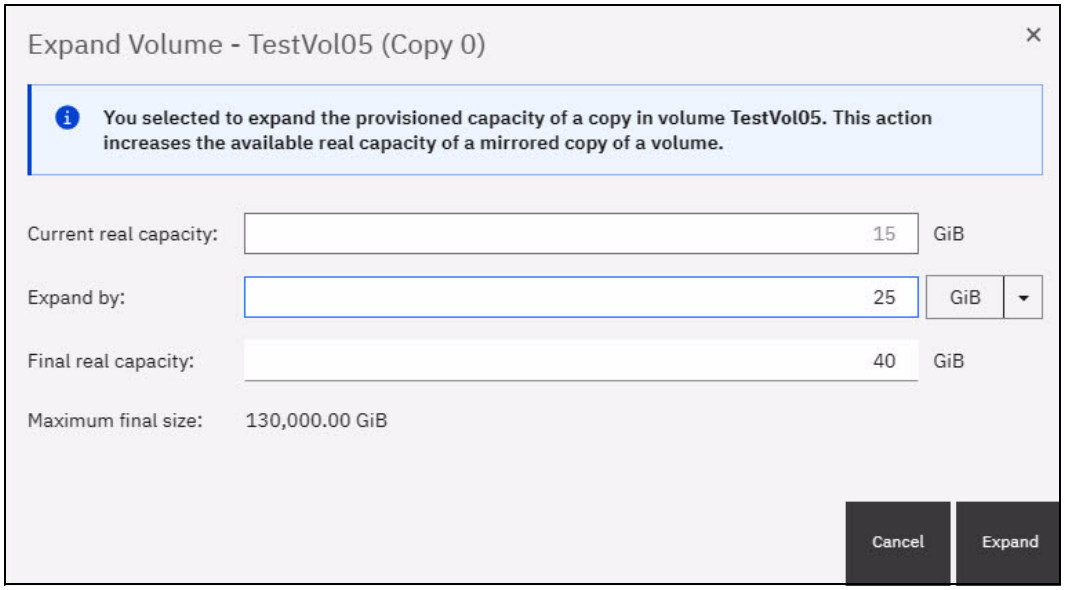


Figure 6-49 Specifying the expanded volume size

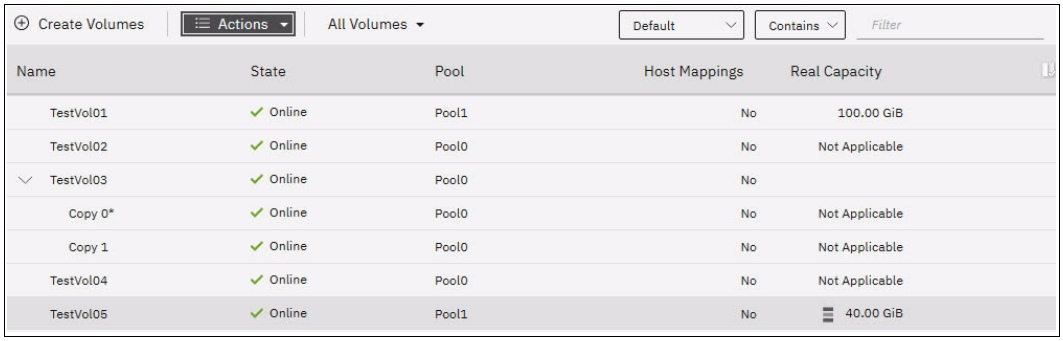


Figure 6-50 Expanded volume size

Modifying capacity savings

To modify capacity savings options for a volume, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Modify Capacity Savings**, as shown in Figure 6-51.

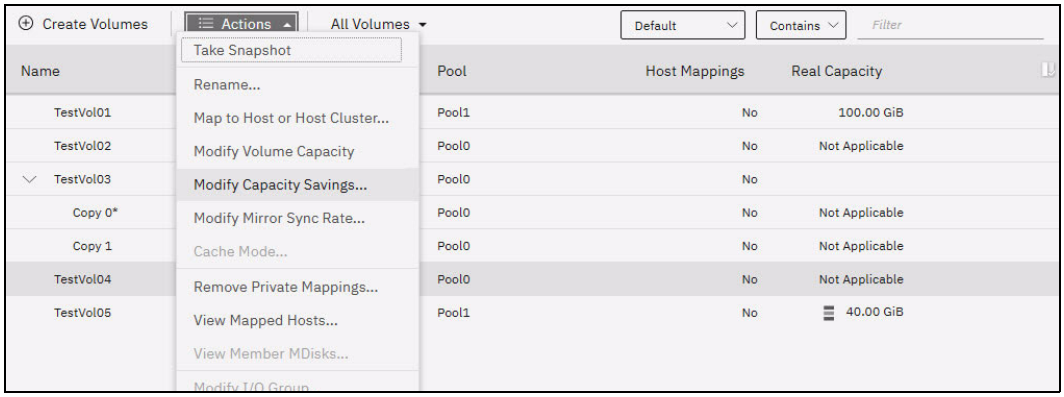


Figure 6-51 Modify Capacity Savings menu item

2. Select the capacity savings method that you want, as shown in Figure 6-52.

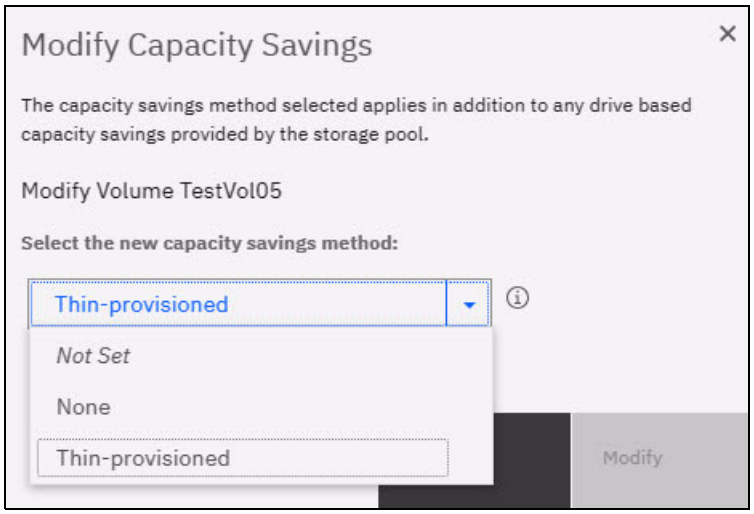


Figure 6-52 Capacity savings options for a volume

3. For volumes that are configured in a DRP, the methods are shown in Figure 6-53.

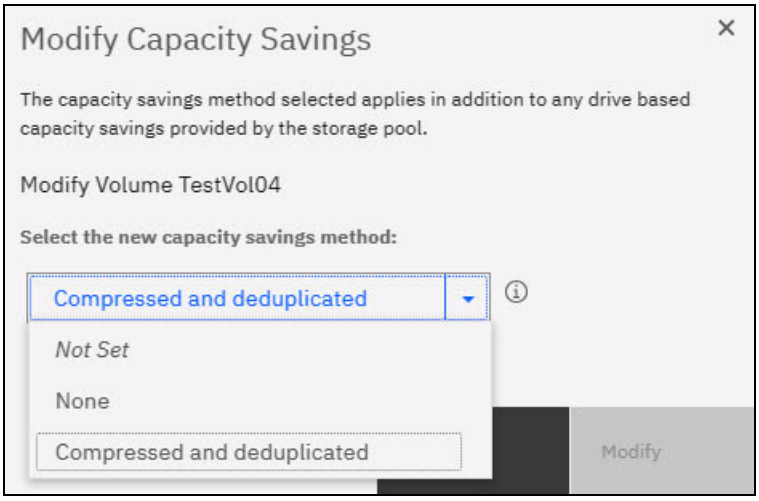


Figure 6-53 Enabling deduplication on a volume

After you configure the capacity savings options of a volume, click **Modify** to apply them.

Modifying the mirror sync rate

This action is available only for mirrored volumes. To modify the mirror sync rate of a volume, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Modify Mirror Sync Rate**, as shown in Figure 6-54.

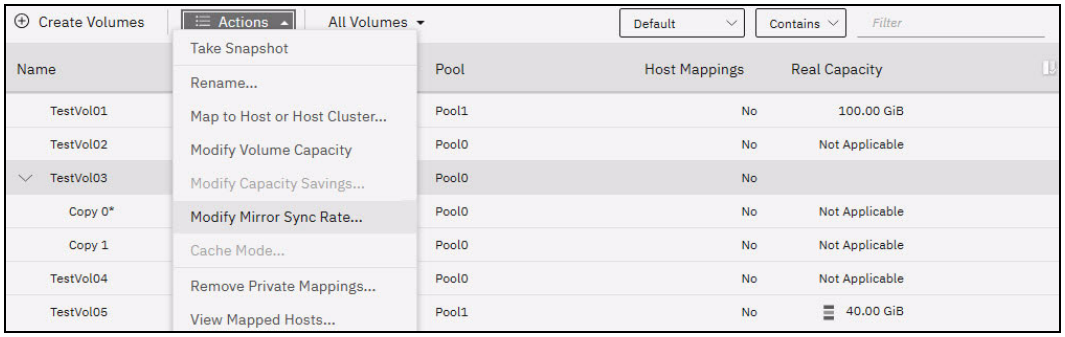


Figure 6-54 Modify Mirror Sync Rate menu item

2. Select the mirror sync rate from the list. Available values are 0 KBps - 64 MBps. Click **Modify** to set the mirror sync rate for the volume, as shown in Figure 6-55.

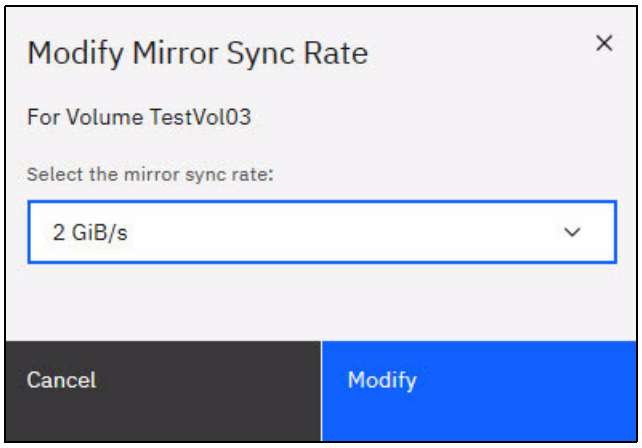


Figure 6-55 Setting the volume mirror sync rate

Changing the volume cache mode

To change the volume cache mode, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Cache Mode** as shown in Figure 6-56 on page 471.

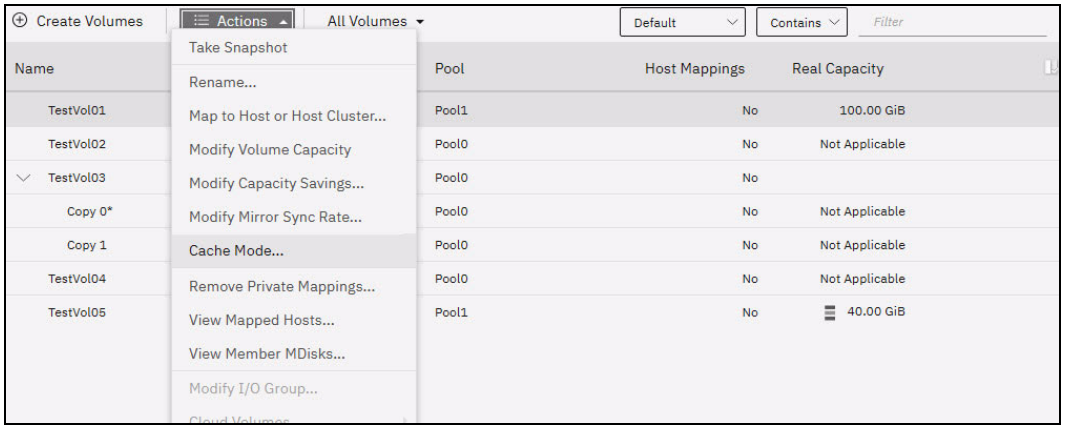


Figure 6-56 Modifying the volume cache mode

2. Select the cache mode that you want for the volume from the drop-down list and click **OK**, as shown in Figure 6-57.

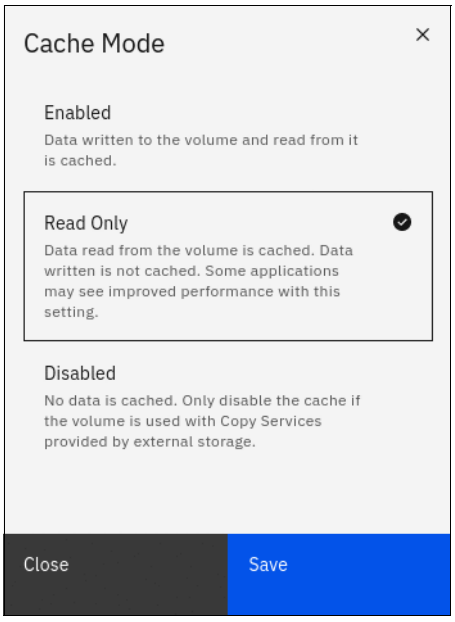


Figure 6-57 Setting the volume cache mode

6.6.5 Deleting a volume

When you try to delete a volume, the system verifies whether it is a part of a host mapping, FlashCopy mapping, or Remote Copy relationship. If any of these mappings exists, the delete attempt fails unless the **`-force`** parameter is specified on the corresponding remove commands. If volume protection is enabled and if the system detects any I/O activity to the volume within the configured time frame, a delete fails even if the **`-force`** parameter is specified. The **`-force`** parameter overrides the volume dependencies, not the volume protection setting.

To delete a volume, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Delete**, as shown in Figure 6-58.

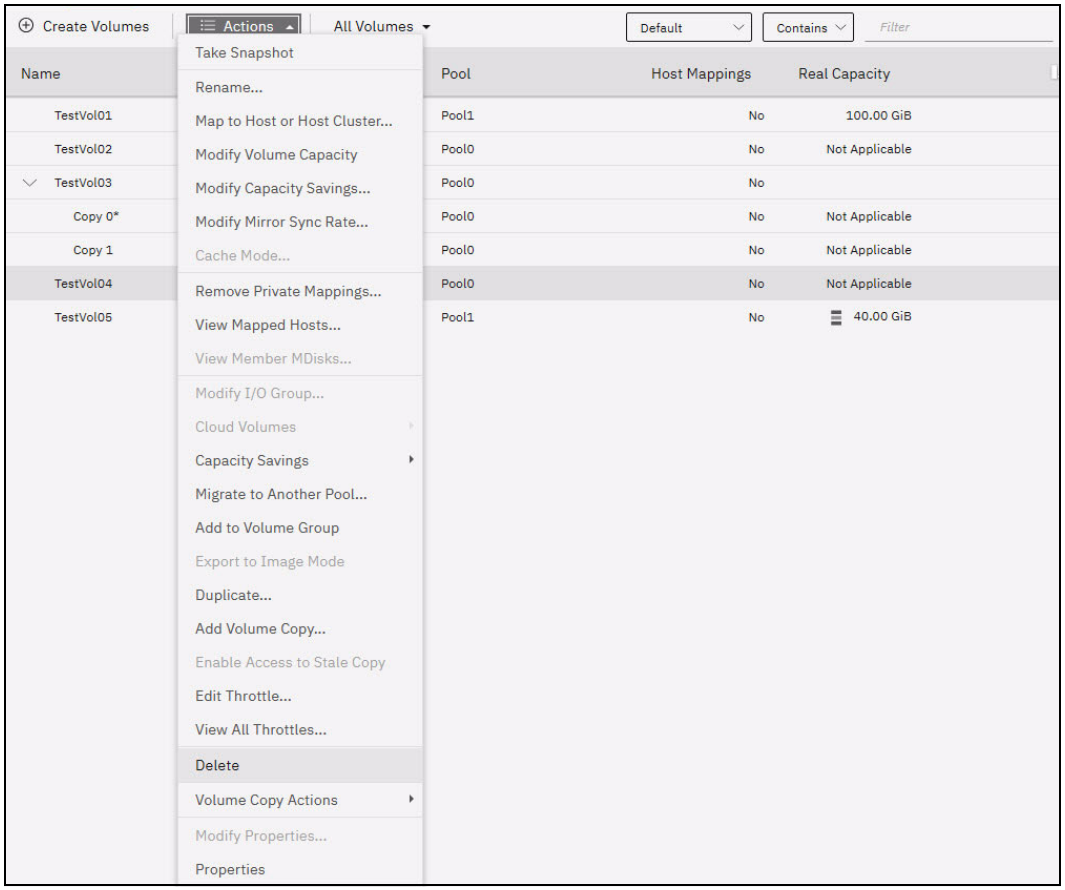


Figure 6-58 Volume Delete menu item

2. Review the list of volumes that are selected for deletion and provide the number of volumes that you intend to delete, as shown in Figure 6-59 on page 473. Click **Delete** to remove the volume from the system configuration.

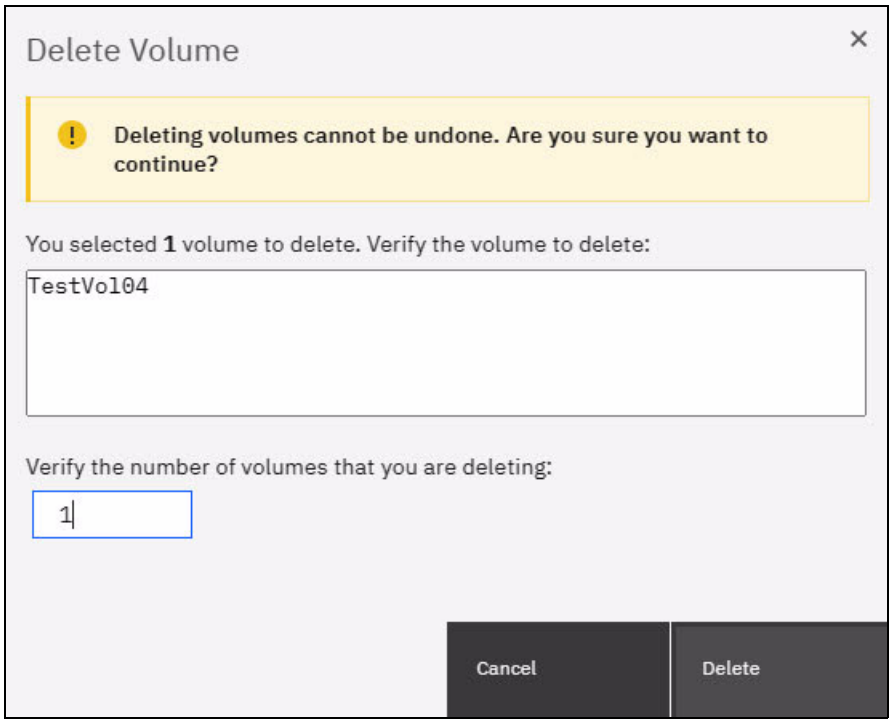


Figure 6-59 Confirming the volume deletion

6.6.6 Mapping a volume to a host

To make a volume available to a host or cluster of hosts, it must be mapped to the host or host cluster. A volume can be mapped to the host at creation time or later.

To map a volume to a host or cluster, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Map to Host or Host Cluster**, as shown in Figure 6-60.

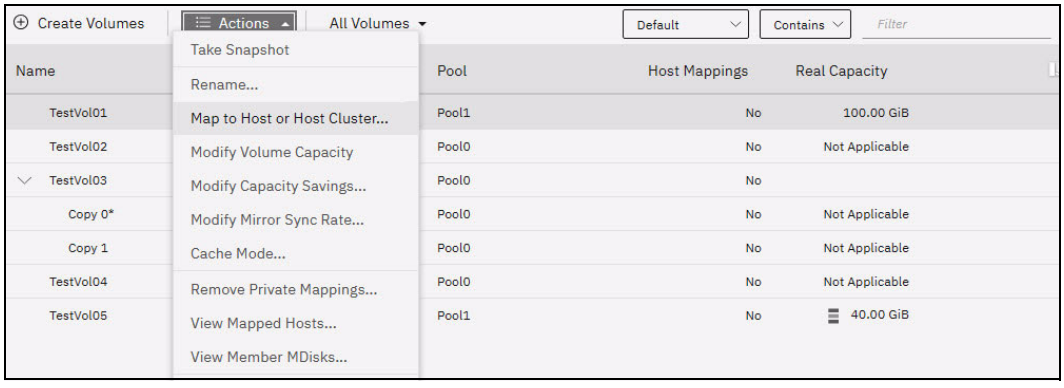


Figure 6-60 Volume mapping menu item

Tip: An alternative way of opening the **Actions** menu is to select a volume and right-click.

2. When the Create Mapping window opens, select whether to create a mapping to a host or to a host cluster. The list of objects of the suitable type is displayed. Select to which hosts or host clusters the volume should be mapped.

   You can allow the storage system to assign the SCIS LUN ID to the volume by selecting the **System Assign** option, or select **Self Assign** and provide the LUN ID yourself.

   Click **Next** to proceed to the next step.

   In Figure 6-61, a single volume is mapped to a host and the storage system assigns the SCSI LUN IDs.

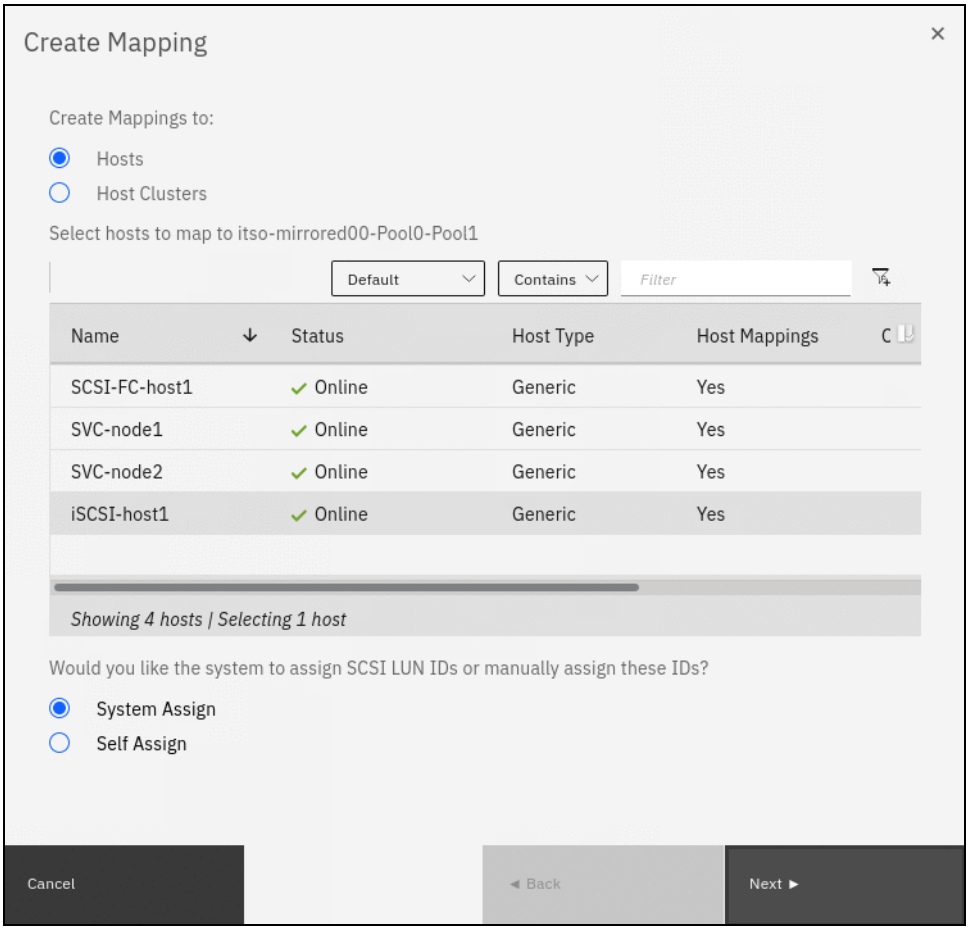


Figure 6-61 Mapping a volume to a host

3. A summary window opens and shows all the volume mappings for the selected host. The new mapping is highlighted, as shown in Figure 6-62. Review the future configuration state and click **Map Volumes** to map the volume.

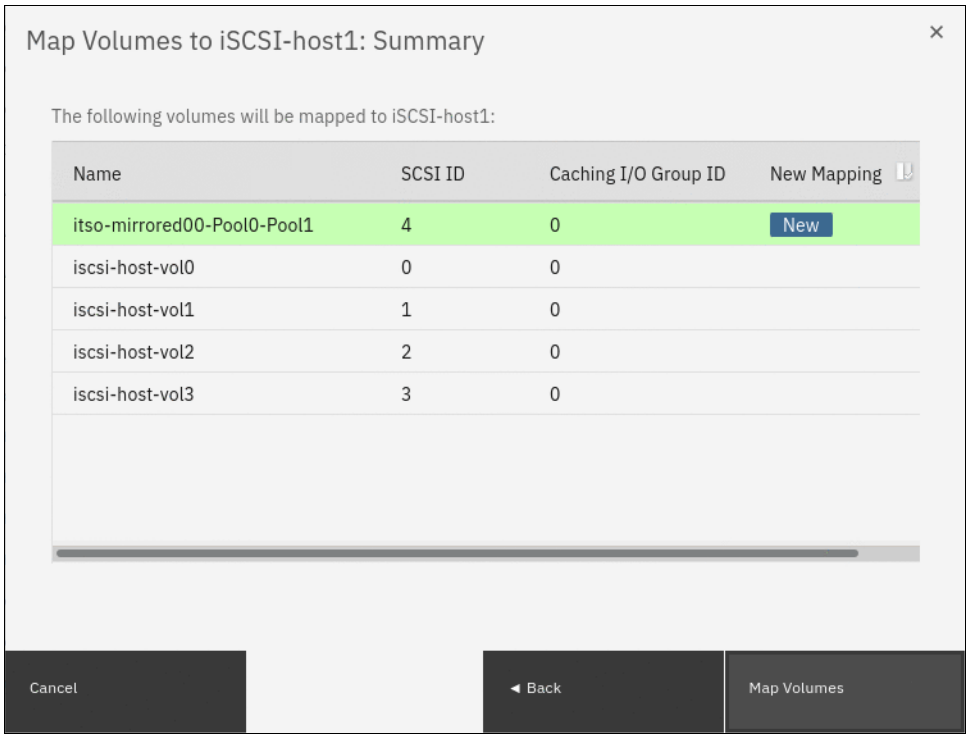


Figure 6-62 Mapping a volume to a host: Summary

4. After the hosts are mapped, you can list the volumes that are mapped to the host by selecting **Hosts** → **Mappings**, as shown in Figure 6-63.

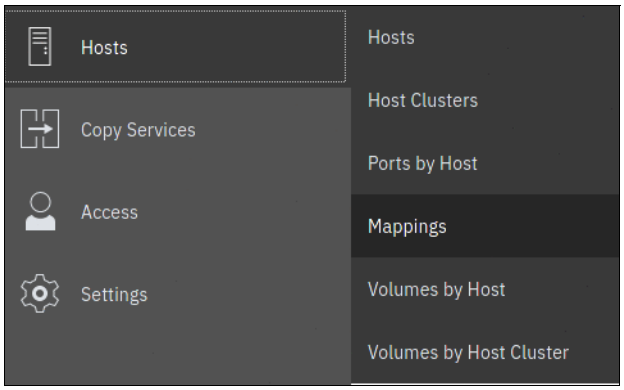


Figure 6-63 Accessing the Hosts Mapping view

A window opens with a list of volumes that are mapped to all hosts, as shown in Figure 6-64.

Host Name	SCSI ID	Volume Name	UID	I/O Group ID	I/O Group Name
iSCSI-host1	1	iscsi-host-vol1	600507640084031DD800000000000061	0	io_grp0
iSCSI-host1	4	itso-mirrored00-Pool0-Pool1	600507640084031DD80000000000006A	0	io_grp0
iSCSI-host1	0	iscsi-host-vol0	600507640084031DD800000000000060	0	io_grp0
SCSI-FC-host1	2	iscsi-cluster-vol	600507640084031DD800000000000065	0	io_grp0

Figure 6-64 List of volume to host mappings

5. To see volumes that are mapped to clusters instead of hosts, open the pull-down menu to change the value from **Private Mappings** to **Shared Mappings**.

> **Note:** You can use the filter to display only the hosts or volumes that you want to see.

The host can now access the mapped volume. For more information about discovering the volumes on the host, see Chapter 5, “Using storage pools” on page 365.

6. To remove the volume to host mapping, select **Hosts** → **Mappings.**
7. Select the volume or volumes, right-click, and click **Unmap Volumes**, as shown in Figure 6-65.

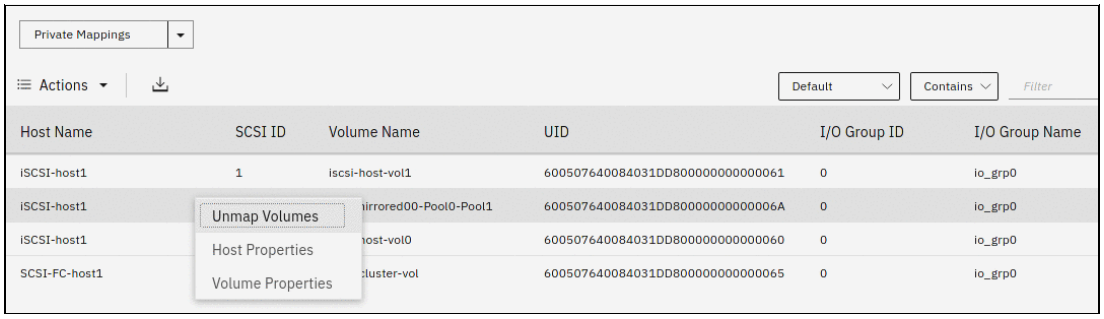


Figure 6-65 Unmap Volumes menu item

8. In the Delete Mapping window, enter the number of volumes that you intend to unmap, as shown in Figure 6-66 on page 477. This action is as a security measure that minimizes changes that result from an accidental unmap of an invalid volume.

> **Note:** Removing volume to host mapping makes the volume unavailable to the host. Make sure that the host is prepared for the operation. An improperly run volume unmap operation might cause data unavailability or loss.

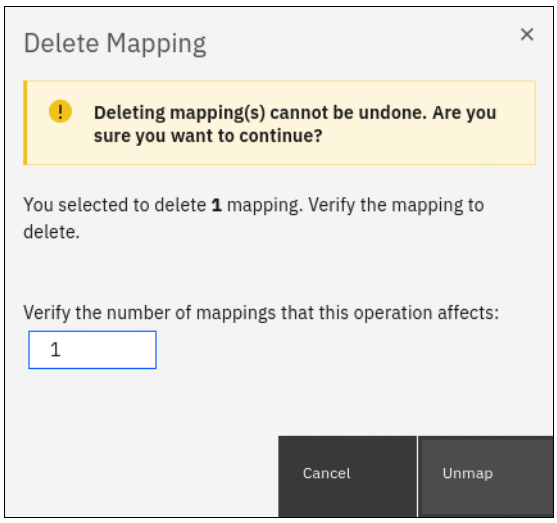


Figure 6-66 Volume unmap confirmation window

Click **Unmap** to complete the operation. Volume mapping is removed and is no longer displayed in the volume map view, as shown in Figure 6-67.

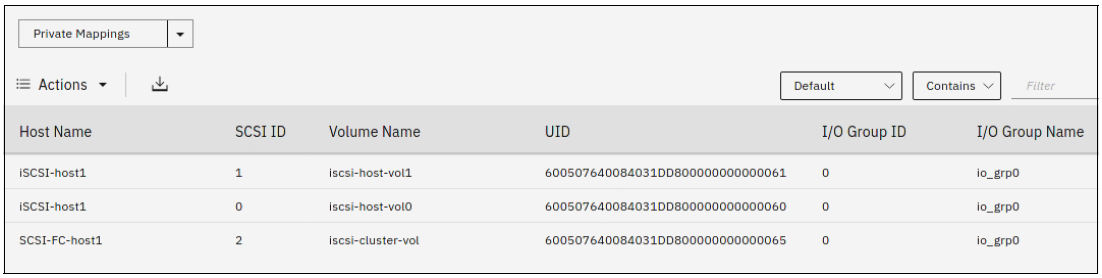


Figure 6-67 Volume mapping removed

6.6.7 Modify I/O group or Nondisruptive Volume Move

Moving volumes between I/O groups is a task that is sometimes needed for workload balancing or migration between clustered enclosures. Because the caching I/O group is the enclosure that mediates the I/O between the host system and the storage, moving a volume to another I/O group also shifts the resource consumption of CPU, memory, and front-end host and back-end storage traffic.

Generally, in enclosure-based systems you want to align the I/O group with the enclosure that contains the pool in which the volume is stored.

Note: Specific conditions prevent the changing of the I/O group dynamically with NDVM for a volume. If the volume uses data reduction in a DRP, or if a volume is a member of a FlashCopy map and is in an RC relationship, the first command in the sequence, **`addvdiskaccess`**, fails.

Modify the I/O group of a volume without host mappings

To modify the I/O group, complete the following steps:

1. To start the process, select the volumes, right-click and then select **Modify I/O Group**. See Figure 6-68 on page 478.

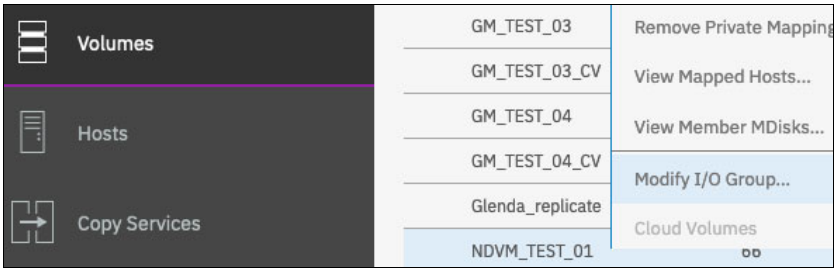


Figure 6-68 Modify I/O Group menu item

If no host mappings exist for the volumes, the operation immediately displays the target I/O group selection dialog box, as shown in Figure 6-69.

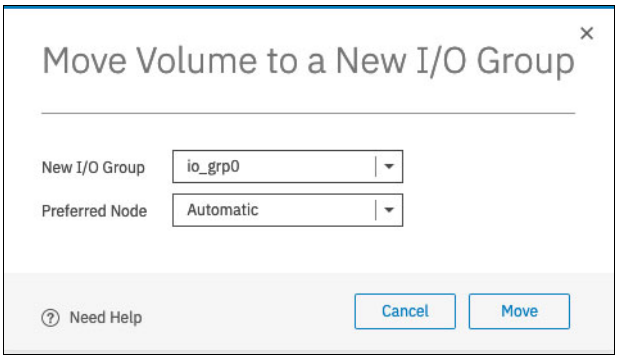


Figure 6-69 I/O Group selection dialog box

2. Select the new I/O group and preferred node and click **Move** to move the volume to the new I/O group and preferred node. This GUI action runs the following commands:
   - **`addvdiskaccess -iogrp`** *<new i/o group> <volume>*

     Adds the specified I/O group to the set of I/O groups in which the volume can be made accessible to hosts.
   - **`movevdisk -iogrp`** *<new i/o group> <volume>*

     Moves the preferred node of the volume to the new (target) caching I/O group.
   - **`rmvdiskaccess -iogrp`** *<new i/o group> <volume>*

     Removes the old (source) I/O group from the set of I/O groups in which the volume can be made accessible to hosts.

Note: Zoning must be configured between the host and the new I/O group. Also, ensure that all hosts to which the volume is mapped discovers the new paths to the volume.

Modifying the I/O group of a volume with host mappings

If the volume is mapped to a host, the GUI detects the host mapping and starts a wizard, as shown in Figure 6-70 on page 479.

Complete the following steps to modify the I/O group of a mapped volume:

1. Verify that all hosts that use the volume are zoned to the target I/O group. From the Move Volume to a New I/O Group page, click **Next**.

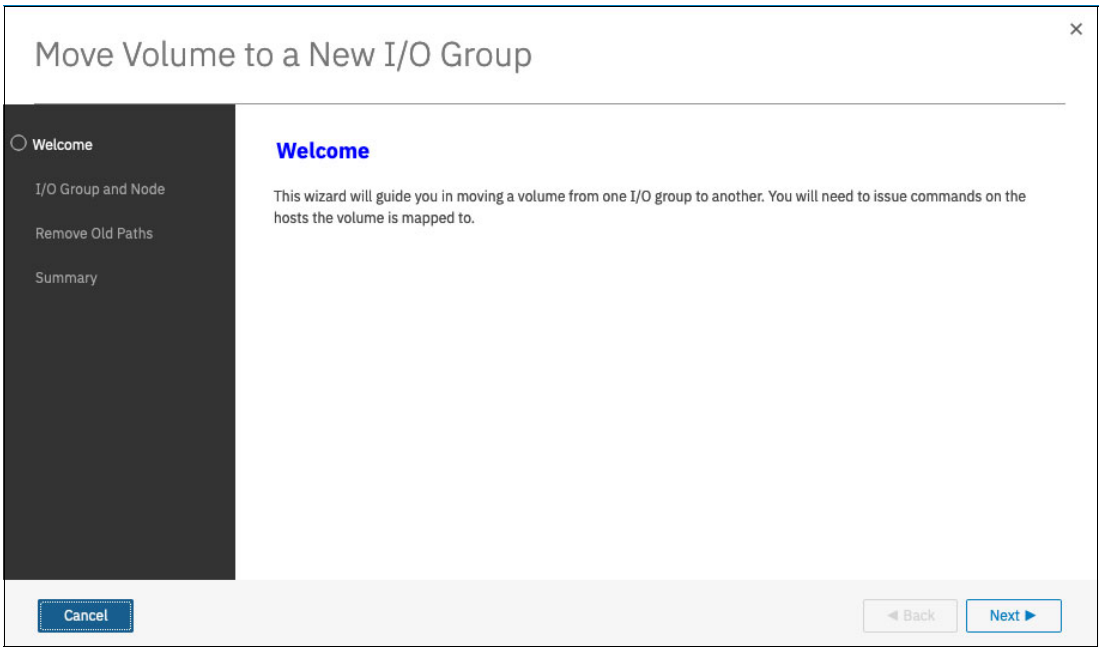


Figure 6-70 Modify the I/O Group for a mapped volume wizard: Welcome

2. Select the new, target I/O group and preferred node, and click **Apply** and **Next** (Figure 6-71). The GUI runs the following commands:
   - **`addvdiskaccess -iogrp`** *<new i/o group> <volume>*

     Adds the specified I/O group to the set of I/O groups in which the volume can be made accessible to hosts.
   - **`movevdisk -iogrp`** *<new i/o group> <volume>*

     Moves the preferred node of the volume to the new (target) caching I/O group.

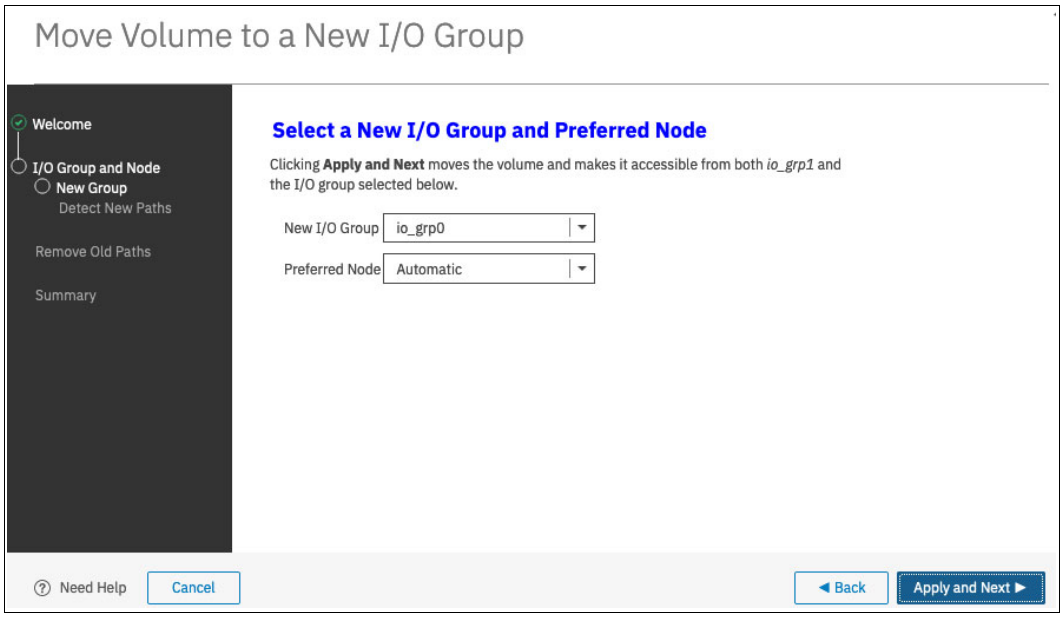


Figure 6-71 Modify the I/O Group for a mapped volume wizard: I/O group selection window

A window opens and confirms the successful completion of the commands, as shown in Figure 6-72.

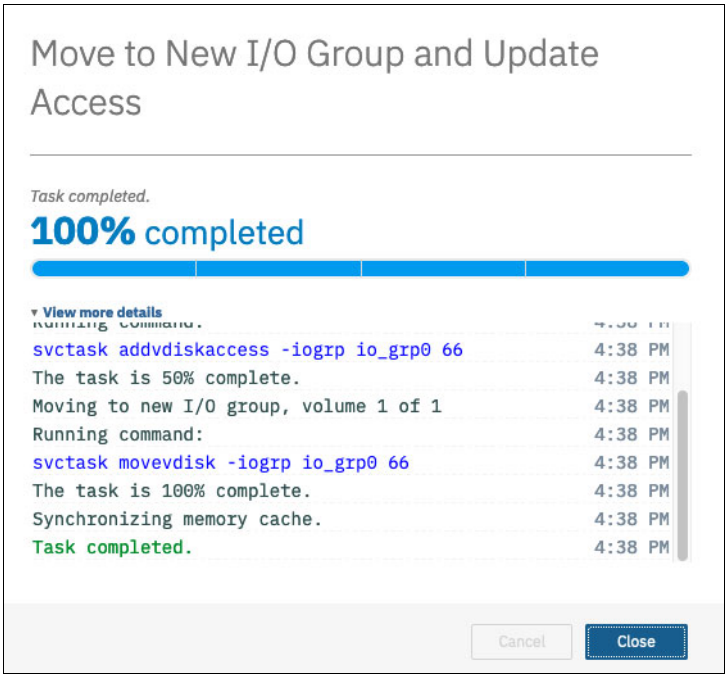


Figure 6-72 Modify the I/O Group for a mapped volume wizard: First stage completed (details)

3. Click **Close** to proceed to the validation window, as shown in Figure 6-73. Click **Need Help** if you need to see information about how to prepare the host for the volume move. After the host is ready for the volume path change, select the box that confirms that path validation was performed on the host. Click **Apply and Next**.

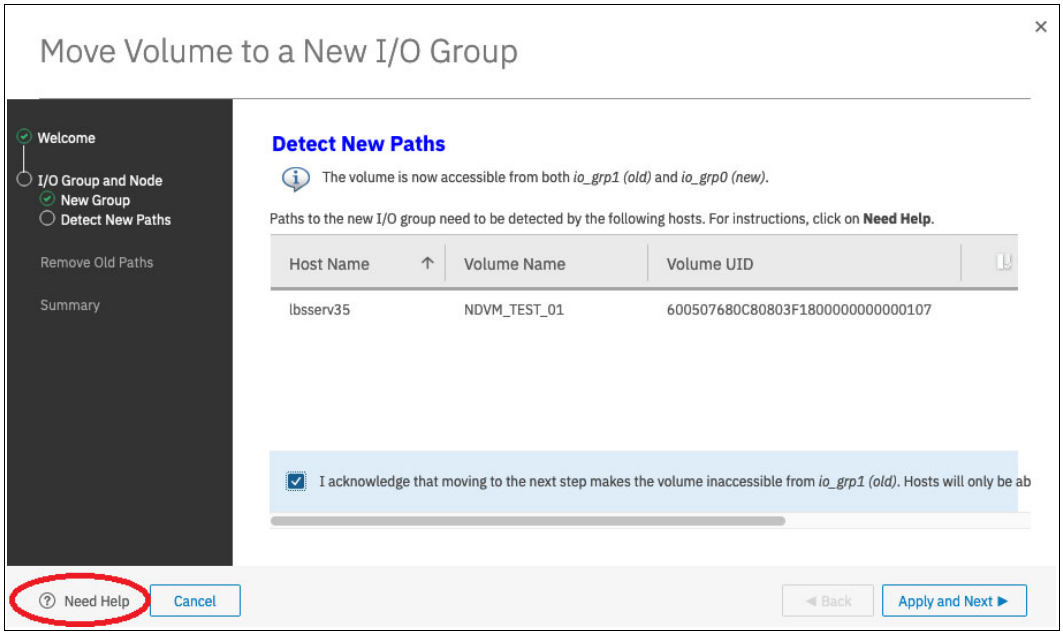


Figure 6-73 Modify I/O Group for a mapped volume wizard: Validation

Note: Failure to ensure that the host discovered the new paths to all the volumes might result in this process being disruptive and cause the host to lose access to the moved volume or volumes.

4. Verify the removal of access to the old I/O group is successful. See Figure 6-74.

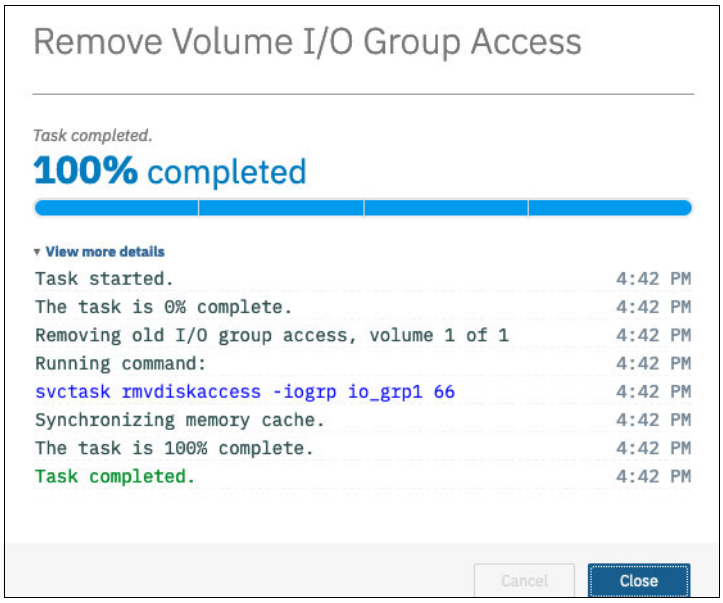


Figure 6-74 Modify I/O Group for a mapped volume wizard: Second stage completes (details)

5. Click **Close** to proceed to the final window of the wizard, as shown in Figure 6-75. Hosts cannot access the volume through the old I/O group. Depending on the operating system, you might also need a restart to remove the old paths.

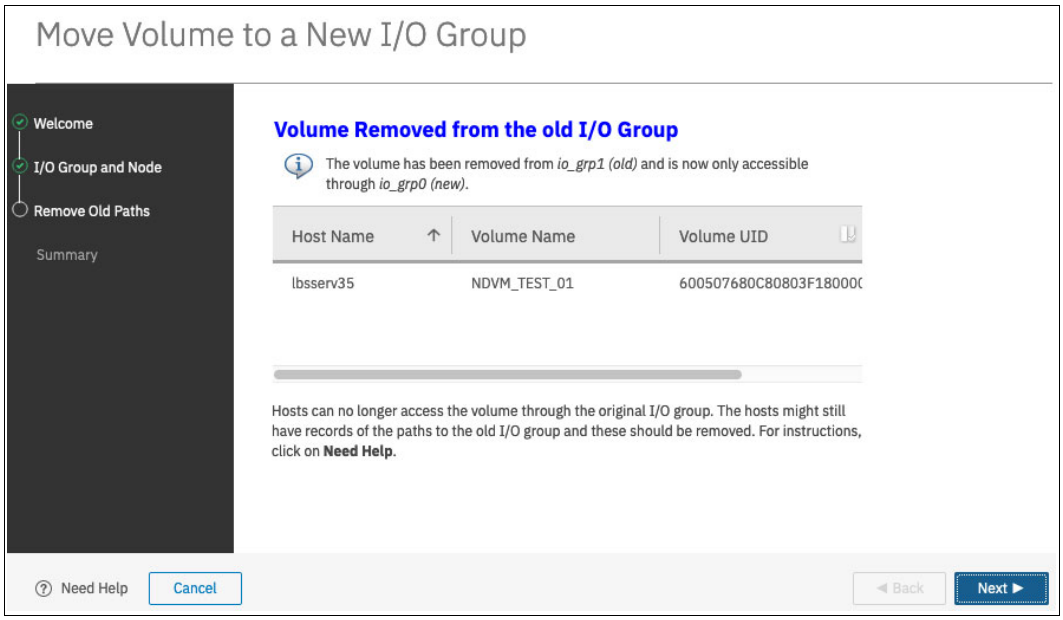


Figure 6-75 Modify I/O Group for a mapped volume wizard: Operation complete

6.6.8 Migrating a volume to another storage pool

IBM Storage Virtualize enables online volume migration with no application downtime. Volumes can be moved between storage pools without affecting business workloads that are running on these volumes.

Migrating a volume is done by the use of a volume migration feature or by creating a volume copy.

Volume migration by using the migration feature

Volume migration is a low-priority process that does not affect the performance of the IBM Storage Virtualize system. However, as subsequent volume extents are moved to the new storage pool, the performance of the volume is determined more by the characteristics of the new storage pool.

Note: You cannot move a volume copy that is compressed to an I/O group that contains a node that does not support compressed volumes.

To migrate a volume to another storage pool, complete the following steps:

1. In the **Volumes** menu, highlight the volume that you want to migrate. Select **Actions** → **Migrate to Another Pool**, as shown in Figure 6-76.

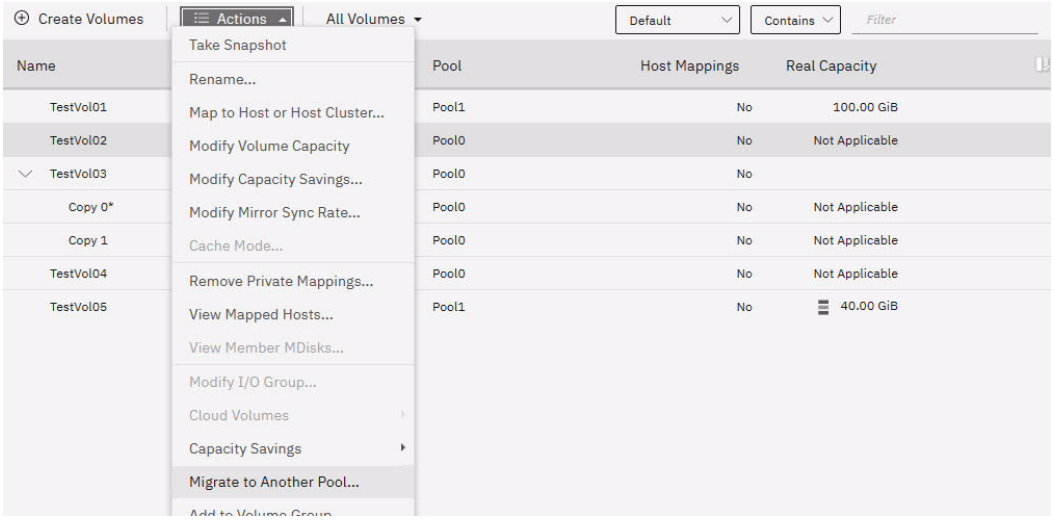


Figure 6-76 Migrate Volume Copy: Selecting a menu item

2. The Migrate Volume Copy window opens. If your volume consists of more than one copy, select the copy that you want to migrate to another storage pool, as shown in Figure 6-77. If the selected volume consists of one copy, the volume copy selection window is not displayed.

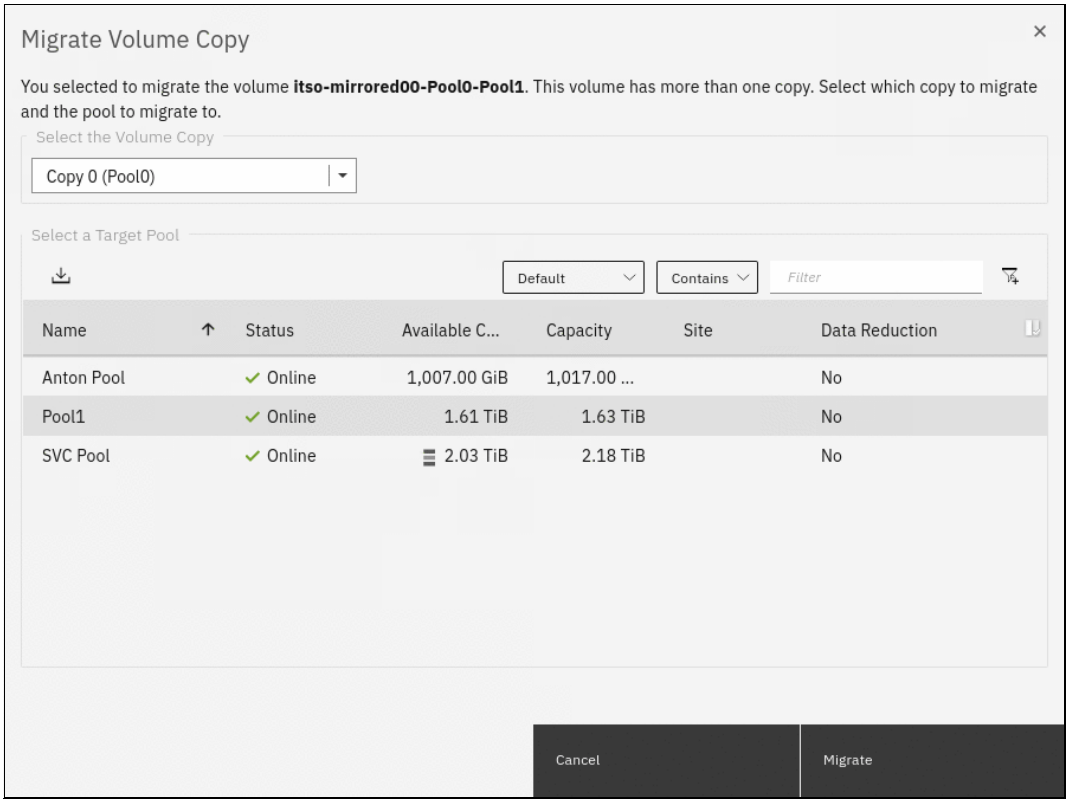


Figure 6-77 Migrate Volume Copy: Selecting the volume copy

3. Select a Target Pool from the list of all pools that are a valid migration copy target for the selected volume copy. Click **Migrate**. See Figure 6-77.
4. The time that it takes for the migration process to complete depends on the size of the volume. The status of the migration can be monitored by selecting **Monitoring** → **Background Tasks**, as shown in Figure 6-78.

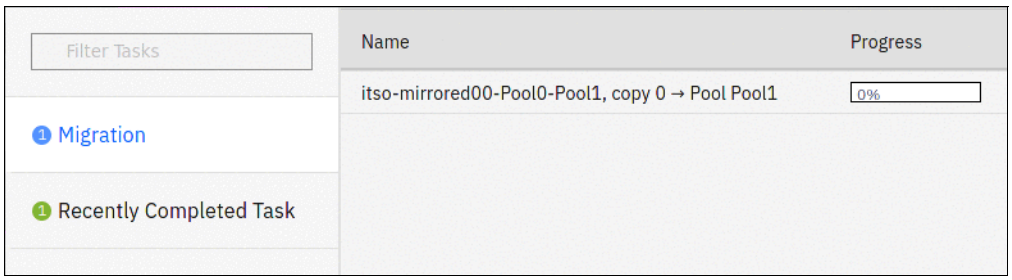


Figure 6-78 Monitoring the volume migration progress

After the migration task completes, the completed migration task is visible in the Recently Completed Task window of the **Background Tasks** menu, as shown in Figure 6-79 on page 484.



Figure 6-79 Volume migration complete

In the **Volumes** → **Volumes** menu, the volume copy is now displayed in the target storage pool, as shown in Figure 6-80.

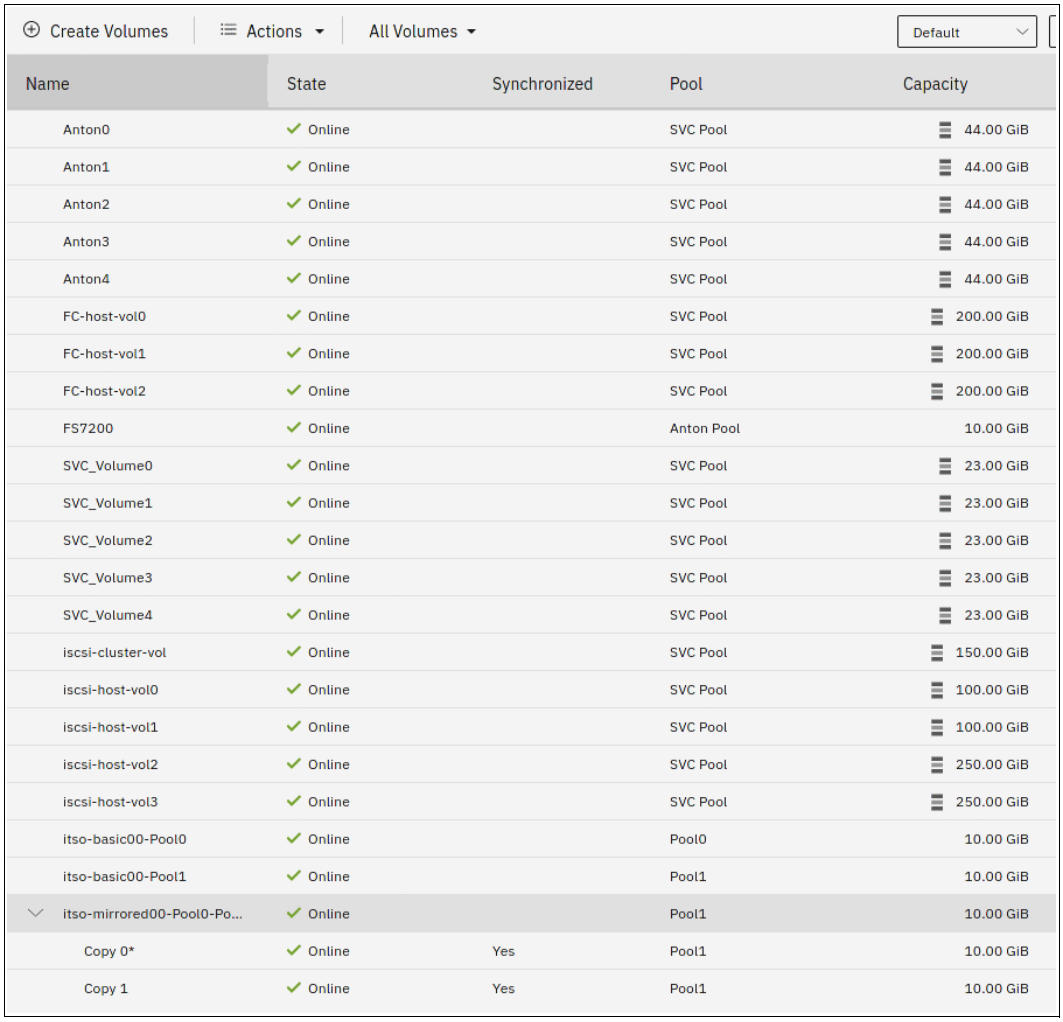


Figure 6-80 Volume copy after migration

The volume copy is now migrated without any host or application downtime to the new storage pool.

Note: Migrating a volume between storage pools with different extent sizes is *not* supported. If you must migrate a volume to a storage pool with a different extent size, use the volume migration by adding a volume copy method.

Volume migration by adding a volume copy

IBM Storage Virtualize supports creating, synchronizing, splitting, and deleting volume copies. A combination of these tasks can be used to migrate volumes between storage pools.

The easiest way to migrate volumes is to use the migration feature that is described in 6.6.8, “Migrating a volume to another storage pool” on page 482. However, in some use cases, the preferred or only method of volume migration is to create a copy of the volume in the target storage pool and then remove the old copy.

Note: You can specify the storage efficiency characteristics of the new volume copy differently than the ones of the primary copy. For example, you can make a thin-provisioned copy of a standard-provisioned volume.

This volume migration option can be used for single-copy volumes only. If you must move a copy of a mirrored volume by using this method, you must delete one of the volume copies first and then create a copy in the target storage pool. This process causes a temporary loss of redundancy while the volume copies synchronize.

To migrate a volume by using the volume copy feature, complete the following steps:

1. Select the volume that you want to move, and select **Actions** → **Add Volume Copy**, as shown in Figure 6-81.

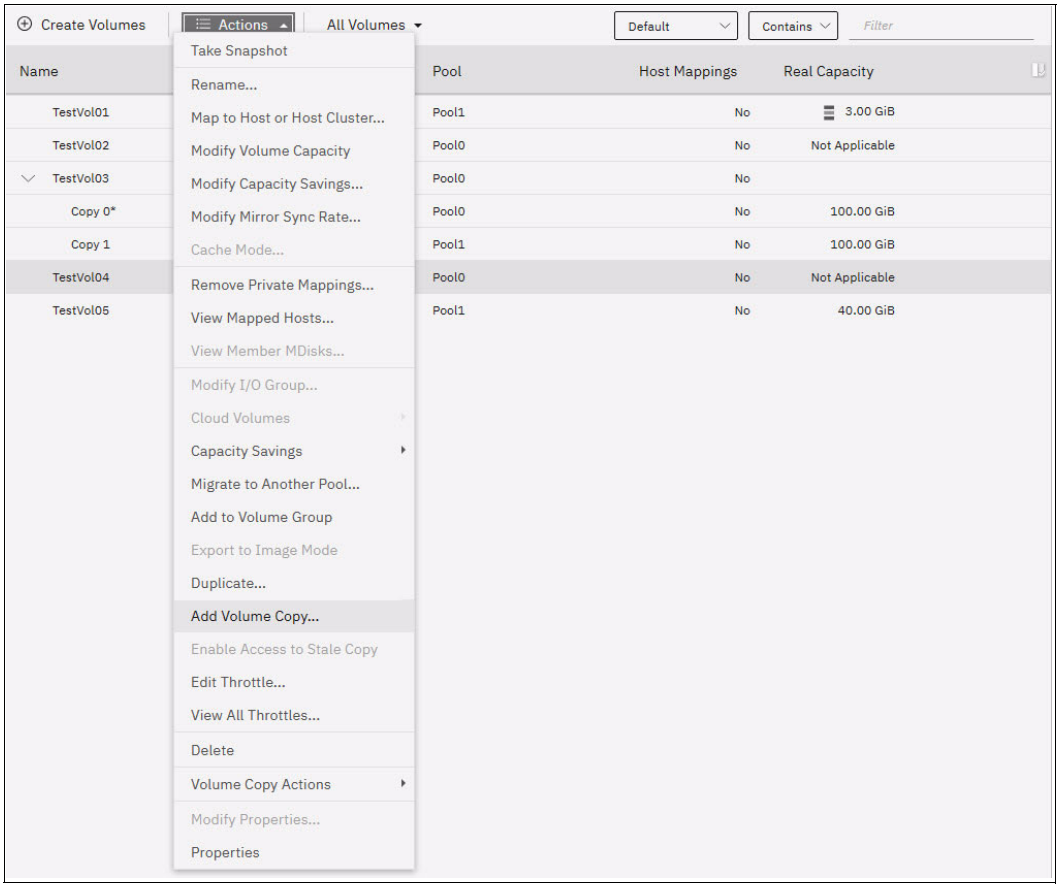


Figure 6-81 Add Volume Copy menu item

2. Create a second copy of your volume in the target storage pool, as shown in Figure 6-82. You can modify the capacity savings options for the new volume copy. In our example, a compressed copy of the volume is created in target pool **Pool1**. The Deduplication option is not available if either of the volume copies is not in a DRP. Click **Add** to proceed.

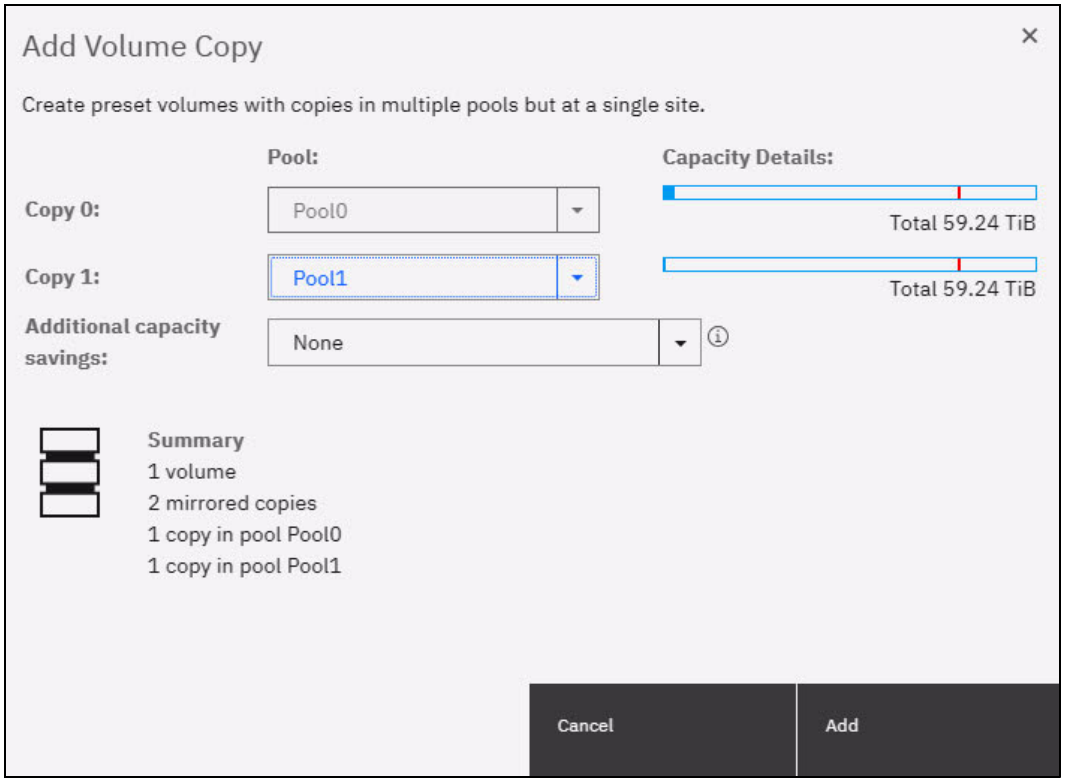


Figure 6-82 Adding a volume copy

Wait until the copies are synchronized, as shown in Figure 6-83.

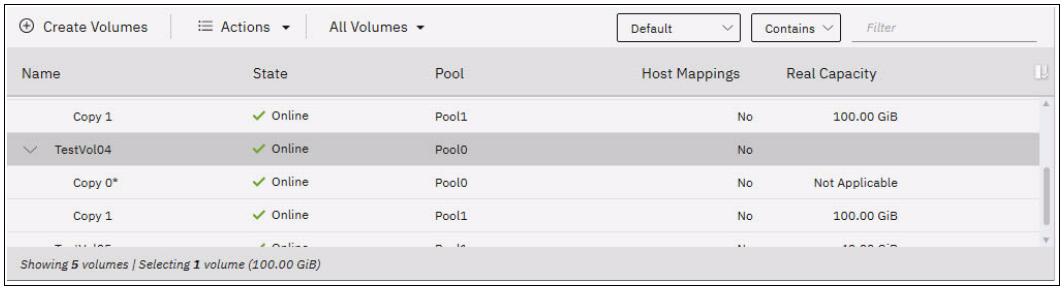


Figure 6-83 Verifying that the volume copies are synchronized

3. Change the roles of the volume copies by making the new copy the primary copy, as shown in Figure 6-84. The current primary copy is displayed with an asterisk next to its name.

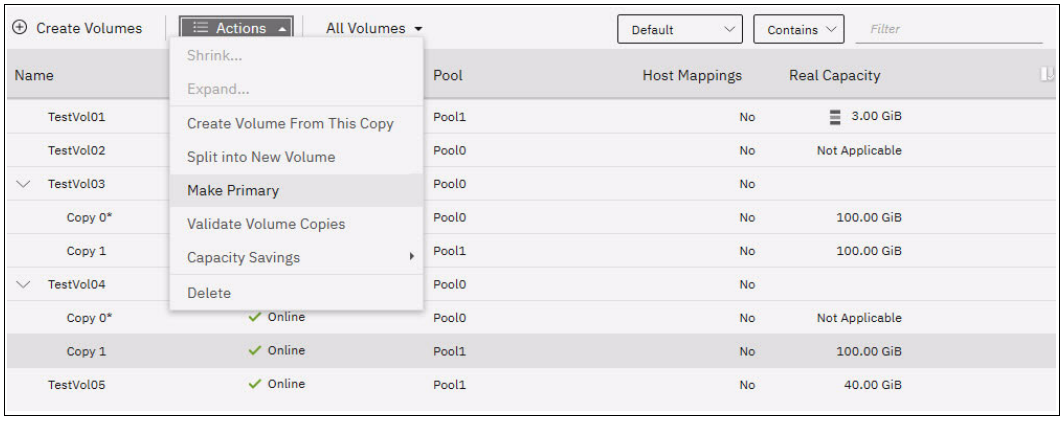


Figure 6-84 Setting the volume copy in the target storage pool as the primary copy

4. Split or delete the volume copy in the source pool, as shown in Figure 6-85.

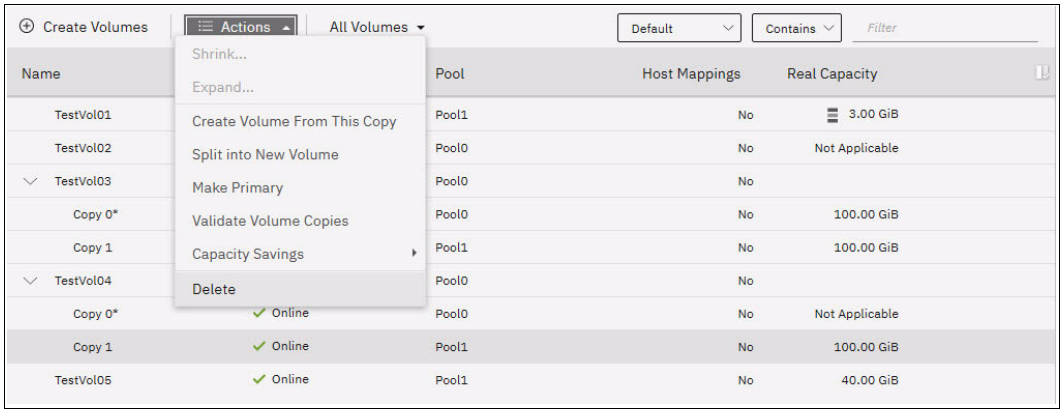


Figure 6-85 Deleting the volume copy in the source pool

5. Confirm the removal of the volume copy, as shown in Figure 6-86.

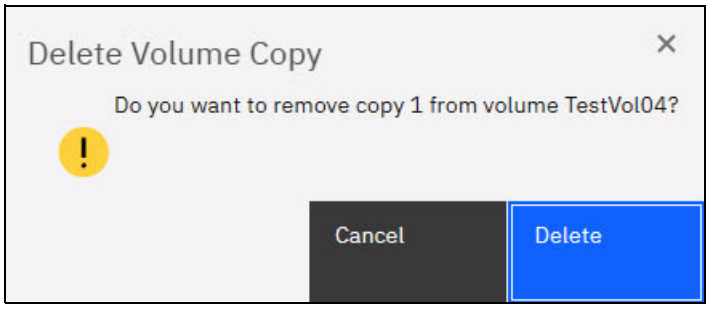


Figure 6-86 Confirming the deletion of a volume copy

6. The Volumes view shows that the volume has a single copy in the target pool, as shown in Figure 6-87.

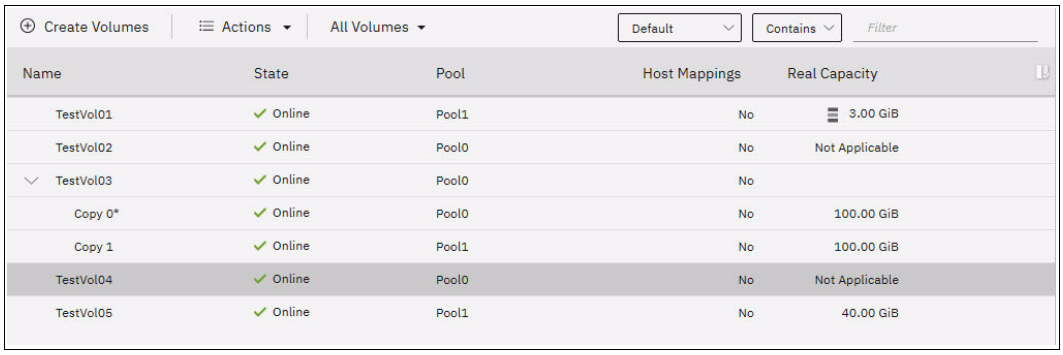


Figure 6-87 Volume copy in the target storage pool

Migrating volumes by using the volume copy feature requires more user interaction, but might be a preferred option for specific use cases. One such example is migrating a volume from a tier 1 storage pool to a lower performance tier 2 storage pool.

First, the volume copy feature can be used to create a copy in the tier 2 pool (steps 1 and 2). All reads are still performed in the tier 1 pool to the primary copy. After the volume copies are synchronized (step 3), all writes are destaged to both pools, but the reads are still done only from the primary copy.

To test the performance of the volume in the new pool, switch the roles of the volume copies to make the new copy the primary (step 4). If the performance is acceptable, the volume copy in tier 1 can be split or deleted. If the tier 2 pool shows unsatisfactory performance, switch the primary volume copy to one that is backed by tier 1 storage.

With this method, you can migrate between storage tiers with a fast and secure back-out option.

6.7 Volume operations by using the CLI

This section describes how to perform various volume configuration and administrative tasks by using the CLI.

For more information about how to set up CLI access, see Appendix A, “Command line interface setup” on page 1195.

6.7.1 Displaying volume information

To display information about all volumes that are defined within the IBM Storage Virtualize environment, run the **lsvdisk** command. To display more information about a specific volume, run the command again and provide the volume name or the volume ID as the command parameter, as shown in Example 6-1.

Example 6-1 The lsvdisk command

```
IBM_Storwize:ITSO:superuser>lsvdisk -delim ' '
id name IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity type FC_id
FC_name RC_id RC_name vdisk_UID fc_map_count copy_count fast_write_state se_copy_count
```

```
RC_change compressed_copy_count parent_mdisk_grp_id parent_mdisk_grp_name formatting
encrypt volume_id volume_name function
0 A_MIRRORED_VOL_1 0 io_grp0 online many many 10.00GB many
6005076400F58004980000000000002 0 2 empty 0 no 0 many many no yes 0 A_MIRRORED_VOL_1
1 COMPRESSED_VOL_1 0 io_grp0 online 1 Pool1 15.00GB striped
6005076400F58004980000000000003 0 1 empty 0 no 1 1 Pool1 no yes 1 COMPRESSED_VOL_1
2 vdisk0 0 io_grp0 online 0 Pool0 10.00GB striped     6005076400F58004980000000000004 0 1
empty 0 no 0 0 Pool0 no yes 2 vdisk0
3 THIN_PROVISION_VOL_1 0 io_grp0 online 0 Pool0 100.00GB striped
6005076400F58004980000000000005 0 1 empty 1 no 0 0 Pool0 no yes 3 THIN_PROVISION_VOL_1
4 COMPRESSED_VOL_2 0 io_grp0 online 1 Pool1 30.00GB striped
6005076400F58004980000000000006 0 1 empty 0 no 1 1 Pool1 no yes 4 COMPRESSED_VOL_2
5 COMPRESS_VOL_3 0 io_grp0 online 1 Pool1 30.00GB striped
6005076400F58004980000000000007 0 1 empty 0 no 1 1 Pool1 no yes 5 COMPRESS_VOL_3
6 MIRRORED_SYNC_RATE_16 0 io_grp0 online many many 10.00GB many
6005076400F58004980000000000008 0 2 empty 0 no 0 many many no yes 6 MIRRORED_SYNC_RATE_16
7 THIN_PROVISION_MIRRORED_VOL 0 io_grp0 online many many 10.00GB many
6005076400F58004980000000000009 0 2 empty 2 no 0 many many no yes 7
THIN_PROVISION_MIRRORED_VOL
8 Tiger 0 io_grp0 online 0 Pool0 10.00GB striped     6005076400F5800498000000000010 0 1
not_empty 0 no 0 0 Pool0 yes yes 8 Tiger
12 vdisk0_restore 0 io_grp0 online 0 Pool0 10.00GB striped
6005076400F5800498000000000000E 0 1 empty 0 no 0 0 Pool0 no yes 12 vdisk0_restore
13 vdisk0_restore1 0 io_grp0 online 0 Pool0 10.00GB striped
6005076400F5800498000000000000F 0 1 empty 0 no 0 0 Pool0 no yes 13 vdisk0_restore1
```

6.7.2 Creating a volume

Running the **mkvdisk** command creates sequential, striped, or image mode volumes. When they are mapped to a host object, these objects are seen as disk drives on which the host can perform I/O operations.

Creating an image mode disk: If you do not specify the **-size** parameter when you create an image mode disk, the entire MDisk capacity is used.

You must know the following information before you start to create the volume:

- In which storage pool the volume will have its extents
- From which I/O group the volume will be accessed
- Which IBM Storage Virtualize node will be the preferred node for the volume
- Size of the volume
- Name of the volume
- Type of the volume
- Whether this volume is to be managed by IBM Easy Tier to optimize its performance

When you are ready to create your striped volume, run the **mkvdisk** command. The command that is shown in Example 6-2 creates a 10 GB striped volume within the storage pool `Pool0` and assigns it to the I/O group `io_grp0`. Its preferred node is node `1`. The volume is given ID 8 by the system.

Example 6-2 The mkvdisk command

```
IBM_Storwize:ITSO:superuser>mkvdisk -mdiskgrp Pool0 -iogrp io_grp0 -size 10 -unit gb -name
Tiger
Virtual Disk, id [8], successfully created
```

To verify the results, run the **lsvdisk** command and provide the volume ID as the command parameter, as shown in Example 6-3.

Example 6-3 The lsvdisk command

```
IBM_Storwize:ITSO:superuser>lsvdisk 8
id 8
name Tiger
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name Pool0
capacity 10.00GB
type striped
formatted no
formatting yes
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076400F580049800000000000010
preferred_node_id 2
fast_write_state not_empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
File system
mirror_write_priority latency
RC_change no
compressed_copy_count 0
access_IO_group_count 1
last_access_time
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
owner_type none
owner_id
owner_name
encrypt yes
volume_id 8
volume_name Tiger
function
throttle_id
throttle_name
IOPs_limit
bandwidth_limit_MB
volume_group_id
volume_group_name
cloud_backup_enabled no
cloud_account_id
cloud_account_name
backup_status off
last_backup_time
restore_status none
backup_grain_size
deduplicated_copy_count 0
```

```
copy_id 0
status online
sync yes
auto_delete no
primary yes
mdisk_grp_id 0
mdisk_grp_name Pool0
type striped
mdisk_id
mdisk_name
fast_write_state not_empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status balanced
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
encrypt yes
deduplicated_copy no
used_capacity_before_reduction
0.00MB
```

The required tasks to create a volume are complete.

6.7.3 Creating a thin-provisioned volume

Example 6-4 shows creating a thin-provisioned volume, which requires the following parameters to be specified:

-rsize This parameter makes the volume a thin-provisioned volume. If this parameter is missing, the volume is created as standard-provisioned.

-autoexpand This parameter specifies that thin-provisioned volume copies automatically expand their real capacities by allocating new extents from their storage pool (optional).

-grainsize This parameter sets the grain size in kilobytes (KB) for a thin-provisioned volume (optional).

Example 6-4 Running the mkvdisk command

```
IBM_Storwize:ITSO:superuser>mkvdisk -mdiskgrp Pool0 -iogrp 0 -vtype striped -size 10 -unit
gb -rsize 50% -autoexpand -grainsize 256
```

```
Virtual Disk, id [9], successfully created
```

This command creates a thin-provisioned volume with 10 GB of virtual capacity in a storage pool that is named `Site1_Pool` and is owned by I/O group `io_grp0`. The real capacity is set to automatically expand until the real volume size of 10 GB is reached. The grain size is set to 256 KB, which is the default.

Disk size: When the **`-rsize`** parameter is used to specify the real physical capacity of a thin-provisioned volume, the following options are available to specify the physical capacity: **`disk_size`**, **`disk_size_percentage`**, and **`auto`**.

Use the **`disk_size_percentage`** option to define initial real capacity by using a percentage of the disk's virtual capacity that is defined by the **`-size`** parameter. This option takes as a parameter an integer, or an integer that is immediately followed by the percent (%) symbol.

Use the **`disk_size`** option to directly specify the real physical capacity by specifying its size in the units that are defined by using the **`-unit`** parameter (the default unit is MB). The **`-rsize`** value can be greater than, equal to, or less than the size of the volume.

The **`auto`** option creates a volume copy that uses the entire size of the MDisk. If you specify the **`-rsize auto`** option, you must also specify the **`-vtype image`** option.

An entry of 1 GB uses 1,024 MB.

6.7.4 Creating a volume in image mode

Use an image mode volume to bring a non-virtualized disk under the control of the IBM Storage Virtualize system. After it is managed by the system, you can migrate the volume to the standard MDisk.

When an image mode volume is created, it directly maps to the unmanaged MDisk from which it is created. Therefore, except for a thin-provisioned image mode volume, the number of LBAs on the volume equals the number of LBAs on the MDisk.

Size: An image mode volume must be at least 512 bytes (the capacity cannot be 0) and always occupies at least one extent.

You must use the **`-mdisk`** parameter to specify an MDisk that has a mode of `unmanaged`. The **`-fmtdisk`** parameter cannot be used to create an image mode volume.

Capacity: If you create a mirrored volume from two image mode MDisks without specifying a **`-capacity`** value, the capacity of the resulting volume is the smaller of the two MDisks. The remaining space on the larger MDisk is inaccessible.

If you do not specify the **`-size`** parameter when you create an image mode disk, the entire MDisk capacity is used.

Running the **`mkvdisk`** command to create an image mode volume is shown in Example 6-5.

Example 6-5 The mkvdisk (image mode) command

```
IBM_2145:ITSO_CLUSTER:superuser>mkvdisk -mdiskgrp ITSO_Pool1 -iogrp 0 -mdisk mdisk25 -vtype
image -name Image_Volume_A
```

```
Virtual Disk, id [6], successfully created
```

As shown in Example 6-5 on page 492, an image mode volume that is named `Image_Volume_A` is created that uses the `mdisk25` MDisk. The MDisk is moved to the storage pool `ITSO_Pool1`, and the volume is owned by the I/O group `io_grp0`.

If you run the **`lsvdisk`** command, it shows a volume that is named `Image_Volume_A` with the type `image`, as shown in Example 6-6.

Example 6-6 The lsvdisk command

```
IBM_2145:ITSO_CLUSTER:superuser>lsvdisk -filtervalue type=image
id name           IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity
type  FC_id FC_name RC_id RC_name vdisk_UID                         fc_map_count copy_count
fast_write_state se_copy_count RC_change compressed_copy_count parent_mdisk_grp_id
parent_mdisk_grp_name formatting encrypt volume_id volume_name    function
6  Image_Volume_A 0           io_grp0       online 5             ITSO_Pool1     1.00GB
image                            6005076801FE80840800000000000021 0            1
empty            0             no        0                     5
ITSO_Pool1            no         no      6         Image_Volume_A
```

6.7.5 Adding a volume copy

You can add a copy to a volume. If volume copies are defined on different MDisks, the volume remains accessible, even when the MDisk on which one of its copies depends becomes unavailable. You can also create a copy of a volume on a dedicated MDisk by creating an image mode copy of the volume. Although volume copies can increase the availability of data, they are not separate objects.

Volume mirroring can be also used as an alternative method of migrating volumes between storage pools.

To create a copy of a volume, run the **`addvdiskcopy`** command. This command creates a copy of the chosen volume in the specified storage pool, which changes a non-mirrored volume into a mirrored one.

The following scenario shows how to create a copy of a volume in a different storage pool. As shown in Example 6-7, the volume initially has a single copy with `copy_id 0` that is provisioned in pool `Pool0`.

Example 6-7 The lsvdisk command

```
IBM_Storwize:ITSO:superuser>lsvdisk 2
id 2
name vdisk0
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name Pool0
capacity 10.00GB
type striped
formatted yes
formatting no
mdisk_id
mdisk_name
FC_id
```

```
FC_name
RC_id
RC_name
vdisk_UID 6005076400F580049800000000000004
preferred_node_id 2
fast_write_state empty
cache readonly
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
File system
mirror_write_priority latency
RC_change no
compressed_copy_count 0
access_IO_group_count 1
last_access_time
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
owner_type none
owner_id
owner_name
encrypt yes
volume_id 2
volume_name vdisk0
function
throttle_id
throttle_name
IOPs_limit
bandwidth_limit_MB
volume_group_id
volume_group_name
cloud_backup_enabled no
cloud_account_id
cloud_account_name
backup_status off
last_backup_time
restore_status none
backup_grain_size
deduplicated_copy_count 0

copy_id 0
status online
sync yes
auto_delete no
primary yes
mdisk_grp_id 0
mdisk_grp_name Pool0
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
```

```
se_copy no
easy_tier on
easy_tier_status balanced
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
encrypt yes
deduplicated_copy no
used_capacity_before_reduction 0.00MB
```

Example 6-8 shows adding the second volume copy by running the **addvdiskcopy** command.

Example 6-8 The addvdiskcopy command

```
IBM_Storwize:ITSO:superuser>addvdiskcopy -mdiskgrp Pool1 -vtype striped -unit gb vdisk0
Vdisk [2] copy [1] successfully created
```

During the synchronization process, you can see the status by running the **lsvdisksyncprogress** command.

As shown in Example 6-9, the first time that the status is checked, the synchronization progress is at 48%, and the estimated completion time is 201018232305. The estimated completion time is displayed in the YYMMDDHHMMSS format. In our example, it is 2020, Oct-18 20:23:05. When the command is run again, the progress status is at 100%, and the synchronization is complete.

Example 6-9 Synchronization

```
IBM_Storwize:ITSO:superuser>lsvdisksyncprogress
vdisk_id vdisk_name copy_id progress estimated_completion_time
2        vdisk0     1       0        201018202305
IBM_Storwize:ITSO:superuser>lsvdisksyncprogress
vdisk_id vdisk_name copy_id progress estimated_completion_time
2        vdisk0     1       100
```

As shown in Example 6-10, the new volume copy (copy_id 1) was added and appears in the output of the **lsvdisk** command.

Example 6-10 The lsvdisk command

```
IBM_Storwize:ITSO:superuser>lsvdisk vdisk0
id 2
name vdisk0
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id many
mdisk_grp_name many
capacity 10.00GB
type many
```

```
formatted yes
formatting no
mdisk_id many
mdisk_name many
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076400F580049800000000000004
preferred_node_id 2
fast_write_state empty
cache readonly
udid
fc_map_count 0
sync_rate 50
copy_count 2
se_copy_count 0
File system
mirror_write_priority latency
RC_change no
compressed_copy_count 0
access_IO_group_count 1
last_access_time
parent_mdisk_grp_id many
parent_mdisk_grp_name many
owner_type none
owner_id
owner_name
encrypt yes
volume_id 2
volume_name vdisk0
function
throttle_id
throttle_name
IOPs_limit
bandwidth_limit_MB
volume_group_id
volume_group_name
cloud_backup_enabled no
cloud_account_id
cloud_account_name
backup_status off
last_backup_time
restore_status none
backup_grain_size
deduplicated_copy_count 0

copy_id 0
status online
sync yes
auto_delete no
primary yes
mdisk_grp_id 0
mdisk_grp_name Pool0
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
```

```
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status balanced
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
encrypt yes
deduplicated_copy no
used_capacity_before_reduction 0.00MB

copy_id 1
status online
sync yes
auto_delete no
primary no
mdisk_grp_id 1
mdisk_grp_name Pool1
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status balanced
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 1
parent_mdisk_grp_name Pool1
encrypt yes
deduplicated_copy no
```

```
used_capacity_before_reduction 0.00MB
```

When adding a volume copy, you can define it with different parameters than the original volume copy. For example, you can create a thin-provisioned copy of a standard-provisioned volume to migrate a thick-provisioned volume to a thin-provisioned volume. The migration can also be done in the opposite direction.

Volume copy mirror parameters: To change the parameters of a volume copy, you must delete the volume copy and redefine it with the new values.

In Example 6-11, the volume name is changed from `VOL_NO_MIRROR` to `VOL_WITH_MIRROR`.

Example 6-11 Volume name changes

```
IBM_Storwize:ITSO:superuser>chvdisk -name VOL_WITH_MIRROR VOL_NO_MIRROR
IBM_Storwize:ITSO:superuser>
```

Using the -autodelete flag to migrate a volume

This section shows how to run the **addvdiskcopy** command with the **-autodelete** flag set. The **-autodelete** flag causes the primary copy to be deleted after the secondary copy is synchronized.

Example 6-12 shows a shortened **lsvdisk** output for a decompressed volume with a single volume copy.

Example 6-12 A decompressed volume

```
IBM_Storwize:ITSO:superuser>lsvdisk UNCOMPRESSED_VOL
id 9
name UNCOMPRESSED_VOL
IO_group_id 0
IO_group_name io_grp0
status online
...

copy_id 0
status online
sync yes
auto_delete no
primary yes
...
compressed_copy no
...
```

Example 6-13 adds a compressed copy with the **-autodelete** flag set.

Example 6-13 Compressed copy

```
IBM_Storwize:ITSO:superuser>addvdiskcopy -autodelete -rsize 2 -mdiskgrp 0 -compressed
UNCOMPRESSED_VOL
Vdisk [9] copy [1] successfully created
```

Example 6-14 on page 499 shows the **lsvdisk** output with another compressed volume (copy 1) and volume copy 0 being set to **auto_delete yes**.

Example 6-14 The lsvdisk command output

```
IBM_Storwize:ITSO:superuser>lsvdisk UNCOMPRESSED_VOL
id 9
name UNCOMPRESSED_VOL
IO_group_id 0
IO_group_name io_grp0
status online
...
compressed_copy_count 2
...

copy_id 0
status online
sync yes
auto_delete yes
primary yes
...

copy_id 1
status online
sync no
auto_delete no
primary no
...
```

When copy 1 is synchronized, copy 0 is deleted. You can monitor the progress of volume copy synchronization by running the **lsvdisksyncprogress** command.

6.7.6 Splitting a mirrored volume

Running the **splitvdiskcopy** command creates an independent volume in the specified I/O group from a volume copy of the specified mirrored volume. In effect, the command changes a volume with two copies into two independent volumes, each with a single copy.

If the copy that you are splitting is not synchronized, you must use the **-force** parameter. If you are attempting to remove the only synchronized copy of the source volume, the command fails. However, you can run the command when either copy of the source volume is offline.

Example 6-15 shows the **splitvdiskcopy** command, which is used to split a mirrored volume. It creates a volume that is named SPLIT_VOL from a copy with ID 1 of the volume that is named VOLUME_WITH_MIRRORED_COPY.

Example 6-15 Splitting a volume

```
IBM_Storwize:ITSO:superuser>splitvdiskcopy -copy 1 -iogrp 0 -name SPLIT_VOL
VOLUME_WITH_MIRRORED_COPY
Virtual Disk, id [1], successfully created
```

As you can see in Example 6-16, the new volume is created as an independent volume.

Example 6-16 The lsvdisk command

```
IBM_Storwize:ITSO:superuser>lsvdisk SPLIT_VOL
id 1
name SPLIT_VOL
IO_group_id 0
IO_group_name io_grp0
status online
```

```
mdisk_grp_id 1
mdisk_grp_name Pool1
capacity 10.00GB
type striped
formatted yes
formatting no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076400F580049800000000000012
preferred_node_id 1
fast_write_state empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
File system
mirror_write_priority latency
RC_change no
compressed_copy_count 0
access_IO_group_count 1
last_access_time
parent_mdisk_grp_id 1
parent_mdisk_grp_name Pool1
owner_type none
owner_id
owner_name
encrypt yes
volume_id 1
volume_name SPLIT_VOL
function
throttle_id
throttle_name
IOPs_limit
bandwidth_limit_MB
volume_group_id
volume_group_name
cloud_backup_enabled no
cloud_account_id
cloud_account_name
backup_status off
last_backup_time
restore_status none
backup_grain_size
deduplicated_copy_count 0

copy_id 0
status online
sync yes
auto_delete no
primary yes
mdisk_grp_id 1
mdisk_grp_name Pool1
type striped
mdisk_id
```

```
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status balanced
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 1
parent_mdisk_grp_name Pool1
encrypt yes
deduplicated_copy no
used_capacity_before_reduction 0.00MB
```

6.7.7 Modifying a volume

Running the **chvdisk** command modifies a single property of a volume. Only one property can be modified at a time. Therefore, changing the volume name and modifying its I/O group requires two invocations of the command.

Tips: Changing the I/O group with which this volume is associated requires a flush of the cache within the nodes in the current I/O group to ensure that all data is written to disk. I/O must be suspended at the host level before you perform this operation.

If the volume has a mapping to any hosts, it is impossible to move the volume to an I/O group that does not include any of those hosts.

This operation fails if insufficient space exists to allocate bitmaps for a mirrored volume in the target I/O group.

If the **-force** parameter is used and the system cannot destage all write data from the cache, the contents of the volume are corrupted by the loss of the cached data.

If the **-force** parameter is used to move a volume that has out-of-sync copies, a full resynchronization is required.

6.7.8 Deleting a volume

To delete a volume, run the **rmvdisk** command. When this command is run on a managed mode volume, any data on the volume is lost, and the extents that made up this volume are returned to the pool of free extents in the storage pool.

If any RC, IBM FlashCopy, or host mappings still exist for the target of **rmvdisk** command, the delete fails unless the **-force** flag is specified. This flag causes the deletion of the volume and any volume to host mappings and copy mappings.

If the volume is being migrated to image mode, the delete fails unless the **-force** flag is specified. The use of the **-force** flag halts the migration and then deletes the volume.

If the command succeeds (without the **-force** flag) for an image mode volume, the write cache data is flushed to the storage before the volume is removed. Therefore, the underlying LU is consistent with the disk state from the point of view of the host that uses the image mode volume (crash-consistent file system). If the **-force** flag is used, consistency is not ensured; that is, the data that the host believes to be written might not be present on the LU.

If any non-destaged data exists in the fast write cache for the target of **rmvdisk** command, the deletion of the volume fails unless the **-force** flag is specified, in which case, any non-destaged data in the fast write cache is deleted.

Example 6-17 shows how to run the **rmvdisk** command to delete a volume from your IBM Storage Virtualize configuration.

Example 6-17 The rmvdisk command

```
IBM_2145:ITSO_CLUSTER:superuser>rmvdisk volume_A
```

This command deletes the `volume_A` volume from the IBM Storage Virtualize configuration. If the volume is assigned to a host, you must use the **-force** flag to delete the volume, as shown in Example 6-18.

Example 6-18 The rmvdisk -force command

```
IBM_2145:ITSO_CLUSTER:superuser>rmvdisk -force volume_A
```

6.7.9 Volume protection

To prevent active volumes or host mappings from being deleted inadvertently, the system supports a global setting that prevents these objects from being deleted if the system detects recent I/O activity to these objects.

To set the time interval for which the volume must be idle before it can be deleted from the system, run the **chsystem** command. This setting affects the following commands:

- **rmvdisk**
- **rmvolume**
- **rmvdiskcopy**
- **rmvdiskhostmap**
- **rmmdiskgrp**
- **rmhostiogrp**
- **rmhost**
- **rmhostport**

These commands fail unless the volume was idle for the specified interval or the **-force** parameter was used.

To enable volume protection by setting the required inactivity interval, run the following command:

```
svctask chsystem -vdiskprotectionenabled yes -vdiskprotectiontime 60
```

The **-vdiskprotectionenabled yes** parameter enables volume protection and the **-vdiskprotectiontime** parameter specifies for how long a volume must be inactive (in minutes) before it can be deleted. In this example, volumes can be deleted only if they were inactive for over 60 minutes.

To disable volume protection, run the following command:

```
svctask chsystem -vdiskprotectionenabled no
```

6.7.10 Expanding a volume

Expanding a volume presents a larger capacity disk to your operating system. Although this expansion can be easily performed by using IBM Storage Virtualize, you must ensure that your operating system supports expansion before this function is used.

Assuming that your operating system supports expansion, you can run the **expandvdisksize** command to increase the capacity of a volume, as shown in Example 6-19.

Example 6-19 The expandvdisksize command

```
IBM_2145:ITSO_CLUSTER:superuser>expandvdisksize -size 5 -unit gb volume_C
```

This command expands the `volume_C` volume, which was 35 GB, by another 5 GB.

To expand a thin-provisioned volume, you can use the **-rsize** option, as shown in Example 6-20. This command changes the real size of the `volume_B` volume to a real capacity of 55 GB. The capacity of the volume is unchanged.

Example 6-20 The lsvdisk command

```
IBM_Storwize:ITSO:superuser>lsvdisk volume_B
id 26
capacity 100.00GB
type striped
.
.
copy_id 0
status online
used_capacity 0.41MB
real_capacity 50.02GB
free_capacity 50.02GB
overallocation 199
autoexpand on
warning 80
grainsize 32
se_copy yes

IBM_Storwize:ITSO:superuser>expandvdisksize -rsize 5 -unit gb volume_B
IBM_Storwize:ITSO:superuser>lsvdisk volume_B
id 26
name volume_B
```

```
capacity 100.00GB
type striped
.
.
copy_id 0
status online
used_capacity 0.41MB
real_capacity 55.02GB
free_capacity 55.02GB
overallocation 181
autoexpand on
warning 80
grainsize 32
se_copy yes
```

Important: If a volume is expanded, its type becomes striped, even if it was previously sequential or in image mode.

If not enough extents are available to expand your volume to the specified size, the following error message is displayed:

```
CMMVC5860E The action failed because there were not enough extents in the
storage pool.
```

6.7.11 HyperSwap volume modification with CLI

The following CLI commands for administering volumes were released in IBM Storage Virtualize V7.6. However, only for HyperSwap volume creation and deletion the GUI uses the new commands **mkvolume** and **rmvolume**:

- **mkvolume**
- **mkimagevolume**
- **addvolumecopy**
- **rmvolumecopy**
- **rmvolume**

In addition, the **lsvdisk** output shows more fields: `volume_id`, `volume_name`, and `function`, which helps to identify the individual VDisks that make up a HyperSwap volume. This information is used by the GUI to provide views that reflect the client's view of the HyperSwap volume and its site-dependent copies, as opposed to the "low-level" VDisks and VDisk Change Volumes.

The following individual commands are related to HyperSwap:

- **mkvolume**

  Creates an empty volume by using storage from a storage pool. The type of volume that is created is determined by the system topology and the number of storage pools that is specified. The volume is always formatted (zeroed). The **mkvolume** command can be used to create the following objects:

  - Basic volume: Any topology
  - Mirrored volume: Standard topology
  - Stretched volume: Stretched topology
  - HyperSwap volume: HyperSwap topology

- **rmvolume**

Removes a volume. For a HyperSwap volume, this process includes deleting the active-active relationship and the change volumes.

The **-force** parameter that is used by **rmvdisk** is replaced by a set of override parameters, one for each operation-stopping condition, which makes it clearer to the user exactly what protection they are bypassing.

- **mkimagevolume**

  Creates an image mode volume. This command can be used to import a volume, which preserves data. It can be implemented as a separate command to provide greater differentiation between the action of creating an empty volume and creating a volume by importing data on an MDisk.

- **addvolumecopy**

  Adds a copy to a volume. The new copy is always synchronized from the existing copy. For stretched and HyperSwap topology systems, this command creates a HA volume. This command can be used to create the following volume types:

  – Mirrored volume: Standard topology
  – Stretched volume: Stretched topology
  – HyperSwap volume: HyperSwap topology

- **rmvolumecopy**

  Removes a copy of a volume. This command leaves the volume intact. It also converts a Mirrored, Stretched, or HyperSwap volume to a basic volume. For a HyperSwap volume, this command includes deleting the active-active relationship and the change volumes.

  This command enables a copy to be identified by its site.

  The **-force** parameter that is used by **rmvdiskcopy** is replaced by a set of override parameters, one for each operation-stopping condition, which makes it clearer to the user what protection they are bypassing.

6.7.12 Mapping a volume to a host

To map a volume to a host, run the **mkvdiskhostmap** command. This mapping makes the volume available to the host for I/O operations. A host can perform I/O operations only on volumes that are mapped to it.

When the host bus adapter (HBA) on the host scans for devices that are attached to it, the HBA discovers all of the volumes that are mapped to its FC ports and their SCSI identifiers (SCSI LUN IDs).

For example, the first disk that is found is generally SCSI LUN 1. You can control the order in which the HBA discovers volumes by assigning the SCSI LUN ID as required. If you do not specify a SCSI LUN ID when mapping a volume to the host, the storage system automatically assigns the next available SCSI LUN ID based on any mappings that exist with that host.

Note: The SCSI-3 standard requires LUN 0 to exist on every SCSI target. This LUN must implement a number of standard commands, including Report LUNs. However, this LUN does not have to provide any storage capacity.

Example 6-21 shows how to map volumes `volume_B` and `volume_C` to the defined host `Almaden` by running the **mkvdiskhostmap** command.

Example 6-21 The mkvdiskhostmap command

```
IBM_Storwize:ITSO:superuser>mkvdiskhostmap -host Almaden volume_B
```

```
Virtual Disk to Host map, id [0], successfully created
IBM_Storwize:ITSO:superuser>mkvdiskhostmap -host Almaden volume_C
Virtual Disk to Host map, id [1], successfully created
```

Example 6-22 shows the output of the **lshostvdiskmap** command, which shows that the volumes are mapped to the host.

Example 6-22 The lshostvdiskmap -delim command

```
IBM_2145:ITSO_CLUSTER:superuser>lshostvdiskmap -delim :
id:name:SCSI_id:vdisk_id:vdisk_name:vdisk_UID
2:Almaden:0:26:volume_B:6005076801AF813F1000000000000020
2:Almaden:1:27:volume_C:6005076801AF813F1000000000000021
```

Assigning a specific LUN ID to a volume: The optional **-scsi scsi_lun_id** parameter can help assign a specific LUN ID to a volume that is to be associated with a host. The default (if nothing is specified) is to assign the next available ID based on the current volume that is mapped to the host.

Specific HBA device drivers stop when they find a gap in the sequence of SCSI LUN IDs, as shown in the following examples:

- Volume 1 is mapped to Host 1 with SCSI LUN ID 1
- Volume 2 is mapped to Host 1 with SCSI LUN ID 2
- Volume 3 is mapped to Host 1 with SCSI LUN ID 4

When the device driver scans the HBA, it might stop after discovering volumes 1 and 2 because no SCSI LUN is mapped with ID 3.

Important: Ensure that the SCSI LUN ID allocation is contiguous.

If you use host clusters, run the **mkvolumehostclustermap** command to map a volume to a host cluster instead (see Example 6-23).

Example 6-23 The mkvolumehostclustermap command

```
BM_Storwize:ITSO:superuser>mkvolumehostclustermap -hostcluster vmware_cluster
UNCOMPRESSED_VOL
Volume to Host Cluster map, id [0], successfully created
```

6.7.13 Listing volumes that are mapped to the host

To show the volumes that are mapped to the specific host, run the **lshostvdiskmap** command, as shown in Example 6-24.

Example 6-24 The lshostvdiskmap command

```
IBM_2145:ITSO_CLUSTER:superuser>lshostvdiskmap -delim , Siam
id,name,SCSI_id,vdisk_id,vdisk_name,wwpn,vdisk_UID
3,Siam,0,0,volume_A,210000E08B18FF8A,60050768018301BF280000000000000C
```

In the output of the command, you can see that only one volume (`volume_A`) is mapped to the host `Siam`. The volume is mapped with SCSI LUN ID 0.

If no hostname is specified by the **lshostvdiskmap** command, it returns all defined host-to-volume mappings.

Specifying the flag before the hostname: Although the **-delim** flag normally comes at the end of the command string, you must specify this flag before the hostname in this case.

You can also run the **lshostclustervolumemap** command to show the volumes that are mapped to a specific host cluster, as shown in Example 6-25.

Example 6-25 The lshostclustervolumemap command

```
IBM_Storwize:ITSO:superuser>lshostclustervolumemap
id name           SCSI_id volume_id volume_name      volume_UID
IO_group_id IO_group_name
0  vmware_cluster 0       9         UNCOMPRESSED_VOL 6005076400F5800498000000000000011 0
io_grp0
```

6.7.14 Listing hosts that are mapped to the volume

To identify the hosts to which a specific volume was mapped, run the **lsvdiskhostmap** command, as shown in Example 6-26.

Example 6-26 The lsvdiskhostmap command

```
IBM_2145:ITSO_CLUSTER:superuser>lsvdiskhostmap -delim , volume_B
id,name,SCSI_id,host_id,host_name,vdisk_UID
26,volume_B,0,2,Almaden,6005076801AF813F1000000000000020
```

This command shows the list of hosts to which the volume `volume_B` is mapped.

Specifying the -delim flag: Although the optional **`-delim`** flag normally comes at the end of the command string, you must specify this flag before the volume name in this case. Otherwise, the command does not return any data.

6.7.15 Deleting a volume to host mapping

Deleting a volume mapping does not affect the volume. Instead, it removes only the host's ability to use the volume. To unmap a volume from a host, run the **`rmvdiskhostmap`** command, as shown in Example 6-27.

Example 6-27 The rmvdiskhostmap command

```
IBM_2145:ITSO_CLUSTER:superuser>rmvdiskhostmap -host Tiger volume_D
```

This command unmaps the volume that is called `volume_D` from the host that is called `Tiger`.

You can also run the **`rmvolumehostclustermap`** command to delete a volume mapping from a host cluster, as shown in Example 6-28.

Example 6-28 The rmvolumehostclustermap command

```
IBM_Storwize:ITSO:superuser>rmvolumehostclustermap -hostcluster vmware_cluster
UNCOMPRESSED_VOL
```

This command unmaps the volume that is called `UNCOMPRESSED_VOL` from the host cluster that is called `vmware_cluster`.

Note: Removing a volume that is mapped to the host makes the volume unavailable for I/O operations. Ensure that the host is prepared for this situation before removing a volume mapping.

6.7.16 Migrating a volume

You might want to migrate volumes from one set of MDisks to another set of MDisks to decommission an old disk subsystem to better distribute load across your virtualized environment, or to migrate data into the IBM Storage Virtualize environment by using image mode.

For more information about migration, see Chapter 8, “Storage migration” on page 635.

Important: After migration is started, it continues until it completes unless it is stopped or suspended by an error condition or the volume that is being migrated is deleted.

As you can see from the parameters that are shown in Example 6-29, before you can migrate your volume, you must determine the name of the volume that you want to migrate and the name of the storage pool to which you want to migrate it.

To list the names of volumes and storage pools, run the **lsvdisk** and **lsmdiskgrp** commands.

The command that is shown in Example 6-29 moves `volume_C` to the storage pool that is named `STGPool_DS5000-1`.

Example 6-29 The migratevdisk command

```
IBM_2145:ITSO_CLUSTER:superuser>migratevdisk -mdiskgrp STGPool_DS5000-1 -vdisk volume_C
```

Note: If insufficient extents are available within your target storage pool, you receive an error message. Ensure that the source MDisk group and target MDisk group have the same extent size.

You can use the optional threads parameter to control the priority of the migration process. The default is `4`, which is the highest priority setting. However, if you want the process to take a lower priority over other types of I/O, you can specify `3`, `2`, or `1`.

You can run the **lsmigrate** command at any time to see the status of the migration process, as shown in Example 6-30.

Example 6-30 The lsmigrate command

```
IBM_2145:ITSO_CLUSTER:superuser>lsmigrate
migrate_type MDisk_Group_Migration
progress 0
migrate_source_vdisk_index 27
migrate_target_mdisk_grp 2
max_thread_count 4
migrate_source_vdisk_copy_id 0

IBM_2145:ITSO_CLUSTER:superuser>lsmigrate
migrate_type MDisk_Group_Migration
progress 76
migrate_source_vdisk_index 27
migrate_target_mdisk_grp 2
max_thread_count 4
migrate_source_vdisk_copy_id 0
```

Progress: The progress is shown in terms of percentage complete. If no output is displayed when running the command, all volume migrations are finished.

6.7.17 Migrating a fully managed volume to an image mode volume

Migrating a fully managed volume to an image mode volume enables the IBM Storage Virtualize system to be removed from the data path. This feature might be useful when the IBM Storage Virtualize system is used as a data mover.

To migrate a fully managed volume to an image mode volume, the following rules apply:

- Cloud snapshots must not be enabled on the source volume.
- The destination MDisk must be greater than or equal to the size of the volume.
- The MDisk that is specified as the target must be in an unmanaged state.

- Regardless of the mode in which the volume starts, it is reported as a managed mode during the migration.
- If the migration is interrupted by a system recovery or cache problem, the migration resumes after the recovery completes.

Example 6-31 shows running the **migratetoimage** command to migrate the data from `volume_A` onto `mdisk10`, and to put the MDisk `mdisk10` into the `STGPool_IMAGE` storage pool.

Example 6-31 The migratetoimage command

```
IBM_2145:ITSO_CLUSTER:superuser>migratetoimage -vdisk volume_A -mdisk mdisk10 -mdiskgrp
STGPool_IMAGE
```

6.7.18 Shrinking a volume

The **shrinkvdisksize** command reduces the capacity that is allocated to the specific volume by the specified amount. You cannot shrink the real size of a thin-provisioned volume to less than its used size. All capacities (including changes) must be in multiples of 512 bytes. An entire extent is reserved, even if it is only partially used. The default capacity unit is MB.

You can use this command to shrink the physical capacity of a volume or to reduce the virtual capacity of a thin-provisioned volume without altering the physical capacity that is assigned to the volume. To change the volume size, use the following parameters:

- For a standard-provisioned volume, use the **-size** parameter.
- For a thin-provisioned volume's real capacity, use the **-rsize** parameter.
- For a thin-provisioned volume's virtual capacity, use the **-size** parameter.

When the virtual capacity of a thin-provisioned volume is changed, the warning threshold is automatically scaled.

If the volume contains data that is being used, do not shrink the volume without backing up the data first. The system reduces the capacity of the volume by removing arbitrarily chosen extents, or extents from those sets that are allocated to the volume. You cannot control which extents are removed. Therefore, you cannot assume that it is unused space that is removed.

Image mode volumes cannot be reduced. To reduce their size, first they must be migrated to fully managed mode.

Before the **shrinkvdisksize** command is used on a mirrored volume, all copies of the volume must be synchronized.

Important: Consider the following guidelines when you are shrinking a disk:

- If the volume contains data or host-accessible metadata (for example, an empty physical volume of an LVM), do not shrink the disk.
- This command can shrink a FlashCopy target volume to the same capacity as the source.
- Before you shrink a volume, validate that the volume is not mapped to any host objects.
- You can determine the exact capacity of the source or master volume by running the **svcinfo lsvdisk -bytes vdiskname** command.

Shrink the volume by the required amount by running the following command:

```
shrinkvdisksize -size disk_size -unit b | kb | mb | gb | tb | pb vdisk_name | vdisk_id
```

Example 6-32 shows running the **shrinkvdisksize** command to reduce the size of volume volume_D by 44 GB.

Example 6-32 The shrinkvdisksize command

```
IBM_2145:ITSO_CLUSTER:superuser>shrinkvdisksize -size 44 -unit gb volume_D
```

6.7.19 Listing volumes that use MDisks

To identify which volumes use space on the specified MDisk, run the **lsmdiskmember** command. Example 6-33 displays a list of volume IDs of all volume copies that use mdisk8. To correlate the IDs that are displayed in this output to volume names, run the **lsvdisk** command.

Example 6-33 The lsmdiskmember command

```
IBM_2145:ITSO_CLUSTER:superuser>lsmdiskmember mdisk8
id copy_id
24 0
27 0
```

6.7.20 Listing MDisks that are used by the volume

To list MDisks that supply space that is used by the specified volume, run the **lsvdiskmember** command. Example 6-34 lists the MDisk IDs of all MDisks that are used by the volume with ID 0.

Example 6-34 The lsvdiskmember command

```
IBM_2145:ITSO_CLUSTER:superuser>lsvdiskmember 0
id
4
5
6
7
```

If you want to know more about these MDisks, you can run the **lsmdisk** command and provide the MDisk ID that is listed in the output of the **lsvdiskmember** command as a parameter.

6.7.21 Listing volumes that are defined in the storage pool

To list volumes that are defined in the specified storage pool, run the **lsvdisk -filtervalue** command. Example 6-35 shows how to use the **lsvdisk -filtervalue** command to list all volumes that are defined in the storage pool that is named Pool0.

Example 6-35 The lsvdisk -filtervalue command: Volumes in the pool

```
IBM_Storwize:ITSO:superuser>lsvdisk -filtervalue mdisk_grp_name=Pool0 -delim ,
id,name,IO_group_id,IO_group_name,status,mdisk_grp_id,mdisk_grp_name,capacity,type,FC_id,FC
_name,RC_id,RC_name,vdisk_UID,fc_map_count,copy_count,fast_write_state,se_copy_count,RC_cha
nge,compressed_copy_count,parent_mdisk_grp_id,parent_mdisk_grp_name,formatting,encrypt,volu
me_id,volume_name,function
```

```
0,A_MIRRORED_VOL_1,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F5800498000000000
00002,0,1,empty,0,no,0,0,Pool0,no,yes,0,A_MIRRORED_VOL_1,
2,VOLUME_WITH_MIRRORED_COPY,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F5800498
00000000000004,0,1,empty,0,no,0,0,Pool0,no,yes,2,VOLUME_WITH_MIRRORED_COPY,
3,THIN_PROVISION_VOL_1,0,io_grp0,online,0,Pool0,100.00GB,striped,,,,,6005076400F58004980000
0000000005,0,1,empty,1,no,0,0,Pool0,no,yes,3,THIN_PROVISION_VOL_1,
6,MIRRORED_SYNC_RATE_16,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F58004980000
0000000008,0,1,empty,0,no,0,0,Pool0,no,yes,6,MIRRORED_SYNC_RATE_16,
7,THIN_PROVISION_MIRRORED_VOL,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F58004
9800000000000009,0,1,empty,1,no,0,0,Pool0,no,yes,7,THIN_PROVISION_MIRRORED_VOL,
8,Tiger,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F580049800000000000010,0,1,e
mpty,0,no,0,0,Pool0,no,yes,8,Tiger,
9,UNCOMPRESSED_VOL,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F5800498000000000
00011,0,1,empty,0,no,1,0,Pool0,no,yes,9,UNCOMPRESSED_VOL,
12,vdisk0_restore,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F58004980000000000
000E,0,1,empty,0,no,0,0,Pool0,no,yes,12,vdisk0_restore,
13,vdisk0_restore1,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F5800498000000000
0000F,0,1,empty,0,no,0,0,Pool0,no,yes,13,vdisk0_restore1,
```

6.7.22 Listing storage pools in which a volume has its extents

To show to which storage pool a specific volume belongs, run the **lsvdisk** command, as shown in Example 6-36.

Example 6-36 The lsvdisk command: Storage pool ID and name

```
IBM_Storwize:ITSO:superuser>lsvdisk 0
id 0
name A_MIRRORED_VOL_1
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name Pool0
capacity 10.00GB
type striped
formatted yes
formatting no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076400F580049800000000000002
preferred_node_id 2
fast_write_state empty
cache readwrite
udid 4660
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
File system
mirror_write_priority latency
RC_change no
compressed_copy_count 0
access_IO_group_count 1
last_access_time
```

```
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
owner_type none
owner_id
owner_name
encrypt yes
volume_id 0
volume_name A_MIRRORED_VOL_1
function
throttle_id 1
throttle_name throttle1
IOPs_limit 233
bandwidth_limit_MB 122
volume_group_id
volume_group_name
cloud_backup_enabled no
cloud_account_id
cloud_account_name
backup_status off
last_backup_time
restore_status none
backup_grain_size
deduplicated_copy_count 0

copy_id 0
status online
sync yes
auto_delete no
primary yes
mdisk_grp_id 0
mdisk_grp_name Pool0
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status measured
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
encrypt yes
deduplicated_copy no
used_capacity_before_reduction0.00MB
```

For more information about these storage pools, run the **lsmdiskgrp** command, as described in Chapter 6, “Volumes” on page 419.

6.7.23 Tracing a volume from a host back to its physical disks

In some cases, you might need to verify which physical disks are used to store the data of a volume. This information is not directly available to the host, but it might be obtained by using a sequence of queries.

Before you trace a volume, you must unequivocally map a logical device that is seen by the host to a volume that is presented by the storage system. The best volume characteristic for this purpose is the volume ID. This ID is available to the operating system in the Vendor Specified Identifier field of page 0x80 or 0x83 (vital product data [VPD]), which the storage device sends in response to SCSI **INQUIRY** command from the host.

In practice, the ID can be obtained from the multipath driver in the operating system. After you know the volume ID, you can use it to identify the physical location of data.

Note: For sequential and image mode volumes, a volume copy is mapped to one MDisk. This configuration is usually not used for striped volumes unless the volume size is less than the extent size. Therefore, a single striped volume uses multiple MDisks in a typical case.

For example, on a Linux host that is running a native multipath driver, you can use the output of the command **multipath -ll** to find the volume ID, as shown in Example 6-37.

Example 6-37 Volume ID returned by the multipath -ll command

```
mpath1 (360050768018301BF2800000000000004)  IBM,2145
[size=2.0G][features=0][hwhandler=0]
\_ round-robin 0 [prio=200][ enabled]
\_ 4:0:0:1 sdd  8:48   [active][ready]
\_ 5:0:0:1 sdt  65:48  [active][ready]
\_ round-robin 0 [prio=40][ active]
\_ 4:0:2:1 sdak 66:64  [active][ready]
\_ 5:0:2:1 sdal 66:80  [active][ready]
```

Note: The volume ID that is shown in the output of `multipath -ll` is generated by the Linux `scsi_id`. For systems that provide the VPD by using page 0x83, such as IBM Storage Virtualize devices, the ID that is obtained from the VPD page is prefixed by the number 3, which is the Network Address Authority (NAA) type identifier.

Therefore, the volume NAA identifier (that is, the volume ID that is obtained by running the SCSI INQUIRY command) starts at the second displayed digit. In Example 6-37, the volume ID starts with digit 6.

After you know the volume ID, complete the following steps:

1. To list volumes that are mapped to the host, run the **lshostvdiskmap** command. Example 6-38 shows the list of volumes that are mapped to host Almaden.

Example 6-38 The lshostvdiskmap command

```
IBM_2145:ITSO_CLUSTER:superuser>lshostvdiskmap -delim , Almaden
id,name,SCSI_id,vdisk_id,vdisk_name,vdisk_UID
2,Almaden,0,26,volume_B,60050768018301BF2800000000000005
2,Almaden,1,27,volume_A,60050768018301BF2800000000000004
2,Almaden,2,28,volume_C,60050768018301BF2800000000000006
```

Look for the VDisk unique identifier (UID) that matches the volume UID that was identified and note the volume name (or ID) for a volume with this UID.

2. To list the MDisks that contain extents that are allocated to the specified volume, run the **lsvdiskmember *vdiskname*** command, as shown in Example 6-39.

Example 6-39 The lsvdiskmember command

```
IBM_2145:ITSO_CLUSTER:superuser>lsvdiskmember volume_A
id
0
1
2
3
4
10
11
13
15
16
17
```

3. For each of the MDisk IDs that were obtained in step 2, run the **lsmdisk mdiskID** command to discover the MDisk controller and LUN information. Example 6-40 shows the output for mdisk0. The output displays the back-end storage controller name and the controller LUN ID to help you to track back to a LUN within the disk subsystem.

Example 6-40 The lsmdisk command

```
IBM_2145:ITSO_CLUSTER:superuser>lsmdisk 0
id 0
name mdisk0
status online
mode managed
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
capacity 128.0GB
quorum_index 1
block_size 512
controller_name ITSO-DS3500
ctrl_type 4
ctrl_WWNN 20080080E51B09E8
controller_id 2
path_count 4
max_path_count 4
ctrl_LUN_# 0000000000000000
UID 60080e50001b0b62000007b04e731e4d00000000000000000000000000000000
preferred_WWPN 20580080E51B09E8
```

```
active_WWPN 2058O080E51BO9E8
fast_write_state empty
raid_status
raid_level
redundancy
strip_size
spare_goal
spare_protection_min
balanced
tier generic_hdd
```

You can identify the back-end storage that is presenting the LUN by using the value of the `controller_name` field that was returned for the MDisk.

7

Hosts

A host system can be defined as any networked computer server, virtual or physical, that provides workloads and services to the storage.

This chapter describes the processes that are required to attach a supported host system to IBM Storage Virtualize storage system through various supported interconnect protocols. The chapter also explains the following key concepts for host attachment:

- Host-clustering
- N_Port ID Virtualization (NPIV) support for a host-to-storage system communication
- Portsets feature introduced for multitenancy:
  - IP portsets feature is part of IBM Storage Virtualize since v8.4.2.
  - FC portsets feature is new to IBM Storage Virtualize since v8.5.0.

This chapter includes the following topics:

- 7.1, “Host attachment overview” on page 518
- 7.2, “Host objects overview” on page 519
- 7.3, “NVMe over Fibre Channel” on page 520
- 7.4, “NVMe over Remote Direct Memory Access” on page 521
- 7.5, “NVMe over TCP” on page 521
- 7.6, “N_Port ID Virtualization support” on page 522
- 7.7, “IP multi-tenancy” on page 530
- 7.8, “Fibre Channel portset” on page 537
- 7.9, “Hosts operations by using the GUI” on page 555
- 7.10, “Performing hosts operations by using the CLI” on page 597
- 7.11, “Host attachment examples” on page 608
- 7.12, “Container Storage Interface drivers” on page 632

IBM i considerations: For information about implementing the IBM Storage Virtualize family and its advanced functions with IBM i, refer to *Performance and Best Practices Guide for IBM Storage FlashSystem and IBM SAN Volume Controller: Updated for IBM Storage Virtualize Version 8.6*, SG24-8543, Appendix A - IBM i considerations.

© Copyright IBM Corp. 2024. All rights reserved.

7.1 Host attachment overview

IBM Storage Virtualize family v8.6.0 supports various open system host types (from IBM and non-IBM vendors).

These hosts can connect to the storage systems through any of the following protocols:

- Fibre Channel Protocol (FCP)
- Fibre Channel over Ethernet (FCoE)
- iSCSI
- SAS
- iSCSI Extensions for Remote Direct Memory Access (RDMA) (iSER)
- Non-Volatile Memory Express (NVMe) over Fibre Channel (FC-NVMe)
- NVMe over Remote Direct Memory Access (NVMe over RDMA)
- NVMe over Transmission Control Protocol (NVMe over TCP)

Hosts that connect to the storage through any type of storage area network (SAN) fabric switches must be zoned correctly, as described in Chapter 2, “Installation and configuration planning” on page 119.

Note: Specific host operating systems can be connected directly to the IBM Storage Virtualize storage system without the use of SAN switches. For more information, see the IBM System Storage Interoperation Center (SSIC).

To enable multiple access paths and enable correct volume presentation, a host system must have a multipathing driver installed.

In addition, the multipathing driver serves the following purposes:

- Protection from:
  - Fabric path failures, including port failures on IBM Storage Virtualize system nodes.
  - A host bus adapter (HBA) failure (if two HBAs are used).
  - Failures if the host is connected through two HBAs across two separate fabrics.
- Load balancing across the host HBAs.

For more information about the native operating system multipath drivers supported for IBM Storage Virtualize systems, see the SSIC.

For more information about how to attach specific supported host operating systems to the storage systems, see this IBM Documentation web page.

Note: If a specific host operating system is not mentioned in the SSIC, contact your IBM representative or IBM Business Partner to submit a special request for support.

7.2 Host objects overview

Before a host can access the storage capacity, it must first be presented to the storage system as a *host object*.

A host object is configured by using the GUI or command line interface (CLI) and must contain the necessary credentials for host-to-storage communications. After this process is completed, storage capacity can be mapped to that host in the form of a volume.

IBM Storage Virtualize v8.6.0 supports configuring the following host objects:

- Host
- Host cluster

A host cluster object groups clustered servers and treats them as a single entity. This configuration allows multiple hosts to access the same volumes through one shared mapping.

Note: Any volume that is mapped to a host cluster is automatically assigned to all of the members in that cluster with the same SCSI ID.

A typical use case for a host cluster object is to group multiple clustered servers with a common operating system, such as IBM PowerHA® and Microsoft Cluster Server, and enable them to have shared access to common volumes.

The following commands are used for host and host cluster objects:

- The following commands provide information about defined hosts and host clusters start:
  - **lshostcluster**
  - **lshostclustermember**
  - **lshostclustervolumemap**
  - **lshost**
  - **lshostiogrp**
  - **lshostiplogin**
  - **lsiscsiauth**
- The following commands define or configure a host object on a storage system:
  - **mkhost.** Defines an individual host by creating a logical host object.
  - **mkhostcluster**. Defines a cluster host object, and enables the addition of specific hosts to the cluster object.
- The following commands remove or delete defined host objects from a storage system:
  - **rmhostclustermember**
  - **rmhostcluster**
  - **rmvolumehostclustermap**
  - **rmhost**
  - **rmhostiogrp**
  - **rmhostport**
- The following commands change defined host objects:
  - **chhostcluster.** Changes the name, type, or site of a host cluster object that is part of a host cluster.
  - **chhost**. Changes the name or type.

- The following commands add a host to a host cluster, to an I/O group, or add a port to a host object:
  - **addhostclustermember**
  - **addhostiogrp**
  - **addhostport**

For more information about each command, see Host commands. The instructions to perform basic tasks on hosts and host clusters are provided in 7.10, “Performing hosts operations by using the CLI” on page 597.

7.3 NVMe over Fibre Channel

IBM Storage Virtualize systems running firmware v8.6.0 can be attached to an NVMe host through NVMe-oF, which uses FCP (FC-NVMe) as its underlying transport protocol. This NVMe-oF gives the target the ability to transfer the data directly from host memory. In addition to FC-NVMe, a host can send commands and data together (first burst), which improves efficiency and provides better performance at distances.

It is now possible to run SCSI and NVMe in parallel. However, for optimal planning, see the links to the configuration limits and restrictions for each IBM Storage Virtualize product family as listed in Table 7-1.

Table 7-1 Configuration and restriction for Storage Virtualize product family

Product	More information
IBM SAN Volume Controller	https://www.ibm.com/support/pages/node/6966894
IBM FlashSystem 9500	https://www.ibm.com/support/pages/node/6966914
IBM FlashSystem 9100 / 9200	https://www.ibm.com/support/pages/node/6966910
IBM FlashSystem 5x05 / 5200	https://www.ibm.com/support/pages/node/6966908

A volume can be mapped only to a host through one protocol. IBM FlashCopy, volume mirroring, Remote Copy (RC), and Data Reduction Pools (DRPs) are all supported by NVMe-oF. IBM Storage Virtualize v8.6.0 also supports HyperSwap for NVMe-oF attached hosts.

Note: IBM Storage Virtualize version v8.6.0 uses Asymmetric Namespace Access (ANA), which enables HyperSwap and Nondisruptive Volume Move (NDVM) support for FC-NVMe hosts. The following features are available for FC-NVMe attached hosts:

- Sites can be defined to facilitate awareness of HyperSwap volume site properties.
- It is possible to map HyperSwap volumes by using multiple I/O groups on the same and different sites.
- Hosts can use I/O through a nonoptimized path, even if the primary site is available.
- The ability to fail over to the secondary site if the primary site is down.

For more information about NVMe, see *IBM Storage and the NVM Express Revolution*, REDP-5437.

7.4 NVMe over Remote Direct Memory Access

IBM Storage Virtualize v8.6.0 can be attached to an NVMe host through NVMe over Remote Direct Memory Access (RDMA). NVMe over RDMA uses RDMA over Converged Ethernet (RoCE) v2 as the transport protocol. RoCE v2 is based on user datagram protocol (UDP).

RDMA is a host-offload, host-bypass technology that allows an application (including storage) to make data transfers directly to and from another application's memory space. The RDMA-capable Ethernet NICs (RNICs), and not the host, manage reliable data transfers between source and destination.

RNICs can use RDMA over Ethernet by way of RoCE encapsulation. RoCE wraps standard InfiniBand payloads with Ethernet or IP over Ethernet frames, and is sometimes called *InfiniBand over Ethernet*. The following main RoCE encapsulation types are available:

- RoCE V1

  This type uses dedicated Ethernet Protocol Encapsulation (Ethernet packets between source and destination MAC addresses by using Ethertype 0x8915).

- RoCE V2:
  - This type uses dedicated UDP over Ethernet Protocol Encapsulation (IP UDP packets by using port 4791 between source and destination IPs; UDP packets are sent over Ethernet by using source and destination MAC addresses)
  - This type is *not* compatible with other Ethernet options, such as RoCE v1.

Note: Unlike RoCE V1, RoCE V2 is routable.

7.5 NVMe over TCP

IBM Storage Virtualize v8.6.0 can be attached to an NVMe host through NVMe over Transmission Control Protocol (TCP). NVMe over TCP is a transport that allows NVMe performance without any constraint to the data center infrastructure.

Demonstration video: For more information, see IBM Storage Spectrum Virtualize 8.6 GUI, including Volume Group Snapshots (with Safeguarded Copy) and Policy Based Replication.

NVMe over TCP needs more CPU resources than protocols using RDMA. Each NVMe/TCP port on FlashSystem supports multiple IPs and multiple VLANs. Generally, NVMe-TCP runs on all switches and is routable.

For operating system support and multipathing, see IBM System Storage Interoperation Center (SSIC)

For more information about using NVMe hosts with IBM FlashSystem, see NVMe over RDMA and NVMe over TCP host attachments.

Note: NVMe over TCP is supported on FlashSystem platforms that are installed with Mellanox CX-4 or CX-6 adapters.

7.6 N_Port ID Virtualization support

IBM FlashSystem storage systems use a pair of distinct control modules that are known as *node canisters* that share active/active access to all volumes within the same I/O group. Each node canister has its own FC worldwide node name (WWNN). Each node canister's network adapter or HBA ports have a set of worldwide port names (WWPNs) that are presented to the fabric.

These ports are used for following purposes:

- Internode communication between storage system nodes
- Back-end controllers communication for external storage virtualization (available only for the IBM FlashSystem storage systems that support this feature)
- Host communications

If a node canister fails or is removed, its path to the host goes offline. As a result, the host's native operating system multipathing software is required to fail over to the set of WWPNs for the node or nodes still online.

As such, it is important to ensure that the multipathing driver is correctly implemented and configured.

When N_Port ID Virtualization (NPIV) mode is enabled, target ports (also known as *host attach ports*) that are dedicated *only* to host communications become available. This configuration efficiently separates the internode communication from the host I/O.

Host attachment ports can be moved between nodes within the same I/O group. The operation is transparent to the host and is beneficial if, for example, a node in an I/O group went offline. Moving the host attach ports to the online node in the same I/O group masks the path failures without requiring the multipathing driver to perform any path recovery.

When NPIV is enabled on the storage system, each physical WWPN reports up to four virtual WWPNs, as listed in Table 7-2.

Table 7-2 IBM Storage Virtualize NPIV ports

NPIV port	Description
Primary port	The WWPN that communicates with back-end storage. It can be used for local and remote node-to-node traffic.
Primary SCSI host attach port	The WWPN that communicates with hosts. It is a target port only. Because it also is the primary port; it is based on this local node's WWNN.
Failover SCSI host attach port	A standby WWPN that communicates with hosts that is brought online only if the partner node within the I/O group goes offline. This WWPN is the same as the primary host attach WWPN of the partner node.
Primary NVMe host attach port	The WWPN that communicates with hosts. It is a target port only. This WWPN is the primary port; therefore, it is based on this local node's WWNN.
Failover NVMe host attach port	A standby WWPN that communicates with hosts that is brought online only if the partner node within the I/O group goes offline. This WWPN is the same as the primary host attach WWPN of the partner node.

Figure 7-1 shows the five WWPNs that are associated with a port when NPIV is enabled.

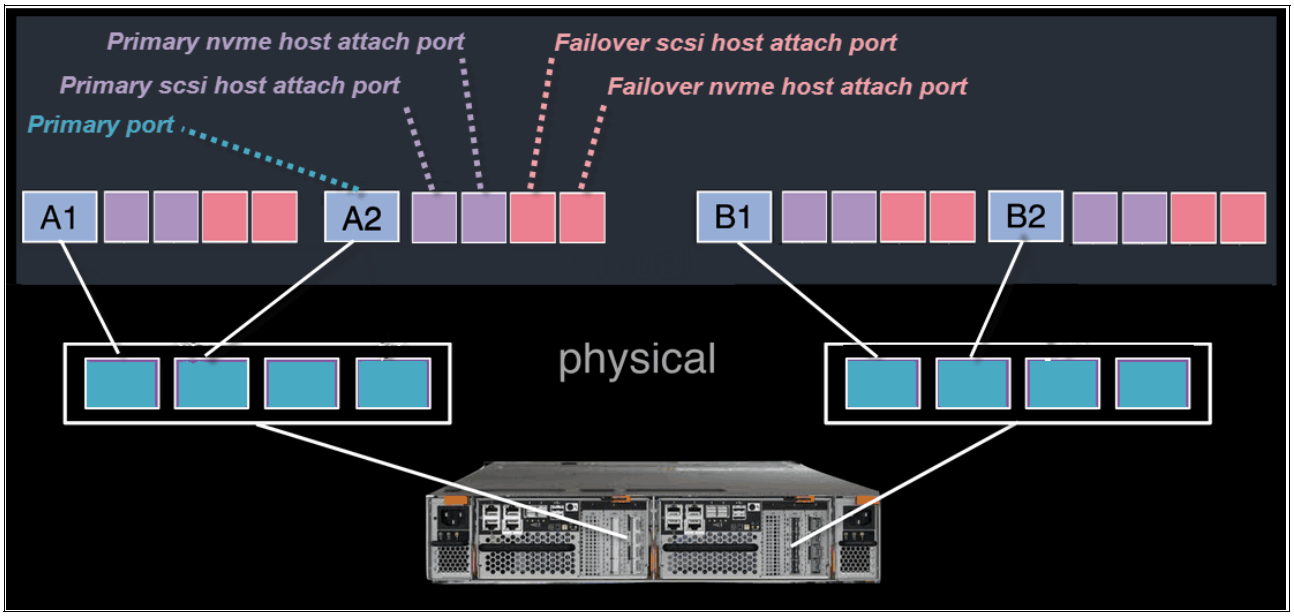


Figure 7-1 Allocation of NPIV virtual WWPN ports per physical port (failover host attach port is not active)

Figure 7-2 shows what occurs when the partner node canister fails. The failover host attach ports on the remaining node canister become active and take on the WWPN of the failed node's primary host attach port.

Note: Figure 7-2 shows only two ports per node canister in detail. However, the same situation applies for all physical ports, including NVMe ports because they use the same NPIV structure, but with the NVMe topology instead of SCSI.

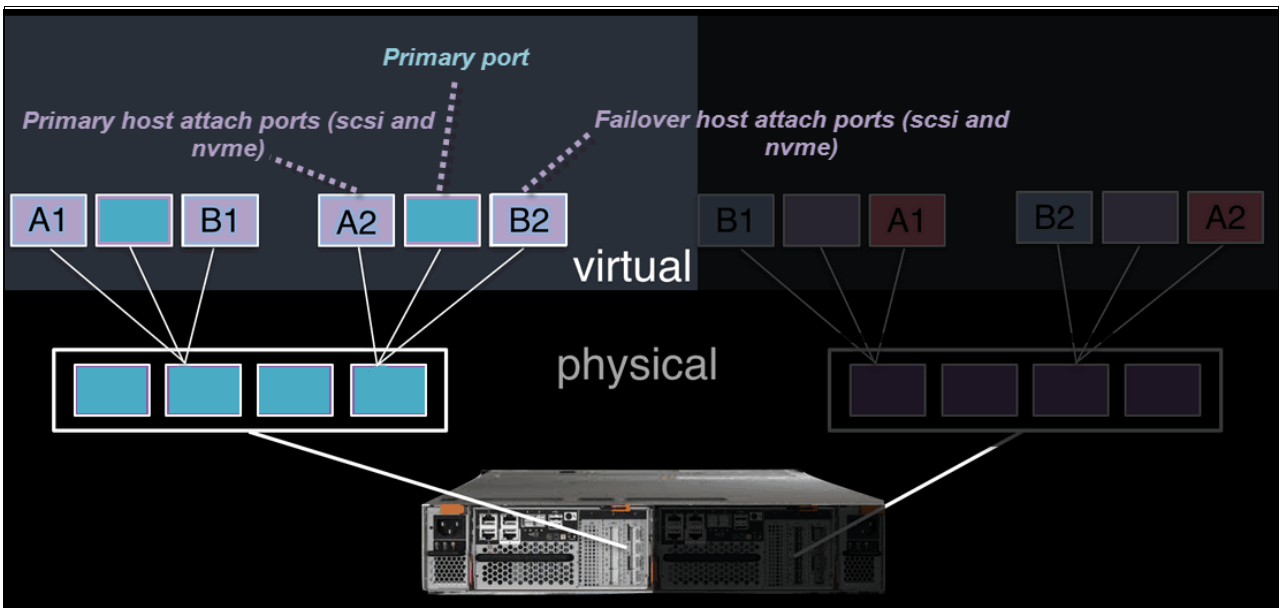


Figure 7-2 Allocation of NPIV virtual WWPN ports per physical port after a node failure

This process occurs automatically when NPIV is enabled on the storage systems. This failover can occur only between the two node canisters in the same I/O group.

The following NPIV mode states are available:

- Disabled: Virtualized host attach ports (NPIV target ports) cannot be used for I/O. Only physical ports are used for I/O. Volumes are presented to the hosts through physical ports only.
- Transitional: Virtualized host attach ports (NPIV target ports) and physical ports are used for I/O. Volumes are presented to the hosts through physical and NPIV target ports.
- Enabled: Only virtualized host attach ports (NPIV target ports) can service host I/O. Volumes are presented to the hosts only through NPIV target ports.

A transitional state enables migration of hosts from previous non-NPIV enabled systems to enabled NPIV systems, which enables a transition state as hosts are rezoned to the *primary* host attach WWPNs.

The process to enable NPIV on a new system is slightly different than on an existing system. For more information, see Enabling NPIV on a new system.

Note: NPIV is supported for FC-based communication only. It is *not* supported for the FCoE or iSCSI protocols.

7.6.1 NPIV prerequisites

Before enabling NPIV, the storage system must meet the following prerequisites:

- The system is running IBM Storage Virtualize version 7.7 or later.
- Both nodes within an I/O group have identical hardware.
- The FC switches to which the system ports are attached support NPIV and have this feature enabled.
- Node connectivity is done according to the information the zoning requirements for N_Port ID virtualization. Both nodes in one I/O group should have their equivalent ports connected to their equivalent fabrics (switch). For example, port 1 of node1 should be on the same fabric as port 1 of node2. For more information, see Zoning requirements for N_Port ID virtualization.

7.6.2 Verifying the NPIV mode state for a new system installation

NPIV is enabled by default on all current IBM FlashSystem storage system running IBM Storage Virtualize 8.4.2 and later. If you are unsure whether NPIV is enabled, complete the following steps to verify that NPIV is enabled:

1. Run the **lsiogrp** command to list the I/O groups that are present in the system, as shown in Example 7-1.

Example 7-1 Listing the I/O groups in the system

```
IBM_IBM FlashSystem:FS9500:superuser>lsiogrp
id name            node_count vdisk_count host_count site_id site_name
0  io_grp0         2          10          0
1  io_grp1         0          0           0
2  io_grp2         0          0           0
3  io_grp3         0          0           0
4  recovery_io_grp 0          0           0
```

Example 7-1 shows one full I/O group with ID 0, two nodes in it, and 10 virtual disks (VDisks). The other I/O groups are empty.

2. Run the **lsiogrp** *<id>* | **grep fctargetportmode** command for the specific I/O group ID to display the **fctargetportmode** setting. If this setting is enabled, as shown in Example 7-2, NPIV host target port mode is enabled. If NPIV mode is disabled, the **fctargetportmode** parameter reports as disabled.

Example 7-2 Checking the NPIV mode by viewing the fctargetportmode field

```
IBM_IBM FlashSystem:FS9500:superuser>lsiogrp 0|grep fctargetportmode
fctargetportmode enabled
```

3. Run the **lstargetportfc** command for a list of the available virtual WWPNs (see Example 7-3). The host_io_permitted and virtualized columns display a status of yes, which indicates that those WWPNs belong to a primary host attach port and should be used when zoning the hosts to the system.

Example 7-3 Listing the virtual WWPNs

```
IBM_IBM FlashSystem:FS9500:superuser>lstargetportfc
id WWPN             WWNN             port_id .. host_io_permitted virtualized protocol
1  500507681011024F 500507681000024F 1          no                no          scsi
2  500507681015024F 500507681000024F 1          yes               yes         scsi
3  500507681019024F 500507681000024F 1          yes               yes         nvme
...
10 500507681014024F 500507681000024F 4          no                no          scsi
11 500507681018024F 500507681000024F 4          yes               yes         scsi
12 50050768101C024F 500507681000024F 4          yes               yes         nvme
```

4. Configure the zones for host-to-storage communication by using the primary host attach ports (virtual WWPNs) as shown in the output of the command (see Example 7-3 on page 525 in which the virtualized ports are marked in bold).

5. If the status of **fctargetportmode** is disabled and this is a new installation, run the **chiogrp** command to set the NPIV mode to the transitional state and then to enabled (see Example 7-4).

Example 7-4 Changing the NPIV mode to enabled

```
IBM_IBM FlashSystem:FS9500:superuser>chiogrp -fctargetportmode transitional 0
IBM_IBM FlashSystem:FS9500:superuser>chiogrp -fctargetportmode enabled 0
```

7.6.3 Enabling NPIV on a system

If IBM FlashSystem storage systems IBM Storage Virtualize Version 7.7.1 or earlier is upgraded to the latest version, the NPIV feature must be enabled manually because it requires changes to host-to-storage zoning,

To enable NPIV mode on a storage system, complete the following steps:

1. Review the SAN fabric layout and zoning rules because NPIV usage has strict requirements. Ensure that equivalent ports are on the same fabric and in the same zone.
2. Check the path count between the hosts and the IBM FlashSystem storage system to ensure that the number of paths is half of the usual supported maximum.

3. Run the **lstargetportfc** command to discover the `primary host attach` WWPNs (virtual WWPNs), as shown in bold in Example 7-5. Those virtualized ports are not enabled for host I/O communication yet (see the `host_io_permitted` column).

Example 7-5 Running the lstargetportfc command to get the primary host WWPNs (virtual WWPNs)

```
IBM_IBM FlashSystem:FS9500:superuser>lstargetportfc
id WWPN             WWNN             port_id owning_node_id current_node_id nportid host_io_permitted virtualized
protocol
1  500507680140A288 500507680100A288 1       1             1               010A00  yes               no          scsi
2  500507680142A288 500507680100A288 1       1                             000000  no                yes         scsi
3  500507680144A288 500507680100A288 1       1                             000000  no                yes         nvme
4  500507680130A288 500507680100A288 2       1             1               010400  yes               no          scsi
5  500507680132A288 500507680100A288 2       1                             000000  no                yes         scsi
6  500507680134A288 500507680100A288 2       1                             000000  no                yes         nvme
7  500507680110A288 500507680100A288 3       1             1               010500  yes               no          scsi
8  500507680112A288 500507680100A288 3       1                             000000  no                yes         scsi
9  500507680114A288 500507680100A288 3       1                             000000  no                yes         nvme
10 500507680120A288 500507680100A288 4       1             1               010A00  yes               no          scsi
11 500507680122A288 500507680100A288 4       1                             000000  no                yes         scsi
12 500507680124A288 500507680100A288 4       1                             000000  no                yes         nvme
...
58 500507680C140009 500507680C000009 4       2             2               010900  yes               no          scsi
59 500507680C180009 500507680C000009 4       2                             000000  no                yes         scsi
60 500507680C1C0009 500507680C000009 4       2                             000000  no                yes         nvme
```

4. Transitional mode for NPIV (Example 7-6) must be enabled to enable host I/O communication and still maintain access to hosts that use hardware-defined ports (not in bold in Example 7-5 on page 526).

Example 7-6 Enabling transitional mode for NPIV

```
IBM_IBM FlashSystem:FS9500:superuser>chiogrp -fctargetportmode transitional 0
IBM_IBM FlashSystem:FS9500:superuser>lsiogrp 0 |grep fctargetportmode
fctargetportmode transitional
```

Alternatively, to activate NPIV in transitional mode by using the GUI, select **Settings** → **System** → **I/O Groups** (see Figure 7-3).

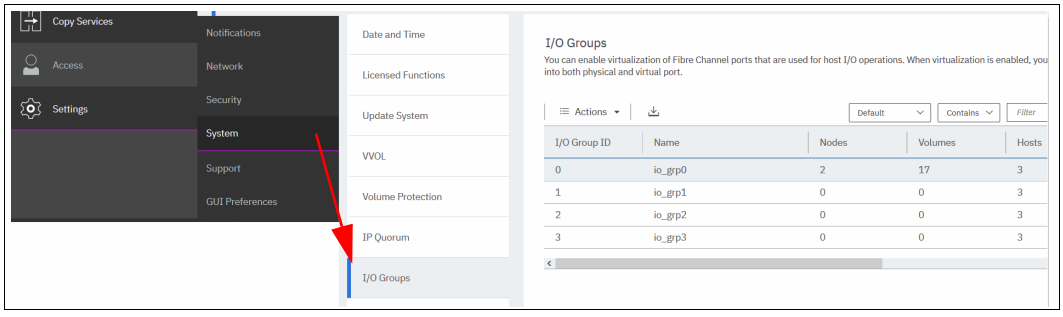


Figure 7-3 I/O Groups menu

then check the current setting by viewing the NPIV column, which shows `disabled` if NPIV is not activated. Select the I/O group to be enabled and select **Actions** → **Change NPIV Settings**, as shown in Figure 7-4.

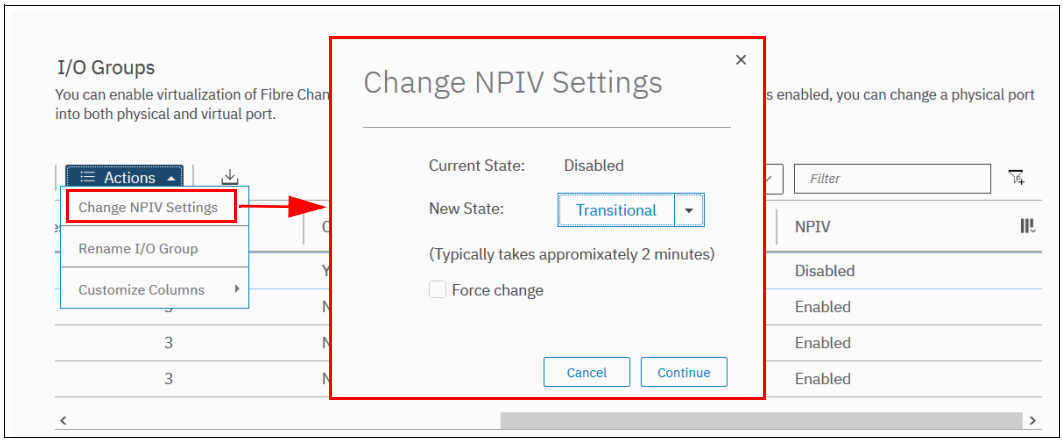


Figure 7-4 Change NPIV Settings windows

Select **Continue** to complete the process and enable NPIV in Transitional Mode.

5. Ensure that host traffic is now allowed on the primary host attach WWPNs (virtual WWPNs), as shown in bold in Example 7-7.

Example 7-7 Host attach WWPNs (virtual WWPNs) permitting host traffic

```
IBM_IBM FlashSystem:FS9500:superuser>lstargetportfc
id WWPN             WWNN             port_id owning_node_id current_node_id nportid host_io_permitted virtualized
protocol
1  500507680140A288 500507680100A288 1       1             1               010A00  yes              no          scsi
2  500507680142A288 500507680100A288 1       1             1               010A02  yes              yes         scsi
3  500507680144A288 500507680100A288 1       1             1               010A01  yes              yes         nvme
4  500507680130A288 500507680100A288 2       1             1               010400  yes              no          scsi
5  500507680132A288 500507680100A288 2       1             1               010401  yes              yes         scsi
6  500507680134A288 500507680100A288 2       1             1               010402  yes              yes         nvme
7  500507680110A288 500507680100A288 3       1             1               010500  yes              no          scsi
8  500507680112A288 500507680100A288 3       1             1               010501  yes              yes         scsi
9  500507680114A288 500507680100A288 3       1             1               010502  yes              yes         nvme
...
58 500507680C140009 500507680C000009 4       2             2               010900  yes              no          scsi
59 500507680C180009 500507680C000009 4       2             2               010901  yes              yes         scsi
60 500507680C1C0009 500507680C000009 4       2             2               010902  yes              yes         nvme
```

6. Add the primary host attach ports (virtual WWPNs) to the host zones, but do not remove the IBM FlashSystem WWPNs. Example 7-8 shows a host zone added to the primary port WWPNs of the IBM FlashSystem nodes.

Example 7-8 Established host zone

```
zone: WINDOWS_HOST_01_IBM_FS9500
                10:00:00:05:1e:0f:81:cc
                50:05:07:68:01:40:A2:88
                50:05:07:68:0C:11:00:09
```

Example 7-9 shows that the primary host attach ports (virtual WWPNs) were added to the example host zone without disrupting its availability.

Example 7-9 Transitional host zone (added host attach ports are in bold)

```
zone: WINDOWS_HOST_01_IBM_FS9500
                10:00:00:05:1e:0f:81:cc
                50:05:07:68:01:40:A2:88
                50:05:07:68:0C:11:00:09
```

```
                    50:05:07:68:01:42:A2:88
                    50:05:07:68:0C:15:00:09
```

7. After activating transitional zoning in the fabrics, validate that the host uses the new NPIV ports for host I/O. Example 7-10 shows the pathing for the host before and after the addition of the new host attach ports through the old IBM Subsystem Device Driver (SDD) Device Specific Module (SDDDSM) multipathing driver. The select count increases on the new paths and stops on the old paths.

Note: SDDDSM, which is a multipathing driver, is not recommended or supported. The recommended multipathing driver for the Microsoft Windows platform is Microsoft Device Specific Module (MSDSM).

Example 7-10 Host device pathing: Before and after

```
C:\Program Files\IBM\SDDDSM>datapath query device

Total Devices : 1

DEV#:   0 DEVICE NAME: Disk3 Part0  TYPE: 2145       POLICY: OPTIMIZED
SERIAL: 60050764008680083800000000000000        LUN SIZE: 20.0GB
============================================================================
Path#            Adapter/Hard Disk        State  Mode       Select     Errors
    0 *   Scsi Port2 Bus0/Disk1 Part0     OPEN   NORMAL    3991778          0
    1 *   Scsi Port2 Bus0/Disk1 Part0     OPEN   NORMAL     416214          0
    2 *   Scsi Port3 Bus0/Disk1 Part0     OPEN   NORMAL      22255          0
    3 *   Scsi Port3 Bus0/Disk1 Part0     OPEN   NORMAL     372785          0
C:\Program Files\IBM\SDDDSM>datapath query device

Total Devices : 1

DEV#:   0  DEVICE NAME: Disk3 Part0  TYPE: 2145       POLICY: OPTIMIZED
SERIAL: 60050764008680083800000000000000        LUN SIZE: 20.0GB
============================================================================
Path#            Adapter/Hard Disk        State  Mode       Select     Errors
    0 *   Scsi Port2 Bus0/Disk1 Part0     OPEN   NORMAL    3991778          2
    1 *   Scsi Port2 Bus0/Disk1 Part0     OPEN   NORMAL     416214          1
    2 *   Scsi Port3 Bus0/Disk1 Part0     OPEN   NORMAL      22255          0
    3 *   Scsi Port3 Bus0/Disk1 Part0     OPEN   NORMAL     372785          2
    4 *   Scsi Port2 Bus0/Disk1 Part0     OPEN   NORMAL      22219          0
    5     Scsi Port2 Bus0/Disk1 Part0     OPEN   NORMAL      95109          0
    6 *   Scsi Port3 Bus0/Disk1 Part0     OPEN   NORMAL          2          0
    7     Scsi Port3 Bus0/Disk1 Part0     OPEN   NORMAL      91838          0
```

Note: Consider the following points:

- Verify that the correct NPIV ports are visible by running the **`lsfabric -host host_id_or_name`** command. If I/O activity is occurring, each host has at least one line in the command output that corresponds to a host port and shows `active` in the activity field:
  - Hosts where no I/O occurred in the past 5 minutes show `inactive` for any login.
  - Hosts that do not adhere to preferred paths might still be processing I/O to primary ports.
- Depending on the host operating system, it can be necessary to rescan the storage and enable the hosts to discover the new paths that are now available.

8. After all hosts are rezoned and the pathing is validated, change the system NPIV to `enabled` mode by running the command that is shown in Example 7-11.

Example 7-11 Enabling the NPIV

```
IBM_IBM FlashSystem:FS9500:superuser>chiogrp -fctargetportmode enabled 0
```

Alternatively, to enable NPIV by using the GUI, go to the I/O Groups window, as shown in Step 4, select the I/O group and then select **Actions** → **Change NPIV Settings** (Figure 7-5).

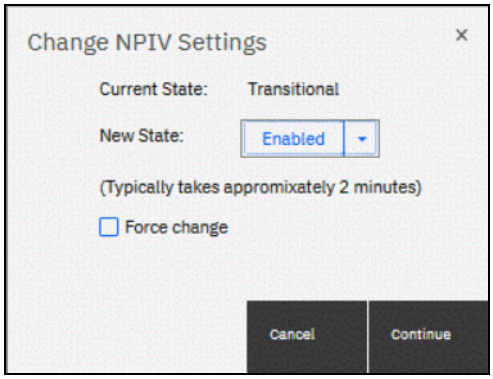


Figure 7-5 Change NPIV Settings window: Selecting Enabled option

Select **Continue** to complete the process and enable NPIV for the I/O Groups.

To complete the NPIV implementation, modify the host zones to remove the old primary attach port WWPNs. Example 7-12 shows the final zone with the host HBA and the IBM FlashSystem virtual WWPNs.

Example 7-12 Final host zone

```
zone: WINDOWS_HOST_01_IBM_FS9100
                10:00:00:05:1e:0f:81:cc
                50:05:07:68:01:42:A2:88
                50:05:07:68:0C:15:00:09
```

Note: If some hosts are still configured to use the physical ports on the IBM SAN Volume Controller system, the system prevents changing **fctargetportmode** from `transitional` to `enabled`, and shows the following error:

```
CMMVC8019E Task could interrupt I/O and force flag not set.
```

7.7 IP multi-tenancy

For IBM Storage Virtualize version 8.4.2 and later, administrators have the ability to define multiple IP addresses. These are defined per port for use by an Ethernet host, including all servers that use iSCSI and iSER host attachment protocols.

This feature includes the following use cases:

- Allowing multiple host tenants to share a single IBM FlashSystem storage system.
- Load balancing and performance tuning by optimizing portsets.
- Scaling out IP addresses and VLANs.
- Implementing a cloud strategy.

This section describes the following operations:

- Creating a portset.
- Assigning the IP addresses to a portset.
- Mapping a portset to a host object.

A *portset* is a grouping of logical addresses that is associated with the specific traffic types. The IBM FlashSystem storage systems support portsets for host attachment, backend storage connectivity, and IP replication traffic.

7.7.1 Limitations and restrictions

Each physical Ethernet port can have a maximum 64 IP addresses for each IP on a unique portset. However, for each port, IP address can be shared between multiple unique portsets for different functions.

Note: Each port can be bound only to a single IP address per portset for a specific Ethernet function, such as host attachment (iSCSI or iSER), backend storage connectivity (iSCSI only), and IP replication.

For cloud environments, each Ethernet port support two IP addresses and VLANs per port for multiple clients that share storage resources on the system.

The specific limitations for configuring IP addresses for host attachment, iSCSI storage virtualization, and IP replication are listed in Table 7-3.

Table 7-3 Configuration limitations

Limit	Description
Maximum number of portset objects	72 per system
Maximum number of IP address objects (includes shared IP address objects)	2048 per system

Limit	Description
Maximum number of IP addresses objects per port	64 (either IPv4 or IPv6)
Maximum number of routable IP addresses objects per port	1 (either IPv4 or IPv6)[a]
Maximum number of IP address objects per node	256 per node
Maximum number of IP addresses objects per node per portset	► Host portsets: Four per node per portset ► Replication portsets: One per node per portset ► Storage portset (Portset 3): Limit to the number of Ethernet ports on the node
Maximum number of iSNS servers	1 IPv4 and 1 IPv6 are supported for portset0 IP addresses only

a The routable IP addresses are used by IP replication function. However, if routable IP addresses are required for host attach and storage, each Ethernet port can be assigned with one routable IP address with a gateway. In most cases, host attach IP addresses can be separated by VLANs or subnets or a combination of both for multi-tenant scenarios.

In addition, if the system uses Emulex or Mellanox HBAs, the limits that are listed in Table 7-4 apply to multiple IP addresses and VLANs.

Table 7-4 Emulex or Mellanox HBAs

Type of HBA	Limit
Emulex	► Maximum of 3 unique VLANs per port ► Maximum of 32 IP addresses per port[a]
Mellanox	For iSER[b] connections ► Maximum of 31 VLANs per port ► Maximum of 31 IP addresses per port with VLAN ► Maximum of 64 IP addresses per port without VLAN For iSCSI only connections ► Maximum of 64 VLANs per ports ► Maximum of 64 IP addresses per ports

[a] If a VLAN is not configured on these ports, the limit is still 32 IP addresses per port. You cannot add IP addresses or VLANs after this limit is reached.

[b] If you use iSER for clustering (node-to-node) connections, you must lower the number of IP addresses per port based on the number that you use for clustering.

Note: iSER host attachment is not supported on IBM FlashSystem 9500, IBM FlashSystem 7300 and IBM SAN Volume Controller SV3. However, it is supported on other IBM Storage Virtualize products.

7.7.2 Prerequisites

In this section, we discuss specific prerequisites that must be met before a portset is created and assigned to an IP address, host object, or IP partnership.

Portset requirements

In general, portsets have the following requirements:

- Each IP address in a portset must be configured on a separate Ethernet port.
- The same ports can share IP addresses across different portsets that use the same IP address for host, storage, and remote-copy traffic. All shared IP addresses must use the same port and have the same VLAN, gateway, and prefix. When IP addresses are shared among multiple portsets, the system creates a logical copy of the IP address and its attributes rather than a new IP address.
- Portsets owned by different ownership groups can share an IP address.
- A port can have up to 64 unique or shared IP addresses. All 64 IP addresses must be IPv4 or IPv6, or a mix of IPv4 and IPv6.
- Each port can be configured with only one unique routable IP address (gateway specified). The routable IP can be shared among multiple portsets.
- Portset 0 is automatically configured when the system is created or updated. It cannot be deleted and serves as the default portset for any IP addresses and host objects that are configured without a specified portset. After an update, all configured host objects are automatically mapped to portset 0.

Host portset requirements

In addition to portset 0, more portsets for host traffic can be created if they adhere to the following requirements:

- Portsets can have a maximum of four IP addresses per node.
- A single portset can contain IPv4, IPv6, or a mix of IPv4 and IPv6 addresses.
- The host must be mapped to a portset that contains at least one IP address from any of nodes on that system.

Replication portset requirements

Replication portsets for IP replication feature the following requirements:

- Replication portset can have maximum of one IP address per node.
- All IP addresses in replication portsets must be IPv4 or IPv6 addresses. IP protocol versions on replication portsets *cannot* be mixed.
- Each IP partnership can be mapped to two portsets, one for each link between systems. For a partnership with a single link, a single portset can be defined in the `portset1` field on the Create Partnership page in the GUI.

  Partnerships with a single link also can be specified for one portset with the **`-link1`** attribute in the **`mkippartnership`** command.

  For a partnership with dual links, a second portset must be defined in the `portset2` field in the GUI. Use the **`-link2`** attribute to specify the second portset for a dual link configuration.
- Portsets replace the requirement for creating remote copy groups for IP partnerships. During software updates, any addresses that are assigned to a remote copy group with an IP partnership are automatically moved to a corresponding portset.

  For example, if remote copy group 1 is defined on the system before the update, IP addresses from that remote copy group are mapped to portset 1 after the update. Similarly, IP address in remote copy group 2 is mapped to portset 2.

Storage portset requirements

The system supports maximum of one portset of storage type `portset3`. The following requirements are specific to storage portsets:

- The maximum of number IP address for a storage portset is equal to the number of Ethernet ports on the node.
- A single portset can contain IPv4, IPv6 IP, or mix of IPv4 and IPv6 addresses.

Ownership group portsets requirements

Portsets can be assigned to ownership groups to restrict access to specific set of users. As a result, restricted users have the following characteristics:

- Can create a portset and IP addresses and assign the hosts to the portset.
- Cannot view portsets that are assigned to a different ownership group.
- Cannot own replication or storage type portset.

7.7.3 Configuring the portset

You can configure the portset by using the GUI or CLI.

Configuring a portset from the GUI

To configure the portsets from the GUI, complete the following steps:

1. Select **Settings** → **Network** → **Portsets** (see Figure 7-6).

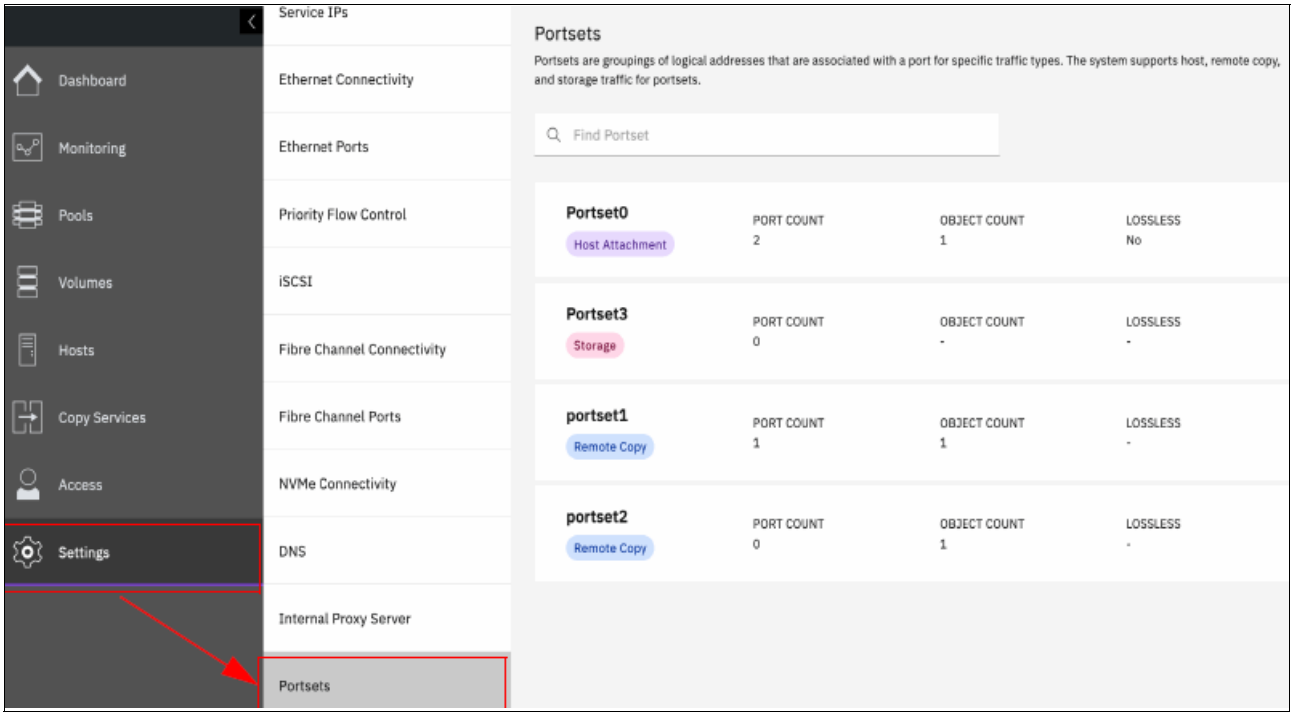


Figure 7-6 Selecting the portset option

2. Select **Create Portsets** (See Figure 7-7).

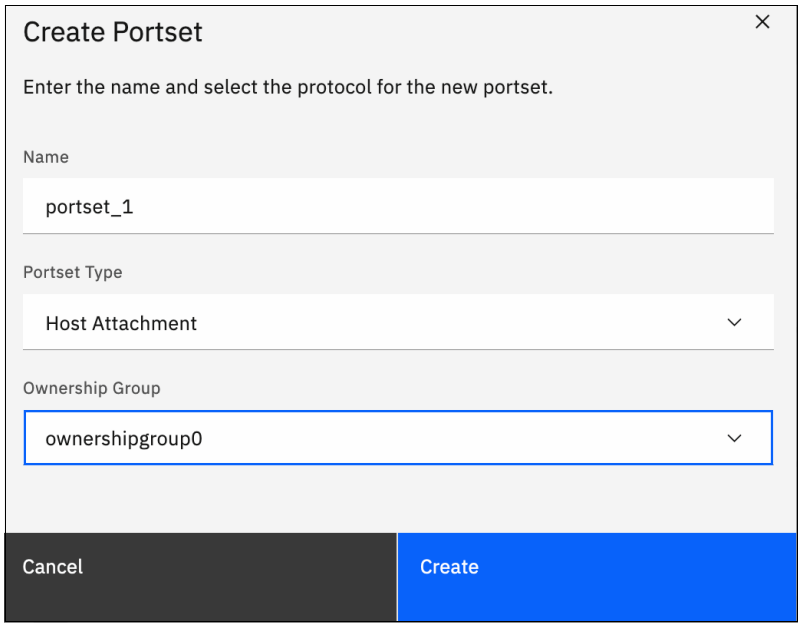


Figure 7-7 Creating a portset

3. On the Create Portset page, enter a name of the portset, and select the one of the following options for the portset type.
   - **Host Attachment**: Indicates that the IP addresses that are added to the portset are used for host attachment only.
   - **Remote Copy**: Indicates that the IP addresses that are added to the portset are used for IP partnerships only.
4. Select the ownership group for the portset. An *ownership group* defines a subset of users and objects within the system. Restricted users are those users who are defined to a specific ownership group and can view or manage only specific resources that are assigned to that ownership group.

   Unrestricted users are not defined to an ownership group and can manage any objects on the system based on their administration role.
5. Click **Create**.

Configuring a portset from the CLI

To create a portset from the CLI, enter the following command:

```
mkportset -name portset_name -type portset_type -ownershipgroup owner_id |
owner_name -porttype fc | ethernet
```

Where *portset_name* is the name of the portset and *portset_type* is `host` or `replication`. The value *owner_id* | *owner_name* indicates the ID or name of the ownership group to which the portset belongs. The *porttype* can be **fc** for FC ports or Ethernet. All are optional values and default host type portset are created (see Figure 7-8).

```
IBM_FlashSystem:FS9110:superuser>mkportset -name portset_2 -type host -ownershipgroup ownershipgroup0
Portset, id [6], successfully created
IBM_FlashSystem:FS9110:superuser>
```

Figure 7-8 Creating portset_2 for node1 in ownershipgroup0

7.7.4 Assigning an IP address to a portset

You can assign an IP address to a portset from the GUI or CLI.

Assigning an IP address to a portset from the GUI

After a portset is configured, an IP address can be assigned to it by completing the following steps:

1. Select **Settings** → **Network** → **Ethernet Ports** (see Figure 7-9).

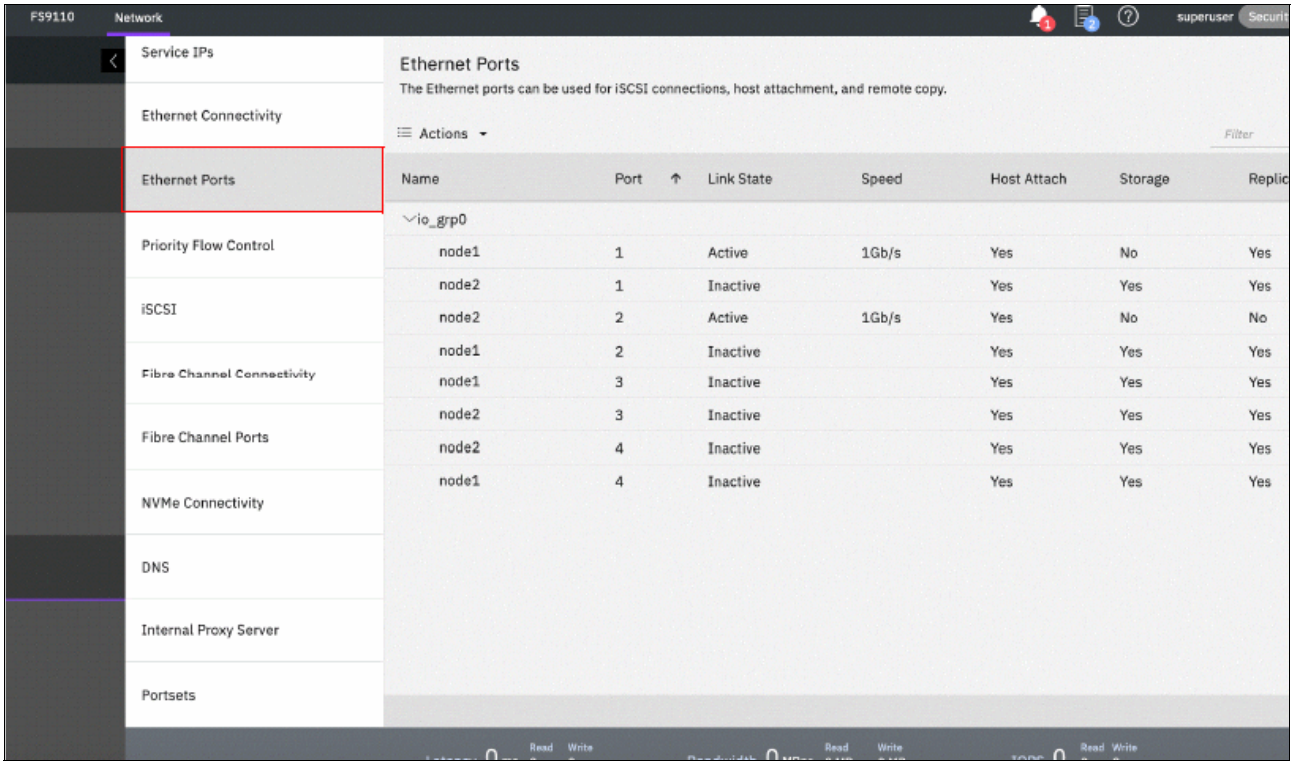


Figure 7-9 Selecting Ethernet ports

2. Right-click the port and select **Manage IP addresses** (see Figure 7-10).

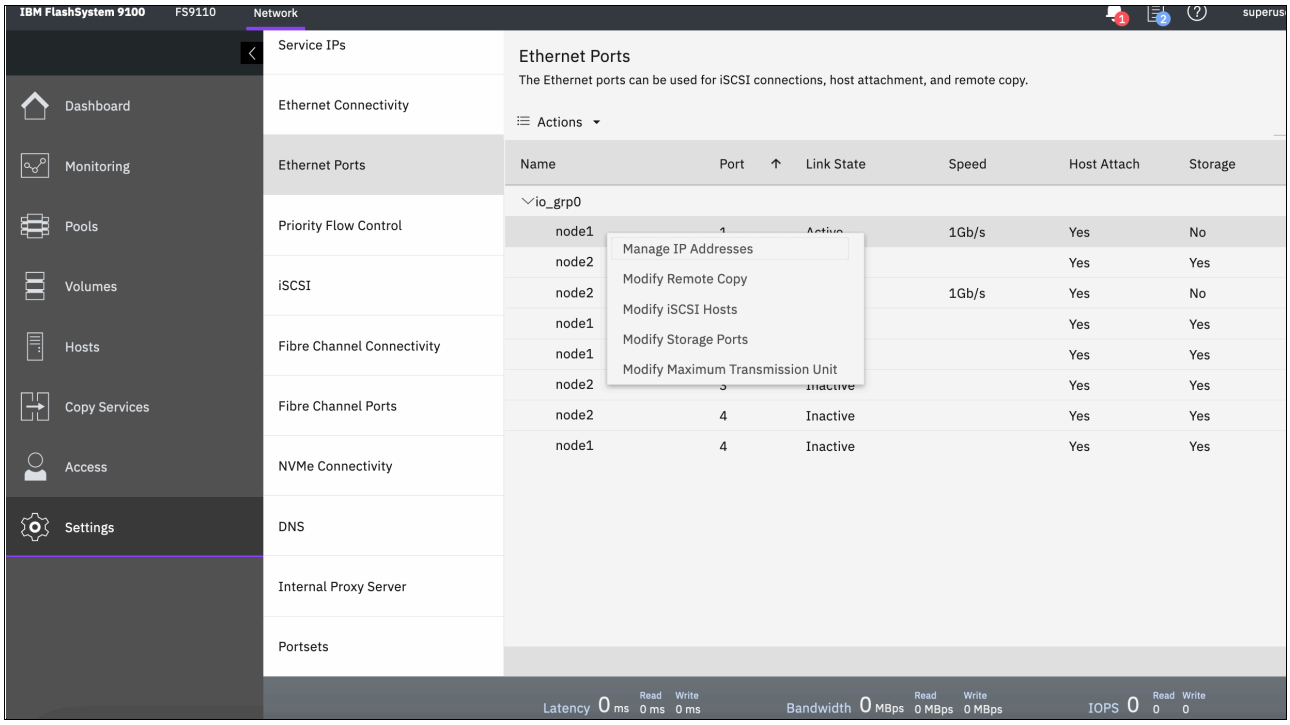


Figure 7-10 Actions available after right-clicking the suitable node and port

3. On the Manage IP Addresses page, select **Add IP Address**. Enter the following relevant information for the IP address that you are adding to the selected port:
   - IP Address: The IP address that is associated with the selected port.
   - Type: Select the IP protocol version of the IP address.
   - Subnet Mask or Prefix: The subnet mask for the IPv4 addresses or the prefix for IPv6 addresses.
   - VLAN: The corresponding VLAN tag to which this IP address belongs.
   - Portset: The name or ID of the portset and ensure that the portset type matches the traffic type that is assigned to the port.
4. Click **Back** to return to the Ethernet Ports page. Verify that the port displays the `Configured` state. If necessary, select another port and add IP addresses to the corresponding portsets.
5. Right-click the port and select **Modify Remote Copy, Modify iSCSI hosts**, or **Modify Storage Ports**. The traffic type for the port must match the traffic for the portset that you created.

Assigning an IP Address to a portset from the CLI

After creating a portset, enter the following command to assign it an IP address:

```
mkip -node node_name -port port_id -portset portset_id | portset_name -ip
x.x.x.x -gw gateway -vlan vlan_id -prefix subnet_prefix
```

Where *node_name* is the name of the node, *port_id* is the port identifier, and *portset_id | portset_name* indicates the ID or name of the portset that was created earlier.

Enter a valid IPv4 or IPv6 address for the **-ip** parameter. This address is assigned to the portset and more can be added to the portset by using the **mkip** command (see Figure 7-11).

```
IBM_FlashSystem:FS9110:superuser>mkip -node node1 -port 2 -portset portset_2 -ip 9.42.162.180 -prefix 24
IP Address, id [4], successfully created
IBM_FlashSystem:FS9110:superuser>
```

Figure 7-11 Portset_2 is assigned IP address 9.42.162.180

7.7.5 Assigning the portset to a host object

A host object can be assigned to a portset by entering the following command in the CLI:

```
mkhost -iscsiname iscsi_name -name host_name -portset portset_id | portset_name
```

Where *iscsi_name* specifies the iSCSI name or IQN to be added to the host, *host_name* is the name of the host, *portset_id* is the numerical portset identifier and *portset_name* is the name of the portset. Enter a portset_id or portset_name.

7.8 Fibre Channel portset

With the newer generation of IBM Storage Virtualize products, more I/O options are available regarding protocols and the number of I/O ports.

IBM Storage Virtualize version 8.6 extends the portset feature to include the support for Fibre Channel (FC) ports known as FC portsets (FC portsets). FC portsets can be used for effective port management and host access over FC or FC-NVMe protocols.

7.8.1 FC portset definition

an FC *portset* is a group of FC I/O ports from Storage Virtualize system over which a host can access storage through FC or FC-NVMe protocol, as shown in Figure 7-12.

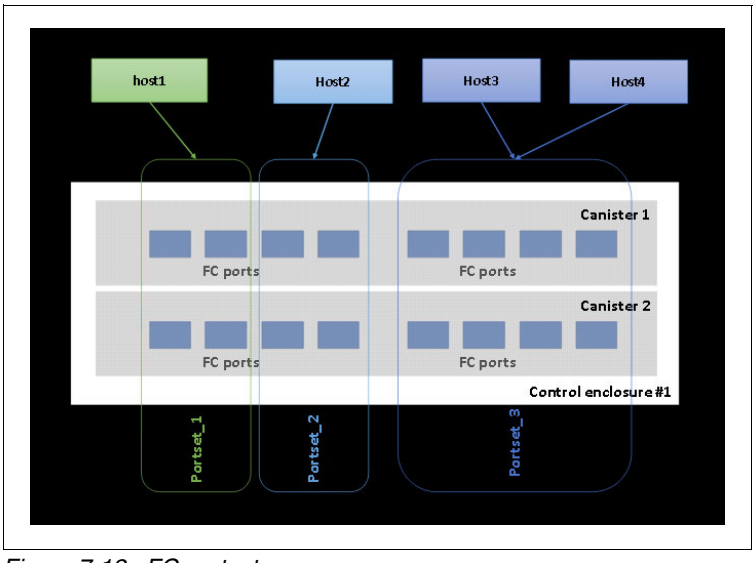


Figure 7-12 FC portsets

When the FC portset feature is used, each host should associate with only one FC portset because a host can access storage through the associated FC portset only. Although Figure 7-12 on page 537 shows each storage FC I/O port in a single FC portset for simplicity, a storage FC I/O port can be part of multiple FC portsets.

Note: The FC portset function still requires suitable zoning in FC switches between host and storage FC I/O ports. FC portsets will not prevent host logins to FC I/O ports not in the FC portset, but an event will be logged until zoning is corrected.

7.8.2 FC portsets usage scenarios

Regarding FCs, host ports are zoned with multiple storage ports to achieve multipathing for storage volume. As the number of host ports and storage ports in a subsystem grow, it becomes complex and it is difficult to control the number of paths to a volume. The use of more than the optimal number of paths per storage volume can result in the following issues:

- Too many host logins to storage
- Skewed I/O workload
- Effects on scalability because of uneven resource use

The FC portset realizes the following benefits:

- Better organization of large FC port counts, especially in newer IBM Storage Virtualize systems, such as FlashSystem 9500, FlashSystem 7300, and SAN Volume Controller SV3.
- Better resource usage for performance and host scaling.
- Minimize suboptimal pathing configuration to storage volume.

Typical FC portset scenarios that feature these benefits including the following examples:

- Configurations with greater than 8 FC ports.
- Configurations with greater than 256 host objects.
- Segregating host and storage ports based on speed; for example, hosts and storage with 16 Gbps speed in one FC portset, and hosts with 32 Gbps speed in a second FC portset.

7.8.3 FC portset behavior for established and new installations

For IBM Storage Virtualize systems running v8.6.0 as a new installation or as part of a firmware upgrade from previous version, a default FC portset is created. The default FC portset is called *portset64*, as shown in Figure 7-13. The default portset includes all the FC ports and all the FC hosts from the IBM Storage Virtualize storage system.

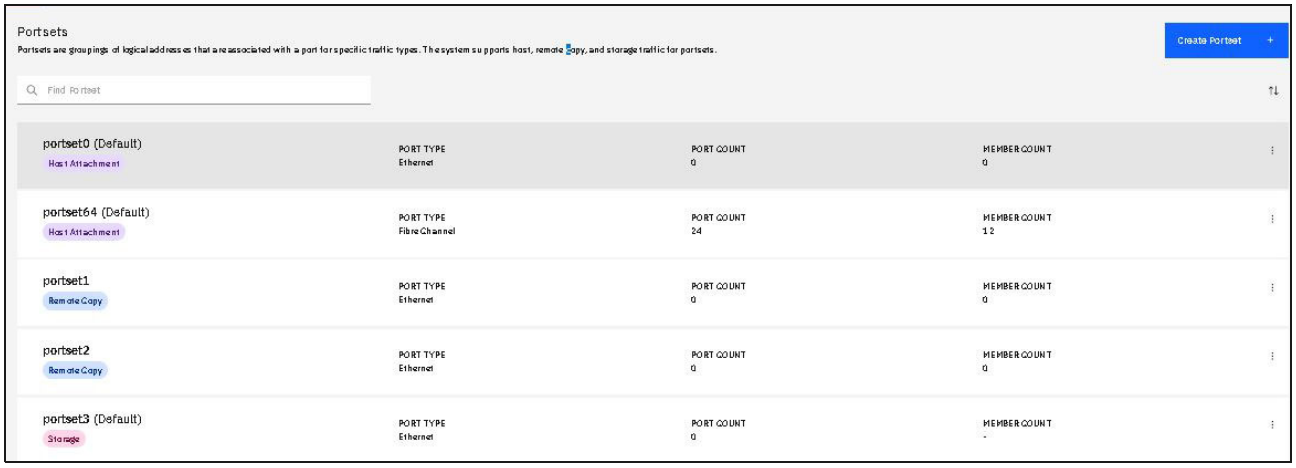


Figure 7-13 Default FC portset

Because the default FC portset, portset64, includes all of the FC ports and FC hosts, the hosts continue to operate as they did before the upgrade. The storage administrator can continue to operate without any special knowledge of FC portsets.

Then storage administrator can define further FC portsets and associate them with FC host objects as needed for the system and workload, as described in 7.8.2, “FC portsets usage scenarios” on page 538.

7.8.4 Configuring the FC portset

To configure the FC portsets from the GUI, complete the following steps:

1. Select **Settings** → **Network** → **Portsets** (see Figure 7-14).

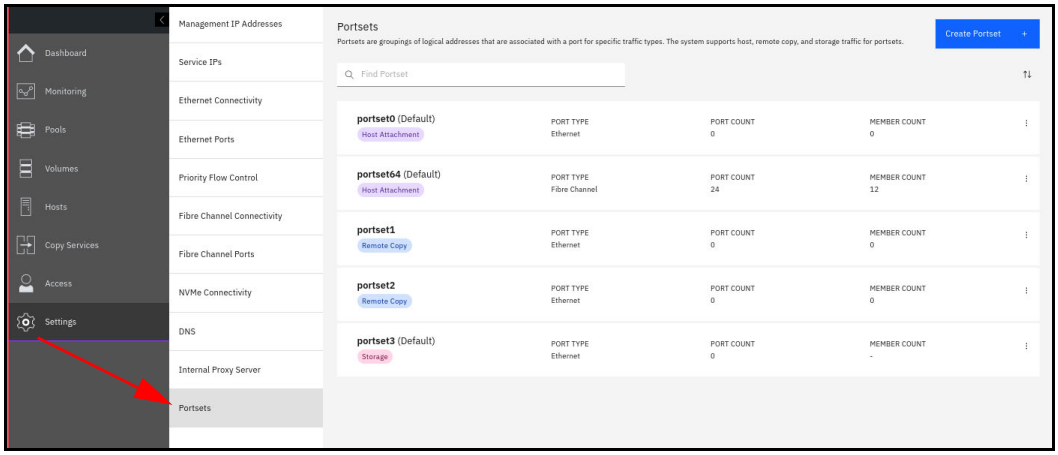


Figure 7-14 Selecting the portset option

1. Select **Create Portsets** (see Figure 7-15). Complete the following steps on the Create Portset page:
   a. Enter a name for the FC portset.
   b. For the Portset Type, select the **Host Attachment** option.
   c. For the Port Type, select **Fibre Channel**, as shown in Figure 7-15.

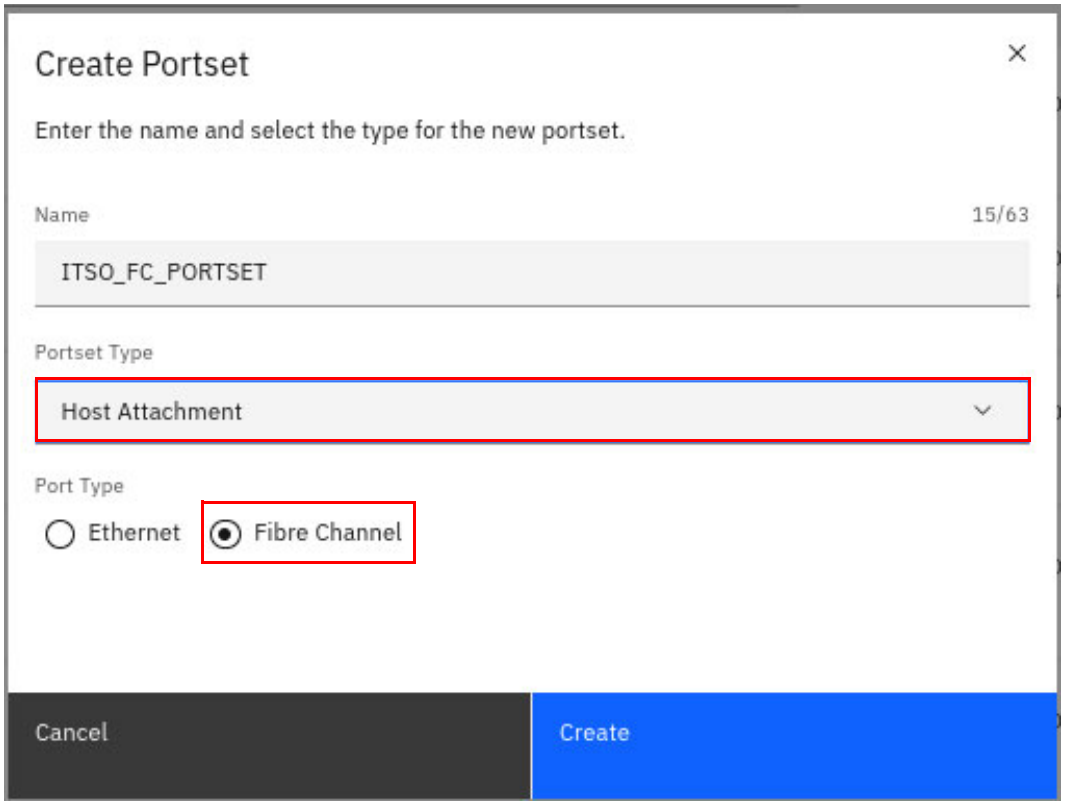


Figure 7-15 Creating FC portset

2. Click **Create**.

Note: As of this writing, the FC portset can be used for host attachments only.

Configuring an FC portset from CLI

To create a portset from the CLI, enter the following command:

```
mkportset -name portset_name -type portset_type -ownershipgroup owner_id |
owner_name -porttype fc | ethernet
```

Where *portset_name* is the name of the portset and *portset_type* is `host` or `replication`. The value *owner_id* | *owner_name* indicates the ID or name of the ownership group to which the portset belongs. The *porttype* can be **fc** for FC ports or **ethernet**. Example 7-13 shows creating an FC portset.

Example 7-13 FC portset creation by way of CLI

```
IBM_FlashSystem:ITSO_FS9500:superuser>mkportset -name ITSO_FC_PORTSET -type host
-porttype fc
Portset, id [5], successfully created
```

7.8.5 Assigning an FC port to an FC portset

An FC port can be added to an FC portset by using its FC I/O port ID

After the FC portset is created, complete the following steps to assign FC ports from the IBM Storage Virtualize storage system by using the GUI:

1. Select **Settings** → **Network** → **Fibre Channel Ports** (see Figure 7-16).

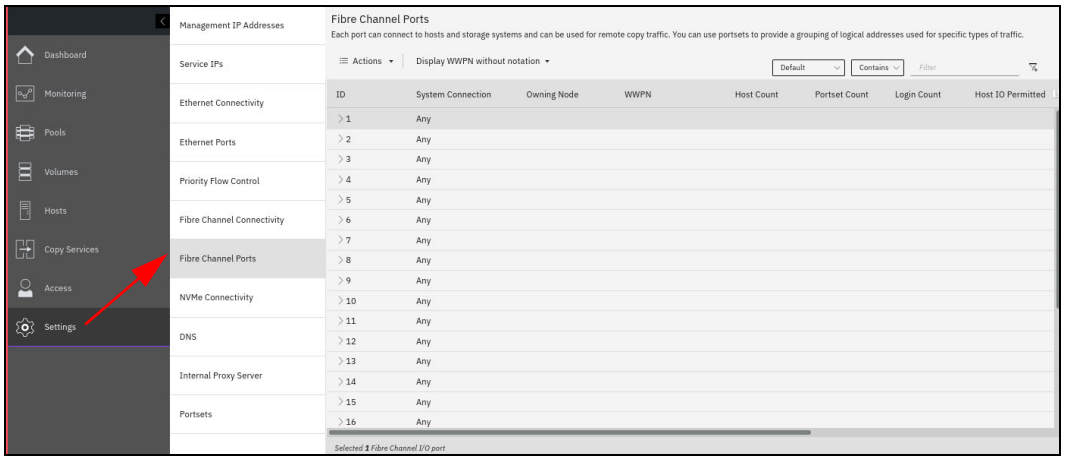


Figure 7-16 Fibre Channel Ports

2. Right-click the wanted port ID and select **Assign Portset** (see Figure 7-17).

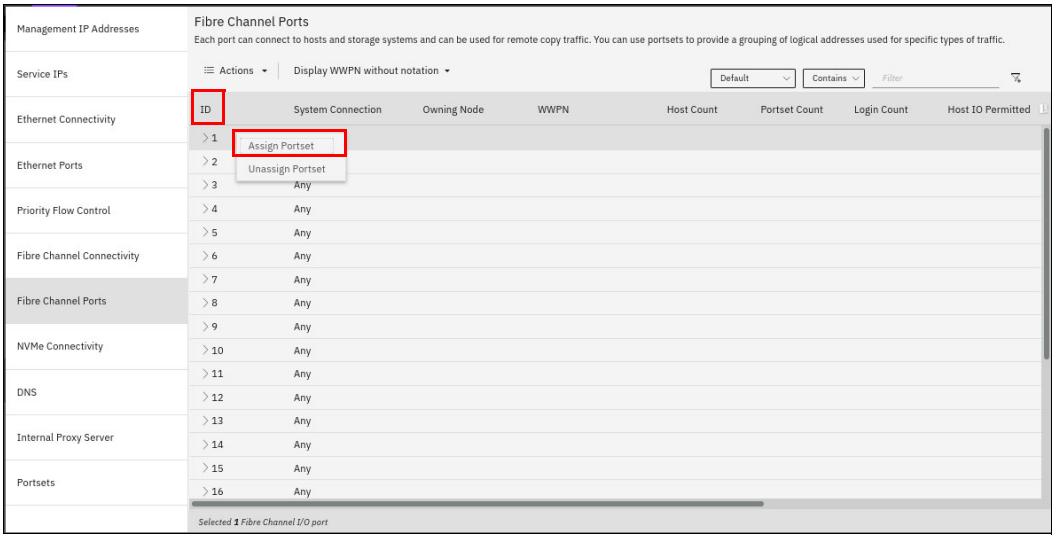


Figure 7-17 Selecting Fibre Channel Port ID

3. Select the checkbox for the FC portsets to which the wanted FC Port ID must be added, as shown in Figure 7-18.

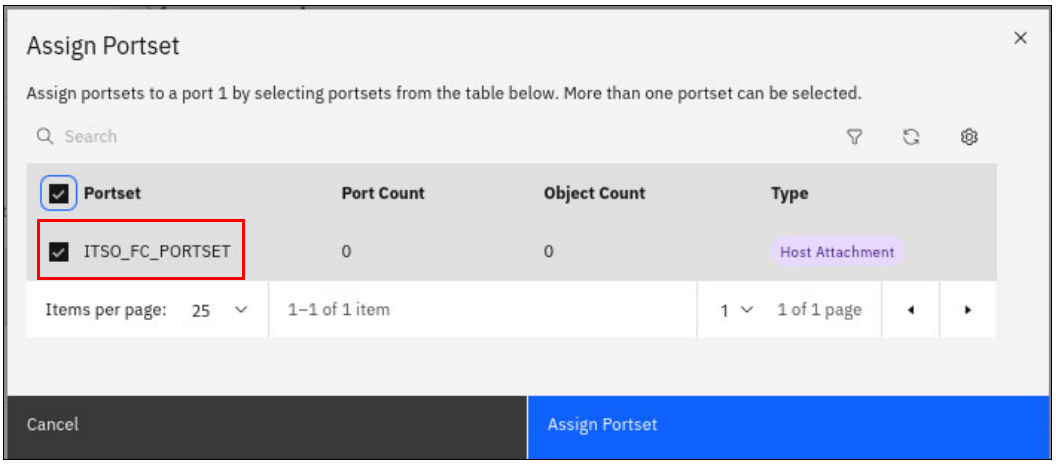


Figure 7-18 Assigning FC portset

Now, the wanted FC I/O port is assigned to the selected FC portset.

When an FC I/O port ID is added to an FC portset, the same FC I/O port ID from all the nodes and node-canisters from the cluster is added to that FC portset. The addition of the FC I/O port ID to an FC portset is done at the cluster level.

Note: The same FC I/O port can be configured in multiple FC portsets. However, each FC portset is limited to four FC I/O ports, except the default portset (`portset64`). As of this writing, the default FC portset (`portset64`) can have up to 64 ports.

Assigning FC ports to an FC portset by using the CLI

An FC port can be associated with an FC portset by using the **`addfcportsetmember`** CLI command and by specifying its *fc_io_port_id*, as shown in Example 7-14.

Example 7-14 Assigning FC port to FC portset

```
IBM_FlashSystem:ITSO_FS9500:superuser>addfcportsetmember -portset ITSO_FC_PORTSET
-fcioportid 1
IBM_FlashSystem:ITSO_FS9500:superuser>
```

Here, FC `portid 1` on each node or node-canister in the cluster is added to FC portset `ITSO_FC_PORTSET`.

7.8.6 Modifying an FC portset for a host

IBM Storage Virtualize systems running firmware v8.6.0 or higher create a default portset (`portset64`). This default portset is associated with all the FC hosts in the storage system. If the storage administrator wants to change the FC portset that is associated with any post, they complete the following steps:

1. Select **Hosts** → **Hosts** (see Figure 7-19).

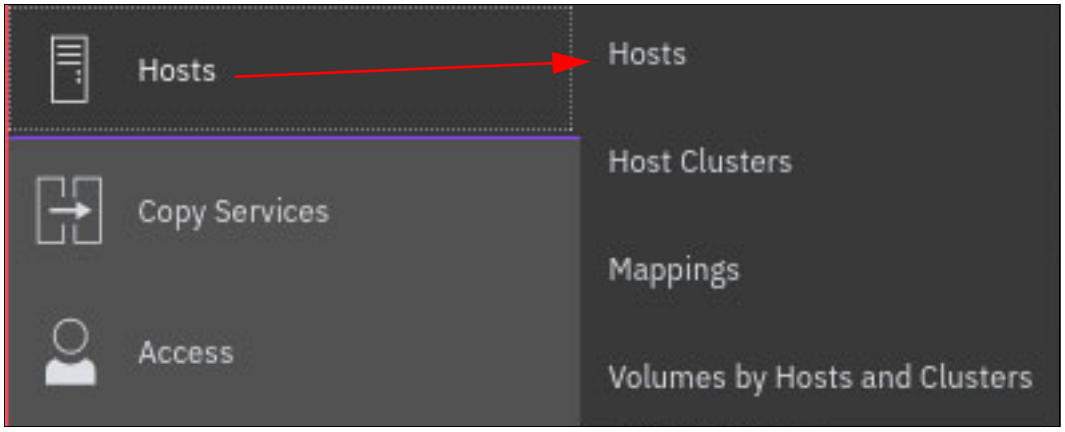


Figure 7-19 Hosts

The UI shows the hosts that are defined in the system, with one of the columns being the default FC portset (see Figure 7-20).

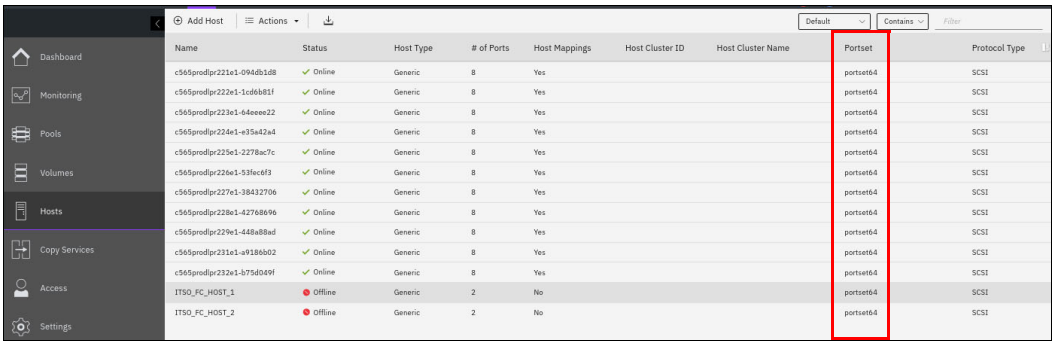


Figure 7-20 Hosts and default FC portset

2. Right-click the wanted host and then click **Properties** as shown in Figure 7-21.

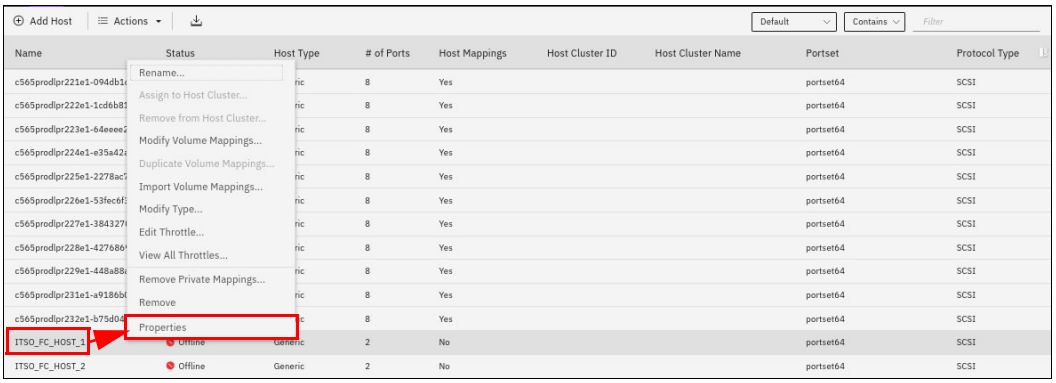


Figure 7-21 Host Properties

The host properties details window shows the associated FC portset, as shown in Figure 7-22.

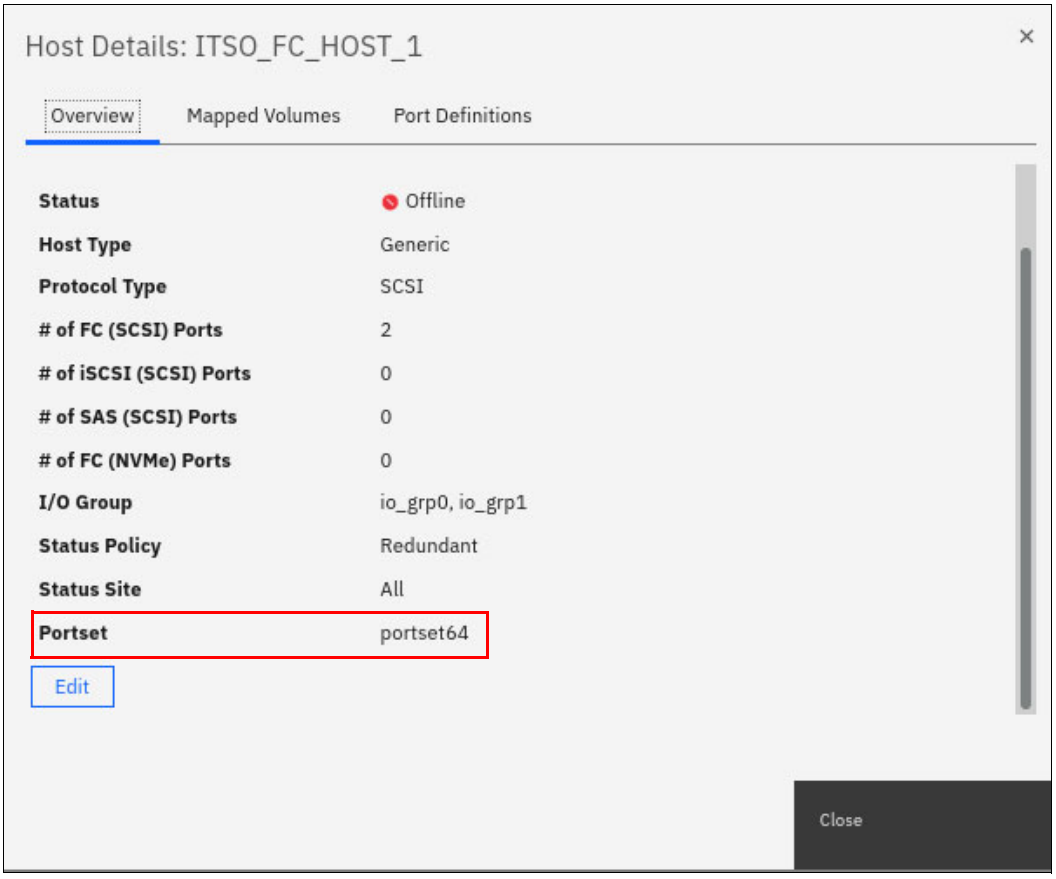


Figure 7-22 Associated FC portset

3. Click **Edit** to change the associated FC portset, as shown in Figure 7-23.

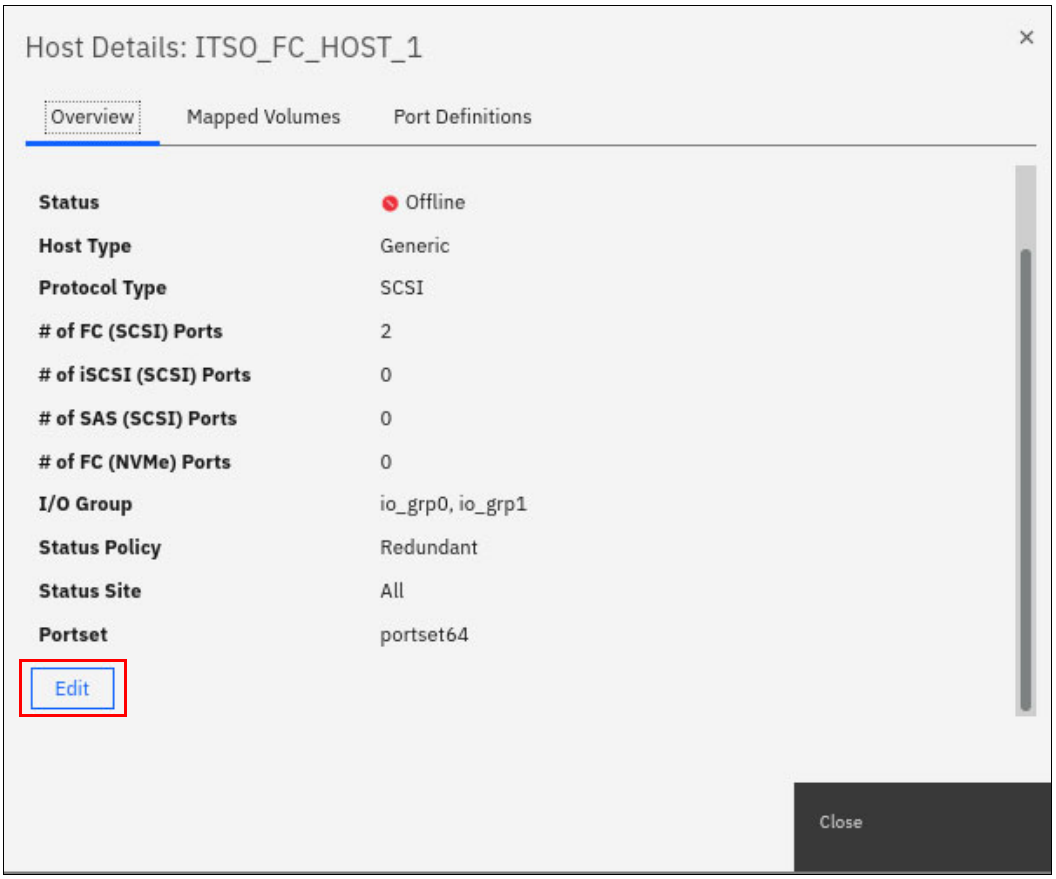


Figure 7-23 Editing the Host properties for FC portset

4. From the Portset drop-down menu, select the wanted FC portset, as shown in Figure 7-24.

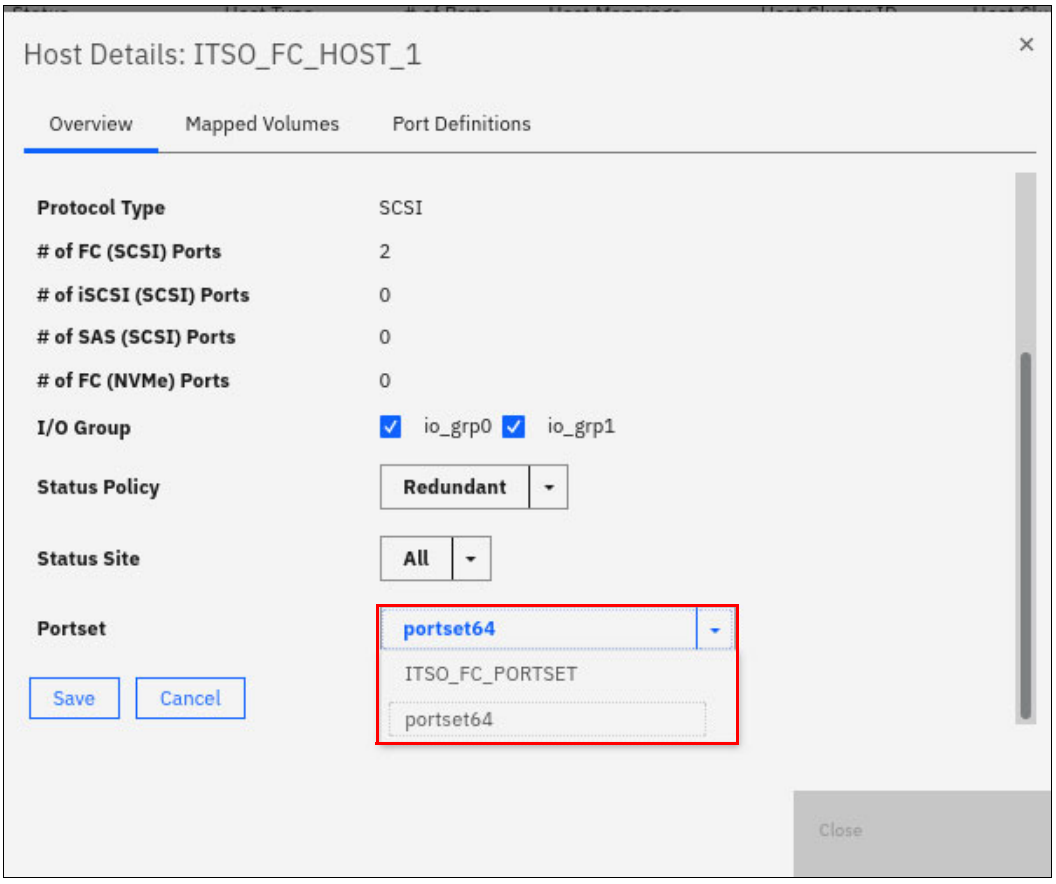


Figure 7-24 Portset dropdown

5. Click **Save**. The specified FC portset is assigned to the wanted host, as shown in Figure 7-25.

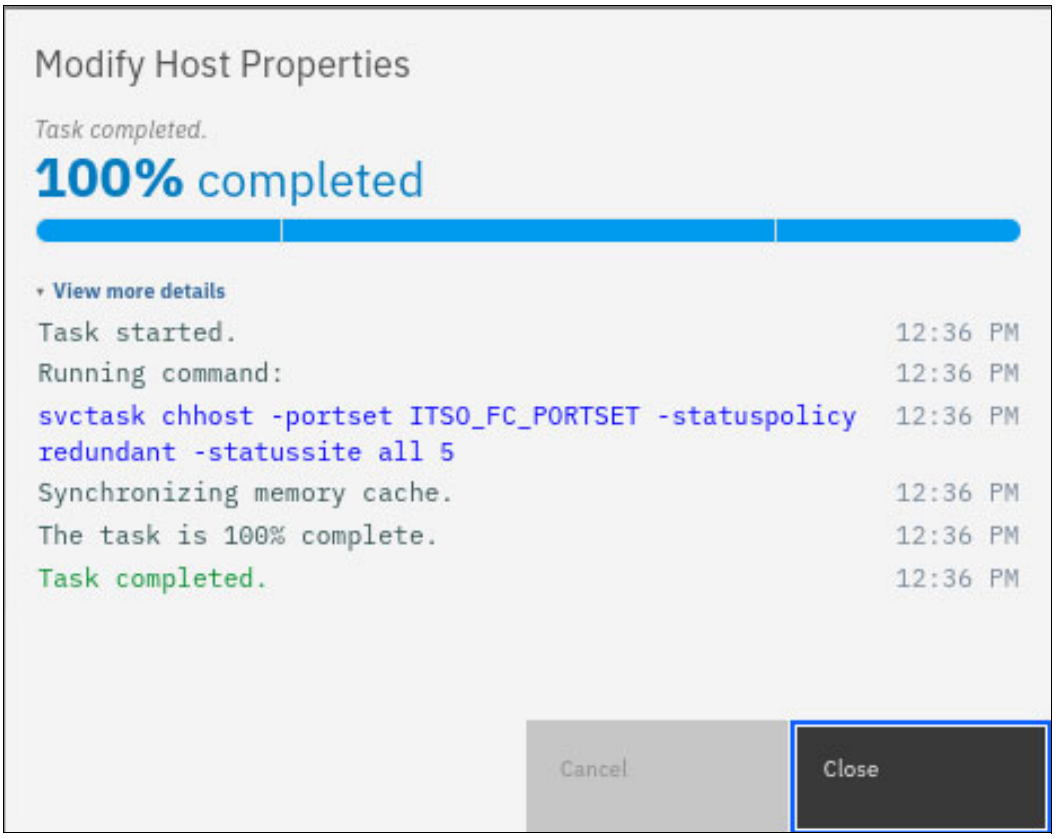


Figure 7-25 FC portset change saved

6. Check the Hosts again (see Step 1 on page 543) to confirm that the wanted FC portset assignment to the requested host was successful, as shown in Figure 7-26.

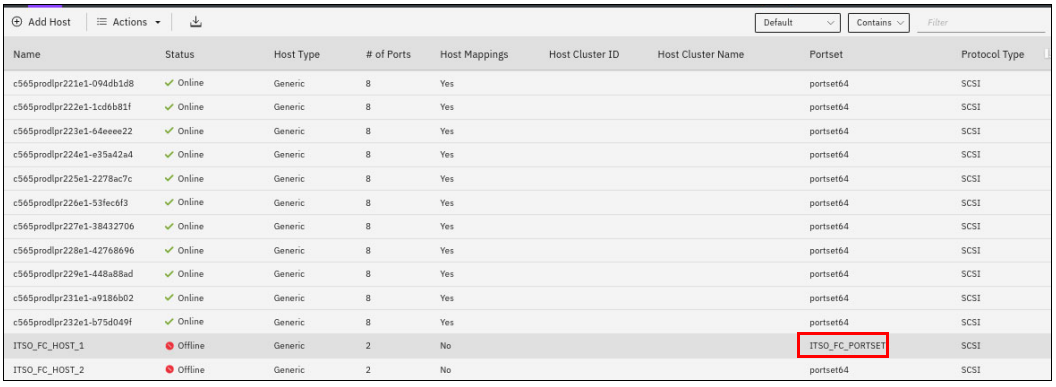


Figure 7-26 New FC portset assigned

Modifying an FC portset for a host by using the CLI

With IBM Storage Virtualize v8.6.0, all FC hosts are associated with the default FC portset, portset64, to start. If the storage administrator wants to change the FC portset that is associated with any host, the **chhost** CLI command is used, as shown in Example 7-15.

Example 7-15 The use of chhost CLI command

```
IBM_FlashSystem:ITSO_FS9500:superuser>chhost -portset ITSO_FC_PORTSET
ITSO_FC_HOST_1
IBM_FlashSystem:ITSO_FS9500:superuser>
```

7.8.7 Removing a port from an FC portset

When an FC port is removed from the FC portset that is based on its *fc_io_port_id,* the same port ID from all of the nodes or node canisters is removed from the FC portset. The removal of an FC port from an FC portset operates at a cluster level.

Warning: If an FC I/O port is removed from the default FC portset, portset64, you cannot add that port back to the default FC portset if the default FC portset has four or more FC I/O ports.

To remove an FC port from FC portset, complete the following steps:

1. Access the FC port as described in Step 1 on page 541.
2. Right-click the wanted FC port and select **Unassign Portset**, as shown in Figure 7-27.

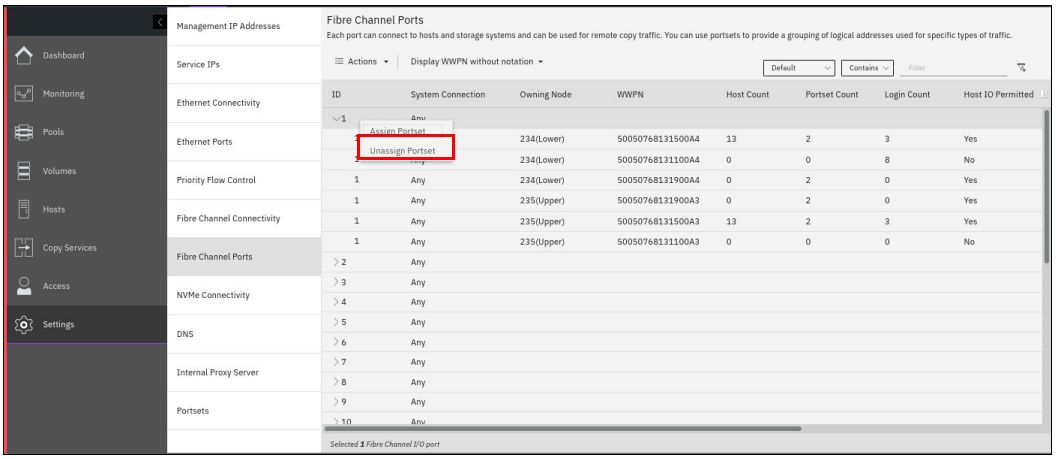


Figure 7-27 Unassign Portset operation

3. Select the wanted FC portset from which to unassign or remove the FC Port, as shown in Figure 7-28.

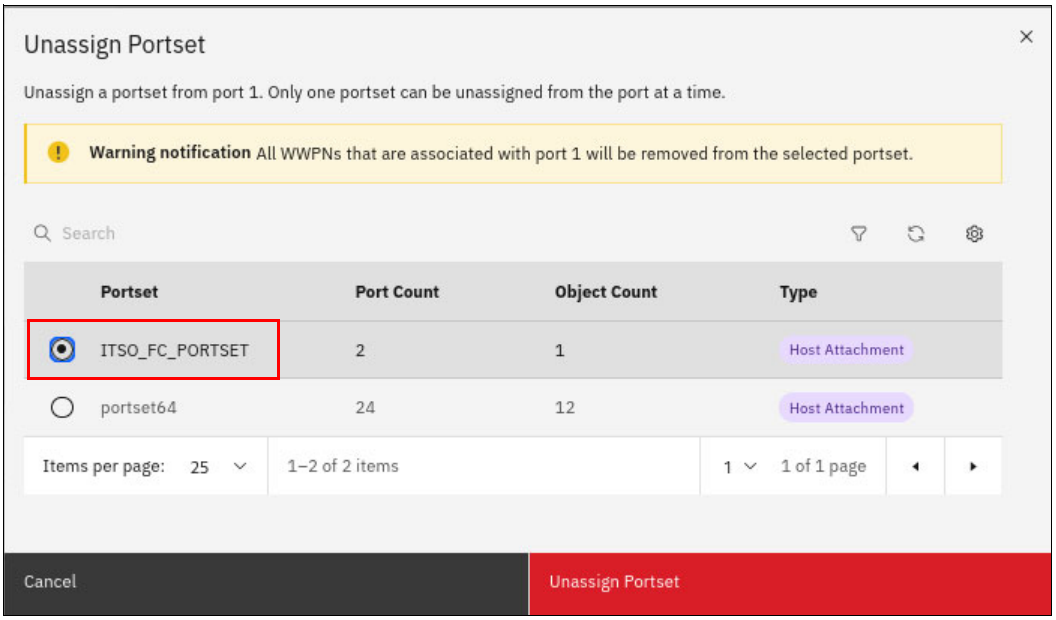


Figure 7-28 Selecting FC portset

4. Click **Unassign Portset**, as shown in Figure 7-29.

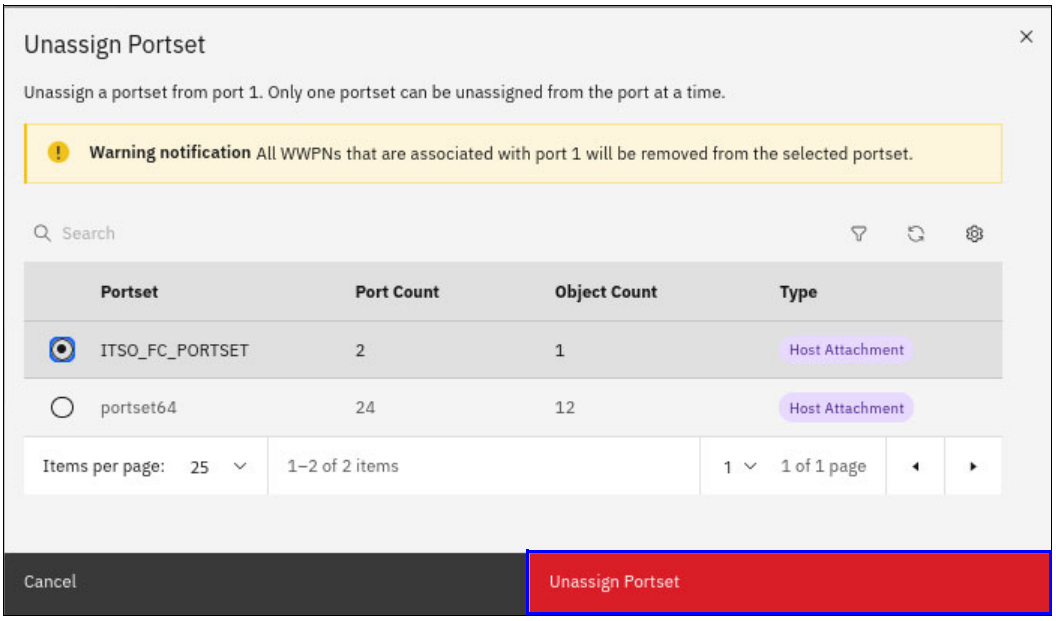


Figure 7-29 Unassign Portset

The intended FC Port is unassigned or removed from the selected FC portset, as shown in Figure 7-30.

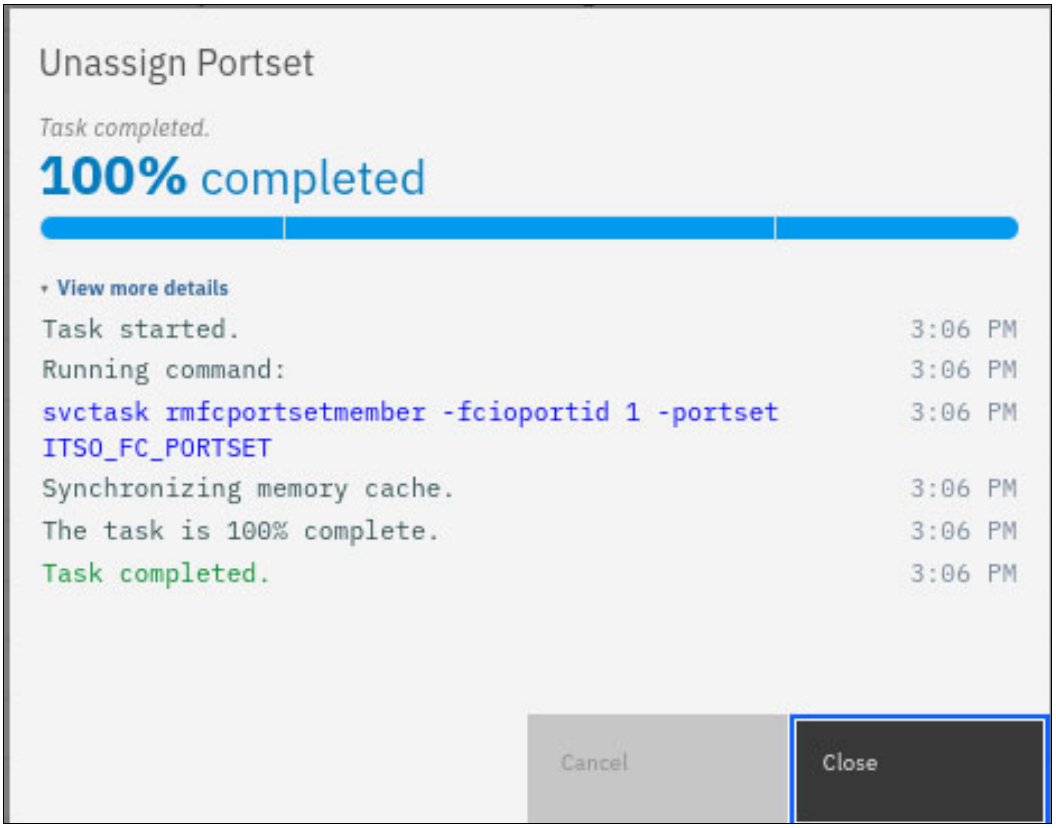


Figure 7-30 Removing FC Port from an FC portset

Note: If active hosts are associated with an FC portset from which a port is being removed, it is recommended you update the host zone to remove WWPN corresponding to the FC Port that is being removed.

Removing a port from an FC portset by using CLI

The storage administrator can remove any FC port from any FC portset by using the CLI, as shown in Example 7-16.

Example 7-16 Removing an FC port from an FC portset

```
IBM_FlashSystem:ITSO_FS9500:superuser> rmfcportsetmember -portset ITSO_FC_PORTSET
-fcioportid 1
IBM_FlashSystem:ITSO_FS9500:superuser>
```

7.8.8 Removing an FC portset

Before removing an FC portset, ensure that no FC host or FC port is associated with the FC portset. an FC portset cannot be removed or deleted if it has port count and host count set.

Note: The IBM Storage Virtualize system does *not* allow the default FC portset, portset64, to be deleted.

To delete an FC portset, follow the steps described:

1. Access the wanted FC portset, as described in Step 1 on page 539.
2. For the selected FC portset, ensure that the PORT COUNT and MEMBER COUNT columns show the value as 0, as shown in Figure 7-31.

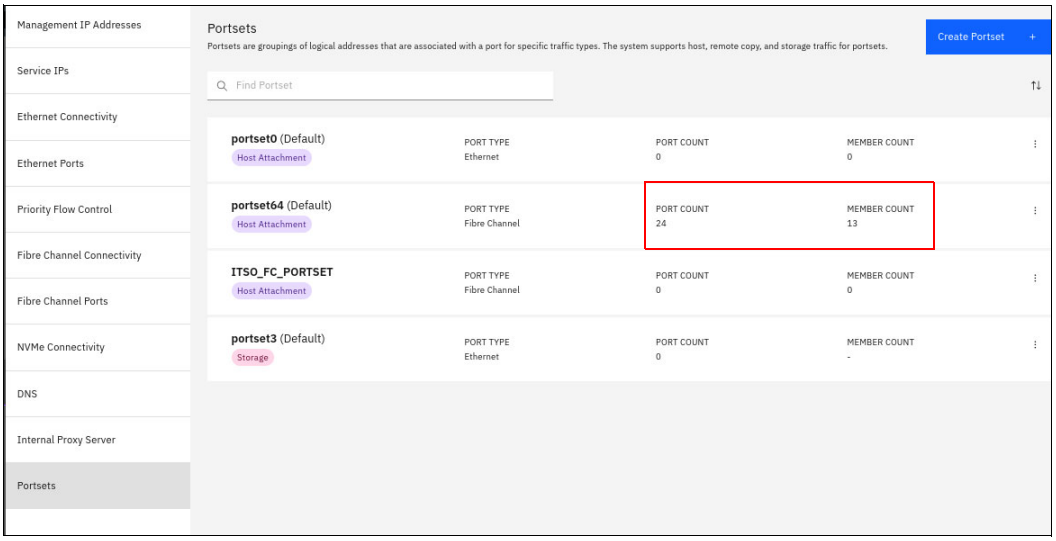


Figure 7-31 FC portset

3. From the FC portset details window, click **Action** → **Delete**, as shown in Figure 7-32.

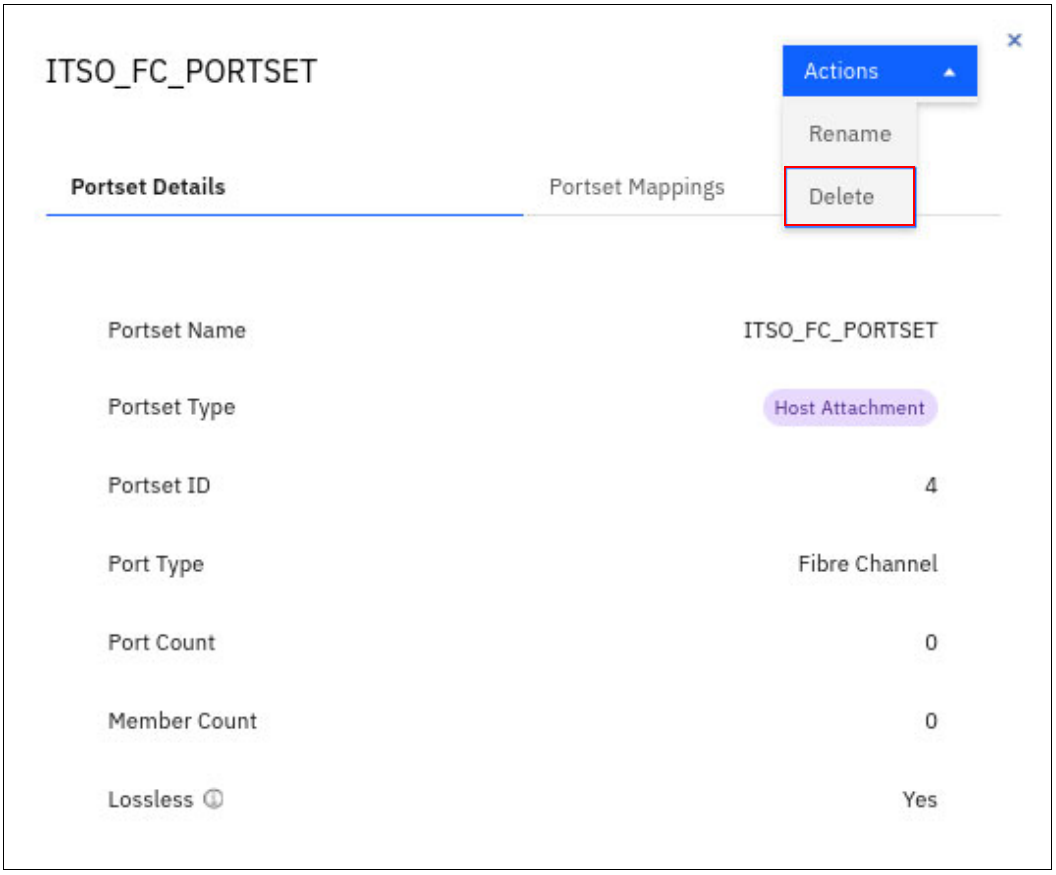


Figure 7-32 FC portset Delete operation

4. Click **Delete** in the Delete Portset window, as shown in Figure 7-33.

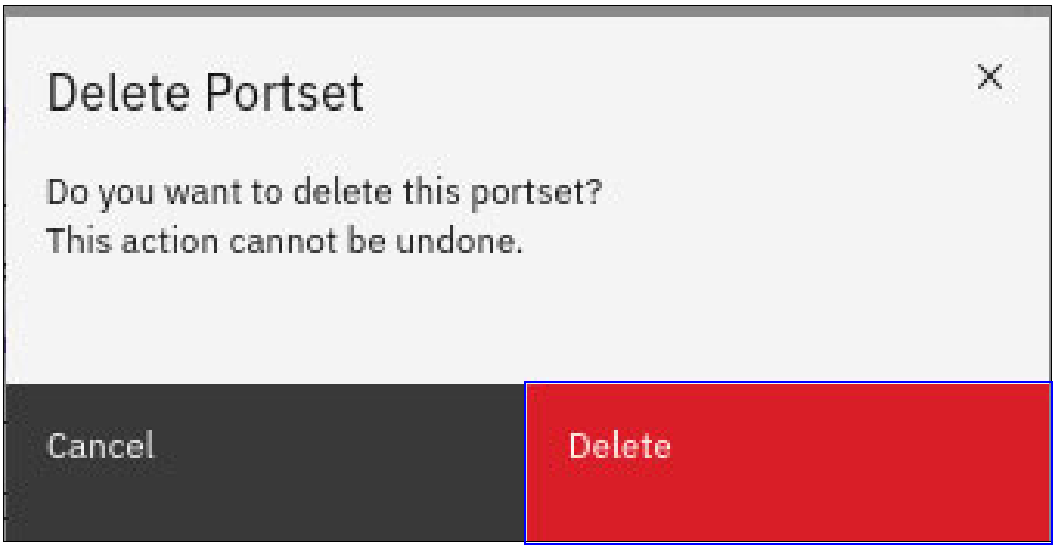


Figure 7-33 FC portset Delete confirmation

5. The selected FC portset is deleted, as shown in Figure 7-34.

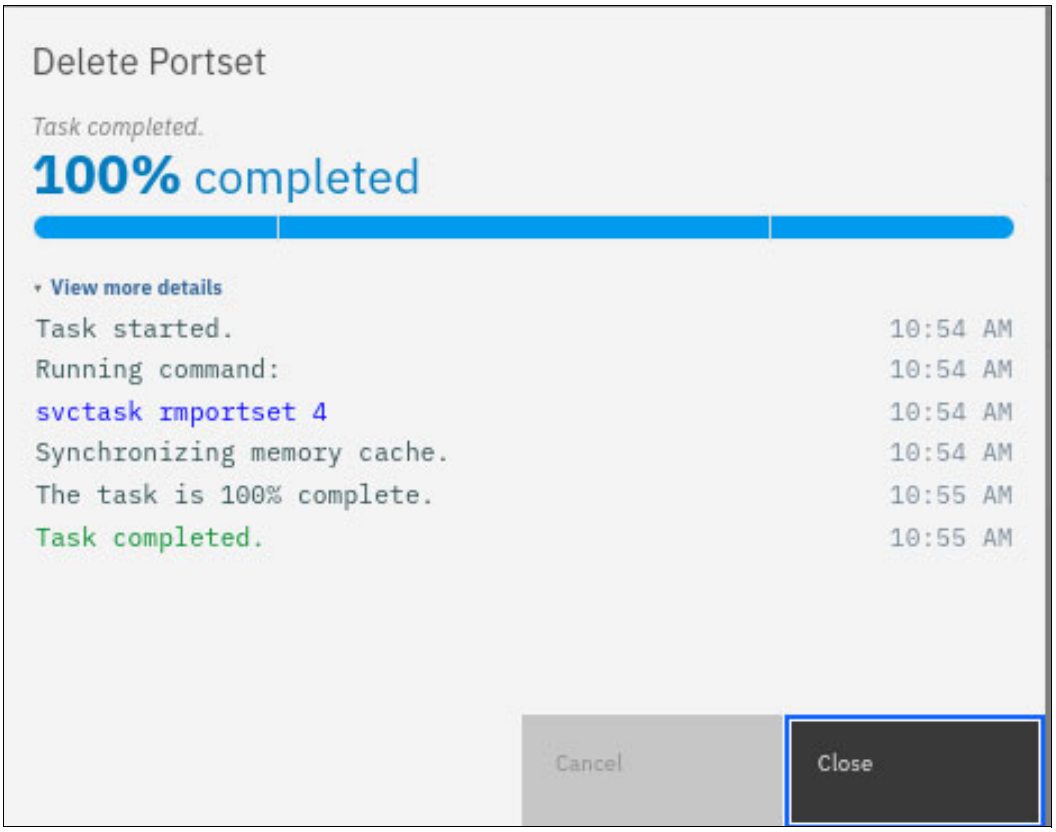


Figure 7-34 FC portset deletion

Removing an FC portset by way of CLI

If the FC portset does not have any host and any FC port that is associated with it, the storage administrator can remove it by using the CLI, as shown in Example 7-17.

Example 7-17 Removing FC portset

```
IBM_FlashSystem:ITSO_FS9500:superuser>rmportset ITSO_FC_PORT_SET_2
IBM_FlashSystem:ITSO_FS9500:superuser>
```

Note: If the FC portset has any host or FC port that is associated with it, the removal is unsuccessful. The following CLI error message is displayed:

```
CMMVC5753E The specified object does not exist or is not a suitable candidate.
```

7.8.9 FC portset misconfiguration and resolution

If a misconfiguration exists between the host zone and the FC portset that is associated with the host, the IBM Storage Virtualize system raises an event, as shown in Figure 7-35.

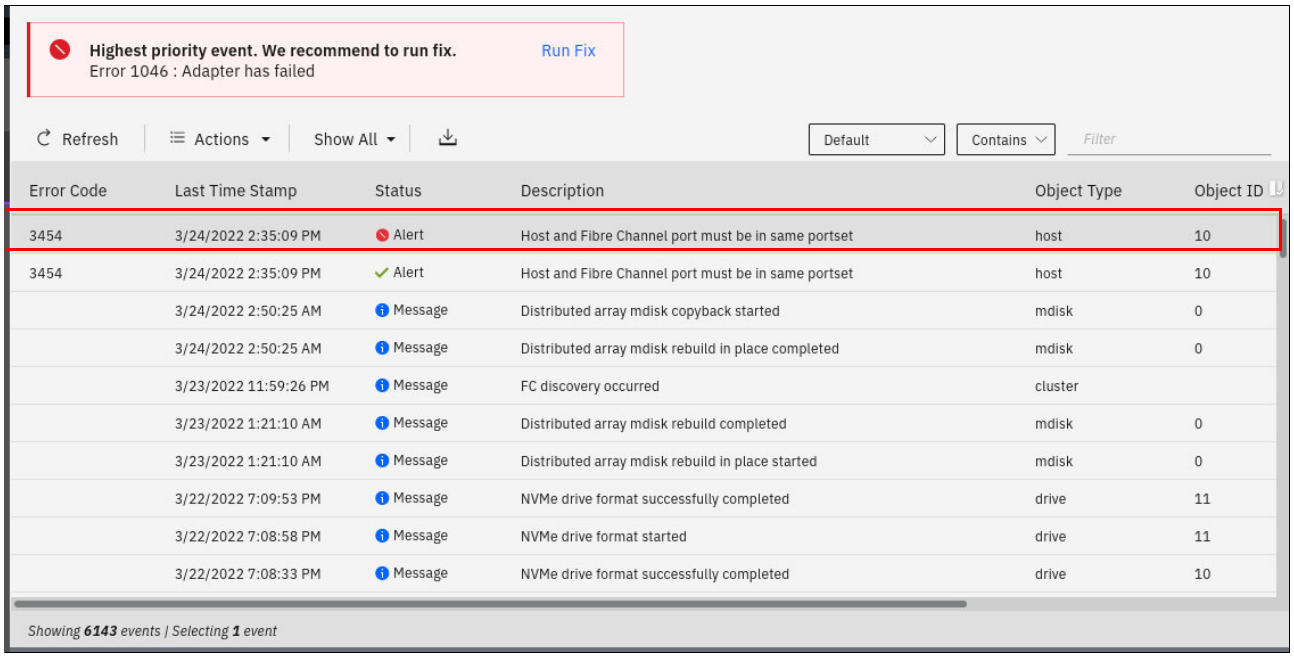


Figure 7-35 FC portset misconfiguration event

Because of this misconfiguration, the I/O over the misconfigured FC I/O ports is blocked for the host, as shown in Example 7-18.

Example 7-18 I/O blocked over misconfigured port

```
IBM_FlashSystem:FORTIFY_FS950:superuser>lsfabric -host 10 -delim :
remote_wwpn:remote_nportid:id:node_name:local_wwpn:local_port:local_nportid:state:name:cluster_name:type
C0507607AA8312B2:641E2F:235:node1:50050768138500A3:21:655101:blocked:c565prodlpr232e1::host
C0507607AA8312B2:641E2F:234:node2:50050768138500A4:21:655401:blocked:c565prodlpr232e1::host
C0507607AA8312B0:631D02:235:node1:50050768138600A3:22:655101:blocked:c565prodlpr232e1::host
C0507607AA8312B0:631D02:234:node2:50050768138600A4:22:655401:blocked:c565prodlpr232e1::host
C0507607AA8312AE:642004:235:node1:50050768137500A3:17:655001:active:c565prodlpr232e1::host
C0507607AA8312AE:642004:234:node2:50050768137500A4:17:655501:active:c565prodlpr232e1::host
C0507607AA8312AC:631F33:235:node1:50050768137600A3:18:655001:active:c565prodlpr232e1::host
C0507607AA8312AC:631F33:234:node2:50050768137600A4:18:655501:active:c565prodlpr232e1::host
```

Note: As of this writing, the I/O request for FC-NVMe host is *not* blocked if the FC I/O port and FC-NVMe host are *not* part of the same FC portset. The **`lsnvmefabric`** CLI command shows a state of `invalid` for such logins. However, I/O continues to work as it did in earlier releases.

To resolve FC portset misconfigurations, ensure that the correct FC I/O ports are included in the FC portset that is associated with the wanted FC host. The storage administrator can modify the FC portset definition to include any missing FC I/O ports, as described in 7.8.5, “Assigning an FC port to an FC portset” on page 541.

After the storage administrator fixes the misconfiguration, the event is marked as `Fixed`, as indicated by the green checkmark that is shown in Figure 7-36.

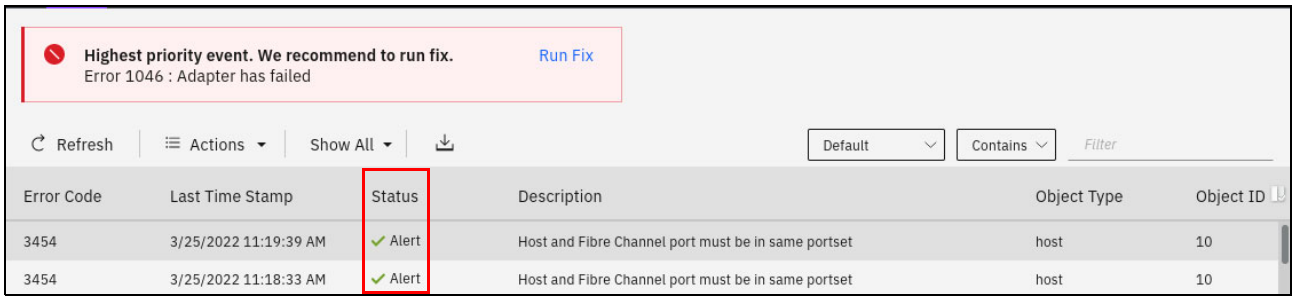


Figure 7-36 Event marked as Fixed

The host I/O that was in the `blocked` state through FC ports as shown in Example 7-18 on page 554 becomes `active`, as shown in Example 7-19.

Example 7-19 I/O active over included FC ports

```
IBM_FlashSystem:FORTIFY_FS9500:superuser>lsfabric -host 10 -delim :
remote_wwpn:remote_nportid:id:node_name:local_wwpn:local_port:local_nportid:state:name:cluster_name:type
C0507607AA8312B2:641E2F:235:node1:50050768138500A3:21:655101:active:c565prodlpr232e1::host
C0507607AA8312B2:641E2F:234:node2:50050768138500A4:21:655401:active:c565prodlpr232e1::host
C0507607AA8312B0:631D02:235:node1:50050768138600A3:22:655101:active:c565prodlpr232e1::host
C0507607AA8312B0:631D02:234:node2:50050768138600A4:22:655401:active:c565prodlpr232e1::host
C0507607AA8312AE:642004:235:node1:50050768137500A3:17:655001:active:c565prodlpr232e1::host
C0507607AA8312AE:642004:234:node2:50050768137500A4:17:655501:active:c565prodlpr232e1::host
C0507607AA8312AC:631F33:235:node1:50050768137600A3:18:655001:active:c565prodlpr232e1::host
C0507607AA8312AC:631F33:234:node2:50050768137600A4:18:655501:active:c565prodlpr232e1::host
```

7.9 Hosts operations by using the GUI

This section describes performing the following host operations by using the IBM Storage Virtualize GUI:

- Creating hosts
- Advanced host administration
- Adding and deleting host ports
- Administering host mappings

7.9.1 Creating hosts

This section describes how to create FC-, iSCSI-, and NVMe-connected host objects by using a GUI. It is assumed that hosts are prepared for attachment and that the host WWPNs, iSCSI initiator names, or NVMe Qualified Names (NQNs) are known.

For more information, see this IBM Documentation web page.

To create a host, complete the following steps:

1. Open the host configuration window by clicking **Hosts** (see Figure 7-37 on page 556).

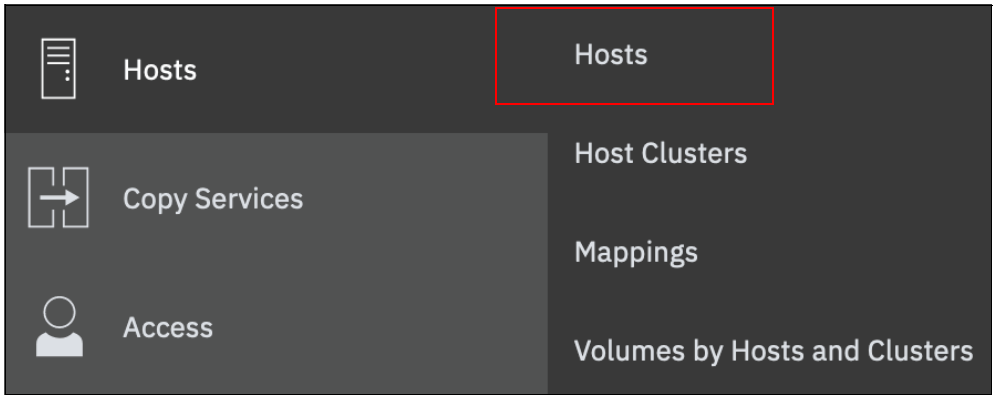


Figure 7-37 Opening the Host window

2. To create a host, click **Add Host**.

For more information about creating specific host types, see the following sections:

- “Creating FC host objects” on page 556
- “Preparing for iSCSI connection” on page 560
- “Creating iSCSI host objects” on page 565
- “rCeating FC NVMe host objects” on page 567
- “Creating NVMe over RDMA host objects” on page 569
- “Creating NVMe over TCP host objects” on page 572

Creating FC host objects

To create FC hosts, complete the following steps:

1. Select **Fibre Channel** in the Host Connections list and complete the configuration fields (see Figure 7-38 on page 557).
2. Enter a host name and click the **Host Port** menu to see a list of all discovered WWPNs (see Figure 7-39 on page 557).

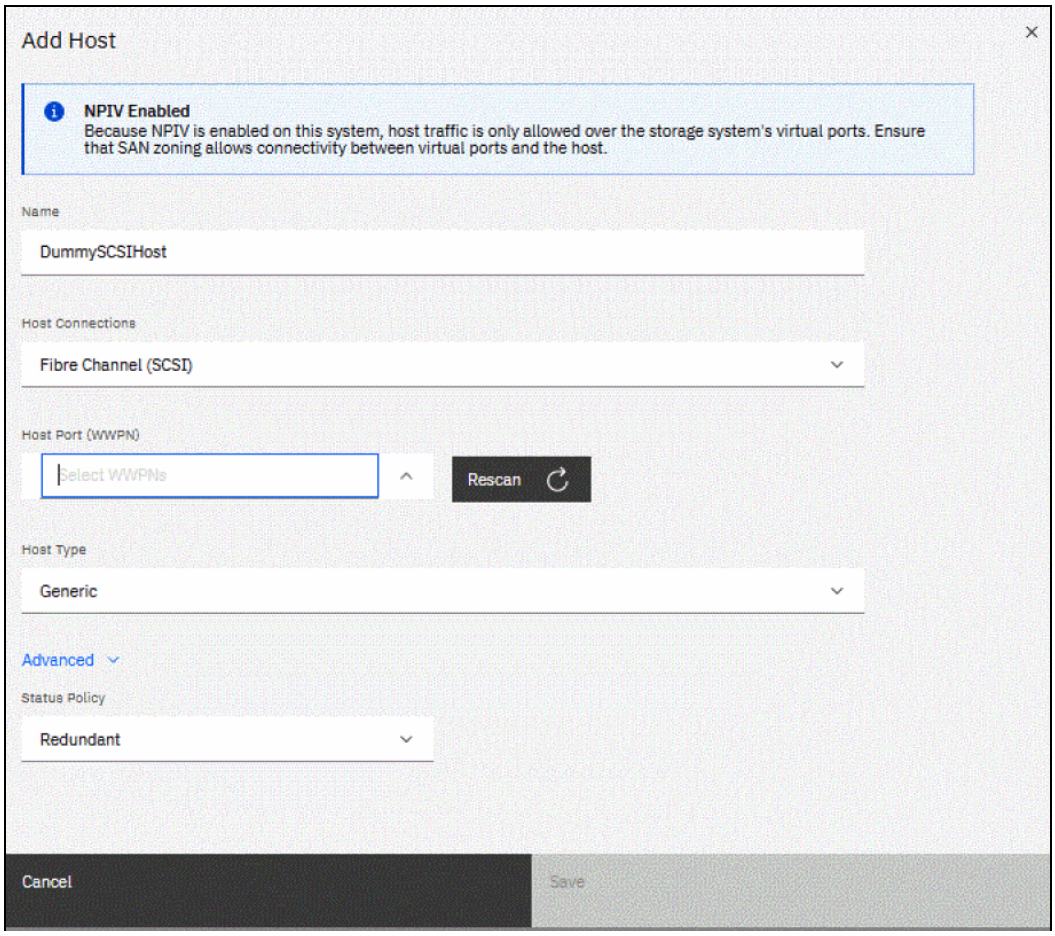


Figure 7-38 Fibre Channel host configuration view

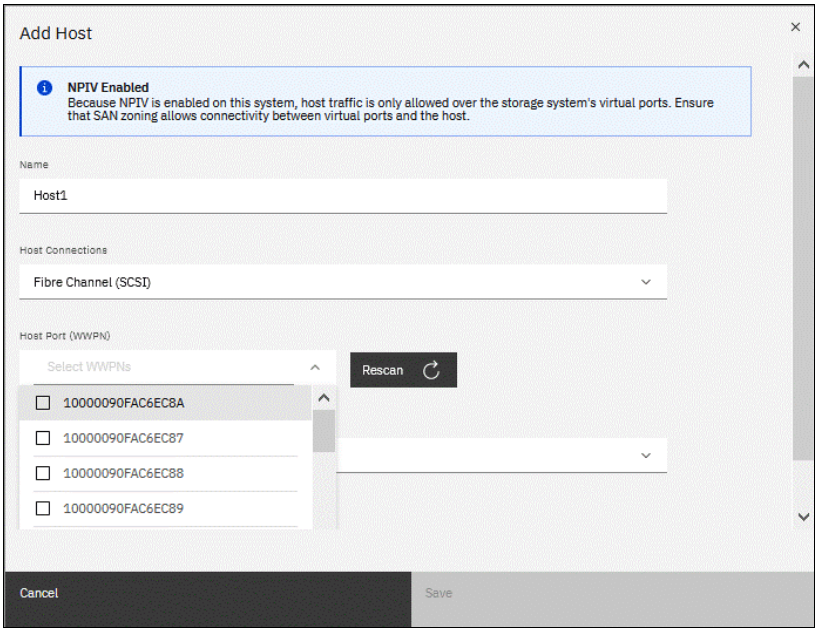


Figure 7-39 Selecting the host WWPNs

3. Select one or more host WWPNs from the list. These WWPNs are visible on the IBM FlashSystem storage system if the hosts are correctly zoned and presented.

   If the host information does not appear in the list, scan for new paths as required by the respective operating system and click the **Rescan** icon that is next to the WWPN box.

   If they still do not appear, check the SAN zoning and ensure that hosts are correctly connected and running. then repeat the scan.

   Creating offline hosts: To create hosts that are offline or not connected, enter the WWPNs manually into the **Host Ports** field to add them to the list.

4. To ports to the host, select the wanted WWPNs from the list and add them to the suitable host.
5. If a Hewlett-Packard UNIX (HP-UX) or Target Port Group Support (TPGS) host are created, click the **Host type** list (see Figure 7-40). then select your host type. If the specific host type is not listed, select **Generic**.

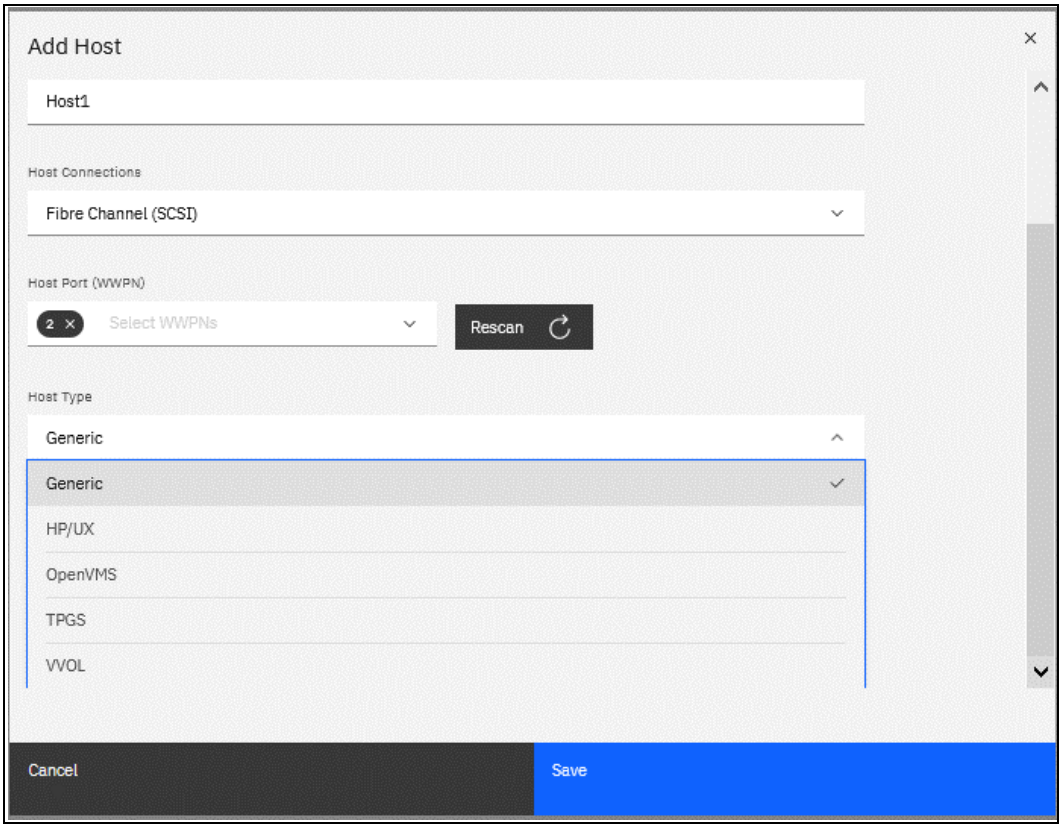


Figure 7-40 Host types selection

6. If setting up object-based access control (OBAC) as described in 12.7, “Encryption” on page 1098, select the **Advanced** section and choose the suitable ownership group from the **Ownership Group** menu, as shown in Figure 7-41 on page 559.

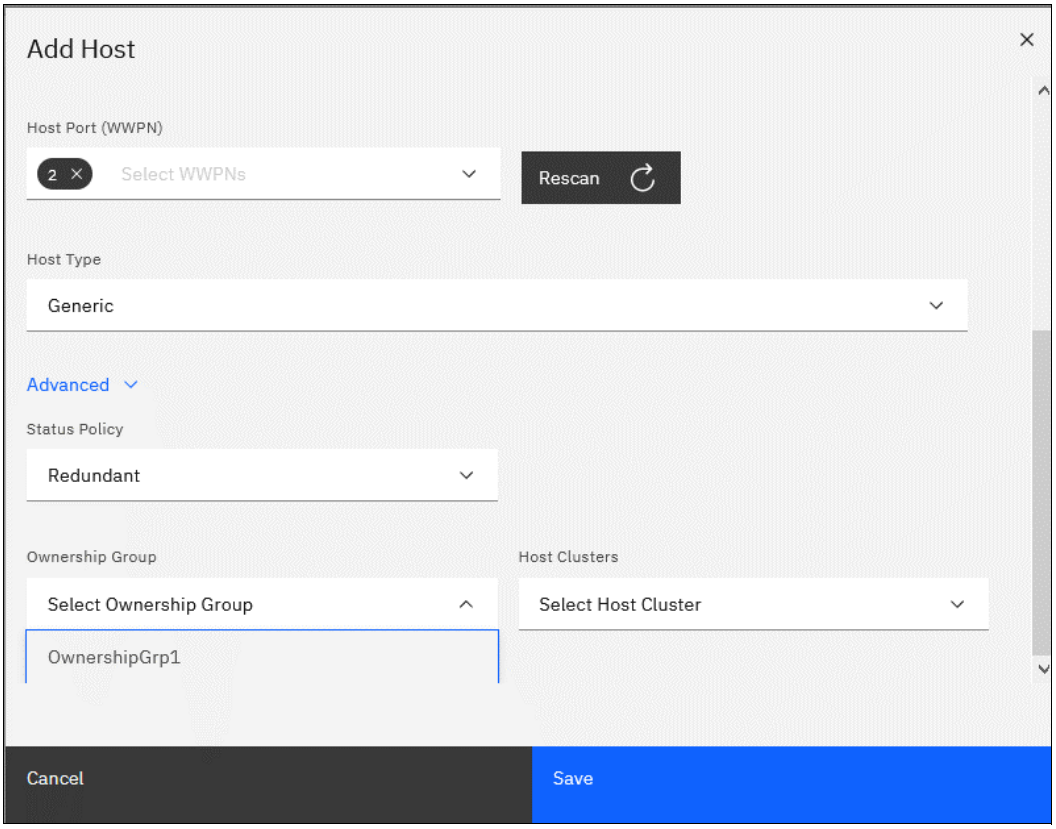


Figure 7-41 Adding a host to an ownership group

Note: If a host cluster object was created, the Host Clusters list appears in the Advanced section, as shown in Figure 7-41. Use this list to add a host to the cluster.

7. Click **Save** to create the host object.
8. Repeat these steps for all of your FC hosts. Figure 7-42 shows the All Hosts view after creating a host.

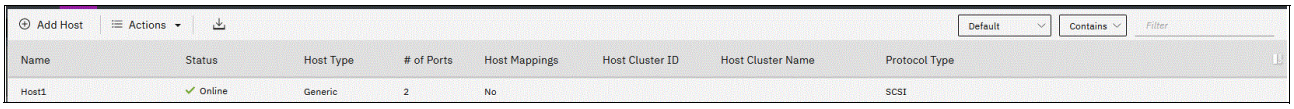


Figure 7-42 Hosts view after creating a host

Volumes can be created and mapped to the defined FC hosts as described in Chapter 6, “Volumes” on page 419.

Preparing for iSCSI connection

Before creating iSCSI host objects on IBM FlashSystem storage systems, ensure that iSCSI connectivity was correctly configured on the host initiators. Configuring the iSCSI connectivity varies based on host system and operating system.

To enable iSCSI connectivity, complete the following steps:

1. Select **Settings** → **Network**, and then click the **iSCSI** tab (see Figure 7-43).

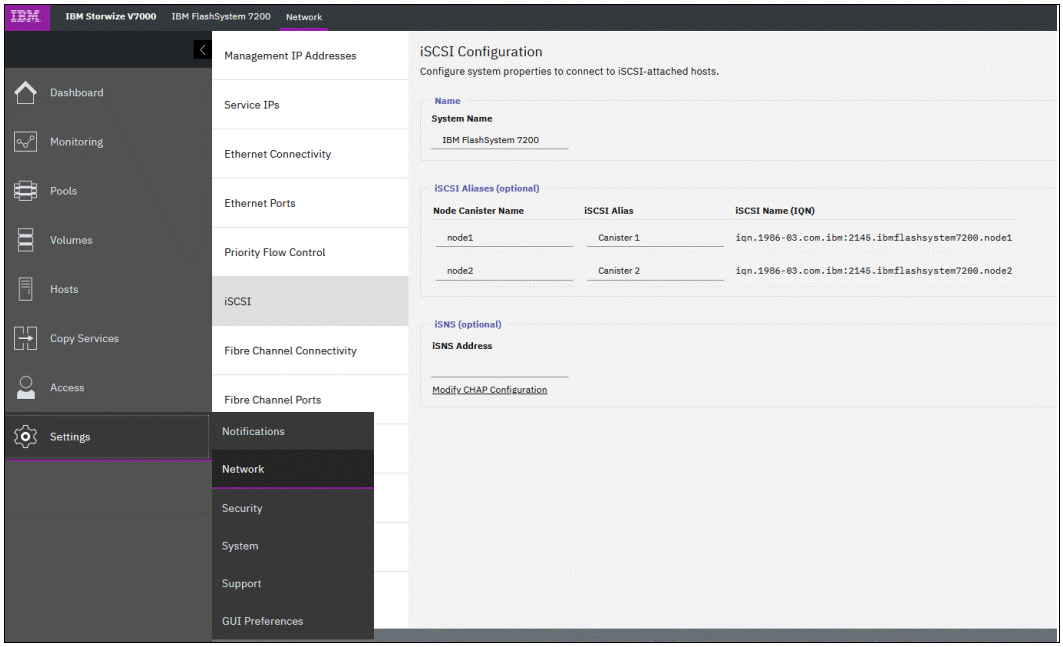


Figure 7-43 Network: iSCSI Configuration view

2. In the iSCSI Configuration window, enter or modify the system name or node names, and provide an optional iSCSI Alias for each node as needed (see Figure 7-44 on page 561).
3. Before continuing, select **Apply Changes** after the prompt to accept any modifications that are made. If suitable to the current environment, the internet Storage Name Service (iSNS) addresses and the Challenge Handshake Authentication Protocol (CHAP) can be configured in the lower left corner of the iSCSI Configuration window.

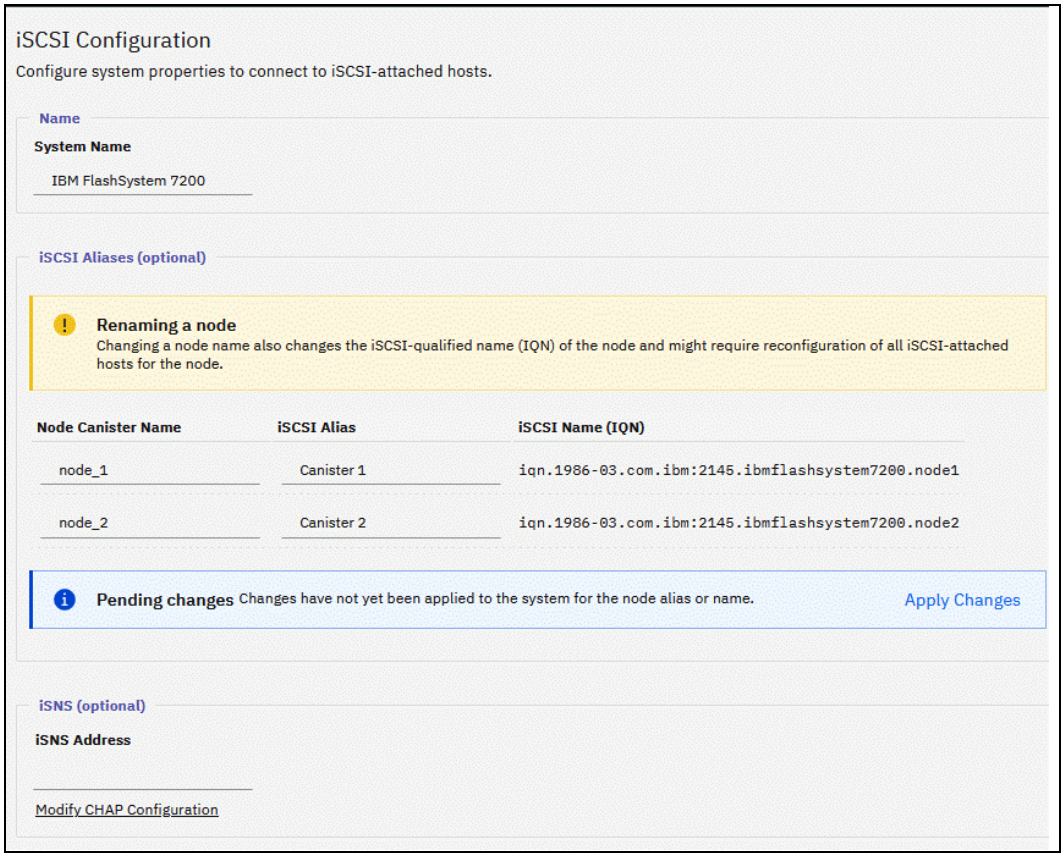


Figure 7-44 iSCSI Configuration modification

Notes:

- The host authentication is optional and is disabled by default. CHAP authentication can be enabled manually and involves sharing a CHAP secret between the cluster and the host. If the correct key is not provided by the host, the IBM FlashSystem storage system does not allow it to perform I/O to volumes.
- IBM FlashSystem systemwide CHAP is a two-way authentication method that uses CHAP to validate both the initiator and target. Changing or removing this systemwide CHAP can be disruptive, as it requires changes to the host configuration. If CHAP is changed without updating the host configuration, it can result in volume or LUN outages.

  For any changes to systemwide CHAP, the host iscsi configuration requires an update to ensure it should only use a valid target (FlashSystem) CHAP, which is being updated. The `node.session.auth.username_in` needs to be assigned as “FlashSystem clustername”, which can be seen with the **`lsnodecanister`** command.

For more information, see Setting up authentication for Linux hosts.

It is possible to have a unique target authentication CHAP for individual host initiators. By using the **`chhost`** command, individual host objects can offer a unique username and chapsecret. However, the target FlashSystem has a common chapsecret and username to offer for initiator authentication. Using the CLI command **`ls`** or **`chsystem`** can help with managing chapsecret, and username is name of the cluster.

When using the SendTarget discovery method with FlashSystem, the IP addresses that are part of the same node and belong to the same portsets are returned in response. This means that if you have 8 nodes, you will need to perform iSCSI SendTarget discovery 8 times to get all of the IP addresses for the target.

4. Configure the Ethernet ports that are to be used for iSCSI communication by selecting **Settings** → **Network** and then clicking the **Ethernet Ports** tab to see a list of the available ports and their corresponding IP addresses. Highlight the port to set the iSCSI IP information and select **Actions** → **Manage IP Addresses** (see Figure 7-45)**.**

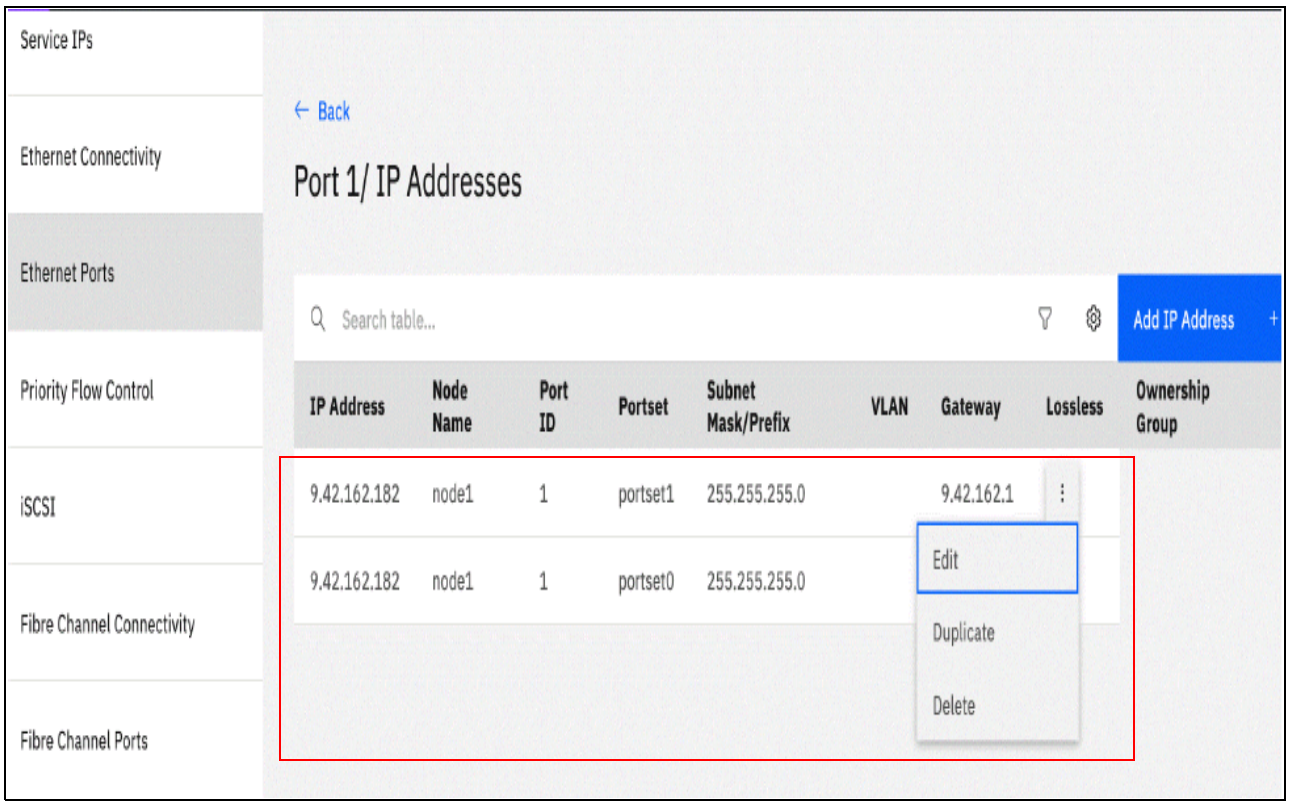


Figure 7-45 Select Manage IP Address from the "Actions" menu

5. Add an IP address by clicking the **Add IP Address** button at the top right of the window. Other options include editing, duplicating, and deleting as shown in Figure 7-46.

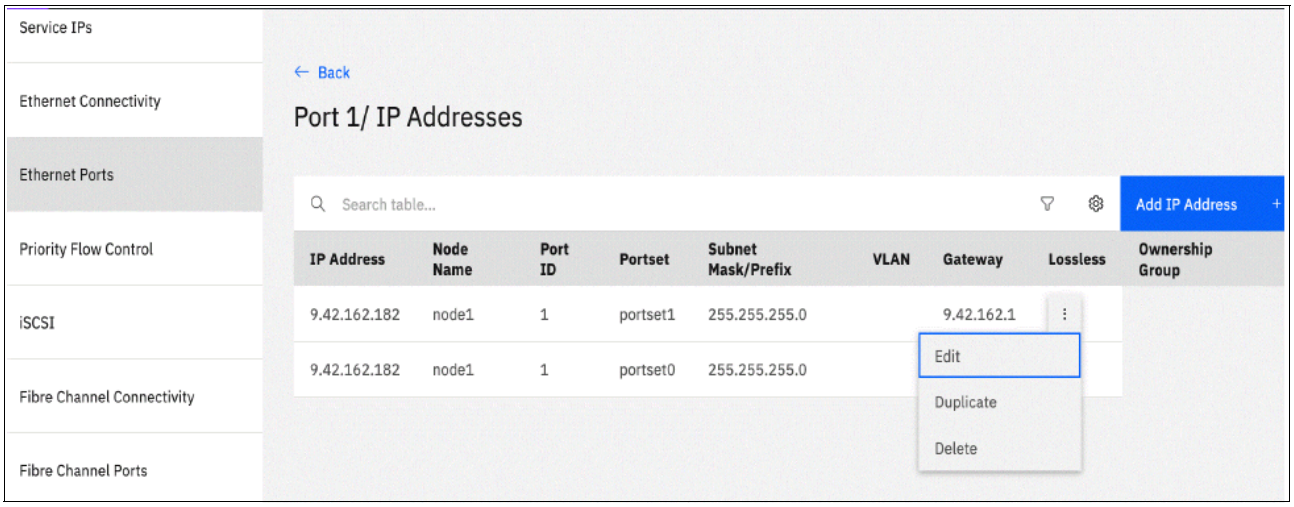


Figure 7-46 Adding, deleting, or modifying IP addresses

6. After the ports are configured, multiple options are available from the Actions drop-down menu, as shown in Figure 7-47. For example, to disable any interfaces that are not required for host connections (and might be used for replication only), select the configured port and then select **Actions** (or right-click while hovering over the chosen port) and select **Modify iSCSI Hosts**.

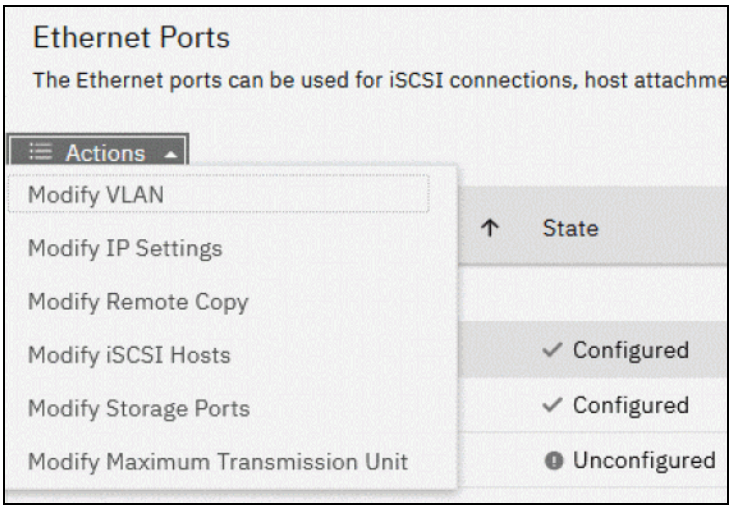


Figure 7-47 Available actions with configured ports

7. Make any necessary changes as prompted in the dialog box (see Figure 7-48) and then click **Modify**.

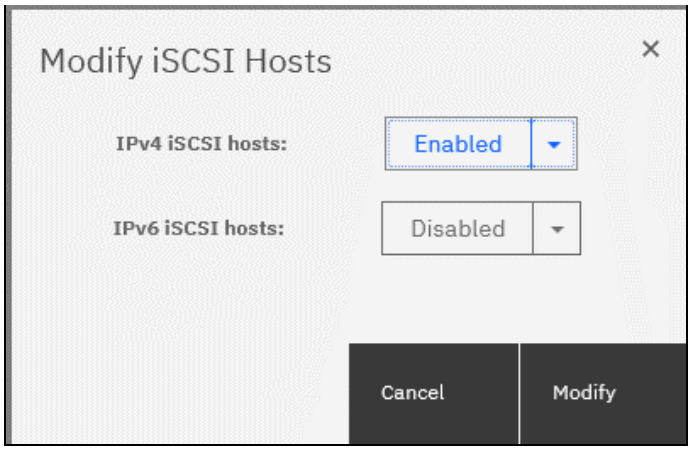


Figure 7-48 Modifying the port for host connectivity

8. A best practice is to isolate iSCSI traffic in a separate subnet or a virtual local area network (VLAN).

   To enable the VLAN, select **Actions** → **Modify VLAN**, as shown in Figure 7-49. The system notification states that at least two ports are affected.

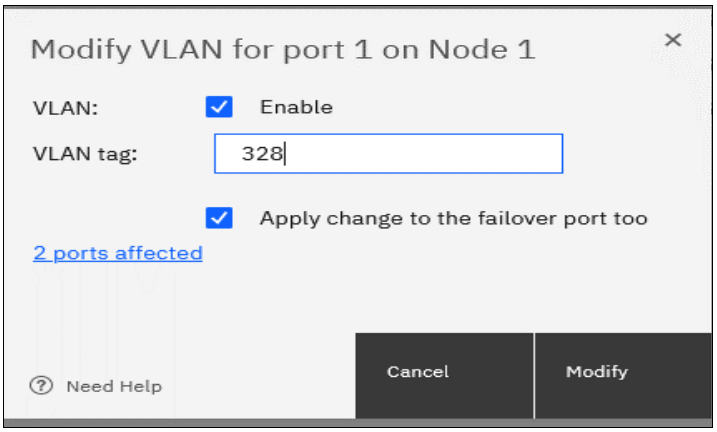


Figure 7-49 VLAN settings modification interface

9. To view the details, click **2 ports affected** (see Figure 7-50). Make any necessary changes and then click **Modify**.

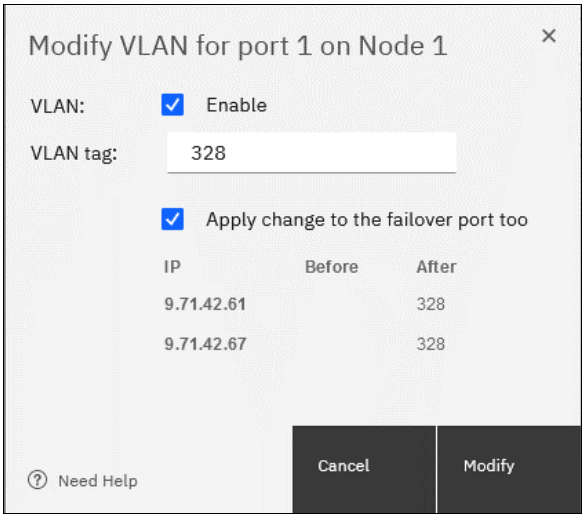


Figure 7-50 VLAN settings: Details

The system is now configured and ready for iSCSI host use. Make a note the initiator iSCSI Qualified Name (IQN) names of storage node canisters (see Figure 7-44 on page 561) because they are necessary to configure access from the host to the storage.

For more information about creating volumes and mapping them to a host, see Chapter 6, “Volumes” on page 419.

Creating iSCSI host objects

When creating an iSCSI-attached host, consider the following points:

- iSCSI IP addresses can fail over to the partner node in the I/O group if a node canister fails. This fail-over mechanism reduces the need for multipathing support in the iSCSI host.
- The IQN of the host is added to an IBM FlashSystem host object in the same manner as adding FC WWPNs. For more information about obtaining the IQN from the host, see the examples in 7.11.3, “iSCSI host connectivity and capacity allocation” on page 613.
- Host objects can have WWPNs and IQNs.
- Standard iSCSI host connection procedures can be used to discover and configure the IBM FlashSystem systems as an iSCSI target.
- The IBM FlashSystem system supports the CHAP authentication methods for iSCSI.
- The name `iqn.1986-03.com.ibm:2076.<cluster_name>.<node_name>` is the IQN for an IBM FlashSystem node. Because the IQN contains the clustered system name and the node name, do *not* change these names after iSCSI is deployed.

  It is possible to check the IQN name and iSCSI configuration in the cluster’s GUI by selecting **Settings** → **Network** → **iSCSI**, as shown in Figure 7-51 on page 566.

  Note: Validate the iSCSI configuration *before* creating iSCSI host objects (configuring the hosts) because some modifications can require a change to the host configuration or that the host object be redefined.

- Each node can be given an iSCSI alias as an alternative to the IQN.

To create iSCSI host objects, complete the following steps:

1. In the left pane, select **Hosts** → **Hosts** → **Add Host** (in the host view) to open the host creation window (see Figure 7-51). Choose **iSCSI** in **Host Connections** list. Note that because the CHAP authentication trigger is on, the fields for CHAP credentials are shown in the interface. If CHAP is not used, turn off the trigger.

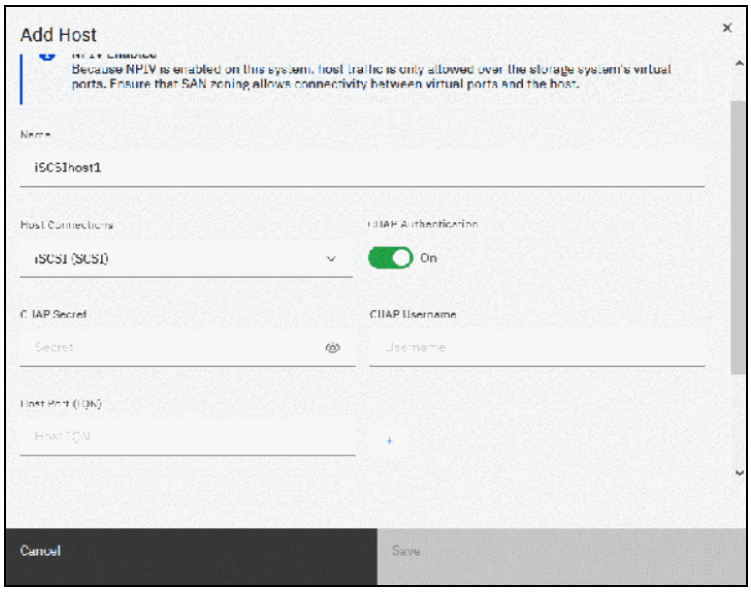


Figure 7-51 Adding the iSCSI host object to the configuration

2. Enter CHAP authentication credentials (if suitable), and then the hostname into the **Name** field. then enter the iSCSI initiator name into the **iSCSI host IQN** field. Click the plus sign (**+**) to add initiator names to the host.
3. To connect to an HP-UX or TPGS host, click the **Host type** field (it might be necessary to scroll down the window), and then select the correct host type. For a VMware Elastic Sky X (ESX) host, select **VVOL**. However, if VMware vSphere virtual volumes (VVOLs) are not used, select **Generic.**
4. Click **Save** to complete the host object creation.
5. Repeat these steps for every iSCSI host that must be created. Figure 7-52 shows the Hosts view window after creating the FC host and iSCSI host.



Figure 7-52 Defined hosts list

rCeating FC NVMe host objects

The process for creating FC NVMe hosts is similar to that of creating SCSI FC hosts, except that it uses host port NQN instead of WWPNs.

To configure an FC-NVMe host, complete the following steps:

1. Go to the Host view by selecting **Hosts** → **Hosts** and then click **Add Host**. In the **Host Connections** menu, select **Fibre Channel (NVMe)**, as shown in Figure 2.

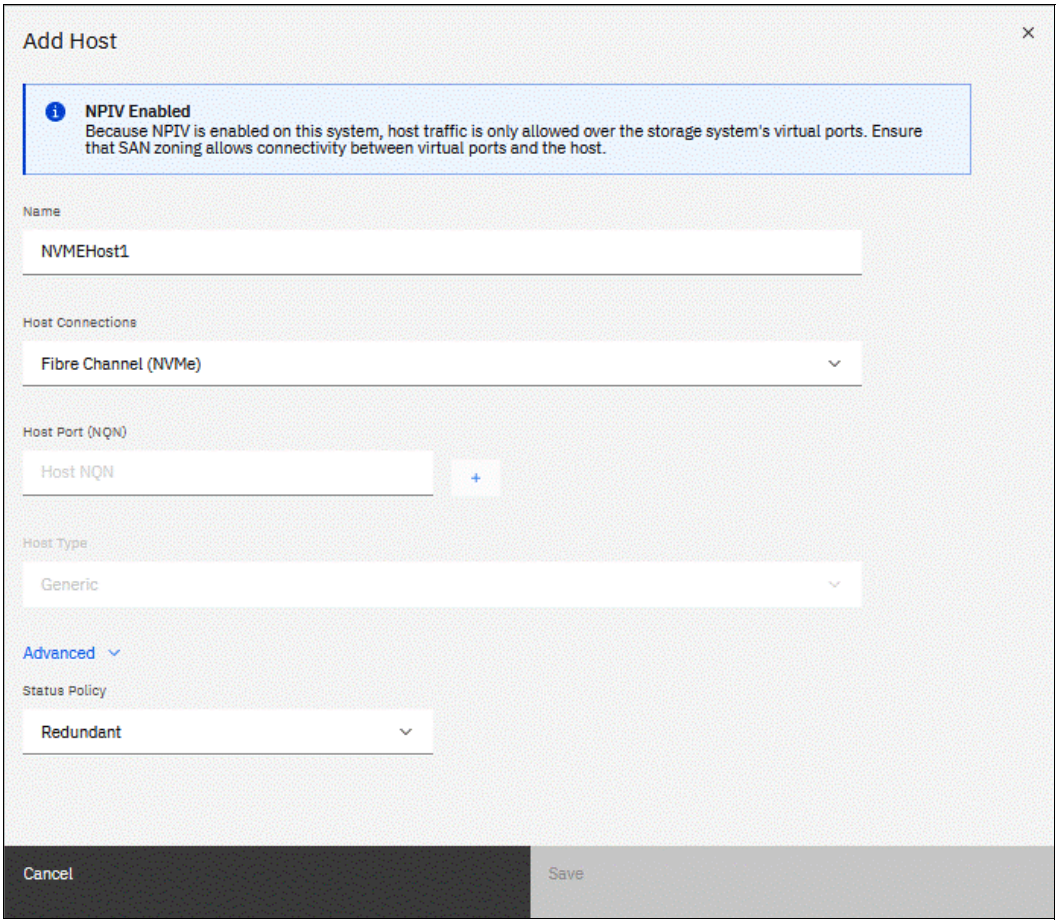


Figure 7-53 Creating an FC-NVMe hos

Note: To see if hosts and IBM Storage Virtualize system are compatible, see the SSIC.

2. tEnter the hostname and NQN of the host, as shown in Figure 7-54 on page 568. (For more information about how to obtain the host NQN, see 7.11.4, “FC NVMe over Fabric host connectivity example” on page 616.) Click the **+** button next to the field to add multiple NQNs.

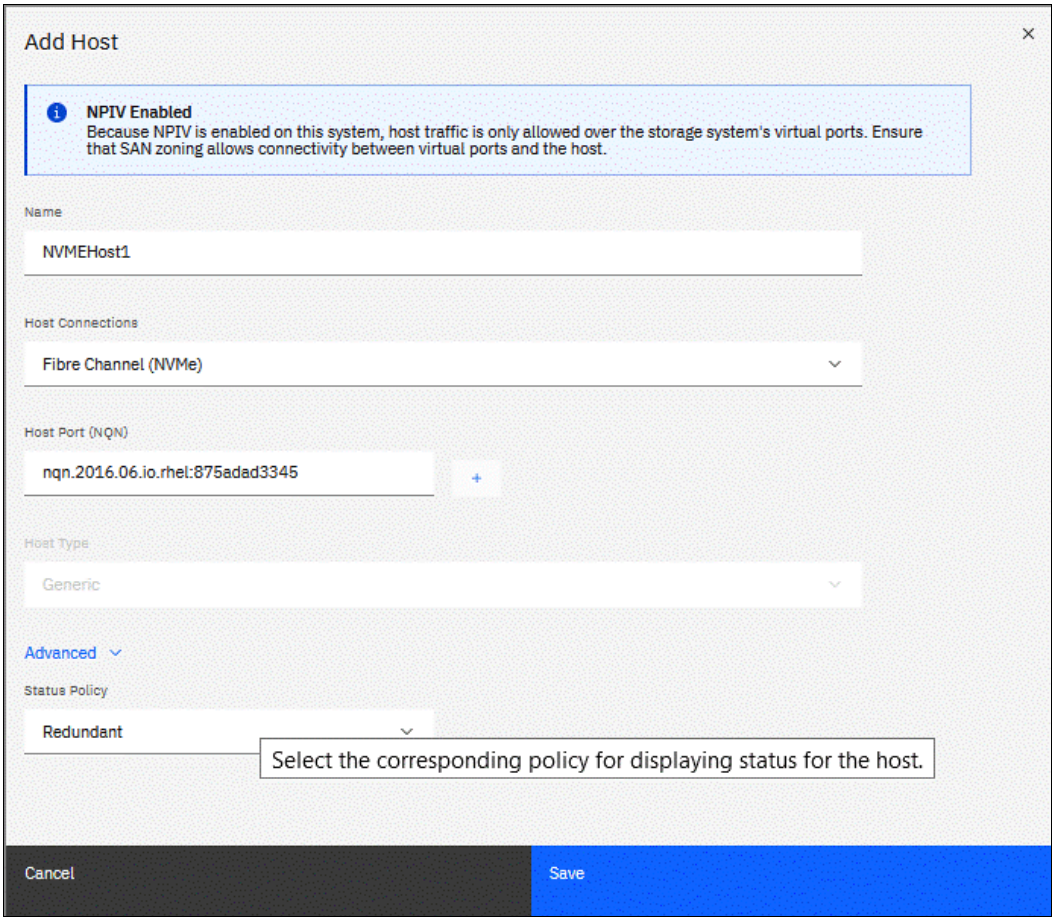


Figure 7-54 Defining the NQN

3. Click **Save**. The host appears in the defined host list, as shown in Figure 7-55.

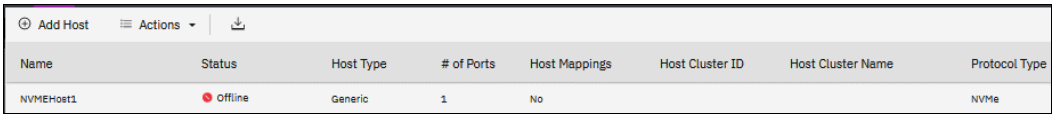


Figure 7-55 NVMe host created

Note: As shown in Figure 7-55, hosts that are not yet connected to the system or are offline can be added by using their known NQN. In this case, their status is `offline` until they are connected or turned on.

4. The storage system I/O group NQN must be configured on the host so it can access the mapped capacity. If the zoning is done correctly, automatic discovery can be started from the host to find the NQN of the I/O group.

To discover the I/O group NQN, run the **lsiogrp** command (see Example 7-20).

Example 7-20 The lsiogrp command

```
IBM_IBM FlashSystem:GLTLoaner:superuser>lsiogrp 0
id 0
name io_grp0
node_count 2
vdisk_count 8
host_count 1
flash_copy_total_memory 20.0MB
flash_copy_free_memory 20.0MB
remote_copy_total_memory 20.0MB
remote_copy_free_memory 20.0MB
mirroring_total_memory 20.0MB
mirroring_free_memory 20.0MB
raid_total_memory 350.0MB
raid_free_memory 310.2MB
maintenance no
compression_active no
accessible_vdisk_count 8
compression_supported yes
max_enclosures 20
encryption_supported yes
flash_copy_maximum_memory 2048.0MB
site_id
site_name
fctargetportmode enabled
compression_total_memory 0.0MB
deduplication_supported yes
deduplication_active no
nqn nqn.1986-03.com.ibm:nvme:2145.000002042140049E
```

The host can now be configured to use the IBM FlashSystem storage system as a target.

Note: For more information about a compatibility matrix and supported hardware, see this IBM Documentation web page and the SSIC.

Creating NVMe over RDMA host objects

The process for creating RDMA NVMe hosts is similar to that of creating FC NVMe hosts, except that it uses host port NQN instead of WWPNs.

Note: To see whether your hosts and IBM Storage Virtualize system are compatible, see the SSIC.

To configure an NVMe over RDMA host, complete the following steps:

1. Go to the Host view by selecting **Hosts** → **Hosts** and then click **Add Host**. In the **Host connections** menu, select **RDMA (NVMe)**, as shown Figure 7-56.

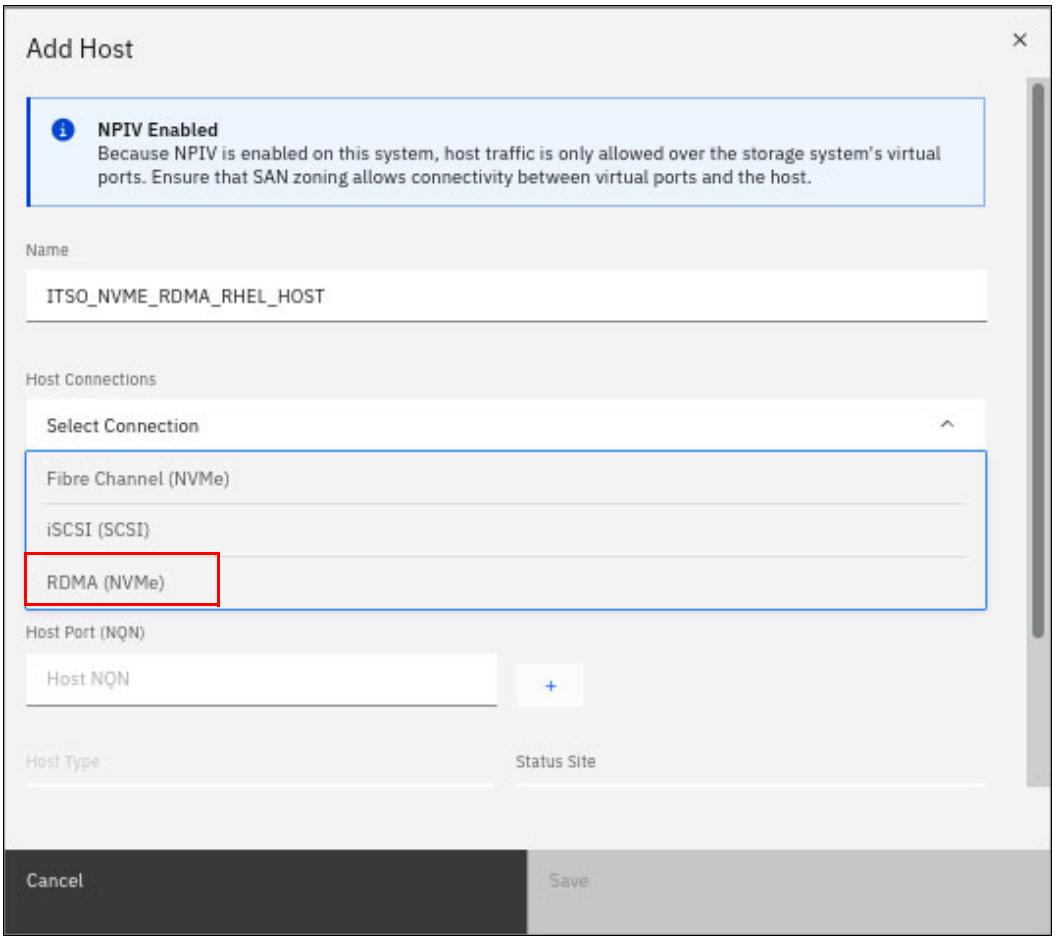


Figure 7-56 Creating an NVMe over RDMA host

2. Enter the hostname and NQN of the host, as shown in Figure 7-57. For more information about how to obtain the host NQN, see 7.11.5, “NVMe over RDMA host connectivity example” on page 622.

   Click the **+** button next to the field to add multiple NQNs.

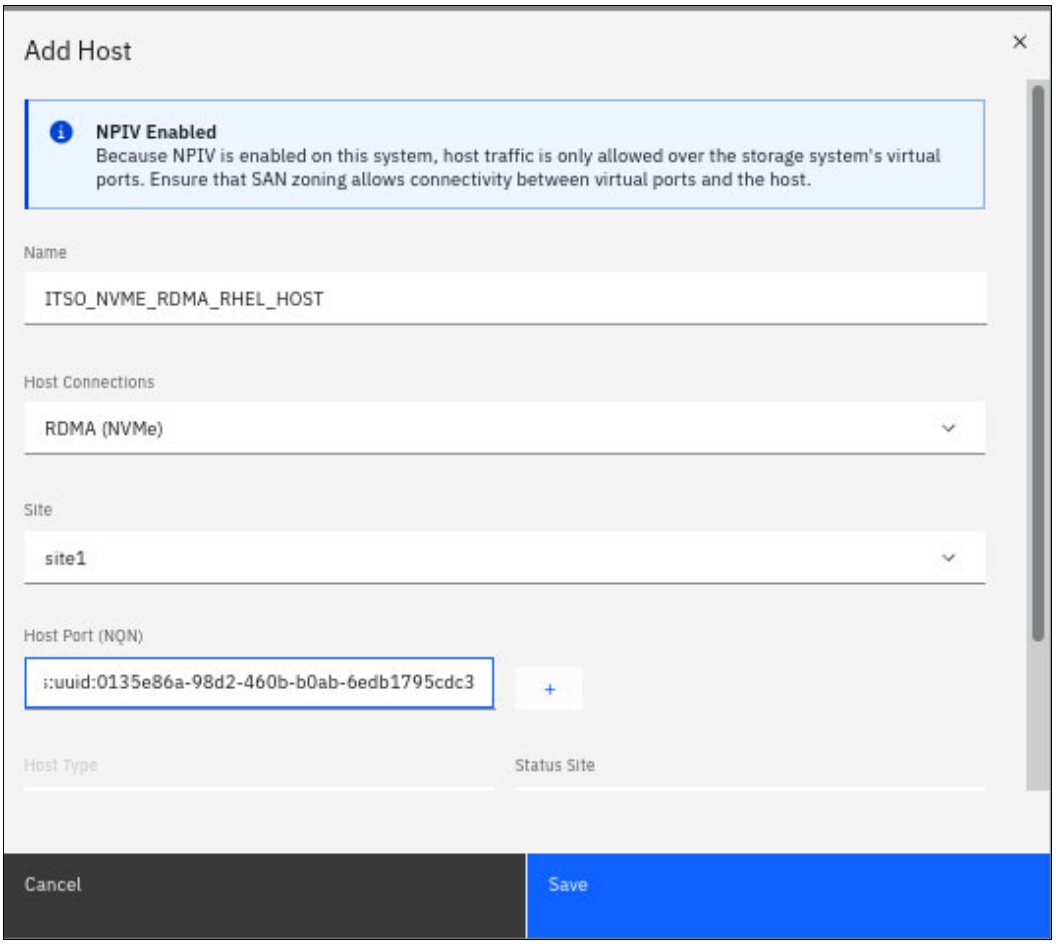


Figure 7-57 Defining NQN for RDMA over NVMe host

3. Click **Save**. The host appears in the defined host list, as shown in Figure 7-58.



Figure 7-58 RDMA over NVMe host created

Note: As shown in Figure 7-58, hosts that are not yet connected to the system or are offline can be added by using their known NQN. In this case, their status is offline until they are configured and connected, as described in 7.11.5, “NVMe over RDMA host connectivity example” on page 622.

Creating NVMe over TCP host objects

The process for creating TCP NVMe hosts is similar to that of creating FC NVMe hosts, except that it uses host port NVMe Qualified Name (NQN) instead of WWPNs..

Note: To see whether your hosts and IBM Storage Virtualize system are compatible, see the SSIC.

To configure an NVMe over TCP host, complete the following steps:

1. Go to the Host view by selecting **Hosts** → **Hosts** and then click **Add Host**. In the **Host connections** menu, select **TCP (NVMe)**, as shown Figure 7-59.

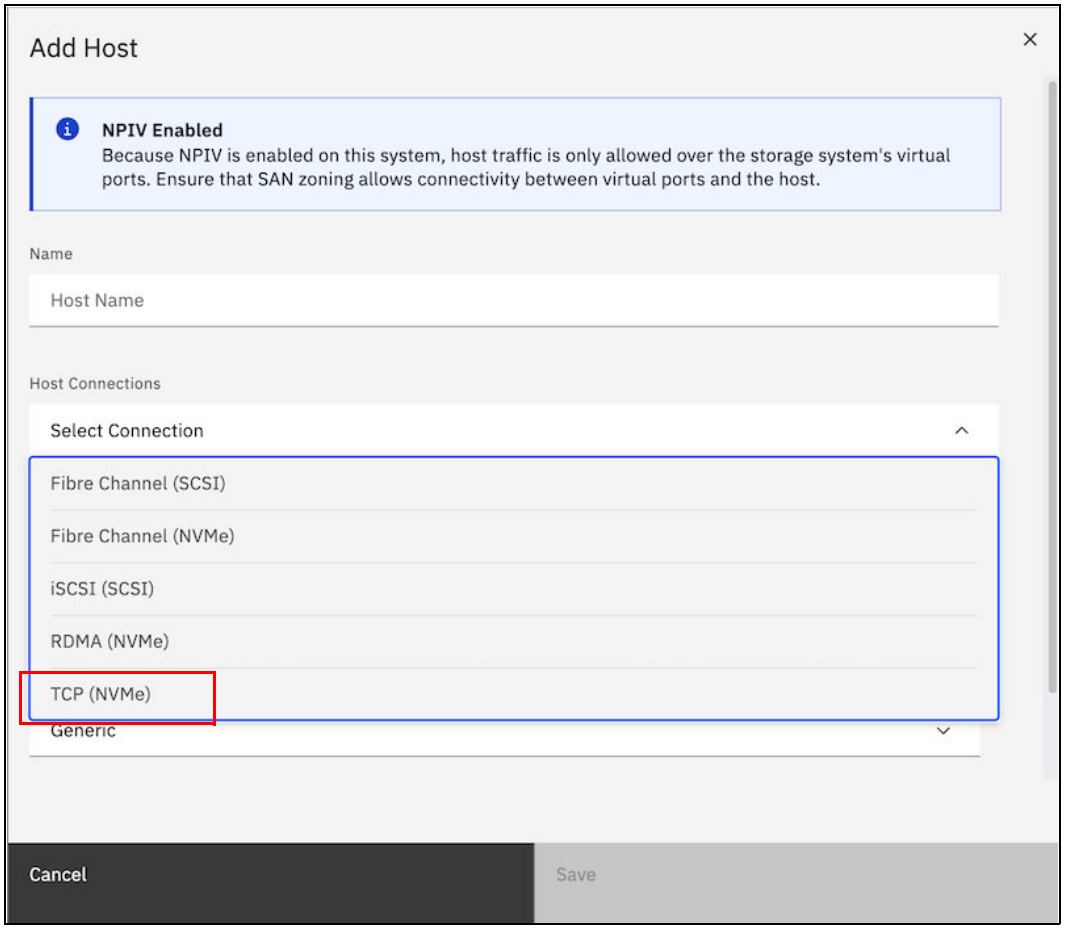


Figure 7-59 Creating an NVMe over TCP host

2. Enter the hostname and NQN of the host, as shown in Figure 7-60. For more information about how to obtain the host NQN, see 7.11.6, “NVMe over TCP host connectivity example” on page 627.

   Click the **+** button next to the field to add multiple NQNs.

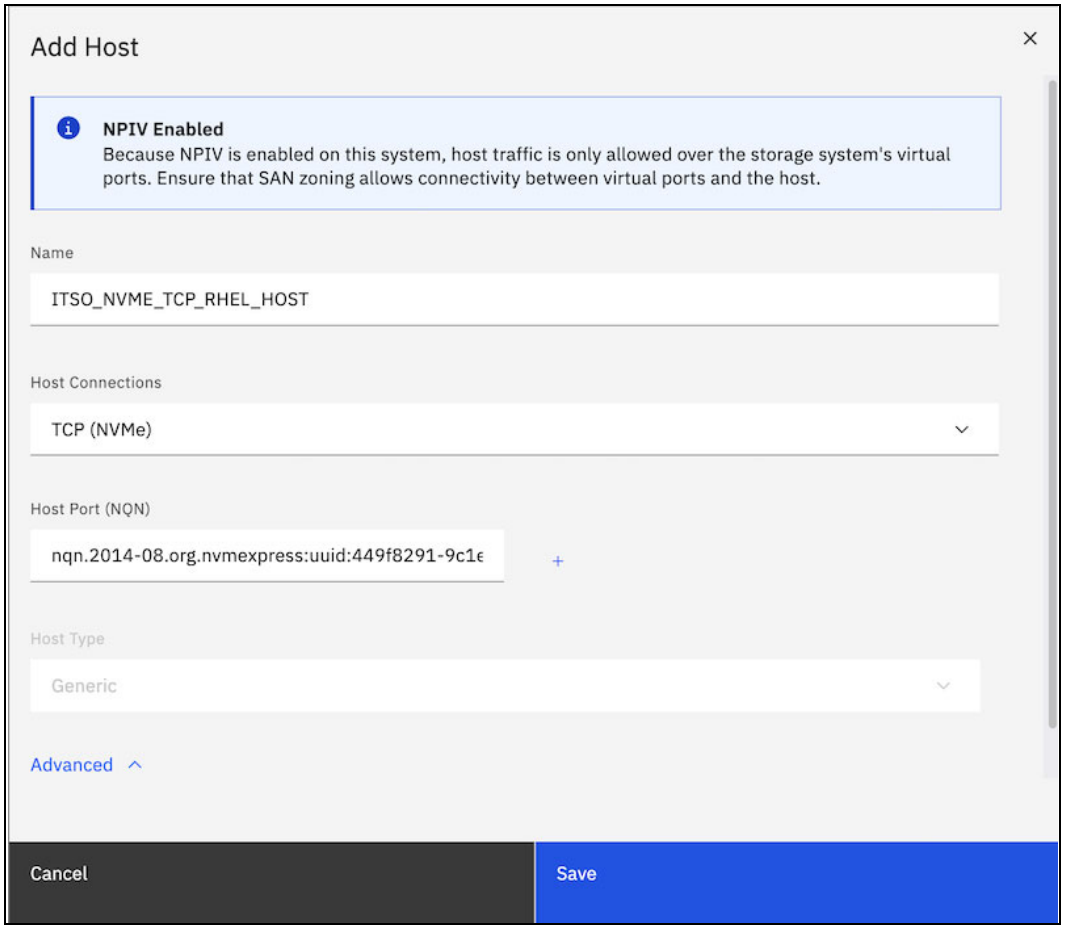


Figure 7-60 Defining NQN for TCP over NVMe host

3. Click **Save**. The host appears in the defined host list, as shown in Figure 7-61.



Figure 7-61 TCP over NVMe host created

Note: As shown in Figure 7-61, hosts that are not yet connected to the system or are offline can be added by using their known NQN. In this case, their status is offline until they are configured and connected, as described in 7.11.6, “NVMe over TCP host connectivity example” on page 627.

7.9.2 Host clusters

A host cluster object enables multiple individual hosts to be grouped and treated as a single entity.

The host cluster object is useful for hosts that are clustered on operating system levels, such as Microsoft Clustering Server, IBM PowerHA, Red Hat Cluster Suite, and VMware ESX. By defining a host cluster object, a user can map one or more volumes to this host cluster object.

As a result, a volume or set of volumes can be mapped and accessed by all individual host objects that are included in the host cluster object. The volumes are mapped by using the same SCSI ID to each host that is part of the host cluster by running a single command.

Although a host can be a part of a host cluster object, volumes are still assigned to an individual host in a *non-shared manner*. It is possible to create one policy that pre-assigns a standard set of SCSI IDs for volumes for shared use and another that generates a set of SCSI IDs for individual non-shared assignments to hosts.

For example, SCSI IDs 0–100 can be designated for individual host assignment while SCSI IDs 101 and greater can be designated for used by host cluster. By using such a policy, specific volumes are not shared, while common volumes for the host cluster can be shared.

Creating a host cluster

Complete the following steps to create a host cluster (it is assumed that individual hosts were created):

1. From the menu on the left side, select **Hosts** → **Host Clusters** (see Figure 7-62).

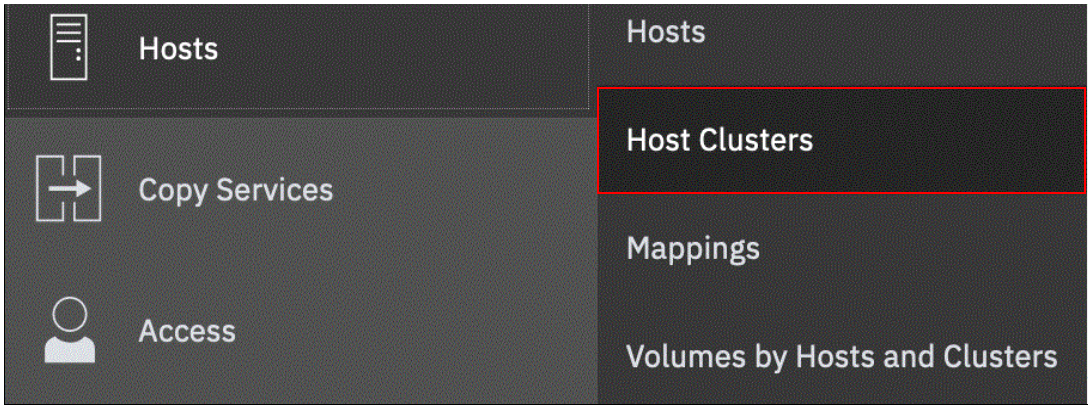


Figure 7-62 Host clusters menu

2. Click **Create Host Cluster** to open the wizard that is shown in Figure 7-63.

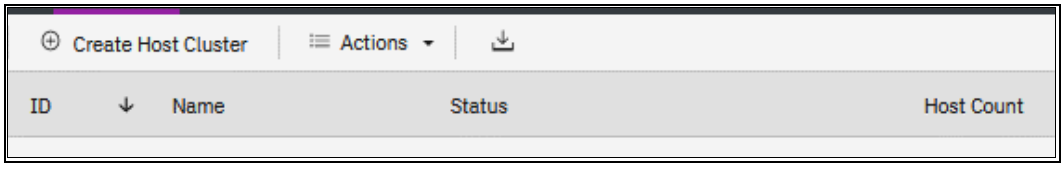


Figure 7-63 Create Host Cluster window

3. Enter a cluster name, and the suitable ownership group, if applicable. then highlight the individual hosts to be included in the cluster object by pressing the **Ctrl** or **Shift** key and selecting them, as shown in Figure 7-64. Click **Next**.

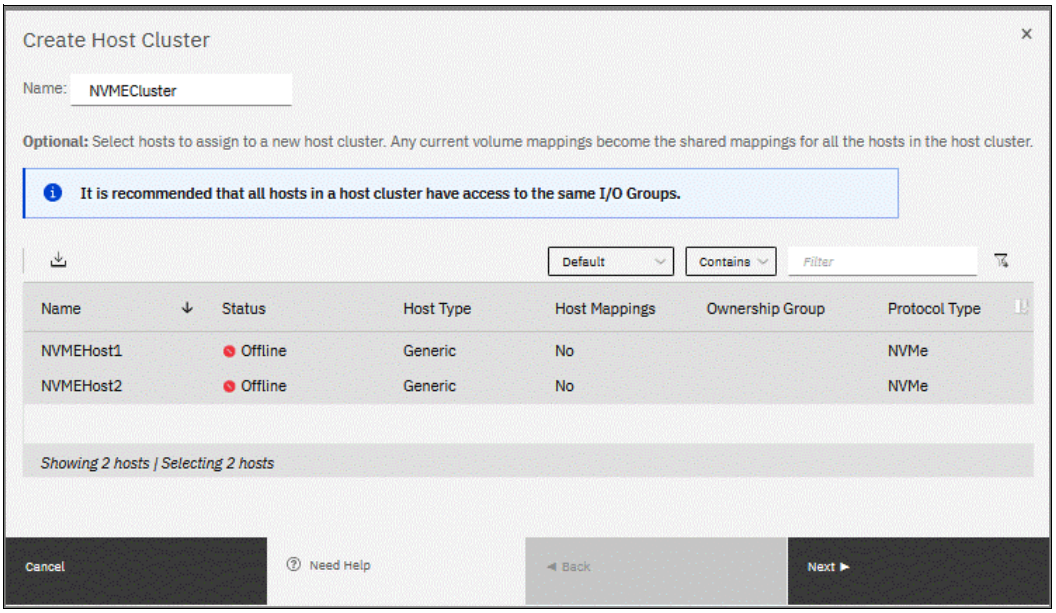


Figure 7-64 Host Cluster details definition

4. Verify that the host selection is correct in the Create Host Cluster Summary window and then click **Make Host Cluster** (see Figure 7-65).

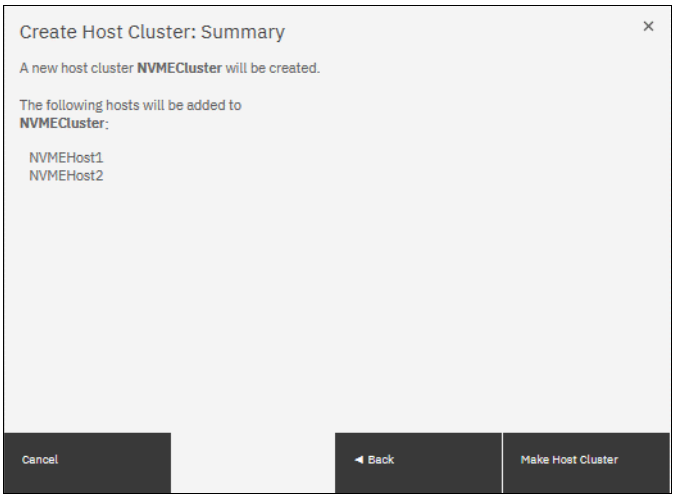


Figure 7-65 Create Host Cluster: Summary

5. When the task is completed, the newly created cluster is visible in the Host Clusters view (see Figure 7-66).

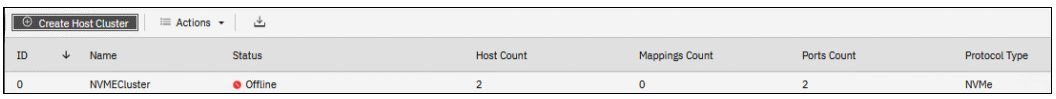


Figure 7-66 Host Clusters view

Note: The host cluster status depends on its member hosts. One offline or degraded host sets the host cluster status as `Degraded`. If all member hosts are offline, the cluster status is set to `Offline`.

Multiple options for cluster configuration and management are available from the Host Clusters view. These options can be accessed by selecting a cluster and clicking **Actions** (see Figure 7-67).

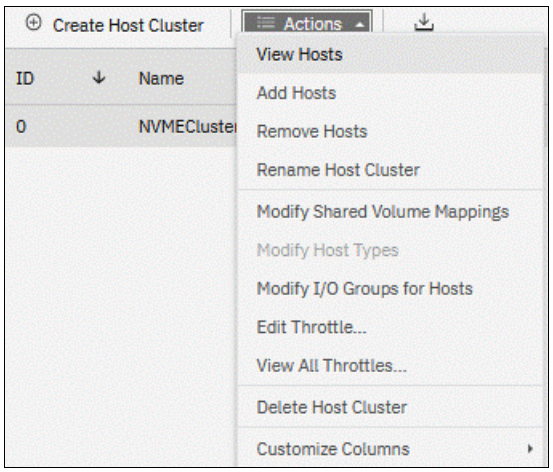


Figure 7-67 Host Clusters Actions menu

From the Actions menu, the following tasks are available:

- View Hosts: View the hosts status within the cluster.
- Add Hosts or Remove Hosts: Add or remove hosts from the cluster.
- Rename Host Cluster: Rename the host cluster.
- Modify Shared Volume Mappings: Add or remove volumes that are mapped to all hosts in the cluster while maintaining the same SCSI ID for all hosts.
- Modify Host Types: Change the host type; for example, from generic to VVOLs.
- Modify I/O Groups for Hosts: Assign or restrict volume access to specific I/O groups.
- Edit Throttle: Restrict the megabytes per second (MBps) or I/O operations per second (IOPS) bandwidth for the host cluster.
- View All Throttles: Show all throttling settings, and allow for changing, deleting, or refining throttle settings.
- Delete Host Cluster: Delete a host cluster.
- Customize Columns: Modify which columns are displayed that show the properties of the host cluster.

For more information about these actions, see 7.9.4, “Actions on host clusters” on page 588.

7.9.3 Actions on hosts

This section describes host administration, including host modification, host mappings, and deleting hosts. The basic host creation process is described in 7.9.1, “Creating hosts” on page 555.

To see a list of available actions that can be taken on hosts, select **Hosts** → **Hosts** view and right-click one of the hosts or click **Actions**. (see Figure 7-68).

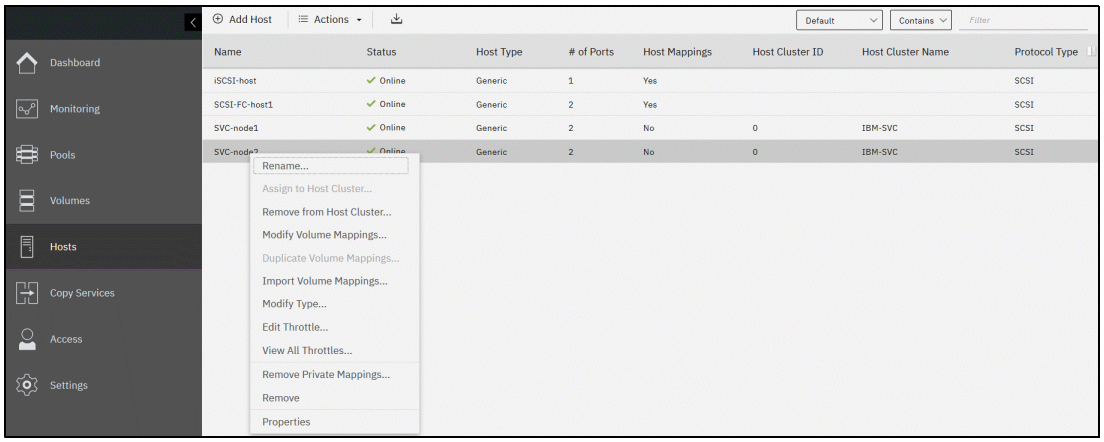


Figure 7-68 Actions on hosts

The following actions are available and described in this section:

- Renaming a host
- Assigning or removing a host to or from a host cluster
- Modifying volume mappings
- Duplicating and importing mappings
- Modifying the host type
- Viewing and editing throttles
- Removing private mappings from a host
- Removing a host
- Viewing IP logins
- Viewing the host properties

Renaming a host

To rename a host, complete the following steps:

1. Select the host, right-click it and then select **Rename**.
2. Enter a new name and click **Rename** (see Figure 7-69). Clicking **Reset** reverts the changes to the original hostname.

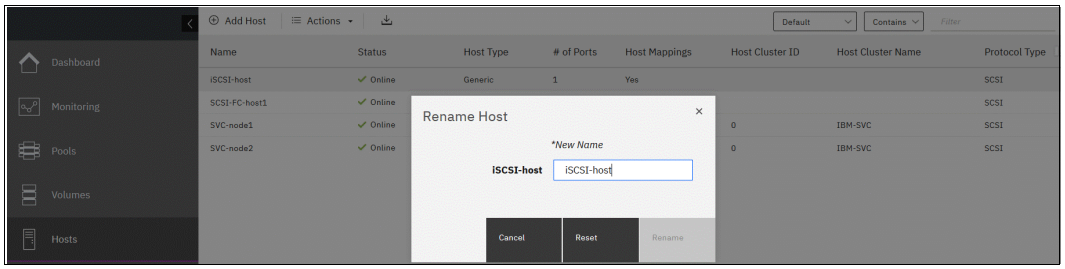


Figure 7-69 Renaming a host

Assigning or removing a host to or from a host cluster

To assign a host to a cluster, complete the following steps:

1. Right-click the host or a set of hosts that you want to add and select **Assign to Host Cluster**. To select multiple objects, press and hold the **Ctrl** key and click each host to be assigned. (or, press and hold the **Shift** key and click the first and last objects to be selected).
2. Select the cluster to which the host is to be added (see Figure 7-70) and click **Next**.

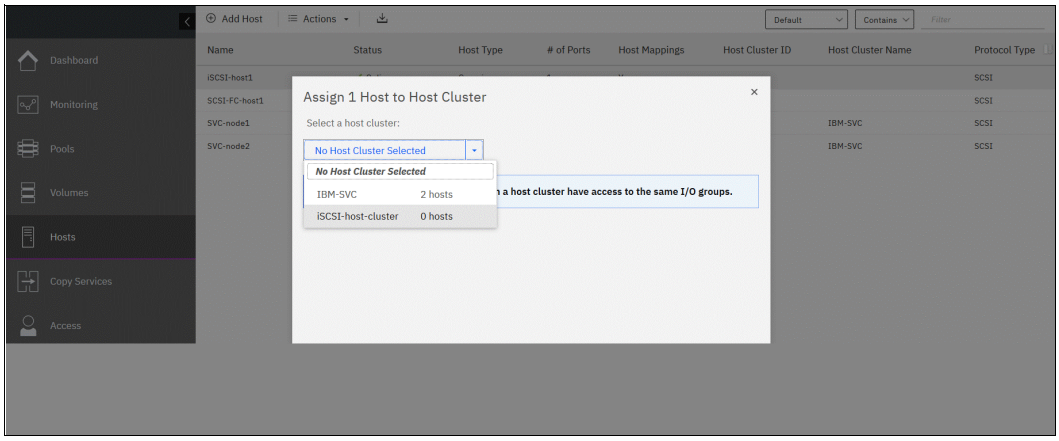


Figure 7-70 Assigning a host to a cluster

3. The IBM Storage Virtualize system checks for SCSI ID conflicts. In a host cluster, all hosts must include the same SCSI IDs for a mapped volume. For example, a single volume cannot be mapped with `SCSI ID 0` to one host and with `SCSI ID 1` to another host.

   If no SCSI ID conflict is detected, the system provides a list of configuration settings for verification (see Figure 7-71). Click **Assign** to complete the operation. When the operation completes, the host are included in all host cluster volume mappings.

> **Note:** The Assign to Host Cluster action is active and visible only for a host that does not belong to the cluster and if at least one host cluster object exists.

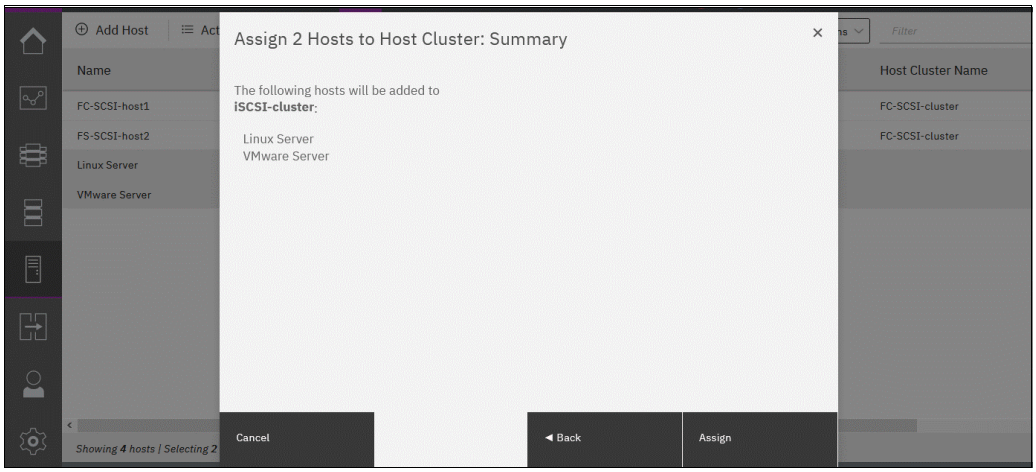


Figure 7-71 Assign host to host cluster confirmation

If a host has a private volume mapping that uses the same SCSI IDs as a host cluster shared mapping, a SCSI ID conflict appears (see Figure 7-72). In this case, the host cannot be assigned to the host cluster. First, the ID conflict must be resolved by removing the private host volume mappings or by changing the assigned SCSI IDs for conflicting mappings.

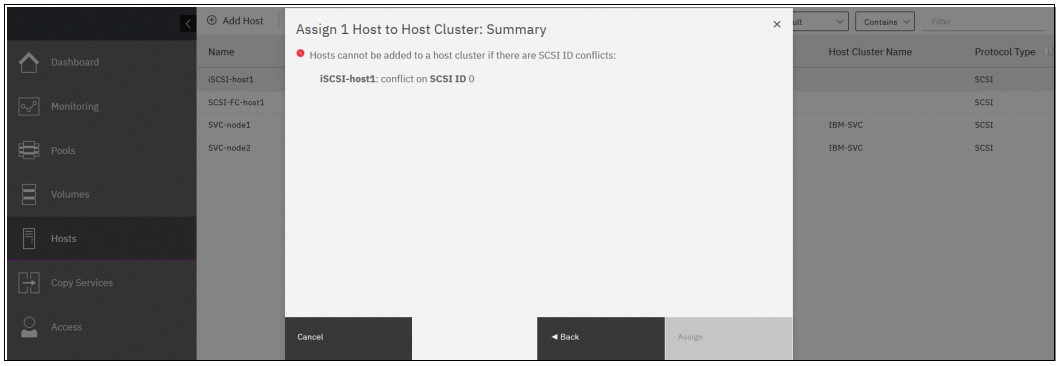


Figure 7-72 Conflict between private and shared volume mappings

To remove a host from a host cluster, complete the following steps:

1. Select a host or a group of hosts, right-click them and then click **Remove from Host Cluster.**
2. Complete the required information in the dialog window (see Figure 7-73). Verify the list of hosts to be removed and determine what to do with the host mappings. Click **Remove Hosts**.

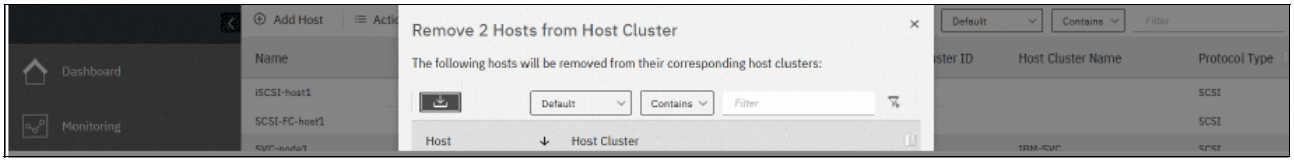


Figure 7-73 Removing a host from a host cluster

Modifying volume mappings

Private volume mappings can be modified for a single host by using the Modify Volume Mappings action.

To modify volume mappings, complete the following steps:

1. Right-click a host and select **Modify Volume Mappings**.

2. Verify the correct private volume mappings are displayed in the pop-up window (see Figure 7-74).

> **Note:** Host cluster shared mappings are not shown in this view. Only host private mappings are listed. To modify the shared host cluster mappings, use another GUI view, as described 7.9.4, “Actions on host clusters” on page 588.

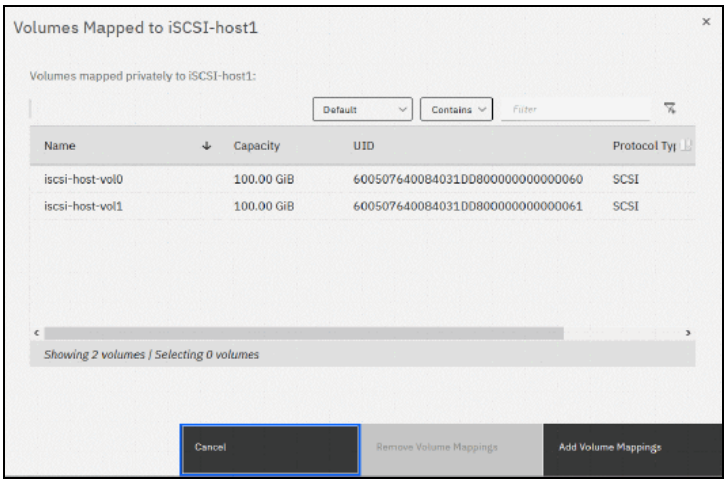


Figure 7-74 Modifying the host volume mappings

3. Select the volume mappings to be deleted, and click **Remove Volume Mappings**. Verify the changes in the next window and complete the removal procedure.
4. Click **Add Volume Mappings** to add a private mapping to the volume. A list of available volumes is presented (see Figure 7-75). Any volume that has a private mapping to this host, or has a shared mapping with a host cluster that includes this host, does not appear in the list.

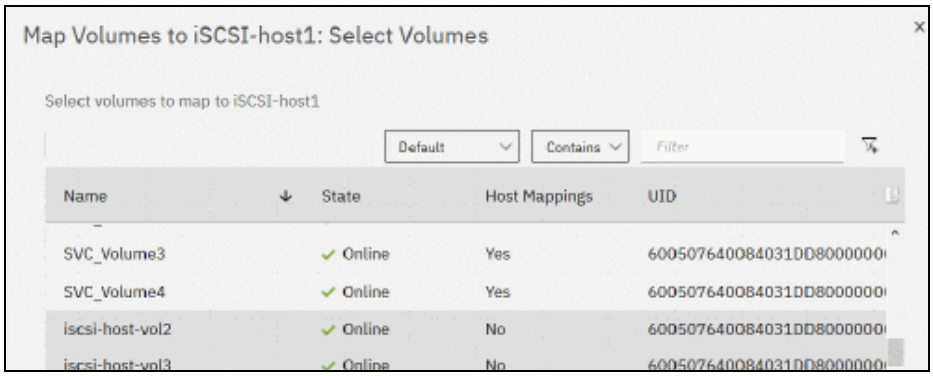


Figure 7-75 Adding private mappings

If the volume is mapped to another host or host cluster, the word Yes appear next to it in the Host Mappings column. Any attempt to map that volume to the host generates a warning message (see Figure 7-76). However, the mapping can be added if access is coordinated on the host side.

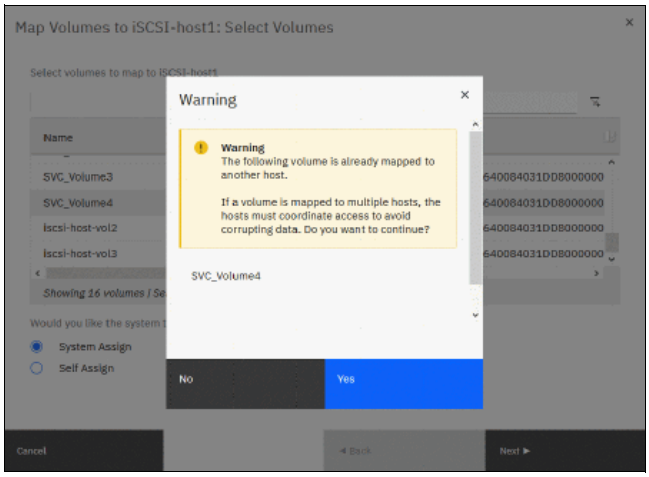


Figure 7-76 Warning that a mapping to another host exists

The storage system automatically assigns the next available SCSI ID for new mappings by default. However, the SCSI logical unit number (LUN) IDs can be assigned manually by clicking **Self Assign** (Figure 7-77). In Figure 7-74 on page 580, only two mappings are shown for this host. However, Figure 7-77 shows three mappings because the third mapping is a shared host cluster mapping, which was not shown in previous views.

Note: The SCSI ID of the volume can be changed only before it is mapped to a host. Changing it later is a disruptive operation because the volume must be unmapped from the host and mapped again with a new SCSI ID.

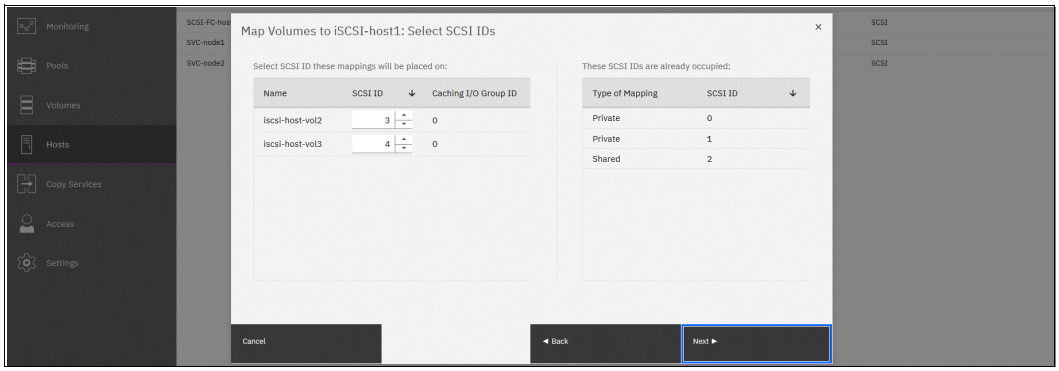


Figure 7-77 Choosing SCSI IDs manually

5. When the assignments are complete, click **Next** to verify the prepared changes and then click **Map Volumes** to complete the operation.

Duplicating and importing mappings

Volumes that are assigned to one host can be quickly and easily mapped to another host object.

This process can be accomplished by using one of the following methods:

- Duplicating the mappings from the selected host to the new host object.
- Selecting a new host and importing host mappings from another host object.

Note: Consider the following points:

- When duplicating or importing mappings, all mappings (private and shared) are copied. The shared mappings of an old host become the private mappings of a new host.
- Mappings can be duplicated only for a host that does not have volumes that are mapped.
- Mappings can be imported only to a host with no mappings.

To duplicate the mappings, complete the following steps:

1. Right-click the host that to be duplicated (source host) and click **Duplicate Volume Mappings**.
2. After the Duplicate Mappings window opens, select a target host to which all of the source host volumes are to be mapped (see Figure 7-78). In this example, the only target candidate is a host that has no mappings.

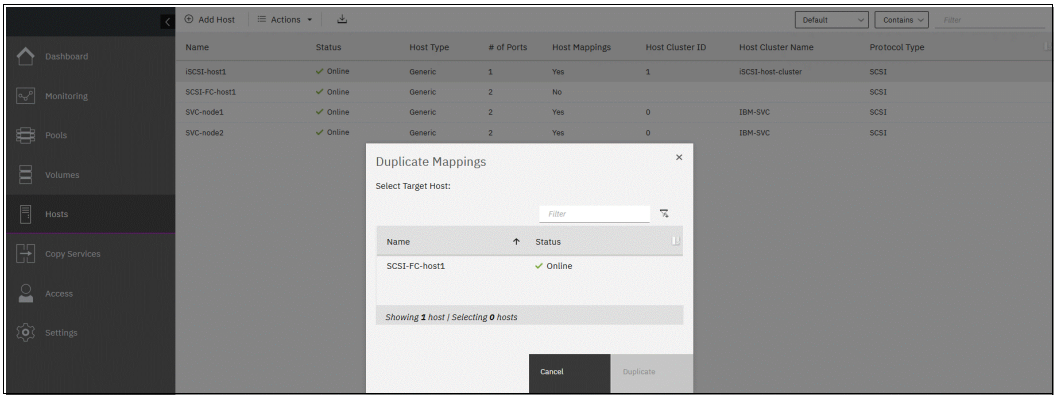


Figure 7-78 Duplicate Mappings window

3. Select a target host and click **Duplicate**. After the operation completes, the target host has the same volume mappings as the source host. Private and shared mappings are duplicated. Mappings on the source host also remain, and can be deleted manually if necessary.

To import hosts mappings to a new host from an existing host, complete the following steps:

1. Right-click the new host (ensure that it has no volumes mapped) and select **Import Volume Mappings**. If private or shared mappings exist for the host, this action is inactive (disabled) in an Actions menu.
2. Select the source host from which you want to import the volume mappings from the Import Mappings window (see Figure 7-79) and click **Import**.

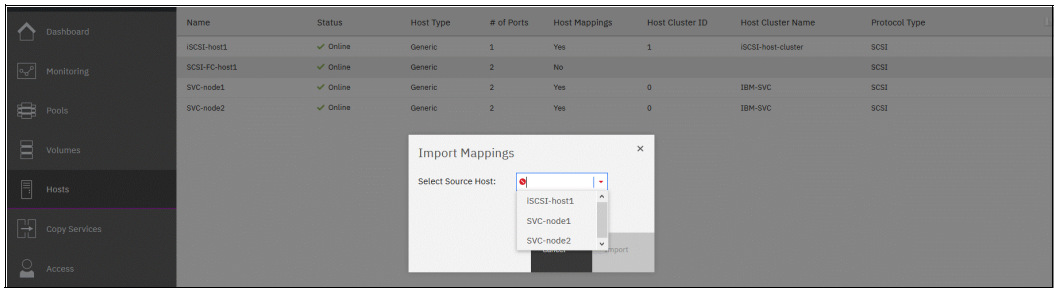


Figure 7-79 Import volume mappings source host selection

3. After the task completes, the target host features all of the same volume mappings as the source host. Shared mappings in which the source host participates are imported as private. Mappings on the source host also remain, and they can be deleted manually if necessary.

> **Note:** Mappings can be imported only from a source host that is in the same ownership group as the target host. If they are not, the import fails prompting the error message:
>
> `The command failed because the objects are in different ownership groups.`

Modifying the host type

The host type is often specified during the host creation process. However, it can be changed by using the **Modify Type** action.

To change the host type, complete the following steps:

1. Select the host or hosts to be modified, right-click and then select **Modify Type**.
2. From the Modify Type dialog window, select one of the available host types (Figure 7-80 on page 584):
   - Generic. The default host type. It is used in most cases, and for all NVMe hosts.
   - Generic (hidden secondary volumes). If this host type is set, all RC relationship secondary volumes are unavailable to that host.
   - HP/UX, OpenVMS, TPGS. Set when IBM Documentation requires the setting for the suitable host operating system types.
   - VVOL. Set if the host is configured to work with VVOLs.

   For more information, see Host attachment overview.

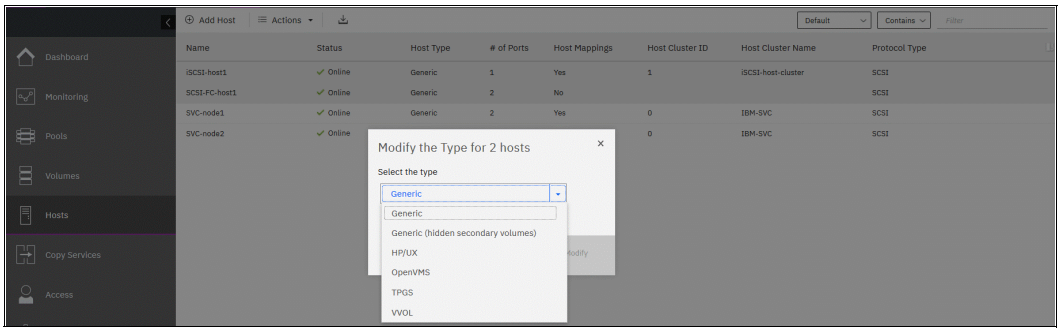


Figure 7-80 Changing the host type

3. Click **Modify**.

Viewing and editing throttles

A throttle is a mechanism that dynamically controls resource allocation during I/O operations for individual hosts or host clusters. If a throttle is defined, the system processes the I/O or delays it to free up resources for more critical I/O.

A host throttle sets the limit for combined read and write I/O to all mapped volumes. Other hosts that are accessing the same set of volumes are not affected by a host throttle.

To create a host throttle, or change or remove a host throttle, complete the following steps:

1. Select the suitable host or hosts, right-click, and select **Edit Throttle**.
2. In the Edit Throttle for Host dialog box (see Figure 7-81), specify the **IOPS limit**, **Bandwidth limit**, or both. Click **Create** to create a host throttle, change the throttle limit and then click **Save** to edit a throttle, or click **Remove** to delete a host throttle.

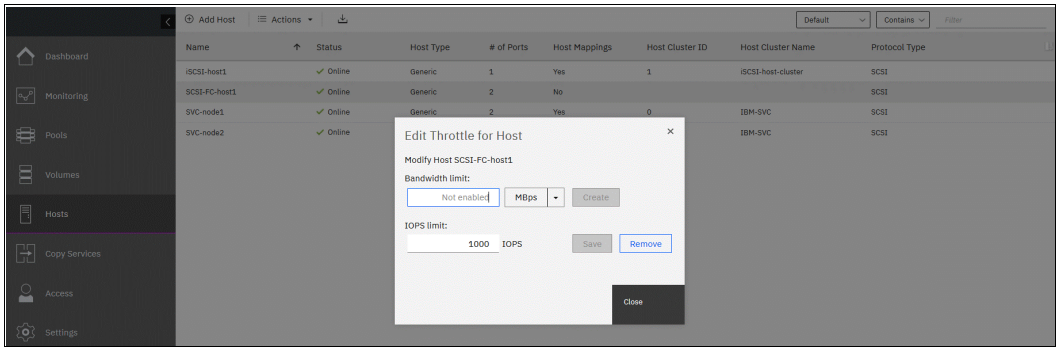


Figure 7-81 Edit Throttle for Host dialog

3. When finished editing or creating, click **Close**.

To view and edit all the throttles configured on the system, right-click any of the hosts and select **View All Throttles** (Figure 7-82). Switch between throttle types by clicking the drop-down menu next to the **Actions** menu. It is also possible to change the view to see all the system's throttles on a single list.

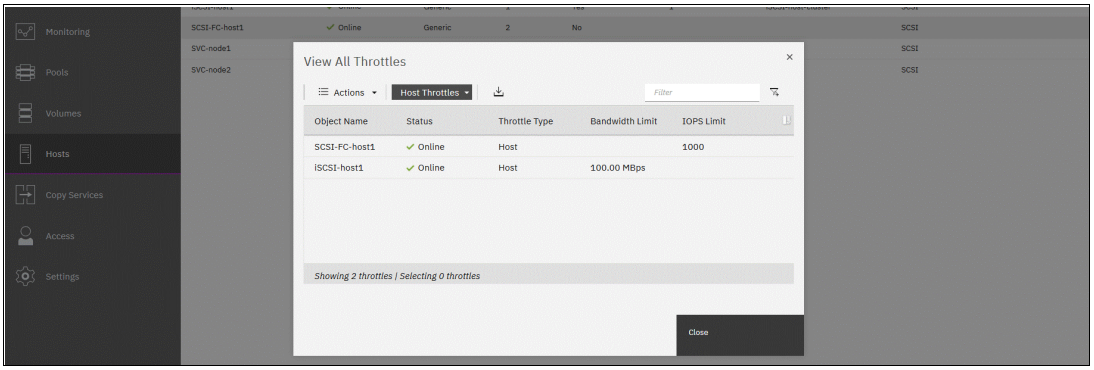


Figure 7-82 View All Throttles window

From this view, delete or edit any throttle by right-clicking it in the list and selecting the suitable action.

Removing private mappings from a host

To remove all host private mappings, complete the following steps:

1. Right-click a host that needs its mapping to be removed, and select **Remove Private Mappings**.
2. If a host is assigned to a cluster, a window opens with a warning that shared mappings will not be removed. Click **Yes** if you want to continue.
3. In the next window, confirm this action by entering the number of volume mappings to be removed (see Figure 7-83) and click **Remove**.

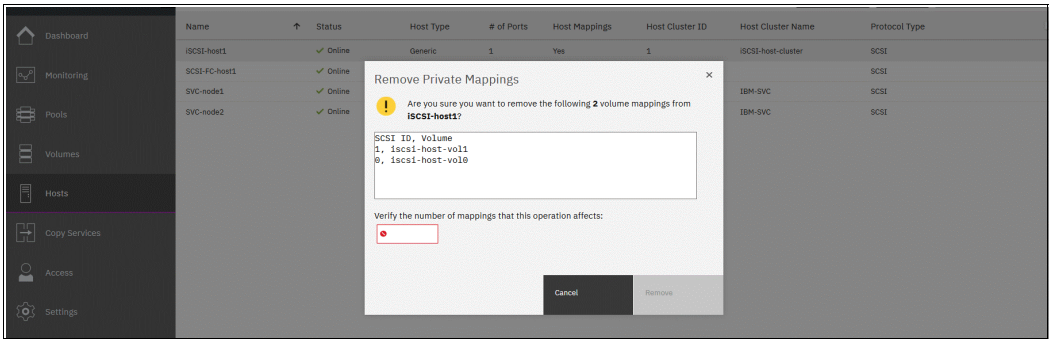


Figure 7-83 Confirming the number of mappings to be removed

Note: Consider the following points:

- Only private mappings are removed. Shared host cluster volume mappings remain.
- After clicking **Remove**, the host loses access to the unmapped volumes. Be sure to run the required procedures on the host operating system, such as unmounting the file system, taking the disk offline, or disabling the volume group, before removing the volume mappings from the host object on the GUI.

Removing a host

To remove a host object, complete the following steps:

1. Select the host or multiple hosts to be removed, right-click them and then select **Remove**.
2. Confirm that the correct list of hosts is displayed in the window by entering the number of hosts to be removed and then click **Remove** (see Figure 7-84).

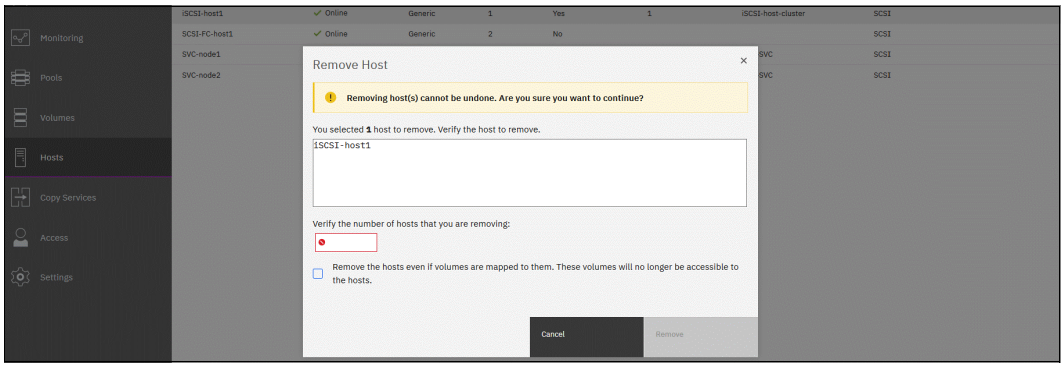


Figure 7-84 Confirming the removal of the host

3. If the host to be removed has volumes mapped to it, select the **Remove the hosts even if volumes are mapped to them** option in the lower part of the window. When this option is selected, all volume mappings of this host are deleted, and the host is removed.

Viewing IP logins

Right-click an iSCSI or iSER host to open the **IP Login Information** window and check the state of the host logins (see Figure 7-85). The drop-down menu in the upper section of the window can be used to switch between the IQNs of the host.

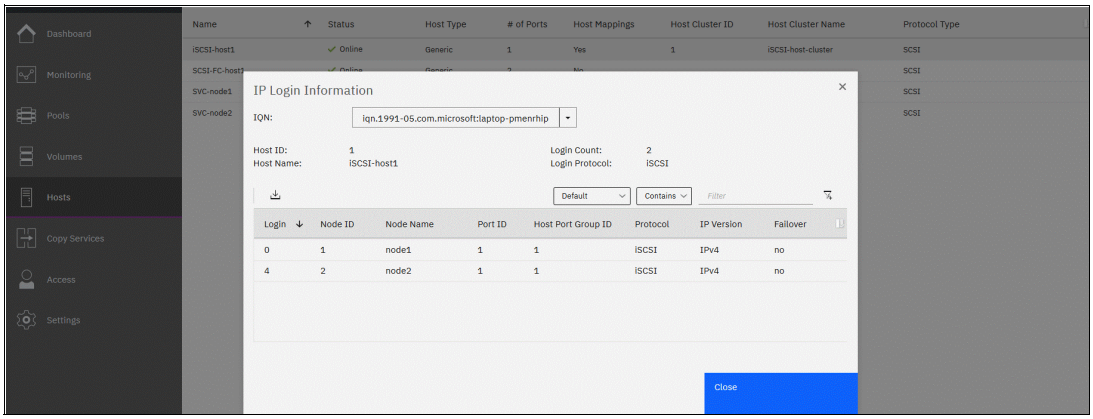


Figure 7-85 Viewing the IP login information

Viewing the host properties

To view the properties of a host object, complete the following steps:

1. Right-click a host and select **Properties**.
2. The Host Details window (see Figure 7-86) opens and three tabs are shown: Overview, Mapped Volumes, and Port Definitions:
   - In the Overview window, click **Edit** to change hostname or host type, select and clear the associate host I/O groups, and then modify the host status policy and status site.
   - The Mapped Volumes tab lists all the volumes that mapped to the host. Private and shared mappings are shown.

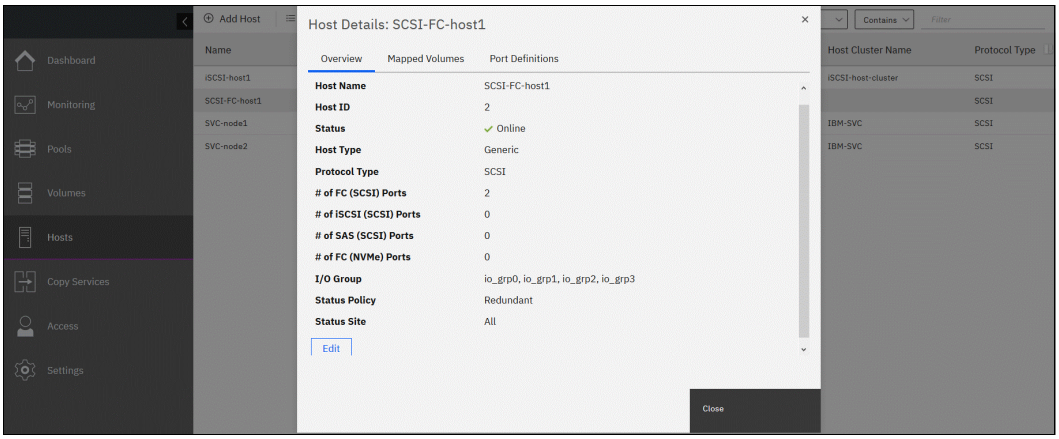


Figure 7-86 Host properties overview

   - The Port Definitions tab enables the user to view all the ports that belong to a specific host, add ports or remove any assigned ports (see Figure 7-87). This tab also includes the NQN of the NVMe host.

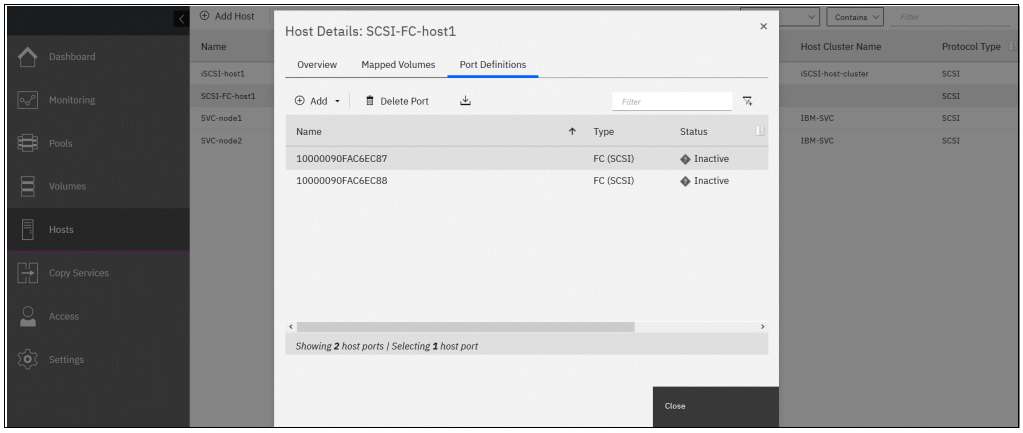


Figure 7-87 Listing port definitions

7.9.4 Actions on host clusters

This section describes actions that can be performed on a host cluster object by using the **Hosts** → **Host Clusters** menu. For more information about the Host Cluster feature and the actions that are required for host cluster creation, see 7.9.2, “Host clusters” on page 574.

Select **Hosts** → **Host Clusters** for a list of configured host clusters and their major parameters, such as cluster status, number of hosts in a cluster, and number of shared mappings.

Right-clicking any of the clusters or selecting one or several clusters and clicking the **Actions** drop-down menu opens the list of available actions (see Figure 7-88).

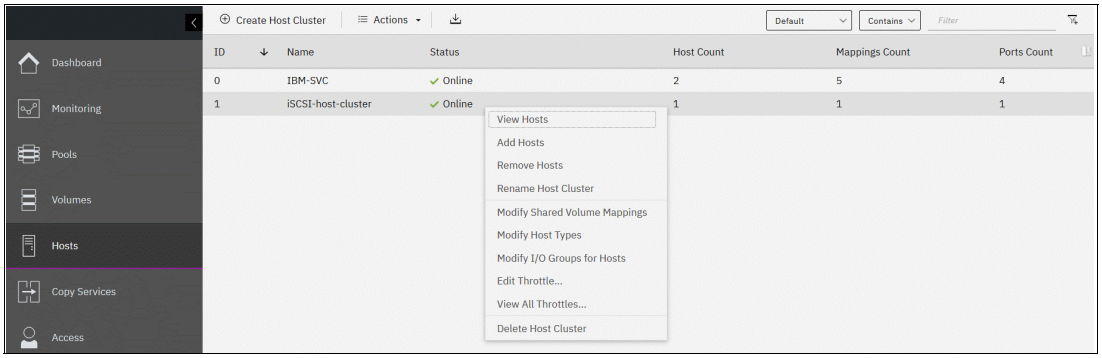


Figure 7-88 Actions that are available on a host cluster object

View Hosts action

Click **View Host** to see a list of hosts that are assigned to a host cluster (see Figure 7-89). Click **Next** and **Previous** to switch to other clusters in the cluster list.

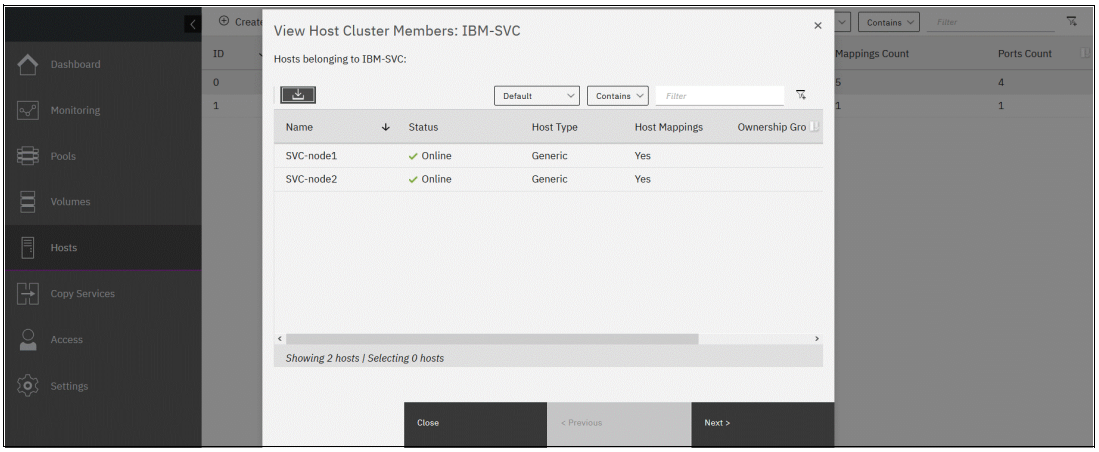


Figure 7-89 View Host Cluster Members window

Add Hosts action

Clicking **Add Hosts** opens a dialog box that displays all stand-alone hosts not assigned to any clusters (see Figure 7-90). Select the specific host to be added to the cluster and click **Next**.

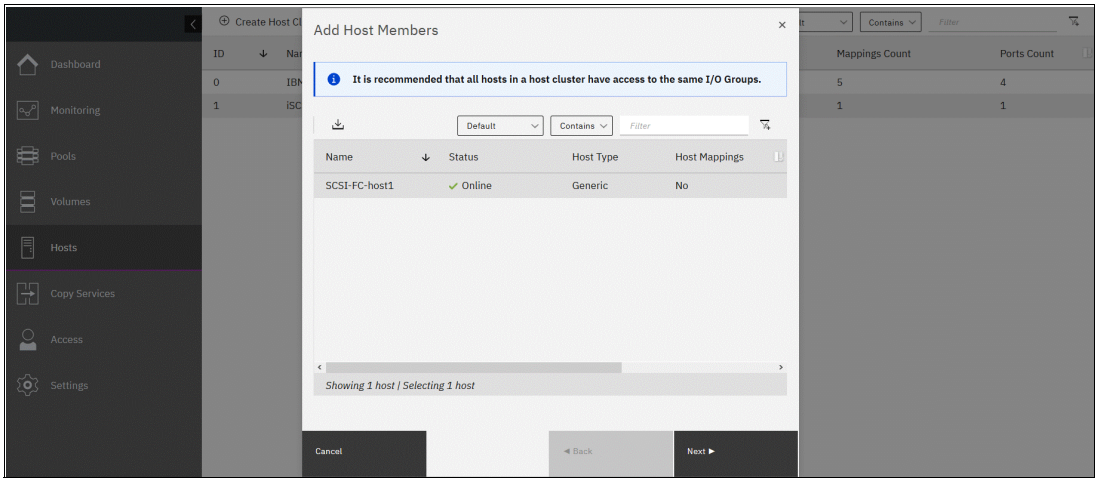


Figure 7-90 Adding a host member

A prompt appears stating that the shared host cluster mappings are applied to the added host, which gains access to all volumes mapped to the host cluster (see Figure 7-91).

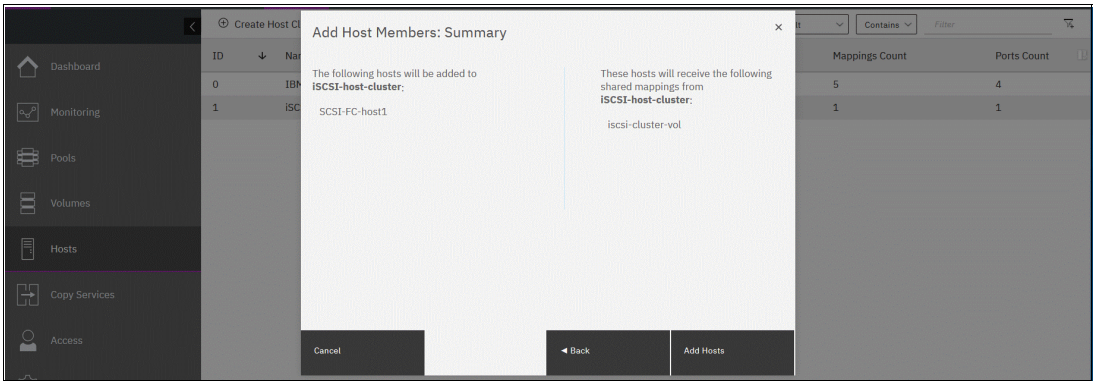


Figure 7-91 Confirming the addition of a host to a cluster

If the changes are correct, click **Add Hosts** to complete the operation.

Remove Hosts action

To remove a host or hosts from a cluster and convert them to stand-alone hosts, right-click the cluster and select **Remove Hosts (**see Figure 7-92).

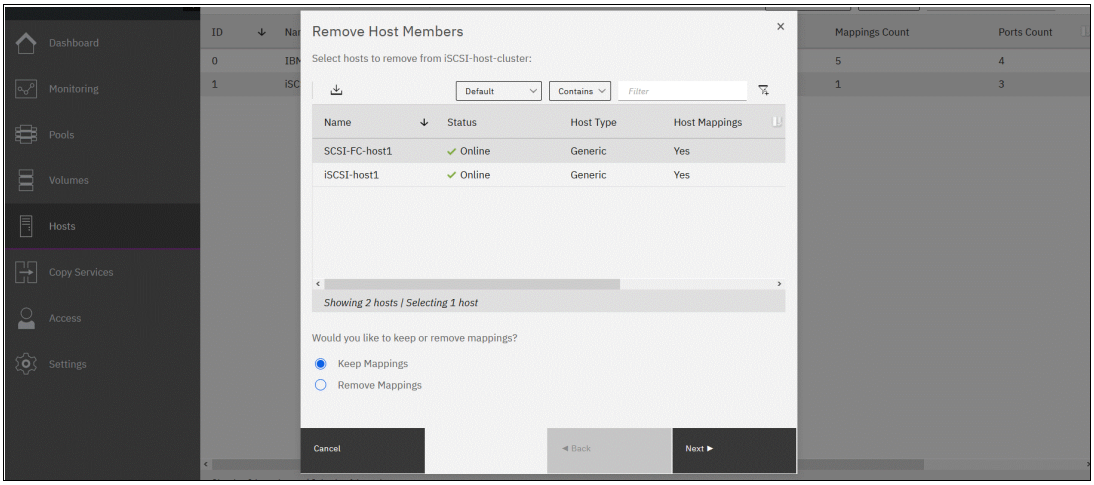


Figure 7-92 Remove Host from Cluster

The dialog box provides two options:

1. A removed host can keep all the shared cluster mappings as private mappings and retain access to volumes.
2. All shared cluster mappings can be removed from the host.

Select the suitable action and click **Next.** Verify the changes and then click **Remove Hosts** to complete the procedure.

Rename Host Cluster action

This action changes the host cluster object name.

Modify Shared Volume Mappings action

This action can be used to create shared mappings for a host cluster or to modify a host cluster.

To add or remove a shared mapping, complete the following steps:

1. Right-click the host cluster and select **Modify Shared Volume Mappings**.
2. A window appears displaying all the shared mappings that exist for the selected cluster (Figure 7-93 on page 591). From this view, select one or more of the shared mappings to be removed and then click **Remove Volume Mappings**.

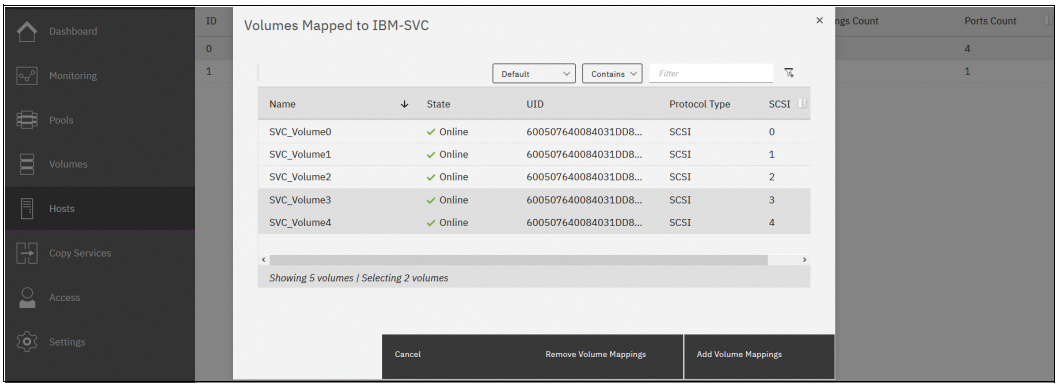


Figure 7-93 Existing shared mappings

3. If new shared mappings must be created, click **Add Volume Mappings** to open the next window (see Figure 7-94). A list appears that displays all of the volumes that are not yet mapped to the selected cluster.

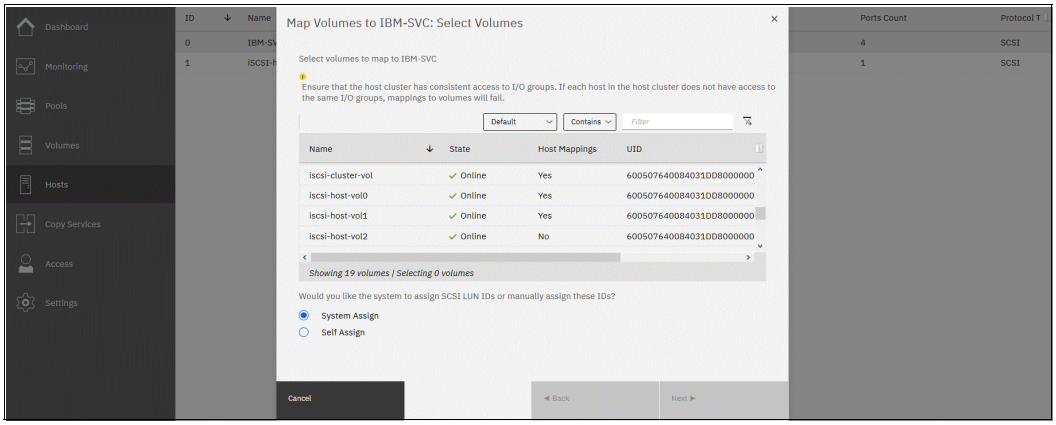


Figure 7-94 Adding shared mappings

4. The system assigns the next available SCSI ID for new mappings automatically by default. However, SCSI LUN IDs can be assigned manually by clicking **Self Assign** (see Figure 7-95).

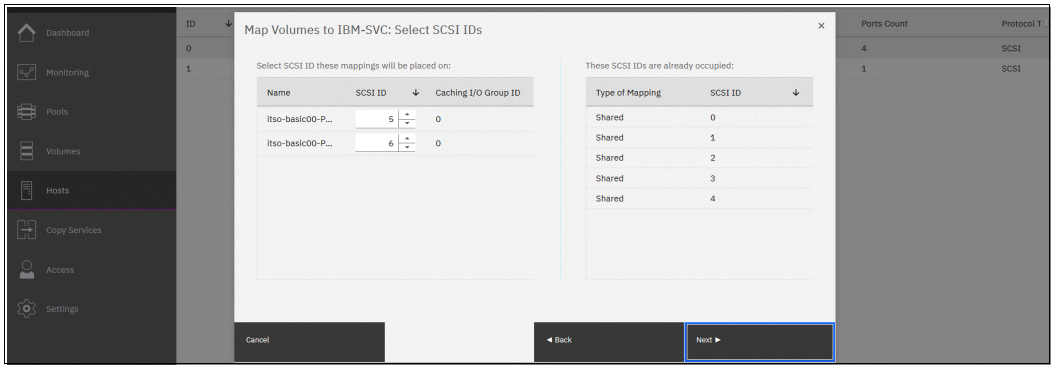


Figure 7-95 Assigning a SCSI ID to mapped volumes manually

5. Click **Next** and a window opens and asks for verification that the changes are correct (see Figure 7-96). Click **Back** to return and change the SCSI IDs or volumes that are being mapped, click **Cancel** to stop the task, or click **Map Volumes** to complete the operation.

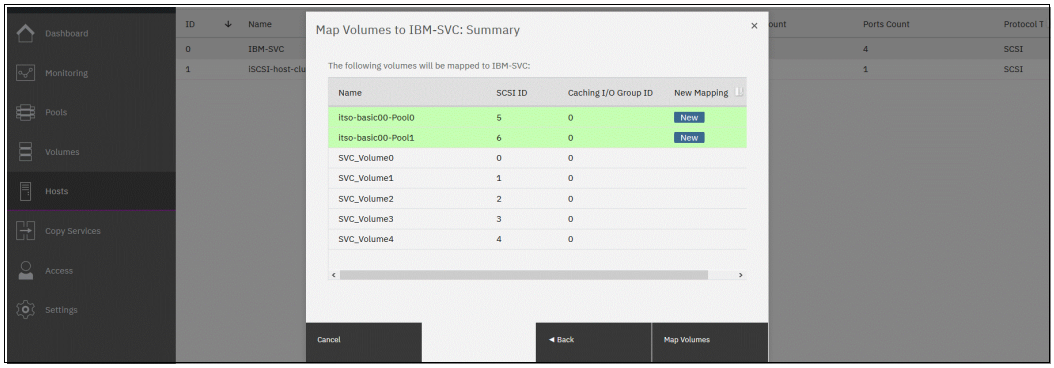


Figure 7-96 Summary of added shared mappings

Modifying host types

This action enables a user to change the host type for all members of a host cluster. The procedure is similar to changing a type on a separate host, as described in “Modifying the host type” on page 583, except that the changes are applied across all hosts that are assigned to the cluster.

Modify I/O groups for hosts action

All hosts are assigned to all I/O groups by default. However, from the Host Clusters window, it is possible to change the list of I/O groups that are associated with a specific host.

The Modify I/O groups for hosts action for a host cluster object changes the I/O group assignment for all hosts who are members of this cluster (see Figure 7-97).

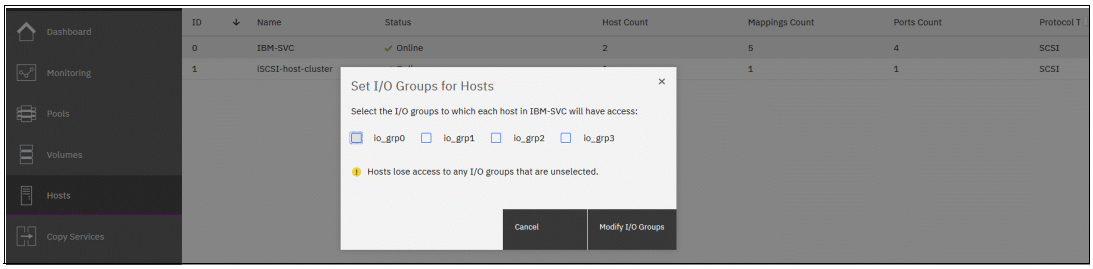


Figure 7-97 Setting the I/O groups for hosts

Viewing and editing all throttles

As with the individual hosts (see “Viewing and editing throttles” on page 584) it is possible to similarly create and edit the throttle parameters for a host cluster by using the Edit Throttle action.

Note: A throttle that is created for a host cluster is applied across all hosts within that cluster.

To create or change a host cluster throttle, complete the following steps:

1. Right-click a host cluster object and select **Edit Throttle**.
2. If any of the hosts in the cluster include defined individual throttles, those throttles must be removed or a warning appears (see Figure 7-98). Click **Remove Throttles**, or click **Cancel** to leave the individual throttles and stop the host throttle creation process.

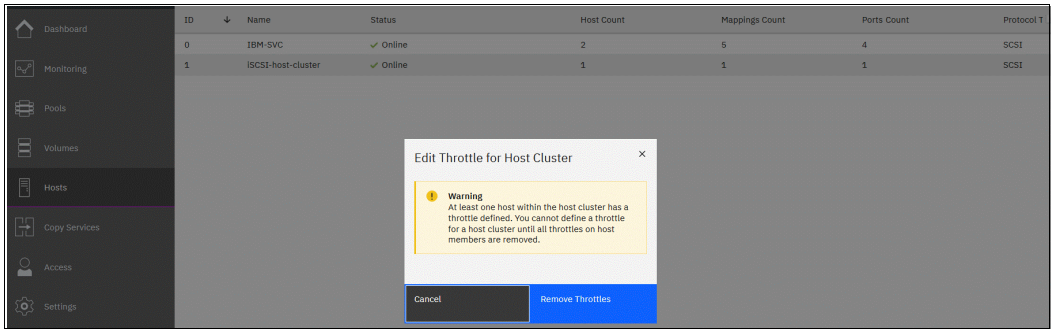


Figure 7-98 Warning that host throttles exist

3. If no individual throttles exist, set or edit I/O or data rate limits in the pop-up window (see Figure 7-99). Click **Create** to create a throttle, or click **IOPS limit** and then click **Save** to change the throttle.

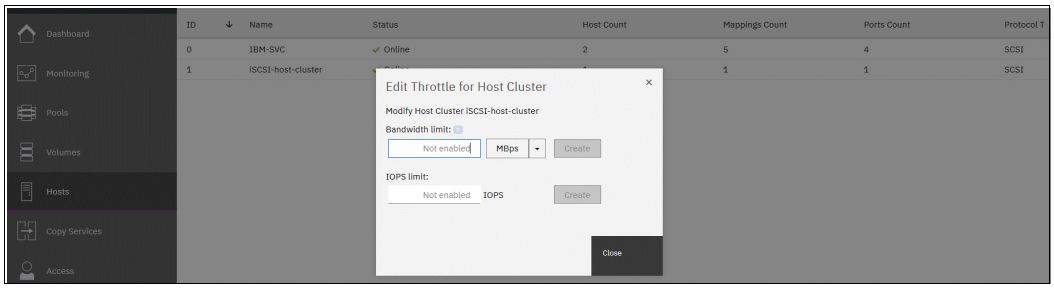


Figure 7-99 Creating a host cluster throttle

The use of the **View All Throttles** action on any host cluster object opens a list of all host cluster throttles that are configured on the object. Switch the display to other types of throttles by clicking a drop-down menu that is next to the **Actions** menu. The view can also be changed to see all the system's throttles in one list.

From this view, any of the throttles can be deleted or edited by right-clicking it in the list and selecting the required action (Figure 7-82 on page 585).

Delete Host Cluster

By using the Delete Host Cluster action, a cluster object can be removed so that all hosts assigned to it become stand-alone hosts. After a cluster is removed, the following options are available:

- Keep the volume mappings by converting them from shared to private for each host.
- Remove all shared mappings before deleting the host object.

An example of the Delete Host Cluster window is show in Figure 7-100 on page 594. Hover over the question marks that are next to the suggested removal options for more information.

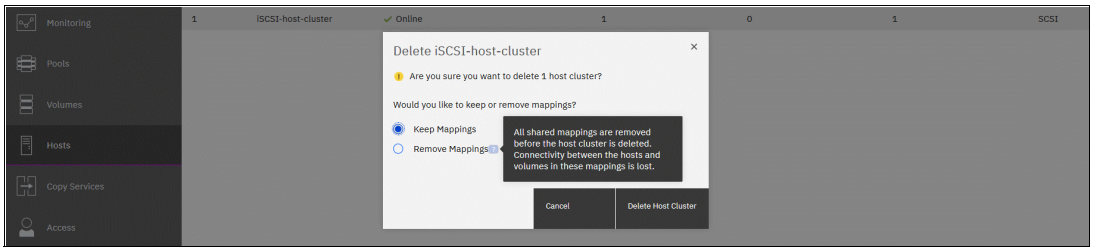


Figure 7-100 Removing a host cluster

7.9.5 Host management views

The Hosts menu provides two other management views: Mappings and Volumes by Hosts and Clusters (Figure 7-101).

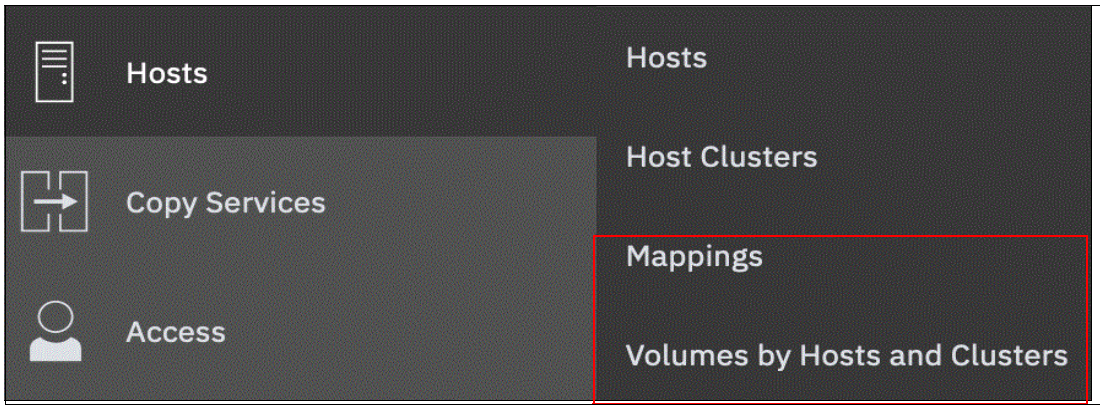


Figure 7-101 Host management views

These actions are the same as the actions that are found in the Hosts and Host Clusters views. However, this method provides a more convenient way to access them depending on the current administration task or the size of system configuration.

Mappings view

To see an overview of the host mappings, select **Hosts** → **Mappings**. This view lists all volume-to-host mappings in the system. It also shows the hosts, mapped volumes, and their SCSI IDs, volume unique identifiers (UIDs), and mapping types. Also displayed is in which I/O group the mapping exists.

To switch between listing only private mappings, only shared mappings, and all host mappings, use the drop-down menu in the upper left. The Private mappings and All Host mappings views show the hosts; switching to Shared mappings shows a list of host clusters and their mappings (see Figure 7-102 and Figure 7-103 on page 595).

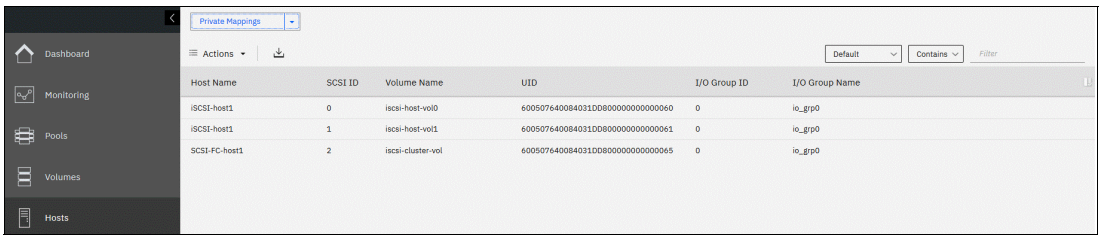


Figure 7-102 Private mappings list

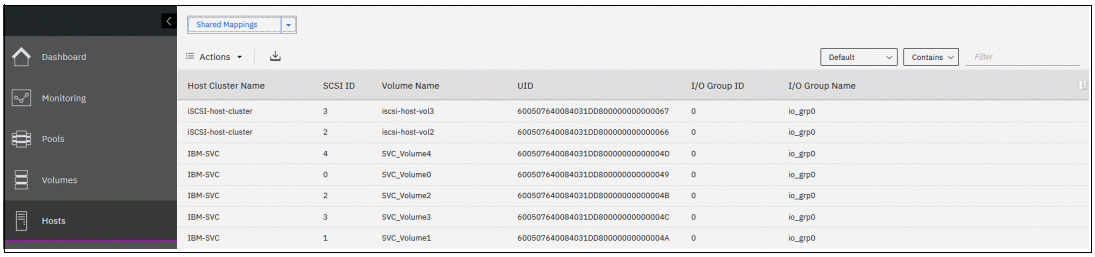


Figure 7-103 Shared mappings list

Click **Actions** (or right-click a mapping in the list) for the following tasks:

- Unmap Volumes
- Host Properties
- Volume Properties

Unmapping a volume

This action removes the mappings for all selected entries. The unmap action is available for shared mappings if in the Shared mappings view (see Figure 7-103) and only for private mappings while in the Private mappings or All Host mappings view.

To remove a volume mapping or mappings, select the records to remove, right-click, and select **Unmap volumes**, or select **Unmap Volumes** from the **Actions** menu (see Figure 7-104.)

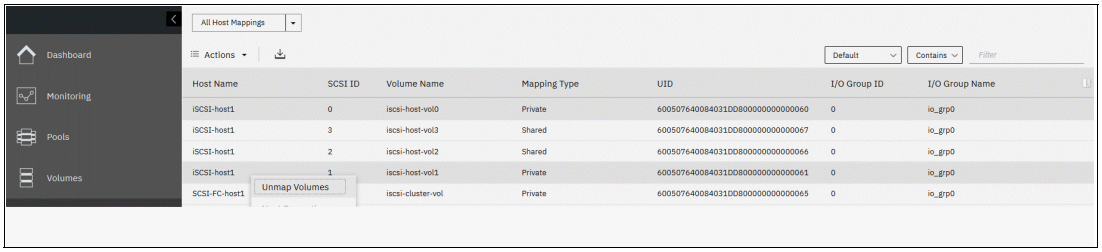


Figure 7-104 Removing two private mappings

In the dialog box, confirm how many volumes are to be unmapped by entering that number into the **Verify** field (see Figure 7-105). then click **Unmap**.

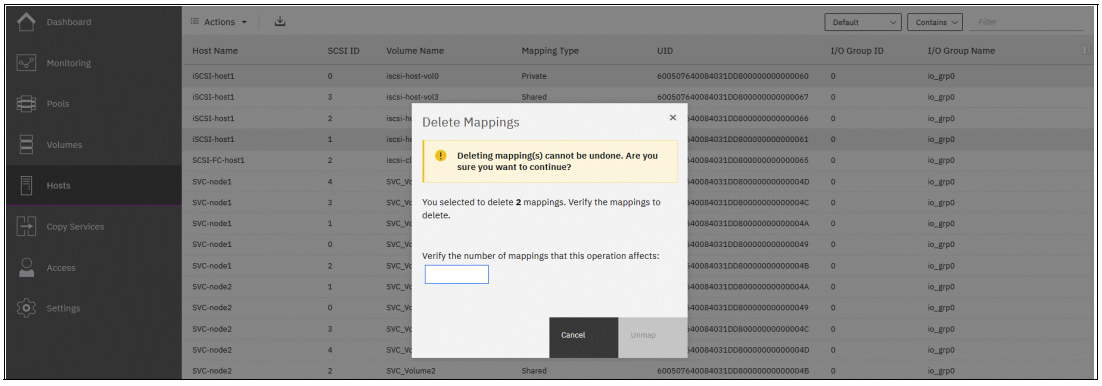


Figure 7-105 Confirming the mapping deletion

Host Properties

Select a single entry and click **Actions** → **Host Properties**. The Host Properties window opens and displays the same contents as described in “Viewing the host properties” on page 587.

Volume Properties

Select an entry and select **Actions** → **Volume Properties**. The Volume Properties view opens and displays the same contents as described in Chapter 7, “Hosts” on page 517.

Volumes by Hosts and Clusters

For a convenient way to manage volumes mapped to a specific host or host cluster, select **Hosts** → **Volumes by Hosts and Clusters**. In contrast to the Mappings view, these views focus on volume management.

The left column shows all configured hosts or host clusters. If many hosts are displayed, enter a specific host name or text string in the filter field to perform a quick search.

Below the list is an Add Host (Create Host Cluster) button, which opens the dialog box when clicked as described in 7.9.1, “Creating hosts” on page 555 and in 7.9.2, “Host clusters” on page 574.

The main window shows a list of volumes and their parameters that are mapped to the selected object (see Figure 7-106). The Volumes by Host view shows volumes that are mapped with both private and shared mappings. The Volumes by Hosts Cluster view shows only volumes with shared mappings.

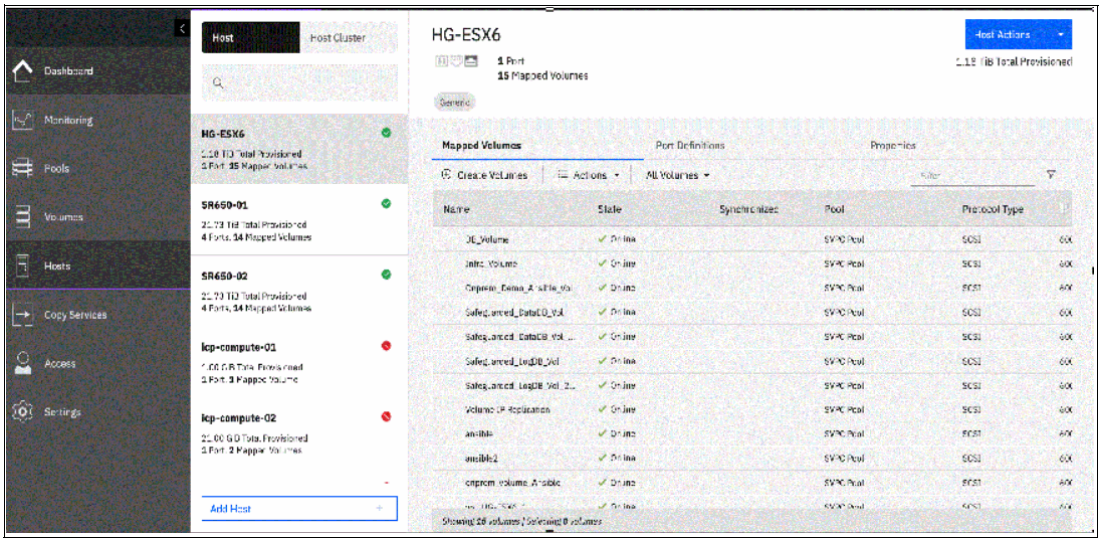


Figure 7-106 Volumes by Host window

Select one of the options from the Volumes menu to filter the volume type:

- All Volumes
- Thin-Provisioned Volumes
- Compressed Volumes
- Deduplicated Volumes

Right-click a volume in the list to open the Volume Actions menu. For more information, see Chapter 6, “Volumes” on page 419.

Create and map a volume by clicking **Create Volumes**.

7.10 Performing hosts operations by using the CLI

This section describes the host-related actions that can be done within the system from the CLI.

7.10.1 Creating a host by using the CLI

This section describes how to create FC and iSCSI hosts by using the CLI. It is assumed that the hosts are prepared for attachment as noted in the guidelines that are available at this IBM Documentation web page.

Creating Fibre Channel hosts

To create an FC host, complete the following steps:

1. Rescan the SAN on the system by running the **detectmdisk** command (Example 7-21).

Example 7-21 Rescanning the SAN

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>detectmdisk
```

Note: The **detectmdisk** command does not return any response.

If zoning was implemented correctly, any new WWPNs are discovered by the system after the **detectmdisk** command is run.

2. List the candidate WWPNs and identify the WWPNs that belong to the new host (see Example 7-22).

Example 7-22 Available WWPNs

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>lsfcportcandidate
fc_WWPN
2100000E1E09E3E9
2100000E1E30E5E8
2100000E1E30E60F
2100000E1EC2E5A2
2100000E1E30E597
2100000E1E30E5EC
```

3. Run the **mkhost** command with the required parameters (see Example 7-23).

Example 7-23 Host creation

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>mkhost -name FC-SCSI-HOST-03 -fcwwpn
2100000E1E30E597:2100000E1E30E5EC
Host, id [3], successfully created
```

Creating iSCSI hosts

Before creating an iSCSI host in IBM FlashSystem systems, determine the IQN address of the host. To find the IQN of the host, see the operating system documentation for that specific host.

To create a host, complete the following steps:

1. Create the iSCSI host by running the **mkhost** command (Example 7-24).

Example 7-24 Creating an iSCSI host by running the mkhost command

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>mkhost -iscsiname
iqn.1994-05.com.redhat:e6ff477b58 -name RHEL-Host-04
Host, id [4], successfully created
```

2. The iSCSI host can be verified by running the **lshost** command (Example 7-25).

Example 7-25 Verifying the iSCSI host by running the lshost command

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>lshost 4
id 4
name RHEL-Host-04
port_count 1
type generic
....
status_site all
iscsi_name iqn.1994-05.com.redhat:e6ff477b58
node_logged_in_count 1
state active
```

Note: When the host is initially configured, the default authentication method is set to `no authentication`, and no CHAP secret is set. To set a CHAP secret for authenticating the iSCSI host with the system, run the **chhost** command with the **chapsecret** parameter. To display a CHAP secret for a defined server, run the **lsiscsiauth** command.

This same method can also be used for FC hosts.

Creating FC NVMe hosts

Before creating an NVMe host in IBM Storage Virtualize systems, determine the NQN address of the host. To find the NQN, refer to the operating system documentation that is specific to that host.

Create a host by completing the following steps:

1. Create the FC-NVMe host by running the **mkhost** command (Example 7-26).

Example 7-26 The mkhost command

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>mkhost -name NVMe-Host-01 -nqn
nqn.2014-08.com.redhat:nvme:nvm-nvmehost01-edf223876 -protocol nvme -type
generic
Host, id [6], successfully created
```

2. Verify the FC-NVMe host by running the **lshost** command (Example 7-27).

Example 7-27 The lshost command

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>lshost 6
id 6
name NVMe-Host-01
port_count 1
...
status_site all
nqn nqn.2014-08.com.redhat:nvme:nvm-nvmehost01-edf223876
node_logged_in_count 2
state active
```

Note: If OBAC is set up, use the **-ownershipgroup** parameter when creating a host to add the host to a pre-configured ownership group. Use the ownership group name or ID.

Consider the following example command:

```
mkhost -name NVMe-Host-01 -nqn
nqn.2014-08.com.redhat:nvme:nvm-nvmehost01-edf223876 -protocol nvme -type
generic -ownershipgroup ownershipgroup0
```

Creating RDMA over NVMe hosts

Before creating an NVMe host in IBM Storage Virtualize systems, determine the NQN address of the host. To find the NQN, refer to the operating system documentation that is specific to that host.

Create a host by completing the following steps:

1. Create the NVMe over RDMA host by running the **mkhost** command (Example 7-28).

Example 7-28 The mkhost command

```
mkhost -name ITSO_NVME_RDMA_RHEL_HOST -nqn
nqn.2014-08.org.nvmexpress:uuid:0135e86a-98d2-460b-b0ab-6edb1795cdc3 -portset
portset0 -protocol rdmanvme  -type generic
Host, id [2], successfully created
```

2. Verify the FC-NVMe host by running the **lshost** command (Example 7-29).

Example 7-29 The lshost command

```
IBM_FlashSystem:GDLSVT8B:superuser>lshost 2
id 2
name ITSO_NVME_RDMA_RHEL_HOST
port_count 1
type generic
iogrp_count 1
status online
site_id 1
site_name site1
host_cluster_id
host_cluster_name
protocol rdmanvme
status_policy redundant
status_site all
nqn nqn.2014-08.org.nvmexpress:uuid:0135e86a-98d2-460b-b0ab-6edb1795cdc3
```

```
node_logged_in_count 2
state active
owner_id
owner_name
portset_id 0
portset_name portset0
```

Note: If OBAC is set up, use the **-ownershipgroup** parameter when creating a host to add the host to a pre-configured ownership group. Use the ownership group name or ID.

Consider the following example command:

```
mkhost -name ITSO_NVME_RDMA_RHEL_HOST -nqn
nqn.2014-08.org.nvmexpress:uuid:0135e86a-98d2-460b-b0ab-6edb1795cdc3 -portset
portset0 -protocol rdmanvme  -type generic -ownershipgroup ownershipgroup0
```

Creating NVMe over TCP hosts

Before creating an NVMe host in IBM Storage Virtualize systems, determine the NQN address of the host. To find the NQN, refer to the operating system documentation that is specific to that host.

Create a host by completing the following steps:

1. Create the NVMe over TCP host by running the **mkhost** command (see Example 7-30).

Example 7-30 The mkhost command

```
mkhost -name ITSO_NVME_TCP_RHEL_HOST -nqn
nqn.2014-08.org.nvmexpress:uuid:0135e86a-98d2-460b-b0ab-6edb1795cdc3 -portset
portset0 -protocol rdmatcp -type generic
Host, id [2], successfully created
```

2. Verify the FC-NVMe over TCP host by running the **lshost** command (see Example 7-31).

Example 7-31 The lshost command

```
IBM_FlashSystem:GDLSVT8B:superuser>lshost 2
id 2
name ITSO_NVME_TCP_RHEL_HOST
port_count 1
type generic
iogrp_count 1
status online
site_id 1
site_name site1
host_cluster_id
host_cluster_name
protocol tcpnvme
status_policy redundant
status_site all
nqn nqn.2014-08.org.nvmexpress:uuid:0135e86a-98d2-460b-b0ab-6edb1795cdc3
node_logged_in_count 2
state active
owner_id
owner_name
portset_id 0
```

```
portset_name portset0
```

Note: If OBAC is set up, use the **-ownershipgroup** parameter when creating a host to add the host to a pre-configured ownership group. Use the ownership group name or ID.

Consider the following example command:

```
mkhost -name ITSO_NVME_TCP_RHEL_HOST -nqn
nqn.2014-08.org.nvmexpress:uuid:0135e86a-98d2-460b-b0ab-6edb1795cdc3 -portset
portset0 -protocol tcpnvme  -type generic -ownershipgroup ownershipgroup0
```

7.10.2 Host administration by using the CLI

This section describes the advanced host operations that can be implemented from within the CLI.

Mapping a volume to a host

To map a volume, complete the following steps:

1. Run the **mkvdiskhostmap** command (see Example 7-32).

Example 7-32 Mapping a volume

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>mkvdiskhostmap -host RHEL-HOST-01
-scsi 0 RHEL_VOLUME
Virtual Disk to Host map, id [0], successfully created
```

2. Check the volume mapping by running the **lshostvdiskmap** command against that host (see Example 7-33).

Example 7-33 Checking the mapped volume

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>lshostvdiskmap RHEL-HOST-01
id name         SCSI_id vdisk_id vdisk_name  .. mapping_type
0  RHEL-HOST-01 0       109      RHEL_VOLUME .. private
```

Mapping a volume that is already mapped to a different host

To map a volume to a host that is mapped to a different host, complete the following steps:

1. Run the **mkvdiskhost -force** command (see Example 7-34).

Example 7-34 Mapping the same volume to a second host

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>svctask mkvdiskhostmap -force -host
RHEL-Host-06 -scsi 0 RHEL_VOLUME
Virtual Disk to Host map, id [0], successfully created
```

Note: The volume `RHEL_VOLUME` is mapped to both of the hosts by using the same SCSI ID. Typically, that is a requirement for most host-based clustering software, such as Microsoft Clustering Service, IBM PowerHA, and VMware ESX clustering.

2. The volume `RHEL_VOLUME` is mapped to two hosts (`RHEL-HOST-01` and `RHEL-Host-06`), and can be seen by running the **lsvdiskhostmap** command (see Example 7-35 on page 602).

Example 7-35 Ensuring that the same volume is mapped to multiple hosts

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>lsvdiskhostmap RHEL_VOLUME
id name        SCSI_id host_id host_name    .. IO_group_name mapping_type
0  RHEL_VOLUME 0       0       RHEL-HOST-01 .. io_grp0       private
0  RHEL_VOLUME 0       1       RHEL-Host-06 .. io_grp0       private
IBM_IBM FlashSystem:ITSO-FS7200:superuser>
```

Unmapping a volume from a host

To unmap a volume from the host, run the **rmvdiskhostmap** command (see Example 7-36).

Example 7-36 Unmapping a volume from a host

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>rmvdiskhostmap -host RHEL-Host-06 RHEL_VOLUME
```

Important: Before unmapping a volume, ensure that the suitable actions were completed on the host (such as unmounting the file system or removing the volume or volume group). Failure to do so can result in data corruption.

Renaming a host

To rename a host definition, run the **chhost -name** command (see Example 7-37, where the host `RHEL-Host-06` is renamed to `FC_RHEL_HOST`).

Example 7-37 Renaming a host

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>chhost -name FC_RHEL_HOST RHEL-Host-06
```

Removing a host

To remove a host from the IBM FlashSystem system, run the **rmhost** command (Example 7-38).

Example 7-38 Removing a host

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>rmhost RHEL-Host-07
```

Note: Before removing a host from IBM FlashSystem systems, ensure that all of the volumes are unmapped from that host, as shown in Example 7-36.

Host properties

To obtain more information about a host, run the **lshost** command with **hostname** or **host id** as a parameter (see Example 7-39).

Example 7-39 Host details

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>lshost ITSO-VMHOST-01
id 10
name ITSO-VMHOST-01
port_count 2
type generic
mask 111111111111111111111111111111111111111111111111111111111111111111
iogrp_count 4
status offline
site_id
site_name
host_cluster_id 0
```

```
host_cluster_name ITSO-ESX-Cluster-01
protocol scsi
status_policy redundant
status_site all
WWPN 2100000E1E30E597
node_logged_in_count 0
state offline
WWPN 2100000E1E30E5E8
node_logged_in_count 0
state offline
owner_id 0
IBM_IBM
FlashSystem:ITSO-FS7300:superuser>
```

Note: The **status_policy** property has two potential values:

- `Complete`: This default policy is implemented when a host is created. In addition, hosts on systems that are upgraded to a new code level include this policy setting.
- `Redundant`: This policy changes the meaning of `Online` and `Degraded` in the `status` property:
  - `Online` indicates redundant connectivity; that is, enough host ports are logged in to enough nodes so that the removal of a single node or a single host port still enables that host to access all its volumes.
  - `Degraded` indicates non-redundant connectivity; that is, a state in which a single point of failure (SPOF) prevents a host from accessing at least some of its volumes.

These options can be changed only by running the **chhost** command. When the host is created by running **mkhost**, the default policy of `redundant` is set.

7.10.3 Adding and deleting a host port by using the CLI

This section describes adding and deleting a host port to and from the system.

Adding ports to a defined host

To add ports to a defined host, complete the following steps:

- For FC-SCSI host ports:
  a. If the host is connected through SAN with FC, and if the WWPN is zoned to the system, run the **lsfcportcandidate** command to compare it with the information that is available from the server administrator. (see Example 7-40)

Example 7-40 Listing the newly available WWPN

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>lsfcportcandidate
fc_WWPN
2100000E1E09E3E9
2100000E1E30E5E8
2100000E1E30E60F
2100000E1EC2E5A2
```

  b. Use host or SAN switch utilities to verify whether the WWPN matches the information for the new WWPN. If the WWPN matches, run the **addhostport** command to add the port to the host (Example 7-41 on page 604).

Example 7-41 Adding the newly discovered WWPN to the host definition

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>addhostport -hbawwpn
2100000E1E09E3E9:2100000E1E30E5E8 ITSO-VMHOST-01
```

This command adds the WWPNs 2100000E1E09E3E9 and 2100000E1E30E5E8 to the ITSO-VMHOST-01 host.

c. If the new HBA is not connected or zoned, the **lshbaportcandidate** command does not display the WWPN. In this case, the WWPN of the HBA or HBAs can be manually entered and the **-force** flag can be used to create the host (see Example 7-42).

Example 7-42 Adding a WWPN to the host definition by using the -force option

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>addhostport -hbawwpn
2100000000000001 -force  ITSO-VMHOST-01
```

This command forces the addition of the WWPN 2100000000000001 to the host ITSO-VMHOST-01.

Note: WWPNs are not case-sensitive within the CLI.

d. Verify the host port count by running the **lshost** command. Example 7-43 shows that the host ITSO-VMHOST-01 has a port count that updated from 2 to 5 after the two commands in previous examples were run.

Example 7-43 Host with the updated port count

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>lshost
id name           port_count iogrp_count status   site_id site_name
0  ITSO-VMHOST-01 5          4           online
```

- For iSCSI and FC-NVMe host ports:

a. If the host uses iSCSI or FC-NVMe as a connection protocol, the host port ID, either iSCSI IQN or NVMe NQN, is used to add the port. Unlike FC-attached hosts, the available candidate IDs cannot be checked. The host administrator can provide the IQN or NQN.

b. After getting the ID, run the **addhostport** command (see Example 7-44, which shows the command for an iSCSI port).

Example 7-44 Adding an iSCSI port to the defined host

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>addhostport -iscsiname
iqn.1994-05.com.redhat:e6ddffaab567 RHEL-Host-05
```

Example 7-45 shows the FC-NVMe port being added.

Example 7-45 The addhostport command

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>addhostport -nqn
nqn.2016-06.io.rhel:875adad3345 RHEL-Host-08
```

Deleting ports from a defined host

If a host port record must be removed from a host object, run the **rmhostport** command.

To perform the removal procedure, complete the following steps:

1. Ensure that the correct port is being removed by running the **lshost** command (see Example 7-46, which shows a check ID verifying that the WWPN being removed belongs to host ITSO-VMHOST-01).

   Example 7-46 Running the lshost command to check the WWPNs

   ```
   IBM_2145:ITSO-SV3:superuser>lshost ITSO-VMHOST-01
   id 0
   name ITSO-VMHOST-01
   port_count 2
   ...
   WWPN 2100000E1E30E597
   node_logged_in_count 2
   state online
   WWPN 2100000E1E30E5E8
   node_logged_in_count 2
   state online
   ```

2. When WWPN or iSCSI IQN to be deleted is determined, run the **rmhostport** command to delete the host port (see Example 7-47).

   Example 7-47 Running the rmhostport command to remove a WWPN

   ```
   IBM_IBM FlashSystem:ITSO-FS7300:superuser>rmhostport -fcwwpn 2100000E1E30E597
   ITSO-VMHOST-01
   ```

   To remove the iSCSI IQN, run the **rmhostport** command with the **iscsiname** argument (see Example 7-48).

   Example 7-48 Removing the iSCSI port from the host

   ```
   IBM_IBM FlashSystem:ITSO-FS7300:superuser>rmhostport -iscsiname
   iqn.1994-05.com.redhat:e6ddffaab567 RHEL-Host-05
   ```

3. To remove the NVMe NQN, run the **rmhostport** with the **nqn** argument (see Example 7-49).

   Example 7-49 Removing the NQN port from the host

   ```
   IBM_IBM FlashSystem:ITSO-FS7300:superuser>rmhostport -nqn
   nqn.2016-06.io.rhel:875adad3345 RHEL-Host-08
   ```

Note: Multiple ports can be removed at once by using the separator or colon (:) between the port names, as shown in the following example:

```
rmhostport -hbawwpn 210000E08B054CAA:210000E08B892BCD ITSO-VMHOST-02
```

7.10.4 Host cluster operations

This section describes the host cluster operations that can be performed by using the CLI.

Creating a host cluster

To create a host cluster, run the **mkhostcluster** command (see Example 7-50).

Example 7-50 Creating a host cluster

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>mkhostcluster -name ITSO-ESX-Cluster-01
Host cluster, id [0], successfully created.
```

Adding a host to a host cluster

After creating a host cluster, a host or a list of hosts can be added by running the **addhostclustermember** command (see Example 7-51).

Example 7-51 Adding a host or hosts to a host cluster

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>addhostclustermember -host
ITSO-VMHOST-01:ITSO-VMHOST-02 ITSO-ESX-Cluster-01
IBM_IBM FlashSystem:ITSO-FS7200:superuser>
```

In Example 7-51, the hosts `ITSO-VMHOST-01` and `ITSO-VMHOST-02` were added as part of host cluster `ITSO-ESX-Cluster-01`.

Listing the host cluster member

To list the host members that are part of a particular host cluster, run the **lshostclustermember** command (see Example 7-52).

Example 7-52 Listing host cluster members by running the lshostclustermember command

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>lshostclustermember ITSO-ESX-Cluster-01
host_id host_name      status  type    site_id site_name
0       ITSO-VMHOST-01 offline generic
4       ITSO-VMHOST-02 offline generic
IBM_IBM FlashSystem:ITSO-FS7300:superuser>
```

Mapping a volume to a host cluster

To map a volume to a host cluster so that it automatically is mapped to member hosts, run the **mkvolumehostclustermap** command (see Example 7-53).

Example 7-53 Mapping a volume to a host cluster

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>mkvolumehostclustermap -hostcluster
ITSO-ESX-Cluster-01 VMware1
Volume to Host Cluster map, id [0], successfully created
IBM_IBM FlashSystem:ITSO-FS7300:superuser>
```

Note: When a volume is mapped to a host cluster, that volume is mapped to all of the members of the host cluster with the same `SCSI_ID`.

Listing the volumes that are mapped to a host cluster

To list the volumes that are mapped to a host cluster, run the **lshostclustervolumemap** command (see Example 7-54).

Example 7-54 Listing volumes that are mapped to a host

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>lshostclustervolumemap ITSO-ESX-Cluster-01
id name               SCSI_id volume_id volume_name ..  protocol
```

```
0  ITSO-ESX-Cluster-01 0       8         VMware1     ..  scsi
0  ITSO-ESX-Cluster-01 1       9         VMware2     ..  scsi
0  ITSO-ESX-Cluster-01 2       10        VMware3     ..  scsi
```

Note: Run the **lshostvdiskmap** command against each host that is part of the host cluster to verify that the mapping type for the shared volumes are `shared` and that the non-shared volumes are `private`.

Removing a volume mapping from a host cluster

To remove a volume mapping to a host cluster, run the **rmvolumehostclustermap** command In Example 7-55, volume VMware3 is unmapped from the host cluster `ITSO-ESX-Cluster-01`.

Example 7-55 Removing a volume mapping

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>rmvolumehostclustermap -hostcluster
ITSO-ESX-Cluster-01 VMware3
```

Note: Use the **-makeprivate** flag to specify which host or hosts are to acquire private mappings from the volume that is being removed from the host cluster.

Removing a host cluster member

To remove a host cluster member, run the **rmhostclustermember** command (see Example 7-56).

Example 7-56 Removing a host cluster member

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>rmhostclustermember -host ITSO-VMHOST-02
-removemappings ITSO-ESX-Cluster-01
```

In Example 7-56, the host `ITSO-VMHOST-02` was removed as a member from the host cluster `ITSO-ESX-Cluster-01`, along with the associated volume mappings because the **-removemappings** flag was specified.

Removing a host cluster

To remove a host cluster, run the **rmhostcluster** command (see Example 7-57).

Example 7-57 Removing a host cluster

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>rmhostcluster -removemappings
ITSO-ESX-Cluster-01
```

The use of the **-removemappings** removes any shared host mappings to volumes *before* the host cluster is deleted.

Note: To keep the volumes mapped to the host objects even after the host cluster is deleted, use the **-keepmappings** flag instead of **-removemappings** for the **rmhostcluster** command. This process converts the host volume mapping to `private` instead of `shared`.

7.10.5 Adding a host or host cluster to an ownership group

To add a host or a host cluster to an ownership group, run the **chhost** or **chhostcluster** command with the **-ownershipgroup** parameter (see Example 7-58).

Example 7-58 Adding a host cluster to an ownership group

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>chhostcluster -ownershipgroup 1 0
```

Note: Specify the ID of the ownership group to be added the host; then specify the ID of the host or host cluster.

In Example 7-58, the command adds host cluster ID 0 to ownership group ID 1.

Removing a host or host cluster from an ownership group

To remove a host or a host cluster from an ownership group, run the **chhost** or **chhostcluster** command with the **-noownershipgroup** parameter (see Example 7-59).

Example 7-59 Removing a host cluster from an ownership group

```
IBM_IBM FlashSystem:ITSO-FS7300:superuser>chhostcluster -noownershipgroup 0
```

In Example 7-59, this command removes host cluster 0 from the ownership group to which it was previously assigned.

7.11 Host attachment examples

This section demonstrates how to attach a Linux-based host by using the information provided in the previous sections of this chapter.

7.11.1 Prerequisites

The host must be running with supported HBAs and on the supported operating system, which in this example is Red Hat Enterprise Linux (RHEL).

In this case, the operating system release level for RHEL can be verified by running the command shown in Example 7-60.

Example 7-60 RHEL release check

```
20201028-09:50:34 root@redbookvm7-1:~ # cat /etc/redhat-release
Red Hat Enterprise Linux Server release 7.6 (Maipo)
```

7.11.2 Fibre Channel host connectivity and capacity allocation

To collect the necessary data, complete the following steps to configure the host object in the storage system and access to the storage capacity:

1. Obtain the necessary connectivity credentials from the host. In this case, the WWPN of the host HBAs are required. The WWPN can be obtained in RHEL by running the command that is shown in Example 7-61 on page 609.

This example shows that the information about the host's FC HBAs is available in the `/sys/class/fc_host` directory. The host's WWPNs are in the `port_name` file in each `hostN` directory. The WWPNs are in bold in the example and will be used for the host object configuration of the storage system.

Example 7-61 Discovering the hosts' WWPNs

```
20201028-10:39:29 root@redbookvm7-1:~ # cd /sys/class/fc_host

20201028-10:40:02 root@redbookvm7-1:/sys/class/fc_host # ls -la
total 0
drwxr-xr-x.  2 root root 0 Oct 26 14:19 .
drwxr-xr-x. 59 root root 0 Oct 26 14:19 ..
lrwxrwxrwx.  1 root root 0 Oct 28 10:01 host33 ->
../../devices/pci0000:00/0000:00:17.0/0000:13:00.0/host33/fc_host/host33
lrwxrwxrwx.  1 root root 0 Oct 28 10:01 host34 ->
../../devices/pci0000:00/0000:00:17.0/0000:13:00.1/host34/fc_host/host34

20201028-10:45:35 root@redbookvm7-1:/sys/class/fc_host # cat host33/port_name
0x10000090fac6ec87
20201028-10:45:50 root@redbookvm7-1:/sys/class/fc_host # cat host34/port_name
0x10000090fac6ec88
```

2. To configure the host object on the storage system, follow the instructions in "Creating FC host objects" on page 556. If zoning was completed for the host, its WWPN is available in the Host Port (WWPN) list. If the host is not yet zoned, ports can be manually added into the field.
3. After the host object is defined, it is visible in the hosts view, and volumes (VDisks) can be mapped to it, as described in Chapter 2, "Installation and configuration planning" on page 119. Details about the host can be found by double-clicking its entry in the Hosts view.

   After the volumes are mapped to the host, go into the Host Details window in the Mapped Volumes tab (see Figure 7-107 on page 610) to verify that all the information is correct.

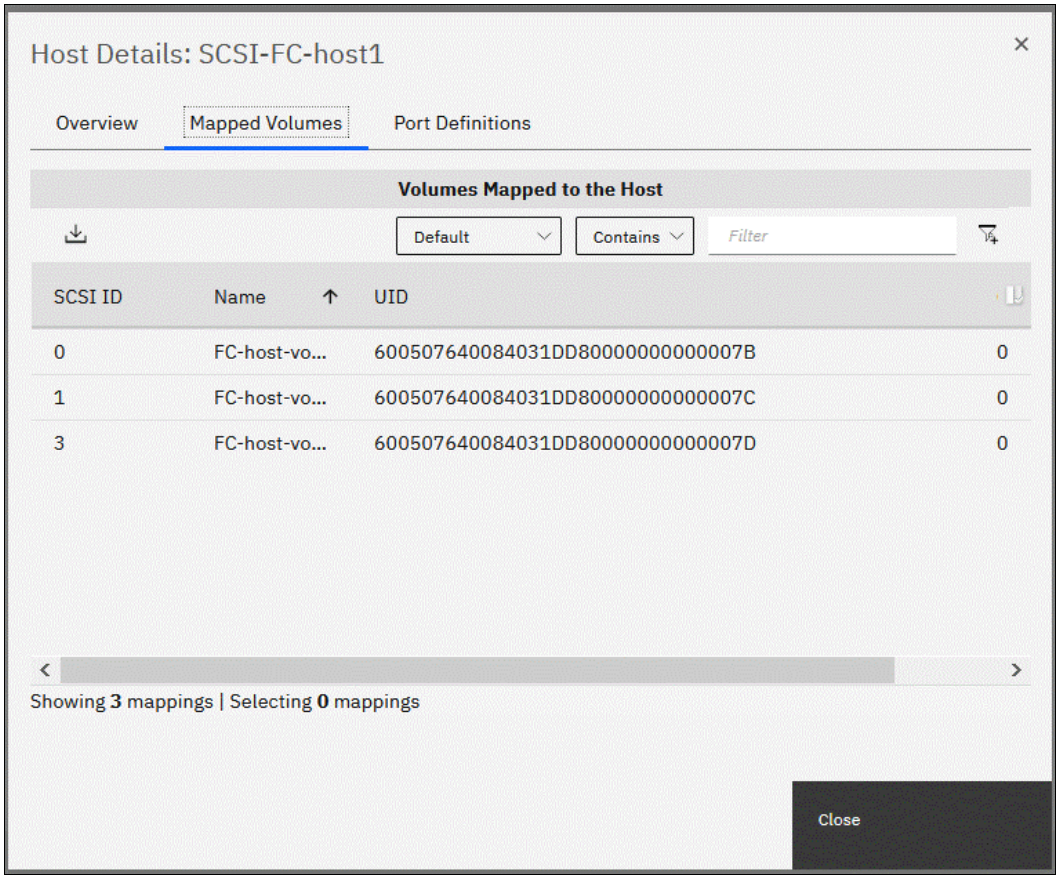


Figure 7-107 Host Details: Mapped volumes

4. Complete the following steps to configure the host side to discover and use the mapped VDisks:

   a. RHEL has its own native multipath driver, which maps the discovered drives and their paths to the `mpath n` device files in `/dev/mapper`. The multipath driver must be correctly configured. For more information, see Attachment requirements for hosts that are running the Linux operating system. To check that the volumes are detected correctly by the host, run the command that is shown in Example 7-62.

Example 7-62 Scanning and rebuilding the multipath

```
20201028-14:19:53 root@redbookvm7-1:/dev # rescan-scsi-bus.sh -r
Syncing file systems
Scanning SCSI subsystem for new devices and remove devices that have
disappeared
Scanning host 0 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
 Scanning for device 0 0 0 0 ...
. . .
20201028-14:10:25 root@redbookvm7-1:/dev # multipath -F

20201028-14:10:30 root@redbookvm7-1:/dev # multipath

20201028-14:14:42 root@redbookvm7-1:/dev # multipath -ll
mpathau (3600507640084031dd80000000000007d) dm-4 IBM     ,2145
size=100G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=enabled
```

```
| |- 33:0:15:3 sdd  8:48   active ready running
| |- 33:0:27:3 sdu  65:64  active ready running
| |- 33:0:28:3 sdaa 65:160 active ready running
| |- 33:0:31:3 sdaf 65:240 active ready running
| |- 34:0:13:3 sdah 66:16  active ready running
| |- 34:0:15:3 sdak 66:64  active ready running
| |- 34:0:1:3  sdv  65:80  active ready running
| `- 34:0:3:3  sdac 65:192 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 33:0:19:3 sdg  8:96   active ready running
  |- 33:0:24:3 sdj  8:144  active ready running
  |- 33:0:25:3 sdm  8:192  active ready running
  |- 33:0:26:3 sdp  8:240  active ready running
  |- 34:0:20:3 sdan 66:112 active ready running
  |- 34:0:26:3 sdaq 66:160 active ready running
  |- 34:0:29:3 sdat 66:208 active ready running
  `- 34:0:31:3 sdaw 67:0   active ready running
mpathat (3600507640084031dd80000000000007c) dm-3 IBM     ,2145
size=100G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=enabled
| |- 33:0:19:1 sdf  8:80   active ready running
| |- 33:0:24:1 sdi  8:128  active ready running
| |- 33:0:25:1 sdl  8:176  active ready running
| |- 33:0:26:1 sdo  8:224  active ready running
| |- 34:0:20:1 sdam 66:96  active ready running
| |- 34:0:26:1 sdap 66:144 active ready running
| |- 34:0:29:1 sdas 66:192 active ready running
| `- 34:0:31:1 sdav 66:240 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 33:0:15:1 sdc  8:32   active ready running
  |- 33:0:27:1 sds  65:32  active ready running
  |- 33:0:28:1 sdy  65:128 active ready running
  |- 33:0:31:1 sdad 65:208 active ready running
  |- 34:0:13:1 sdag 66:0   active ready running
  |- 34:0:15:1 sdaj 66:48  active ready running
  |- 34:0:1:1  sdt  65:48  active ready running
  `- 34:0:3:1  sdz  65:144 active ready running
mpathas (3600507640084031dd80000000000007b) dm-2 IBM     ,2145
size=100G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=enabled
| |- 33:0:15:0 sdb  8:16   active ready running
| |- 33:0:27:0 sdq  65:0   active ready running
| |- 33:0:28:0 sdw  65:96  active ready running
| |- 33:0:31:0 sdab 65:176 active ready running
| |- 34:0:13:0 sdae 65:224 active ready running
| |- 34:0:15:0 sdai 66:32  active ready running
| |- 34:0:1:0  sdr  65:16  active ready running
| `- 34:0:3:0  sdx  65:112 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 33:0:19:0 sde  8:64   active ready running
  |- 33:0:24:0 sdh  8:112  active ready running
  |- 33:0:25:0 sdk  8:160  active ready running
  |- 33:0:26:0 sdn  8:208  active ready running
  |- 34:0:20:0 sdal 66:80  active ready running
  |- 34:0:26:0 sdao 66:128 active ready running
```

```
  |- 34:0:29:0 sdar 66:176 active ready running
  `- 34:0:31:0 sdau 66:224 active ready running
```

b. The **rescan-scsi-bus.sh -r** command (see Example 7-62 on page 610) rescans for new devices and can be necessary if any changes were made to the storage on the SAN.

   The **multipath -F** command flushes the configuration of the multipath driver. then the **multipath** command builds a new configuration for new devices and paths.

   The **multipath -ll** command provides information about path states and to which the `mpath n` device capacity was mapped for each mapped VDisk (see the UUID (universally unique identifier) without digit `3` at the beginning).

c. To start using the capacities that are provided as logical volumes, use *only* the `/dev/mapper/mpath n` device for access. For example, to use the VDisk `600507640084031dd80000000000007c` as a logical volume manager (LVM) physical volume, use the name of the RHEL device (which was mapped by the multipath driver and can be obtained by running the **multipath -ll** command).

   Example 7-62 on page 610 shows the output. `mpathat` with `UID 3600507640084031dd80000000000007c` is marked bold in the example.

d. This new physical volume can be added to the volume group of the host, and logical volumes can be created or extended and configured for any application on the host, as shown in Example 7-63.

Example 7-63 Creating a physical volume in LVM for further use

```
20201028-14:39:10 root@redbookvm7-1:/dev/mapper # pvcreate
/dev/mapper/mpathat
  Physical volume "/dev/mapper/mpathat" successfully created.
20201028-14:39:55 root@redbookvm7-1:/dev/mapper # pvs
  PV                   VG   Fmt  Attr PSize   PFree
  /dev/mapper/mpathat       lvm2 ---  100.00g 100.00g
  /dev/sda2            rhel lvm2 a--  <15.00g      0
```

Summary

Consider the following when provisioning capacity from the storage system to the host:

- On the storage system:
  - Create the host object definition with all the necessary credentials.
  - Map volumes to the defined host object to introduce capacity to the host.
- On the host:
  - Make sure the multipathing driver is configured (usually, the native multipathing driver or device mapper are configured and running in some operating system). It is used to map all paths for the specific volume (VDisk) to the one device because the specifics of the protocol system see each path as a separate device even for the one volume (VDisk).
  - Set up the LVM layer if you plan to use it for more flexibility.
  - Set the file system level, depending on the application.

7.11.3 iSCSI host connectivity and capacity allocation

The iSCSI protocol uses an initiator from host side to send SCSI commands to the storage system's target devices. Therefore, it is necessary to prepare the correct environment on the host side and configure the storage system as described in "Creating iSCSI host objects" on page 565.

This section demonstrates a RHEL host configuration and how to obtain access to the dedicated volumes (VDisks) on the storage system.

For more information about preparing a RHEL host for SCSI connectivity, see this IBM Documentation web page. Select the specific system; then go to **Configuring** → **Host Attachment** → **iSCSI Ethernet host attachment**.

Complete the following steps:

1. Install the iSCSI initiator on the RHEL host by running the **yum** command (see Example 7-64).

   Example 7-64 Installing iscsi-initiator-utils

   ```
   20201028-18:13:51 root@redbookvm7-1:/mnt/disc # yum install
   iscsi-initiator-utils
   ```

2. Now, the iSCSI initiator should be configured, and the connection credentials should be set in the `/etc/iscsi` files. Check or define IQN in `/etc/iscsi/initiatorname.iscsi` (see Example 7-65).

   Example 7-65 Checking the initiator's IQN

   ```
   20201028-19:04:34 root@redbookvm7-1:/etc/iscsi # cat initiatorname.iscsi
   InitiatorName=iqn.1994-05.com.redhat:f3de6ef11811
   ```

3. Restart the iSCSI initiator service if the IQN was modified.
4. After the host is ready and the iSCSI initiator is configured, define a host object on the storage system, as described in "Creating iSCSI host objects" on page 565.

   Ensure that the IQN is set correctly in the Host Details window in the Port Definitions tab (see Figure 7-108 on page 614).

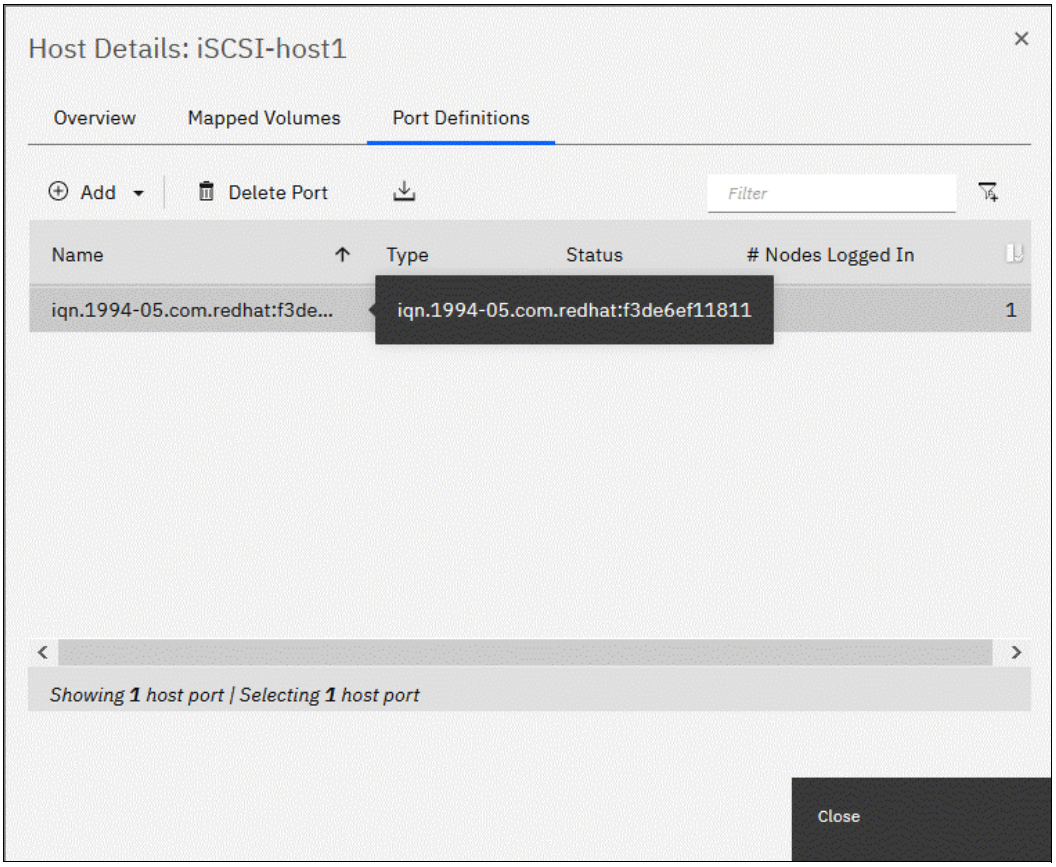


Figure 7-108 Host Details: Port Definitions tab

5. Map the dedicated volumes (VDisks) to the host object.
6. After the iSCSI host object is configured in the storage system and the volumes (VDisks) are mapped to it, the iSCSI targets must be discovered from the host. This discovery can be done by using one of the following methods:
   - Using send targets
   - iSNS

   Record the IP address under the IP column of the Ethernet Ports configuration tab that is shown in Figure 7-109 on page 615, which was configured for an iSCSI connection as described in “Creating iSCSI host objects” on page 565. It is used to find the IQN of our target and for further logins.

   Example 7-66 shows discovery by using the send targets method.

Example 7-66 iSCSI targets discovery

```
20201028-19:04:40 root@redbookvm7-1:/etc/iscsi # iscsiadm --mode discovery
--type sendtargets --portal 9.71.42.61
9.71.42.61:3260,1 iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node1
20201028-19:15:53 root@redbookvm7-1:/etc/iscsi # iscsiadm --mode discovery
--type sendtargets --portal 9.71.42.67
9.71.42.67:3260,1 iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node2
```

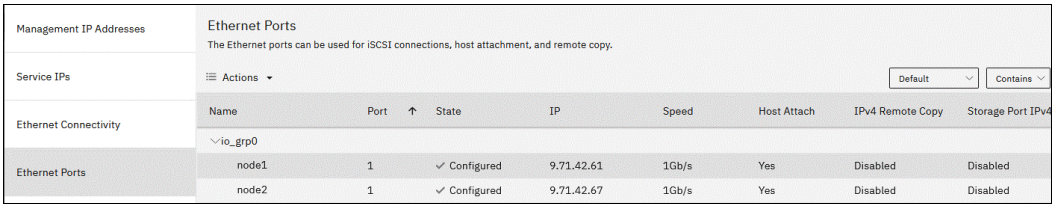


Figure 7-109 Ethernet Ports Configuration tab

For the iSNS discovery method, complete the following steps:

1. Update the configuration file `/etc/iscsi/iscsid.conf` and provide the connection credentials of the iSNS server. Place them into the following variable if it is available in the environment:

   ```
   isns.address = <iSNS server IP address>
   isns.port = <iSNS server port>
   ```

2. Restart the iSCSI initiator service to make the configuration active.
3. Run **`iscsiadm --mode discovery --type isns`** to generate the list of all iSCSI targets that are registered with the iSNS server.
4. Access the volumes (VDisk) space, which was mapped on the storage system to the host object, by logging in to the discovered targets (see Example 7-67).

Example 7-67 Logging in to the discovered targets/storage

```
20201028-19:16:09 root@redbookvm7-1:/etc/iscsi # iscsiadm --mode node --target
iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node1 --portal 9.71.42.61 --login
Logging in to [iface: default, target: iqn.1986-03.com.ibm:2145.ibmIBM
FlashSystem7200.node1, portal: 9.71.42.61,3260] (multiple)
Login to [iface: default, target: iqn.1986-03.com.ibm:2145.ibmIBM
FlashSystem7200.node1, portal: 9.71.42.61,3260] successful.
20201028-19:17:56 root@redbookvm7-1:/etc/iscsi # iscsiadm --mode node --target
iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node2 --portal 9.71.42.67 --login
Logging in to [iface: default, target: iqn.1986-03.com.ibm:2145.ibmIBM
FlashSystem7200.node2, portal: 9.71.42.67,3260] (multiple)
Login to [iface: default, target: iqn.1986-03.com.ibm:2145.ibmIBM
FlashSystem7200.node2, portal: 9.71.42.67,3260] successful.
```

5. After logging in successfully, ensure that the native multipath driver on the RHEL host was installed and configured correctly per the example with the FC connection that is described in 7.11.2, “Fibre Channel host connectivity and capacity allocation” on page 608, and check the output by running **`multipath -ll`**.

Example 7-68 shows an example of the output.

Example 7-68 Multipathing driver/device mapper output

```
20201028-19:39:22 root@redbookvm7-1:/etc/iscsi # multipath -ll
mpathaw (3600507640084031dd800000000000060) dm-6 IBM     ,2145
size=100G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| `- 35:0:0:0  sdbn 68:16  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  `- 36:0:0:0  sdbr 68:80  active ready running
mpathaz (3600507640084031dd800000000000067) dm-9 IBM     ,2145
size=250G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=active
```

```
| `- 36:0:0:3  sdbu 68:128 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  `- 35:0:0:3  sdbq 68:64  active ready running
mpathay (3600507640084031dd800000000000066) dm-8 IBM     ,2145
size=250G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| `- 35:0:0:2  sdbp 68:48  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  `- 36:0:0:2  sdbt 68:112 active ready running
mpathax (3600507640084031dd800000000000061) dm-7 IBM     ,2145
size=100G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| `- 36:0:0:1  sdbs 68:96  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  `- 35:0:0:1  sdbo 68:32  active ready running
```

Record the names of the devices that are marked in bold in Example 7-68 on page 615 because they are used for more configuration tasks, such as LVM physical volume creation or file system creation and mounting, which are in `/dev/mapper/`.

Record the UID number after devices names without the first digit (`3`) because they correspond to the UID of the volume (VDisk) on the storage system.

Summary

Although the example in this section is specifically for RHEL host connectivity, the main principals can be followed when configuring connectivity through iSCSI for any other operating systems.

In summary, the following tasks are necessary for host to storage iSCSI connectivity:

1. Install the iSCSI initiator software on the host.
2. Configure the iSCSI initiator software according to the requirements for the storage system target and the host's OS.
3. Get the host IQNs.
4. Define the host object with iSCSI connectivity by using host IQNs.
5. Record and check the Ethernet ports IP addresses on the storage system, which are configured for iSCSI connectivity.
6. Discover the iSCSI targets by using the storage system IP address that was obtained in step 5.
7. Log in to the storage system iSCSI targets.
8. Check and configure the native multipath driver to confirm the volumes on the host.

7.11.4 FC NVMe over Fabric host connectivity example

NVMe-oF uses different fabrics for transport by using the NVMe protocol. In this example, we use an FC fabric for our NVMe connectivity from the RHEL host to IBM FlashSystem systems.

Start by defining the necessary connectivity information and configuring the host and system.

As with iSCSI connectivity, an NVMe-oF initiator and target must be defined and configured so that the connection works.

Collect the information for connectivity from the host to the system by completing the following steps:

1. Run the command that is shown in Example 7-69 to discover the WWPNs of the host as FC-NVMe connectivity is achieved through FC.

Example 7-69 Obtaining host WWPNs

```
[root@flashlnx4 fc_host]# cat /sys/class/fc_host/host*/port_name
0x10000090faf20bc0
0x10000090faf20bc1
```

2. Discover the NVMe FC ports for the system by running the command that is shown in Example 7-70 and determine which to use. The FC-NVMe connectivity dedicated port is a virtualized port; therefore, NPIV must be enabled.

Example 7-70 Discovering the NVMe FC ports

```
IBM_IBM FlashSystem:FS9500-1:redbook>lstargetportfc|grep -i nvme
id  WWPN             WWNN             port_id owning_node_id current_node_id nportid host_io_permitted virtualized
protocol
3   50050768101901E5 50050768100001E5 1       1             1               080E02  yes               yes         nvme
6   50050768101A01E5 50050768100001E5 2       1             1               020102  yes               yes         nvme
9   50050768101B01E5 50050768100001E5 3       1             1               020102  yes               yes         nvme
12  50050768101C01E5 50050768100001E5 4       1                             000000  yes               yes         nvme
15  5005076810290 1E5 5005076810000 1E5 5       1             1               330242  yes               yes         nvme
18  50050768102A01E5 50050768100001E5 6       1             1               340242  yes               yes         nvme
21  50050768102B01E5 50050768100001E5 7       1                             000000  yes               yes         nvme
24  50050768102C01E5 50050768100001E5 8       1                             000000  yes               yes         nvme
51  50050768101901DF 50050768100001DF 1       2             2               080F02  yes               yes         nvme
54  50050768101A01DF 50050768100001DF 2       2             2               021002  yes               yes         nvme
57  50050768101B01DF 50050768100001DF 3       2             2               021002  yes               yes         nvme
60  50050768101C01DF 50050768100001DF 4       2                             000000  yes               yes         nvme
63  50050768102901DF 50050768100001DF 5       2             2               330342  yes               yes         nvme
66  5005076810 2A01DF 5005076810 0001DF 6       2             2               340342  yes               yes         nvme
```

3. Zone the host with at least one NVMe dedicated port. In Example 7-70, the host is zoned to the ports that are marked in bold.

4. On the host, make sure that the driver is ready to provide NVMe connectivity. In this example, we use an Emulex HBA (see Example 7-71).

Example 7-71 Checking NVMe support for the lpfc driver

```
[root@flashlnx4 fc_host]# cat /etc/modprobe.d/lpfc.conf
options lpfc lpfc_enable_fc4_type=3
```

If the `lpfc.conf` is absent or does not contain the string that is marked in bold in the example, create it and populate it with the string. then restart the lpfc driver by running **`modprob`** commands (first, remove the driver, and then add it back).

Note: Reinitiating the lpfc driver by running the **`modprob`** command changes the NQN of the host.

5. Check that `nvme-cli` and `nvmefc-connect` are installed on the host, as shown in Example 7-72.

Example 7-72 Checking the nvme-cli and nvme-connect availability

```
root@flashlnx4 nvme]# rpm -qa|grep nvme
nvme-cli-1.6-1.el7.x86_64
nvmefc-connect-12.6.61.0-1.noarch
```

Install these packages if they are not installed.

6. Obtain the NQN (see Example 7-73) from the host because it is used to define the host objects on the system.

Example 7-73 Obtaining the NQN

```
[root@flashlnx4 nvme]# cat /etc/nvme/hostnqn
nqn.2014-08.org.nvmexpress:uuid:0c3f53f4-8161-49c6-aaeb-a98d8e5f572c
```

7. Create a host object on the system by using the host NQN, as described in “rCeating FC NVMe host objects” on page 567. Check that the host object has the correct NQN set (see Figure 7-110).

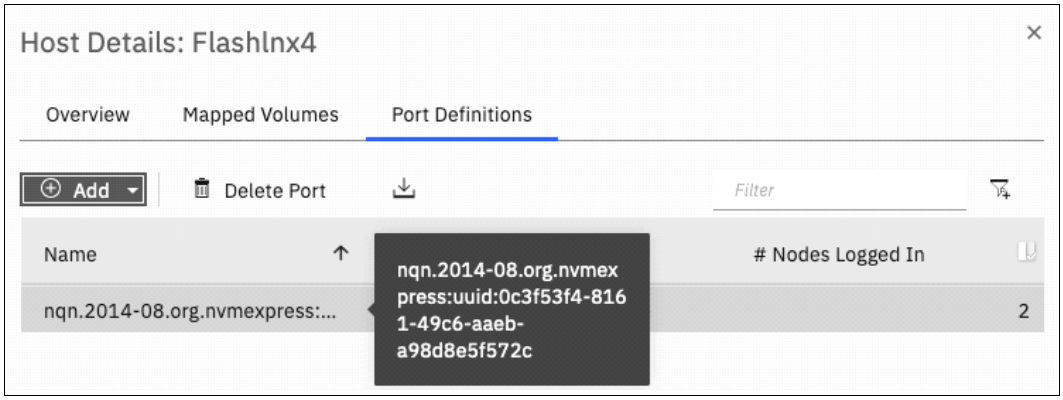


Figure 7-110 Checking the host object NQN on the system

8. On the system, map the volumes to the host object.
9. If the zoning was done correctly, the host object is created on the system and the necessary host utilities and drivers are configured. Verify the target ports that the host can see (see Example 7-74). This information is used in the discovery and connection process.

Example 7-74 Verifying the remote/target ports and information about the FC-NVMe connection

```
[root@flashlnx4 nvme]# cat /sys/class/scsi_host/*/nvme_info

NVMe Initiator Enabled
XRI Dist lpfc0 Total 6144 I/O 5894 ELS 250
NVMe LPORT lpfc0 WWPN x10000090faf20bc0 WWNN x20000090faf20bc0 DID x330040 ONLINE
NVMe RPORT       WWPN x500507605e8c3443 WWNN x500507605e8c3440 DID x333e40 TARGET DISCSRVC ONLINE
NVMe RPORT       WWPN x500507605e8c3463 WWNN x500507605e8c3440 DID x333f40 TARGET DISCSRVC ONLINE
NVMe RPORT       WWPN x50050768102901e5 WWNN x50050768100001e5 DID x330242 TARGET DISCSRVC ONLINE

NVMe Statistics
LS: Xmt 0000000031 Cmpl 0000000031 Abort 00000000
LS XMIT: Err 00000000  CMPL: xb 00000000 Err 00000000
Total FCP Cmpl 000000000035d907 Issue 000000000035d90a OutI/O 0000000000000003
```

```
   abort 00000001 noxri 00000000 nondlp 00000000 qdepth 00000000 wqerr 00000000 err 00000000
FCP CMPL: xb 00000001 Err 00000005

NVMe Initiator Enabled
XRI Dist lpfc1 Total 6144 I/O 5894 ELS 250
NVMe LPORT lpfc1 WWPN x10000090faf20bc1 WWNN x20000090faf20bc1 DID x340040 ONLINE
NVMe RPORT       WWPN x500507605e8c3453 WWNN x500507605e8c3440 DID x343e40 TARGET DISCSRVC ONLINE
NVMe RPORT       WWPN x500507605e8c3473 WWNN x500507605e8c3440 DID x343f40 TARGET DISCSRVC ONLINE
NVMe RPORT       WWPN x50050768102a01df WWNN x50050768100001df DID x340342 TARGET DISCSRVC ONLINE

NVMe Statistics
LS: Xmt 0000000030 Cmpl 0000000030 Abort 00000000
LS XMIT: Err 00000000  CMPL: xb 00000000 Err 00000000
Total FCP Cmpl 000000000035d6c3 Issue 000000000035d6c6 OutI/O 0000000000000003
   abort 00000001 noxri 00000000 nondlp 00000000 qdepth 00000000 wqerr 00000000 err 00000000
FCP CMPL: xb 00000001 Err 00000005
```

Tip: If the remote ports (RPORTs), which are presented from the system, are not visible, verify that the zoning was done correctly for the virtualized NVMe ports on the system.

10. Discover and connect to the storage resources, which requires information from the `nvme_info` file, such as the WWNN and WWPN of the local port (host port) and RPORT (storage port). This information can be cumbersome to collect and input manually; therefore, the script that is shown in Example 7-75 can be used to automate the process. The commands for **`nvme-cli`** are in bold.

Example 7-75 Script for FC-NVMe discovery and connection

```
[root@flashlnx4 tmp]# cat /tmp/disco_connect.bash
#!/bin/bash
#gather list of valid FC adapters by listing /sys/class/fc_host
for HOST in `ls -1 /sys/class/fc_host`;do
  host_wwpn=`grep LPORT /sys/class/scsi_host/${HOST}/nvme_info |awk '{print $5}' |sed -e 's/x//'`
  host_wwnn=`grep LPORT /sys/class/scsi_host/${HOST}/nvme_info |awk '{print $7}' |sed -e 's/x//'`
  #iterate through the list of available targets on each FC adapter
  for LINE in `grep RPORT /sys/class/scsi_host/${HOST}/nvme_info|awk '{print $4":"$6}'|sed -e's/x//g'`;do
    target_wwpn=`echo ${LINE}|cut -d: -f1`
    target_wwnn=`echo ${LINE}|cut -d: -f2`
    echo "Performing Discovery and Connection with hostwwpn: ${host_wwpn} hostwwnn: ${host_wwnn}
targetwwpn: ${target_wwpn} targetwwnn: ${target_wwnn}"
    nvme discover --transport=fc --traddr=nn-0x${target_wwnn}:pn-0x${target_wwpn}
--host-traddr=nn-0x${host_wwnn}:pn-0x${host_wwpn}
    #grab the host nqn from /etc/nvme/hostnqn
    NQN=`cat /etc/nvme/hostnqn`
    nvme connect --transport=fc --traddr=nn-0x${target_wwnn}:pn-0x${target_wwpn}
--host-traddr=nn-0x${host_wwnn}:pn-0x${host_wwpn} -n ${NQN}
  done
done
```

After the systems are successfully discovered and connected, record the ports that marked in bold in Example 7-76.

Example 7-76 Discovery and connection script output

```
[root@flashlnx4 tmp]# . /tmp/disco_connect.bash
Performing Discovery and Connection with hostwwpn: 10000090faf20bc0 hostwwnn:
20000090faf20bc0 targetwwpn: 500507605e8c3443 targetwwnn: 500507605e8c3440
```

```
Discovery Log Number of Records 1, Generation counter 0
=====Discovery Log Entry 0======
trtype:  fibre-channel
adrfam:  fibre-channel
subtype: nvme subsystem
treq:    not required
portid:  2
trsvcid: none
subnqn:
nqn.2017-12.com.ibm:nvme:mt:9840:guid:5005076061D30D60:cid:0000020061D16202
traddr:  nn-0x500507605e8c3440:pn-0x500507605e8c3443
Performing Discovery and Connection with hostwwpn: 10000090faf20bc0 hostwwnn:
20000090faf20bc0 targetwwpn: 0x500507605e8c3463 targetwwnn: 500507605e8c3440

Discovery Log Number of Records 1, Generation counter 0
=====Discovery Log Entry 0======
trtype:  fibre-channel
adrfam:  fibre-channel
subtype: nvme subsystem
treq:    not required
portid:  10
trsvcid: none
subnqn:
nqn.2017-12.com.ibm:nvme:mt:9840:guid:5005076061D30D60:cid:0000020061D16202
traddr:  nn-0x500507605e8c3440:pn-0x500507605e8c3463
Performing Discovery and Connection with hostwwpn: 10000090faf20bc0 hostwwnn:
20000090faf20bc0 targetwwpn: 0x50050768102901e5 targetwwnn: 50050768100001e5

Discovery Log Number of Records 1, Generation counter 0
=====Discovery Log Entry 0======
trtype:  fibre-channel
adrfam:  fibre-channel
subtype: nvme subsystem
treq:    unrecognized
portid:  4
trsvcid: none
subnqn:  nqn.1986-03.com.ibm:nvme:2145.00000204228003CA
traddr:  nn-0x50050768100001e5:pn-0x50050768102901e5
Performing Discovery and Connection with hostwwpn: 10000090faf20bc1 hostwwnn:
20000090faf20bc1 targetwwpn: 500507605e8c3453 targetwwnn: 500507605e8c3440

Discovery Log Number of Records 1, Generation counter 0
=====Discovery Log Entry 0======
trtype:  fibre-channel
adrfam:  fibre-channel
subtype: nvme subsystem
treq:    not required
portid:  6
trsvcid: none
subnqn:
nqn.2017-12.com.ibm:nvme:mt:9840:guid:5005076061D30D60:cid:0000020061D16202
traddr:  nn-0x500507605e8c3440:pn-0x500507605e8c3453
Performing Discovery and Connection with hostwwpn: 10000090faf20bc1 hostwwnn:
20000090faf20bc1 targetwwpn: 500507605e8c3473 targetwwnn: 500507605e8c3440
```

```
Discovery Log Number of Records 1, Generation counter 0
=====Discovery Log Entry 0======
trtype:  fibre-channel
adrfam:  fibre-channel
subtype: nvme subsystem
treq:    not required
portid:  14
trsvcid: none
subnqn:
nqn.2017-12.com.ibm:nvme:mt:9840:guid:5005076061D30D60:cid:0000020061D16202
traddr:  nn-0x500507605e8c3440:pn-0x500507605e8c3473
Performing Discovery and Connection with hostwwpn: 10000090faf20bc1 hostwwnn:
20000090faf20bc1 targetwwpn: 50050768102a01df targetwwnn: 50050768100001df

Discovery Log Number of Records 1, Generation counter 0
=====Discovery Log Entry 0======
trtype:  fibre-channel
adrfam:  fibre-channel
subtype: nvme subsystem
treq:    unrecognized
portid:  4101
trsvcid: none
subnqn:  nqn.1986-03.com.ibm:nvme:2145.00000204228003CA
traddr:  nn-0x50050768100001df:pn-0x50050768102a01df
```

Check the list of NVMe devices that are visible from the host, as shown in Example 7-77.

Example 7-77 NVMe devices list that is visible from the host

```
[root@flashlnx4 tmp]# nvme list
Node             S/N                   Model
Namespace Usage                      Format           FW Rev
---------------- -------------------- ----------------------------------------
--------- -------------------------- ---------------- --------
/dev/nvme6n1     204228003c            IBM     2145
78        132.07  GB / 137.44  GB    512   B +  0 B   8.4.0.0
/dev/nvme7n1     204228003c            IBM     2145
78        132.07  GB / 137.44  GB    512   B +  0 B   8.4.0.0
```

11. Run the **multipath** command to find newly connected volumes.
12. Run **multipath -ll** to see the paths and information about the volumes, as shown in Example 7-78.

Example 7-78 Output of the multipath -ll command

```
multipath -ll
...
eui.2800000000000053005076081080a000f dm-4 NVMe,IBM     2145
size=128G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=1 status=active
| `- 6:0:1:0  nvme6n1 259:0 active ready running
`-+- policy='service-time 0' prio=1 status=enabled
  `- 7:0:1:0  nvme7n1 259:1 active ready running
```

13. Record the name of the volume and UID that are listed in bold for use in future actions, such as adding a partition or implementing a volume as part of an LVM. In Figure 7-111, the same volume is shown on the system.

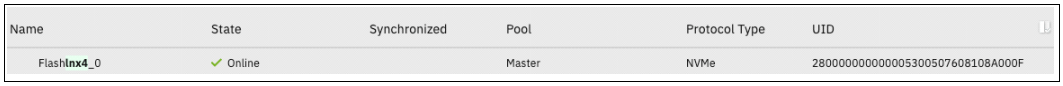


Figure 7-111 Volume as it is seen on the system

Summary

This section provided an example of the use of a RHEL host and an IBM FlashSystem 9100 system. Although other operating system distributions might have specific steps for configuration, the main ideas and principles are the same. If it is necessary to connect to the storage through FC-NVMe, consider the following points:

- Ensure that the host is ready and meets the requirements for FC-NVMe connectivity, such as the following examples:
  - HBA supports FC-NVMe
  - The drivers are configured for NVMe connectivity
- Ensure that the system supports the host HBA for FC-NVMe connectivity.
- Obtain the connectivity information from the host.
- Create a host object on the system by using connectivity information from the host.
- Map volumes to the host object.
- Perform discovery and connection from the host, although some hosts operating system can do it automatically.
- Use the obtained storage resource.

7.11.5 NVMe over RDMA host connectivity example

NVMe over RDMA uses different fabrics for transport by using the NVMe protocol. In this example, we use a RDMA for our NVMe connectivity from the RHEL host to IBM FlashSystem.

Start by defining the necessary connectivity information and configuring the host and system.

As with iSCSI connectivity, an NVMe over RDMA initiator and target must be defined and configured so that the connection works.

Collect the information for connectivity from the host to the system by completing the following steps:

1. Ensure that the required host network interface cards (NICs) supporting NVMe over RDMA are available and connected, as shown in Example 7-79.

Example 7-79 Checking host RDMA cards

```
[root@RHEL_HOST:]#  lspci | grep Mellanox
1a:00.0 Ethernet controller: Mellanox Technologies MT27710 Family [ConnectX-4 Lx]
1a:00.1 Ethernet controller: Mellanox Technologies MT27710 Family [ConnectX-4 Lx]
[root@RHEL_HOST:]#
```

2. Install the required RDMA driver and nvme packages on the host, as shown in Example 7-80.

Example 7-80 Installing the required RDMA packages on host

```
[root@RHEL_HOST:]# yum -y install rdma-core rdma-core-devel nvme-cli mstflint
libibverbs-utils
```

3. Find RDMA NICs and associated network interfaces as shown in Example 7-81.

Example 7-81 Finding RDMA NICs

```
[root@RHEL_HOST:]#  ls /sys/class/infiniband/*/device/net | grep -A1 mlx5
/sys/class/infiniband/mlx5_0/device/net:
enp26s0f0
--
/sys/class/infiniband/mlx5_1/device/net:
enp26s0f1
```

4. Load the required RDMA and NVMe drivers as shown in Example 7-82.

Example 7-82 Loading the required RDMA and NVMe drivers

```
[root@RHEL_HOST:]# modprobe mlx5_core
[root@RHEL_HOST:]# modprobe mlx5_ib
[root@RHEL_HOST:]# modprobe nvme-rdma
```

5. Ensure that the IP addresses to RDMA NICs are assigned statically or using Dynamic Host Configuration Protocol (DHCP), as shown in Example 7-83.

Example 7-83 Assigning IP addresses to RDMA NICs

```
[root@RHEL_HOST:]# ifconfig enp26s0f0
enp26s0f0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 10.0.1.20  netmask 255.255.255.0  broadcast 10.0.1.255
        inet6 fe80::37a5:5d76:aecd:28fd  prefixlen 64  scopeid 0x20<link>
        ether 0c:42:a1:b9:85:fa  txqueuelen 1000  (Ethernet)
        RX packets 95  bytes 7410 (7.2 KiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 1769  bytes 108312 (105.7 KiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

[root@RHEL_HOST:]# ifconfig enp26s0f1
enp26s0f1: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 10.0.2.20  netmask 255.255.255.0  broadcast 10.0.2.255
        inet6 fe80::db9b:fdd8:85cc:ad7f  prefixlen 64  scopeid 0x20<link>
        ether 0c:42:a1:b9:85:fb  txqueuelen 1000  (Ethernet)
        RX packets 65  bytes 5306 (5.1 KiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 214  bytes 14706 (14.3 KiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

Note: IBM generally recommends a Message Transfer Unit (MTU) size of 9000. However, as of this writing, the 100 GbE network cards that are in IBM Storage Virtualize systems support an FC MTU of only 1500.

6. Ensure that the necessary RDMA drivers for the suitable RDMA NICs are loaded, as shown in Example 7-84.

Example 7-84 Ensuring RDMA drivers are loaded

```
[root@RHEL_HOST:]# lsmod | grep mlx
mlx5_ib               376832  0
ib_uverbs             159744  2 rdma_ucm,mlx5_ib
ib_core               393216  13
rdma_cm,ib_ipoib,rpcrdma,ib_srpt,nvme_rdma,iw_cm,ib_iser,ib_umad,ib_isert,rdma_ucm
,ib_uverbs,mlx5_ib,ib_cm
mlx5_core            1384448  1 mlx5_ib
mlxfw                  32768  1 mlx5_core
pci_hyperv_intf        16384  1 mlx5_core
tls                   102400  1 mlx5_core
psample                20480  1 mlx5_core
```

7. Obtain the host nqn, as shown in Example 7-85.

Example 7-85 Obtaining the host nqn

```
[root@RHEL_HOST:]# cat /etc/nvme/hostnqn
nqn.2014-08.org.nvmexpress:uuid:0135e86a-98d2-460b-b0ab-6edb1795cdc3
```

8. Define the host in the IBM Storage Virtualize systems that are associating with the suitable portset that corresponds to the target ports in IBM Storage Virtualize system, as shown in Example 7-86.

Example 7-86 Defining NVMe over RDMA host

```
mkhost -name ITSO_NVME_RDMA_RHEL_HOST -nqn
nqn.2014-08.org.nvmexpress:uuid:0135e86a-98d2-460b-b0ab-6edb1795cdc3 -portset
portset0 -protocol rdmanvme  -type generic
Host, id [2], successfully created
```

9. Discover the targets from the host, as shown in Example 7-87.

Example 7-87 Discovering the RDMA targets over NVMe from host

```
[root@RHEL_HOST:]# nvme discover -t rdma -a 10.0.1.10

Discovery Log Number of Records 8, Generation counter 0
=====Discovery Log Entry 0======
trtype:  rdma
adrfam:  ipv4
subtype: nvme subsystem
treq:    unrecognized
portid:  15369
trsvcid: 4420
subnqn:  nqn.1986-03.com.ibm:nvme:2145.00000204E0400116
traddr:  10.0.1.10
rdma_prtype: roce-v2
rdma_qptype: connected
rdma_cms:    rdma-cm
rdma_pkey: 0x0000
=====Discovery Log Entry 1======
trtype:  rdma
adrfam:  ipv4
```

```
subtype: nvme subsystem
treq:    unrecognized
portid:  15370
trsvcid: 4420
subnqn:  nqn.1986-03.com.ibm:nvme:2145.00000204E0400116
traddr:  10.0.2.11
rdma_prtype: roce-v2
rdma_qptype: connected
rdma_cms:    rdma-cm
rdma_pkey: 0x0000
=====Discovery Log Entry 2======
trtype:  rdma
adrfam:  ipv4
subtype: nvme subsystem
treq:    unrecognized
portid:  17161
trsvcid: 4420
subnqn:  nqn.1986-03.com.ibm:nvme:2145.00000204E0400116
traddr:  10.0.1.14
rdma_prtype: roce-v2
rdma_qptype: connected
rdma_cms:    rdma-cm
rdma_pkey: 0x0000
=====Discovery Log Entry 3======
trtype:  rdma
adrfam:  ipv4
subtype: nvme subsystem
treq:    unrecognized
portid:  17162
trsvcid: 4420
subnqn:  nqn.1986-03.com.ibm:nvme:2145.00000204E0400116
traddr:  10.0.2.15
rdma_prtype: roce-v2
rdma_qptype: connected
rdma_cms:    rdma-cm
rdma_pkey: 0x0000
```

10. Connect to the target from the host by using the **nvme** CLI command, as shown in Example 7-88.

Example 7-88 Connecting to target from host using nvme cli

```
[root@RHEL_HOST:]# nvme connect-all -t rdma -a 10.0.2.11
```

11. Ensure that the host can see the storage subsystem, as shown in Example 7-89.

Example 7-89 List storage subsystem as seen by the host

```
[root@RHEL_HOST:]# nvme list-subsys
nvme-subsys0 - NQN=nqn.1986-03.com.ibm:nvme:2145.00000204E0400116
\
 +- nvme0 rdma traddr=10.0.1.10 trsvcid=4420 live
 +- nvme1 rdma traddr=10.0.2.11 trsvcid=4420 live
 +- nvme2 rdma traddr=10.0.1.14 trsvcid=4420 live
 +- nvme3 rdma traddr=10.0.2.15 trsvcid=4420 live
```

12. From Storage Virtualize system, map volume to the host, as described in 6.6.6, “Mapping a volume to a host” on page 473.
13. Ensure that the native nvme multipathing is set to `N` because RHEL-based operating systems use the device-mapper based multipath I/O option, as shown in Example 7-90.

Example 7-90 Native nvme multipath off for RHEL

```
[root@RHEL_HOST:]# systool -m nvme_core -A multipath
Module = "nvme_core"

    multipath           = "N"
```

Note: For NVMe over RDMA host connectivity, the native multipathing must be set to `N` for RHEL-based operating systems only. For more information about multipathing options for various supported operating systems, see the SSIC.

14. Ensure device mapper multipath daemon is running on the host, as shown in Example 7-91.

Example 7-91 Checking multipath daemon status

```
[root@RHEL_HOST:]# systemctl status multipathd
? multipathd.service - Device-Mapper Multipath Device Controller
   Loaded: loaded (/usr/lib/systemd/system/multipathd.service; enabled; vendor
preset: enabled)
   Active: active (running) since Thu 2022-03-31 21:02:09 CEST; 4 days ago
  Process: 7975 ExecStartPre=/sbin/multipath -A (code=exited, status=0/SUCCESS)
  Process: 7972 ExecStartPre=/sbin/modprobe -a scsi_dh_alua scsi_dh_emc
scsi_dh_rdac dm-multipath (code=exited, st>
 Main PID: 7977 (multipathd)
   Status: "up"
    Tasks: 7
   Memory: 14.5M
   CGroup: /system.slice/multipathd.service
           ??7977 /sbin/multipathd -d -s

Mar 31 21:02:09 RHEL_HOST systemd[1]: Starting Device-Mapper Multipath Device
Controller...
Mar 31 21:02:09 RHEL_HOST multipathd[7977]: --------start up--------
Mar 31 21:02:09 RHEL_HOST multipathd[7977]: read /etc/multipath.conf
Mar 31 21:02:09 RHEL_HOST multipathd[7977]: path checkers start up
Mar 31 21:02:09 RHEL_HOST systemd[1]: Started Device-Mapper Multipath Device
Controller.
```

Note: If the multipath is configured to start on the host, it can be configured to start by using the **`mpathconf --enable`** command.

15. Allow the host multipath the storage volume, as shown in Example 7-92.

Example 7-92 Multipathing storage volume on host

```
[root@RHEL_HOST:]# multipath -v4
```

16. Ensure that the volume is now multipathed, as shown in Example 7-93.

Example 7-93 Multipathed volume

```
[root@RHEL_HOST:]# multipath -ll
mpatha (eui.58000000000000084005076081383004) dm-3 NVME,IBM   2145
size=100G features='0' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| `- 2:1:1:13  nvme2n1 259:2 active ready running
|-+- policy='service-time 0' prio=50 status=enabled
| `- 3:17:1:13 nvme3n1 259:3 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| `- 0:0:1:13  nvme0n1 259:0 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  `- 1:16:1:13 nvme1n1 259:1 active ready running
```

Note: As of this writing, a SAN Boot of a host that uses NVMe over RDMA protocol is *not* supported for IBM Storage Virtuallize v8.6.0 systems.

Summary

This section provided an example of the use of a RHEL host and an IBM Storage Virtuallize system. Although other operating system distributions might have specific steps for their configuration, the main ideas and principles are the same.

If it is necessary to connect to the storage through NVMe over RDMA, consider the following points:

- Ensure that the host is ready and meets the requirements for NVMe over RDMA connectivity, such as the following examples:
  - The host NICs support NVMe over RDMA.
  - The drivers are configured for NVMe over RDMA connectivity.
  - The host NICs are suitably configured with storage network card ports through separate network switches or VLAN.
- Obtain the connectivity information from the host.
- Create a host object on the system by using connectivity information from the host.
- Map volumes to the host object.
- Perform discovery and connection from the host.
- Use the obtained storage resource.

7.11.6 NVMe over TCP host connectivity example

NVMe over TCP use the TCP/IP protocol for transport, which enables NVMe communication over standard Ethernet networks instead of dedicated fabrics like PCIe. In this example, we use TCP for NVMe connectivity from the RHEL host to IBM FlashSystem.

Start by defining the necessary connectivity information and configuring the host and system.

As with iSCSI connectivity, an NVMe over TCP initiator and target must be defined and configured so that the connection works.

Collect the information for connectivity from the host to the system by completing the following steps:

1. Ensure that the required host network interface cards (NICs) supporting NVMe over TCP are available and connected as shown in Example 7-94.

Example 7-94 checking host TCP cards

```
[root@RHEL_HOST:]#  lspci | grep Mellanox
1a:00.0 Ethernet controller: Mellanox Technologies MT27710 Family [ConnectX-4 Lx]
1a:00.1 Ethernet controller: Mellanox Technologies MT27710 Family [ConnectX-4 Lx]
[root@RHEL_HOST:]#
```

2. Install the required driver and nvme packages on the host, as shown in Example 7-95.

Example 7-95 Installing the required packages on host

```
[root@RHEL_HOST:]# yum -y install nvme-cli mstflint libibverbs-utils
```

3. Find NICs and associated network interfaces, as shown in Example 7-96.

Example 7-96 Finding NICs

```
[root@RHEL_HOST:]#  ls /sys/class/infiniband/*/device/net | grep -A1 mlx5
/sys/class/infiniband/mlx5_0/device/net:
enp26s0f0
--
/sys/class/infiniband/mlx5_1/device/net:
enp26s0f1
```

4. Load the required NVMe drivers, as shown in Example 7-97.

Example 7-97 Loading the required NVMe drivers

```
[root@RHEL_HOST:]# modprobe mlx5_core
[root@RHEL_HOST:]# modprobe mlx5_ib
[root@RHEL_HOST:]# modprobe nvme_tcp
```

5. Ensure that the IP addresses to NICs are assigned statically or using Dynamic Host Configuration Protocol (DHCP), as shown in Example 7-98.

Example 7-98 Assigning IP addresses to NICs

```
[root@RHEL_HOST:]# ifconfig enp26s0f0
enp26s0f0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 10.0.1.20  netmask 255.255.255.0  broadcast 10.0.1.255
        inet6 fe80::37a5:5d76:aecd:28fd  prefixlen 64  scopeid 0x20<link>
        ether 0c:42:a1:b9:85:fa  txqueuelen 1000  (Ethernet)
        RX packets 95  bytes 7410 (7.2 KiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 1769  bytes 108312 (105.7 KiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

[root@RHEL_HOST:]# ifconfig enp26s0f1
enp26s0f1: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 10.0.2.20  netmask 255.255.255.0  broadcast 10.0.2.255
        inet6 fe80::db9b:fdd8:85cc:ad7f  prefixlen 64  scopeid 0x20<link>
        ether 0c:42:a1:b9:85:fb  txqueuelen 1000  (Ethernet)
        RX packets 65  bytes 5306 (5.1 KiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 214  bytes 14706 (14.3 KiB)
```

```
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

Note: IBM generally recommends a Message Transfer Unit (MTU) size of 9000. However, as of this writing, the 100 GbE network cards that are in IBM Storage Virtualize systems support an MTU of only 1500.

6. Ensure that the necessary drivers for the suitable NICs are loaded, as shown in Example 7-99.

Example 7-99 Ensuring that drivers are loaded

```
[root@RHEL_HOST:]# lsmod | grep mlx
mlx5_ib               376832  0
ib_uverbs             159744  2 rdma_ucm,mlx5_ib
ib_core               393216  13
rdma_cm,ib_ipoib,rpcrdma,ib_srpt,nvme_rdma,iw_cm,ib_iser,ib_umad,ib_isert,rdma_ucm
,ib_uverbs,mlx5_ib,ib_cm
mlx5_core            1384448  1 mlx5_ib
mlxfw                  32768  1 mlx5_core
pci_hyperv_intf        16384  1 mlx5_core
tls                   102400  1 mlx5_core
psample                20480  1 mlx5_core
```

7. Obtain the host nqn, as shown in Example 7-100.

Example 7-100 Obtaining the host nqn

```
[root@RHEL_HOST:]# cat /etc/nvme/hostnqn
nqn.2014-08.org.nvmexpress:uuid:0135e86a-98d2-460b-b0ab-6edb1795cdc3
```

8. Define the host in the IBM Storage Virtualize systems that are associating with the suitable portset that corresponds to the target ports in IBM Storage Virtualize system, as shown in Example 7-101.

Example 7-101 Defining NVMe over TCP host

```
mkhost -name ITSO_NVME_TCP_RHEL_HOST -nqn
nqn.2014-08.org.nvmexpress:uuid:0135e86a-98d2-460b-b0ab-6edb1795cdc3 -portset
portset0 -protocol tcpnvme  -type generic
Host, id [2], successfully created
```

9. Discover the targets from the host, as shown in Example 7-102.

Example 7-102 Discovering the NVMe targets from host

```
[root@RHEL_HOST:]# nvme discover -t tcp -a 10.0.1.10

Discovery Log Number of Records 8, Generation counter 0
=====Discovery Log Entry 0======
trtype:  tcp
adrfam:  ipv4
subtype: current discovery subsystem
treq:    not specified, sq flow control disable supported
portid:  15369
trsvcid: 8009
subnqn:  nqn.1986-03.com.ibm:nvme:2145.00000204E0400116
traddr:  10.0.1.10
```

```
eflags:  not specified
sectype: none
=====Discovery Log Entry 1======
trtype:  tcp
adrfam:  ipv4
subtype: current discovery subsystem
treq:    not specified, sq flow control disable supported
portid:  15370
trsvcid: 8009
subnqn:  nqn.1986-03.com.ibm:nvme:2145.00000204E0400116
traddr:  10.0.2.11
eflags:  not specified
sectype: none
=====Discovery Log Entry 2======
trtype:  tcp
adrfam:  ipv4
subtype: current discovery subsystem
treq:    not specified, sq flow control disable supported
portid:  17161
trsvcid: 8009
subnqn:  nqn.1986-03.com.ibm:nvme:2145.00000204E0400116
traddr:  10.0.1.14
eflags:  not specified
sectype: none
=====Discovery Log Entry 3======
trtype:  tcp
adrfam:  ipv4
subtype: current discovery subsystem
treq:    not specified, sq flow control disable supported
portid:  17162
trsvcid: 8009
subnqn:  nqn.1986-03.com.ibm:nvme:2145.00000204E0400116
traddr:  10.0.2.15
eflags:  not specified
sectype: none
```

10. Connect to the target from the host by using the **nvme** CLI command, as shown in Example 7-103.

Example 7-103 Connecting to target from host using nvme cli

```
[root@RHEL_HOST:]# nvme connect-all
```

11. Ensure that the host can see the storage subsystem, as shown in Example 7-104.

Example 7-104 List storage subsystem as seen by the host

```
[root@RHEL_HOST:]# nvme list-subsys
nvme-subsys0 - NQN=nqn.1986-03.com.ibm:nvme:2145.00000204E0400116
\
 +- nvme0 tcp traddr=10.0.1.10 trsvcid=8009 live
 +- nvme1 tcp traddr=10.0.2.11 trsvcid=8009 live
 +- nvme2 tcp traddr=10.0.1.14 trsvcid=8009 live
 +- nvme3 tcp traddr=10.0.2.15 trsvcid=8009 live
```

12. From Storage Virtualize system, map volume to the host, as described in 6.6.6, “Mapping a volume to a host” on page 473.

13. Ensure that the native NVMe multipathing is set to `N` because RHEL-based operating systems use the device-mapper based multipath I/O option, as shown in Example 7-105.

Example 7-105 Native nvme multipath off for RHEL

```
[root@RHEL_HOST:]# systool -m nvme_core -A multipath
Module = "nvme_core"

    multipath           = "N"
```

Note: For NVMe over TCP host connectivity, the native multipathing must be set to `N` for RHEL-based operating systems only. For more information about multipathing options for various supported operating systems, see the SSIC.

14. Ensure the device mapper multipath daemon is running on the host, as shown in Example 7-106.

Example 7-106 Checking multipath daemon status

```
[root@RHEL_HOST:]# systemctl status multipathd
? multipathd.service - Device-Mapper Multipath Device Controller
   Loaded: loaded (/usr/lib/systemd/system/multipathd.service; enabled; vendor
preset: enabled)
   Active: active (running) since Thu 2022-03-31 21:02:09 CEST; 4 days ago
  Process: 7975 ExecStartPre=/sbin/multipath -A (code=exited, status=0/SUCCESS)
  Process: 7972 ExecStartPre=/sbin/modprobe -a scsi_dh_alua scsi_dh_emc
scsi_dh_rdac dm-multipath (code=exited, st>
 Main PID: 7977 (multipathd)
   Status: "up"
    Tasks: 7
   Memory: 14.5M
   CGroup: /system.slice/multipathd.service
           ??7977 /sbin/multipathd -d -s

Mar 31 21:02:09 RHEL_HOST systemd[1]: Starting Device-Mapper Multipath Device
Controller...
Mar 31 21:02:09 RHEL_HOST multipathd[7977]: --------start up--------
Mar 31 21:02:09 RHEL_HOST multipathd[7977]: read /etc/multipath.conf
Mar 31 21:02:09 RHEL_HOST multipathd[7977]: path checkers start up
Mar 31 21:02:09 RHEL_HOST systemd[1]: Started Device-Mapper Multipath Device
Controller.
```

Note: Multipath can be configured to start by using the **`mpathconf --enable`** command.

15. Allow the host multipath the storage volume, as shown in Example 7-107.

Example 7-107 Multipathing storage volume on host

```
[root@RHEL_HOST:]# multipath -v4
```

16. Ensure that the volume is now multipathed, as shown in Example 7-108.

Example 7-108 Multipathed volume

```
[root@RHEL_HOST:]# multipath -ll
mpatha (eui.58000000000000840050760813830004) dm-3 NVME,IBM    2145
```

```
size=100G features='0' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| `- 2:1:1:13  nvme2n1 259:2 active ready running
|-+- policy='service-time 0' prio=50 status=enabled
| `- 3:17:1:13 nvme3n1 259:3 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| `- 0:0:1:13  nvme0n1 259:0 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  `- 1:16:1:13 nvme1n1 259:1 active ready running
```

Note: As of this writing, a SAN Boot of a host that uses NVMe over TCP protocol is *not* supported for IBM Storage Virtuallize v8.6.0 systems.

Summary

This section provided an example of the use of a RHEL host and an IBM Storage Virtuallize system. Although other operating system distributions might have specific steps for their configuration, the main ideas and principles are the same.

If it is necessary to connect to the storage through NVMe over TCP, consider the following points:

- Ensure that the host is ready and meets the requirements for NVMe over TCP connectivity, such as the following examples:
  - The host NICs support NVMe over TCP.
  - The drivers are configured for NVMe over TCP connectivity.
  - The host NICs are suitably configured with storage network card ports through separate network switches or VLAN.
- Obtain the connectivity information from the host.
- Create a host object on the system by using connectivity information from the host.
- Map volumes to the host object.
- Perform discovery and connection from the host.
- Use the obtained storage resource.

7.12 Container Storage Interface drivers

IBM Storage Virtualize 8.4.2 introduced support for containers.

A container is a lightweight application package that can be easily moved between environments. Unlike VMs, they contain all the necessary `.bin` and `.lib` files and do not require an operating system to run.

They are typically managed through an automation and orchestration platform, such as Kubernetes or Red Hat OpenShift Container Platform

Because containers are stateless when they are moved or deleted, the corresponding data is lost. However, most applications require a persistent reserve that can be readily accessed between each deployment.

To achieve this functionality, Kubernetes developed an application interface that is called Container Storage Interface (CSI) as a way of presenting a persistent volume claim to

containerized applications. CSI allows dynamic provisioning storage for containers on Kubernetes and Red Hat OpenShift Container Platform that use IBM Storage subsystems.

An open source CSI driver is available from IBM at this GitHub web page.

For more information about the CSI driver and containers, see *Using the IBM Block Storage CSI Driver in a Red Hat*, REDP-5613.

8

Chapter 8

Storage migration

This chapter describes the steps that are involved in migrating data from an external storage controller to an IBM Storage Virtualize based system (IBM FlashSystems, IBM Storwize, and IBM SAN Volume Controller) by using the storage migration wizard.

Note: This chapter covers the storage migration wizard in detail, along with a less detailed description of the enclosure upgrade scenario. However, this chapter does not describe other migration methodologies, such as ones those that use replication or host-based migrations. This chapter also does not cover virtualization of external storage. For more information about these topics, see Chapter 8, “Storage migration” on page 635.

This chapter includes the following topics:

- 8.1, “Introduction” on page 636
- 8.2, “Storage migration overview” on page 637
- 8.3, “Storage migration wizard” on page 640
- 8.4, “Enclosure upgrade migration” on page 656
- 8.5, “Migrating data between systems nondisruptively” on page 657

© Copyright IBM Corp. 2024. All rights reserved.

8.1 Introduction

Migrating data from other storage systems to the IBM Storage Virtualize based system consolidates storage and enables IBM Storage Virtualize features, such as Easy Tier, thin provisioning, compression, encryption, storage replication, cyber-resilient point-in-time copies of volumes, and the easy-to-use GUI to be used across all volumes.

Note: The Safeguarded Copy function is available with software 8.4.2 or later, and is not supported for the IBM FlashSystem 5015, IBM FlashSystem 5035, FlashSystem 5010, FlashSystem 5030, Storwize V5010E, Storwize V5030E, Storwize V7000 Gen2, and Storwize V7000 Gen2+ models.

Storage migration uses the volume mirroring function to enable reads and writes during the migration, which minimizes disruption and downtime. After the migration completes, the external storage controller can be retired.

The FlashSystem supports migration through Fibre Channel (FC) and internet Small Computer Systems Interface (iSCSI) connections.

In addition to migrating data through external virtualization and volume mirroring that is used by the storage migration wizard, scenarios exist in which host-based mirroring is an alternative.

The migration of data within a SAN Volume Controller or from one pool to another is covered in Chapter 5, “Using storage pools” on page 365.

In environments where operating system administrators can perform the migration by using host-side tools, host-based mirroring might reduce the downtime if the new volumes that are presented from the IBM Storage Virtualize system and the established storage system are visible to the host concurrently. However, a restart is recommended to remove any stale entries of old device drivers and devices.

Note: For a demonstration of the storage migration capabilities that are offered with IBM Storage Virtualize, see IBM FlashSystem Product Tour. A login with an IBMid is required.

At the web page, click **Start** to explore the available options. The Explore external virtualization menu demonstrates the steps that are involved in scenarios, such as third-party storage array virtualization and storage migration.

One other important scenario is an enclosure upgrade migration, which is a fairly specialized case that is used specifically in an environment where an IBM Storwize system is upgrading to an IBM FlashSystem system or between different IBM FlashSystem systems. These scenarios can use three capabilities in IBM Storage Virtualize to provide a seamless transition to the new hardware:

- Clustering of the new control enclosure with the existing storage control enclosure (see Table 8-1 on page 637 for more information about the compatible systems).

  For more information, see IBM Storwize and FlashSystem Clustering Interoperability Matrix and Concurrent Compatibility and Code Cross Reference for IBM Storage Virtualize.

- Modifying an I/O Group or performing a Nondisruptive Volume Move (NDVM) to change the caching I/O group for the volumes.
- Volume mirroring to move the data.

Table 8-1 New hardware-clustering options for IBM Storage Virtualize storage systems

Source system type	Clustering options	Limitations and remarks
IBM Storwize V5030E (2072) IBM FlashSystem 5030 (2072) IBM FlashSystem 5035 (2072)	IBM FlashSystem 5030 (2072) IBM FlashSystem 5035 (2072)	► Up to 2 systems/ 4 nodes ► Mod. 212/224 and newer
IBM Storwize V5100 (2077/2078)	IBM Storwize V5100 (2077/2078)	Up to 2 systems/4 nodes
IBM Storwize V7000 (2076) IBM FlashSystem 7200 (2076)	IBM Storwize V7000 (2076) IBM FlashSystem 7200 (2076)	► Up to 4 systems/ 8 nodes ► Mod. 724 and newer
IBM FlashSystem 9100 (9848) IBM FlashSystem 9200 (9848)	IBM FlashSystem 9100 (9848) IBM FlashSystem 9200 (9848)	Up to 4 systems/8 nodes
IBM FlashSystem 5200 (4662)	IBM FlashSystem 5200 (4662)	Up to 4 systems/8 nodes
IBM FlashSystem 7200 (4664)	IBM FlashSystem 7200 (4664)	Up to 4 systems/8 nodes
IBM FlashSystem 7300 (4657)	IBM FlashSystem 7300 (4657)	Up to 4 systems/8 nodes
IBM FlashSystem 9500 (4666)	IBM FlashSystem 9500 (4666)	Up to 2 systems/4 nodes

Clustering as a migration method requires running the same code level on all participating systems. A further limiting factor as of this writing is the transition towards Expert Care because systems with a different service model cannot be clustered.

Moreover, with the IBM Storage Virtualize 8.4.2 code and later, it is possible to migrate data between two IBM Storage Virtualize systems nondisruptively without the need to cluster or externally virtualize. This volume mobility method can also be used to offload some workloads from one IBM Storage Virtualize system cluster to another.

The nondisruptive system migration uses the remote-copy relationship type method to migrate volumes between the source and the target IBM FlashSystem. With this method, the data migration of volumes between systems is done without any application downtime.

8.2 Storage migration overview

To migrate data from a storage controller to the IBM FlashSystem, you must use the built-in external virtualization capability. This capability places externally connected logical units (LUNs) that are under the control of the IBM Spectrum Virtualize system, which acts as a proxy while hosts continue to access them. The volumes are then fully virtualized in the system.

Note: The system does not require a license for its own control and expansion enclosures. However, a license is required for any external systems that are being virtualized based on storage capacity units (SCU)[a] or the number of enclosures. Data can be migrated from storage systems to your system by using the external virtualization function within 90 days of installation of the system without the purchase of a license. After 90 days, any ongoing use of the external virtualization function requires a license.

Set the license temporarily during the migration process to prevent messages that indicate that you are in violation of the license agreement from being sent. When the migration is complete or after 90 days, reset the license to its original limit or purchase more licenses.

a. https://www.ibm.com/docs/en/flashsystem-9x00/8.6.x?topic=software-licensed-functions

Consider the following points about the storage migration process:

- Typically, storage controllers divide storage into many Small Computer System Interface (SCSI) LUNs that are presented to hosts.
- Host access and I/O to the LUNs on the storage controllers must be stopped. This requirement is achieved by removing the FC or iSCSI host mapping and switch zoning on the fabric from the hosts to these storage controllers.
- The original LUNs from these storage controllers are then presented directly to the IBM FlashSystem and not to the hosts. The IBM FlashSystem discovers these external LUNs as *unmanaged* managed disks (MDisks).
- The unmanaged MDisks are *imported* to the IBM FlashSystem as image mode volumes and placed into a temporary storage pool. This storage pool is now a logical container for the LUNs.
- Each MDisk has a one-to-one mapping with an image mode volume. From a data perspective, the image mode volumes represent the LUNs exactly as they were before the import operation. The image mode volumes are on the same physical drives of the external storage controller and the data remains unchanged. The system is presenting active images of the LUNs and acting as a proxy.
- You might need to remove the storage system multipath device driver from the host and reconfigure the host attachment with IBM FlashSystem. However, most current operating systems might not require vendor-specific multipathing drivers and can access the established and new IBM Storage Virtualize-based systems through native multipathing drivers, such as AIX AIXPCM, Linux device mapper, or Microsoft Device Specific Module (MSDSM). The hosts are defined with worldwide port names (WWPNs) or iSCSI Qualified Names (IQNs), and the volumes are mapped to the hosts. After the volumes are mapped, the hosts discover the system's volumes through a host rescan or restart operation.
- After IBM Storage Virtualize volume mirroring operations are started, the image-mode volumes are mirrored to standard striped volumes. Volume mirroring is an online migration task, which means a host can still access and use the volumes during the mirror synchronization process.
- After the mirror operations are complete, the image mode volumes are removed. The external storage system LUNs are now migrated and the now redundant storage can be decommissioned or reused elsewhere.

Important: If you are migrating volumes from another Storwize or IBM FlashSystem family product through external virtualization instead of clustering or replication, the target system *must* be configured in the *replication* layer, and the source system must be configured in the *storage* layer. Otherwise, the source system does not discover the target as a host, and the target does not discover the source as a back-end controller.

The default layer setting for Storwize and IBM FlashSystem family systems is `storage`:

```
chsystem -layer replication
chsystem -layer storage
```

Similarly, the layer setting might need to be changed if you cluster an IBM Storwize system with an IBM FlashSystem enclosure.

8.2.1 Interoperability and compatibility

Interoperability is an important consideration when a new storage system is set up. Before attaching any external storage systems to the IBM FlashSystem, see the IBM System Storage Interoperation Center (SSIC).

At the SSIC site, select **IBM System Storage Enterprise Flash** for IBM FlashSystem 9500 or **IBM System Storage Midrange Disk** for other hardware platforms, such as the IBM Storwize family or IBM FlashSystem 7300 and then, select the suitable **Storage Controller Support** entry for your system as the Storage Model. You can refine your search by selecting the external storage controller that you want to use from the **Storage Controller** menu.

The matrix results indicate the external storage that you want to attach to the system, such as validated firmware levels or support for disks greater than 2 TB.

8.2.2 Prerequisites

Before the storage migration wizard can be started, the external storage controller must be visible to the system. You must also confirm that the restrictions, limits, and prerequisites are met.

Data from the external storage system to the IBM FlashSystem is sent through an iSCSI or Fibre Channel (FC) connection.

Common prerequisites

In VMware environments, Storage vMotion typically is used to move guest data transparently to newly provisioned data stores from the IBM FlashSystem, whereas storage virtualization can be a good option for Raw Device Mapping (RDM).

However, if you have VMware ESXi server hosts and want to migrate by using image mode, you must change the settings on the VMware host so that copies of the volumes can be recognized by the system after the migration completes. To ensure that volume copies can be recognized by the system for VMware ESXi hosts, you must complete one of the following actions:

- Enable the `EnableResignature` setting.
- Disable the `DisallowSnapshotLUN` setting.

For more information about these settings, see the documentation for the VMware ESXi host.

Note: Test the setting changes on a nonproduction server. The LUN has a different unique identifier (UID) after it is imported. It resembles a mirrored volume to the VMware server.

Prerequisites for a Fibre Channel connection

The following prerequisites for an FC connection must be met:

- An FC host interface card/host bus adapter (HIC/HBA) is installed in the node canisters.
- Cable this IBM FlashSystem into the storage area network (SAN) of the external storage that you want to migrate. Ensure that your IBM FlashSystem is cabled into the same SAN as the external storage controller that you are migrating.
- If you use FC, connect the FC cables to the FC ports in *both* canisters of your system, and then to the FC network.

  For more information, see Chapter 2, “Installation and configuration planning” on page 119. Alternatively, attach the external storage controller to the nodes instead of the use of a switched fabric.

Prerequisites for iSCSI connections

The following prerequisites for iSCSI connections must be met:

- Cable the IBM FlashSystem to the external storage system with a redundant switched fabric. Migrating iSCSI external storage requires that the IBM FlashSystem and the storage system are connected through an Ethernet switch. Symmetric ports on *all* nodes of the system must be connected to the same switch and must be configured on the same subnet.
- In addition, modify the Ethernet port attributes to enable the external storage on the Ethernet port to enable external storage connectivity. To modify the Ethernet port for external storage, click **Network** → **Ethernet Ports** and right-click a configured port. Select **Modify Storage Ports** to enable the port for external storage connections.
- Cable the Ethernet ports on the IBM FlashSystem to the fabric in the same way as the external storage system and ensure that they are configured in the same subnet. Optionally, you can use a virtual local area network (VLAN) to define network traffic for the system ports.
- For full redundancy, configure two Ethernet fabrics with separate Ethernet switches. If the source system nodes and the external storage system both have more than two Ethernet ports, an extra redundant iSCSI connection can be established for increased throughput.

8.3 Storage migration wizard

The storage migration wizard simplifies the migration task. The wizard features easy-to-follow windows that guide users through the entire process. The wizard shows you which commands are being run so that you can see exactly what is being performed throughout the process.

Note: The risk of losing data is low when the storage migration wizard is used correctly. However, it is prudent to avoid potential data loss by creating a backup of all the data that is stored on the hosts, storage controllers, and system before the wizard is used.

Everyday, the storage system automatically creates a backup of the control enclosure configuration data, which is stored in a file. This data is replicated across each control node canister within the system, ensuring redundancy. It is crucial to regularly download this file to the management workstation in order to safeguard the data effectively. This backup file is essential in the event of a critical failure that necessitates a system configuration restoration. It is particularly important to back up this file after making any modifications to the system configuration.

Use the following command to back up information about your system configuration, such as volumes, local Metro Mirror information, local Global Mirror information, storage pools, and nodes:

```
# svcconfig backup
```

It is considered a best practice to conduct testing of any configuration changes in a non-production environment prior to proceeding with data migration on production systems.

Complete the following steps to complete the migration by using the storage migration wizard:

1. Select **Pools** → **Import External Storage**, as shown in Figure 8-1. The System Migration window provides access to the storage migration wizard and displays information about the migration progress.

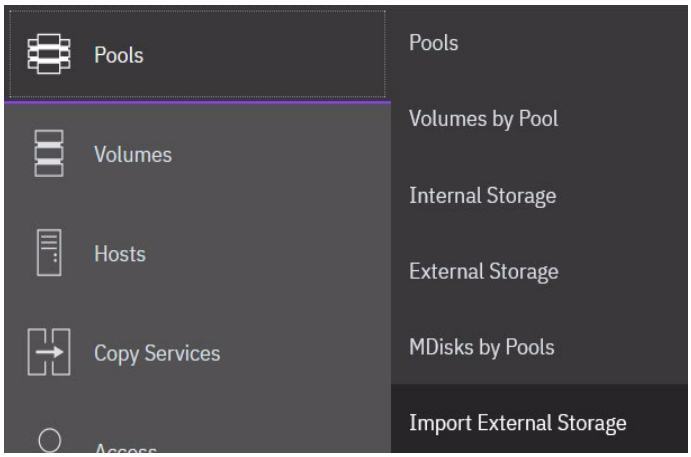


Figure 8-1 Browsing to Storage Migration

2. Click **Start New Import** to begin the storage migration wizard, as shown in Figure 8-2.

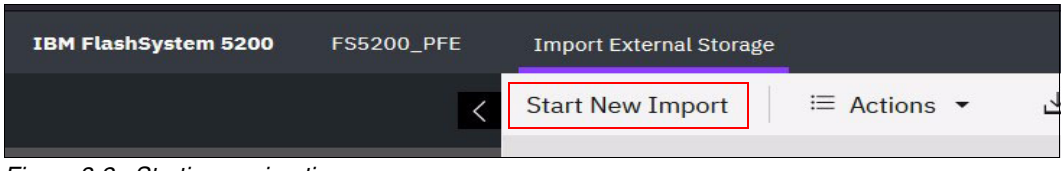


Figure 8-2 Starting a migration

Note: Starting a new migration adds the volume to be migrated to the list that is shown in Figure 8-2. After a volume is migrated, it remains in the list until you finalize the migration.

3. If both FC and iSCSI external systems are detected, a dialog box opens and prompts you to select which protocol is to be used. Select the type of attachment between the system and the external controller from which you want to migrate volumes and click **Next**. If only one type of attachment is detected, this dialog box does not open.

   If the external storage system is not detected, the warning message that is shown in Figure 8-3 is displayed when you attempt to start the migration wizard. Click **Close** and correct the problem before you try to start the migration wizard again.

   Probable causes include incorrect or incomplete SAN zoning. Also, verify that the correct layer is configured; that is, the source must be configured in the storage layer and the target in the replication layer.

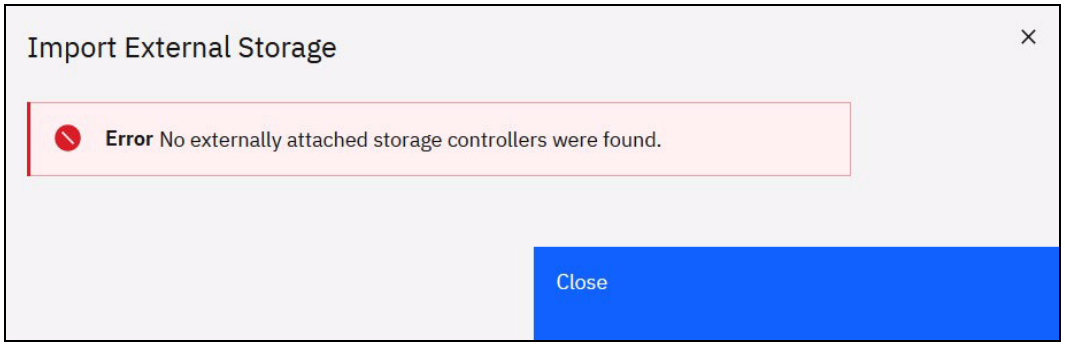


Figure 8-3 Error message if no external storage is detected

4. When the wizard starts, you are prompted to verify the restrictions and prerequisites that are listed in Figure 8-4 on page 643.

   Consider the following restrictions and prerequisites:

   – Storage migration restrictions:

     • Cannot be used to migrate clustered hosts, including clusters of VMware hosts and Virtual I/O Servers (VIOSs)

     • Is not used to migrate SAN boot images

     If you use either of these environments, the migration must be performed outside of the wizard because more steps are required.

     The VMware vSphere Storage vMotion feature might be an alternative for migrating VMware clusters. For more information, see Migration with Storage vMotion.

   – Prerequisites:

     • The system and the external storage controller are connected to the same SAN fabric.

     • If VMware ESXi hosts are involved in the data migration, the VMware ESXi hosts are set to allow volume copies to be recognized.

   For more information about the Storage Migration prerequisites, see 8.2.2, “Prerequisites” on page 639.

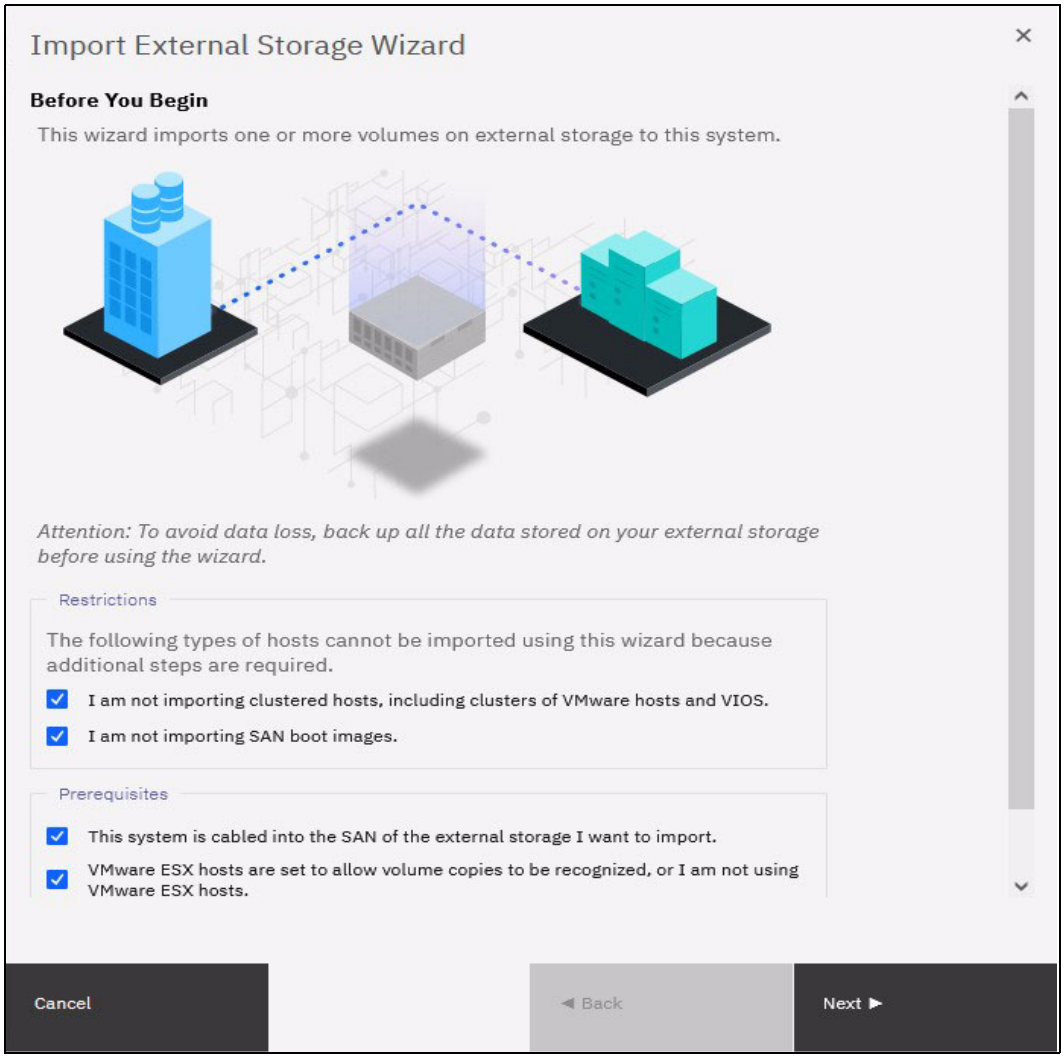


Figure 8-4 Restrictions and prerequisites confirmation

If all restrictions are satisfied and prerequisites are met, select all of the options and click **Next**, as shown in Figure 8-4.

5. Prepare the environment for migration by following the instructions that are shown in Figure 8-5.

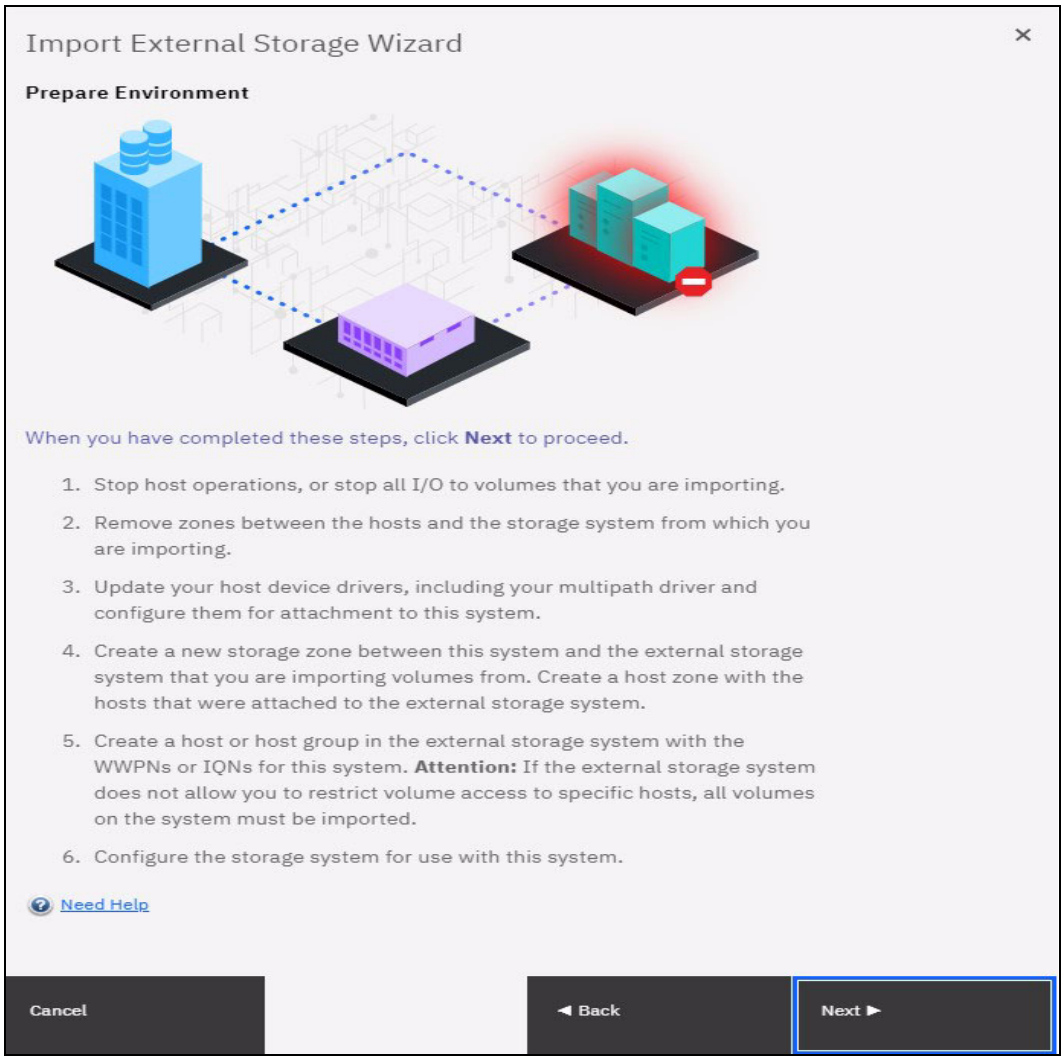


Figure 8-5 Preparing your environment for storage migration

The preparation phase includes the following steps:

a. Before migrating the storage, ensure that all host operations are stopped to prevent applications from generating I/Os to the migrated system. The best way to do so is to unmount the volumes on the host.

b. Remove all zones between the hosts and the controller that you are migrating.

c. Hosts usually do not support concurrent multipath drivers. You might need to remove drivers that are not compatible with the system from the hosts and use the recommended device drivers.

   For more information about supported drivers, see the SSIC.

d. If you are migrating external storage controllers that connect to an IBM FlashSystem that uses FC, ensure that you complete the suitable zoning changes to simplify migration. In fact, an excellent preparatory step is to present a test LUN from the external storage to the IBM FlashSystem before the migration.

For more information about guidelines for the zoning best practices for back-end storage and host zoning, see 2.7.2, “Zoning” on page 131.

Because the IBM FlashSystem now must be seen as a host from the external controller to be migrated, you must define the IBM FlashSystem as a host or host group by using the WWPNs or IQNs on the external system to be migrated. Because some controllers do not support LUN-to-host mapping, they present all the LUNs to the system. In this instance, migrate all of the LUNs.

6. If the previous preparation steps were followed, the IBM FlashSystem is now seen as a host from the controller to be migrated. LUNs can then be mapped to the IBM FlashSystem. Map the external storage controller by following the instructions that are shown in Figure 8-6.

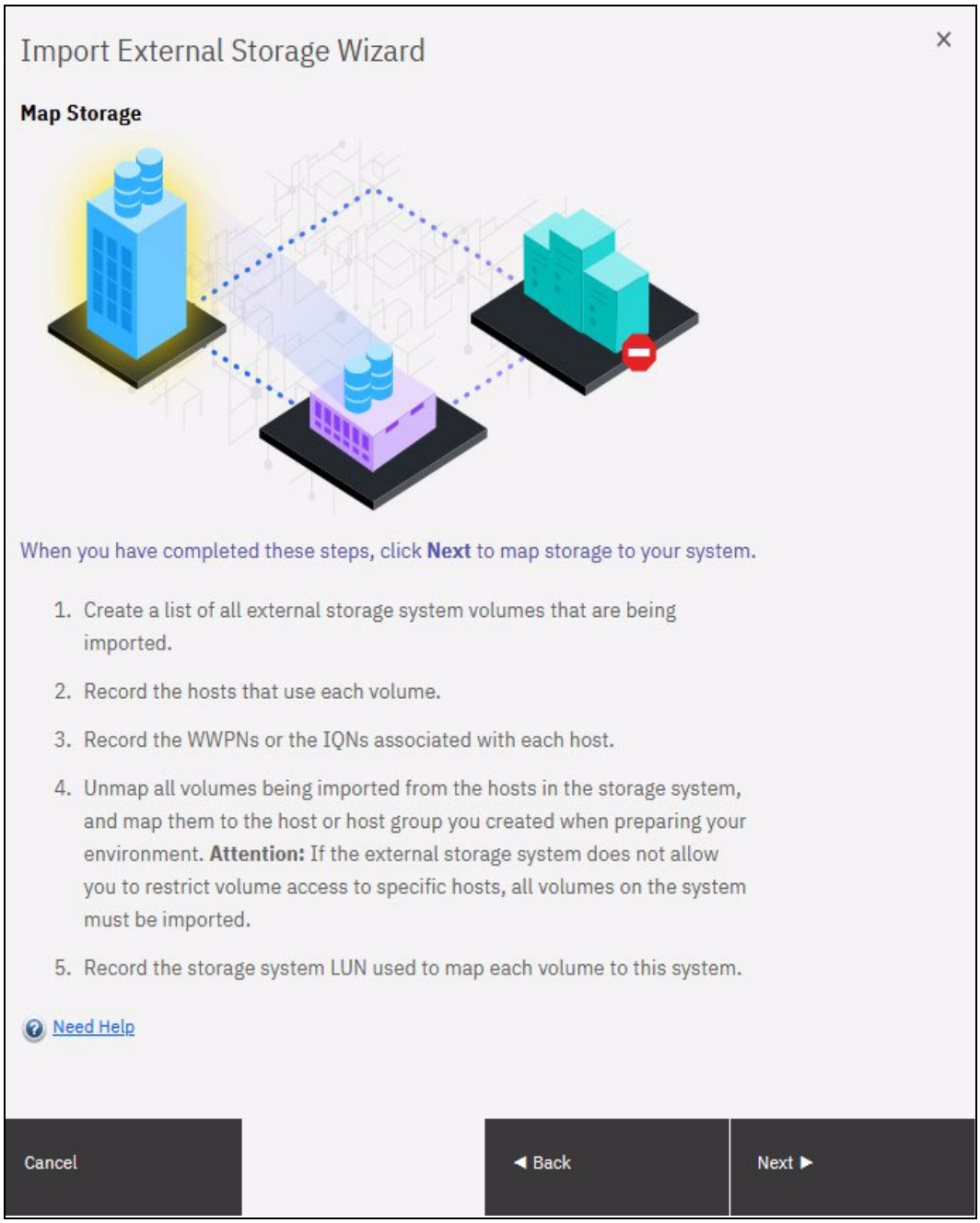


Figure 8-6 Steps to map the LUNs to be migrated

Before you migrate storage, record the hosts and their WWPNs or IQNs for each volume that is being migrated and the SCSI LUN when it is mapped to the system.

Table 8-2 shows an example of a table that is used to capture information that relates to the external storage system LUNs.

Table 8-2 Example table for capturing external LUN information

Volume name or ID	Hosts accessing this LUN	Host WWPNs or IQNs	SCSI LUN when mapped
1 DB2logs	DB2server	21000024FF2...	0
2 Db2data	DB2server	21000024FF2...	1
3 file system	FileServer1	21000024FF2...	2

Note: Make sure to record the SCSI ID of the LUNs to which the host is originally mapped. Some operating systems do not support changing the SCSI ID during the migration.

Click **Next** and wait for the system to discover external devices. The wizard runs a **`detectmdisk`** command, as shown in Figure 8-7.

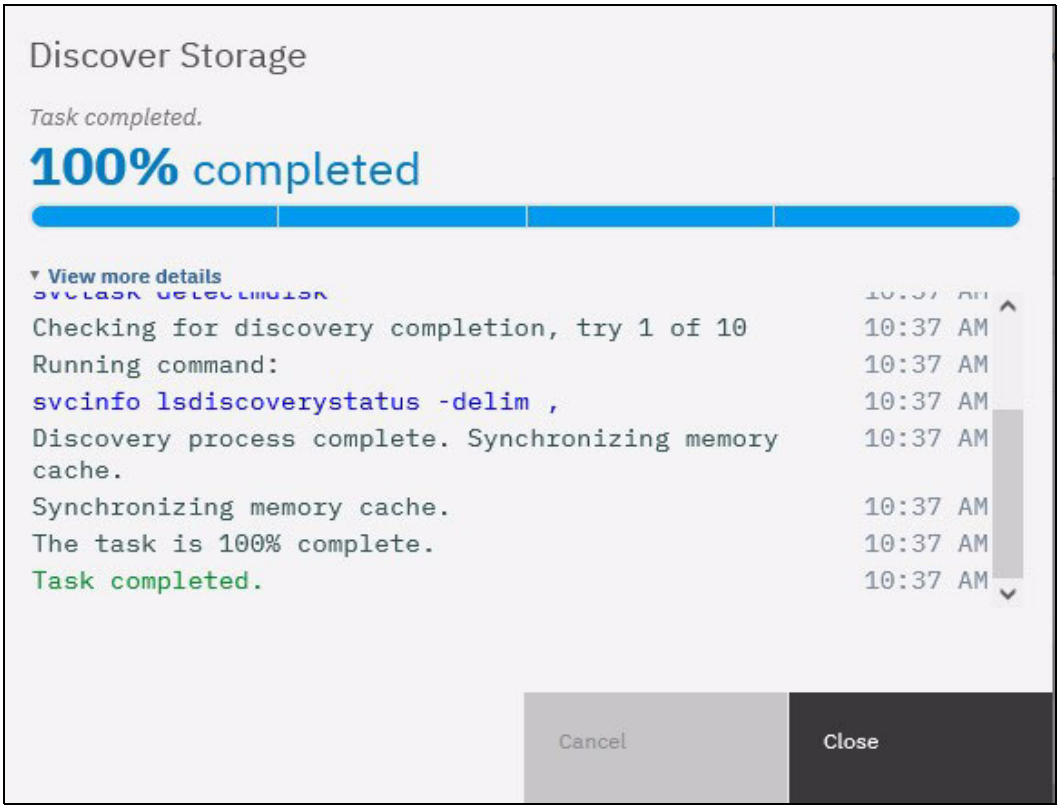


Figure 8-7 Storage Migration external storage discovery detectmdisk command detail

7. The next window shows all the MDisks that were found. If the MDisks to be migrated are not in the list, check your zoning or IP configuration (as applicable) and your LUN mappings. Repeat step 6 on page 645 to trigger the discovery procedure again.

Select the MDisks that you want to migrate, as shown in Figure 8-8.

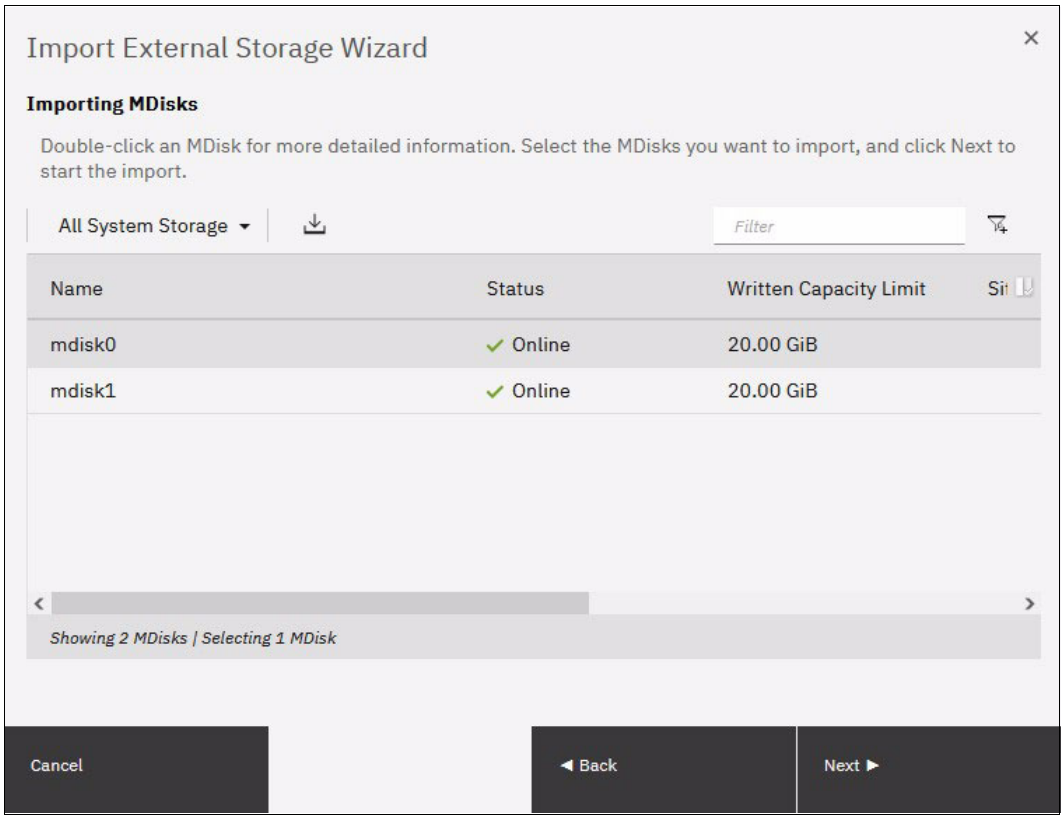


Figure 8-8 Discovering mapped LUNs from external storage

In this example, two MDisks, mdisk0 and mdisk1, were found for migration. Detailed information about an MDisk is available by double-clicking it. To select multiple elements from the table, press **Shift** and then click or **Ctrl** and then click. Optionally, you can export the discovered MDisks list to a comma-separated value (CSV) file for further use by clicking the download icon ( ) to **Export to CSV**.

Note: Select only the MDisks that are applicable to the current migration plan. After step 15 on page 654 of the current migration completes, another migration can be started to migrate any remaining MDisks.

8. Click **Next** and wait for the MDisk to be imported. During this task, the IBM FlashSystem creates a storage pool that is called `MigrationPool_XXXX` and adds the imported MDisk to the storage pool as image mode volumes with the default naming of `{controller}_16digitSequenceNumber (controller2_0000000000000005)...`, as shown in Figure 8-9.

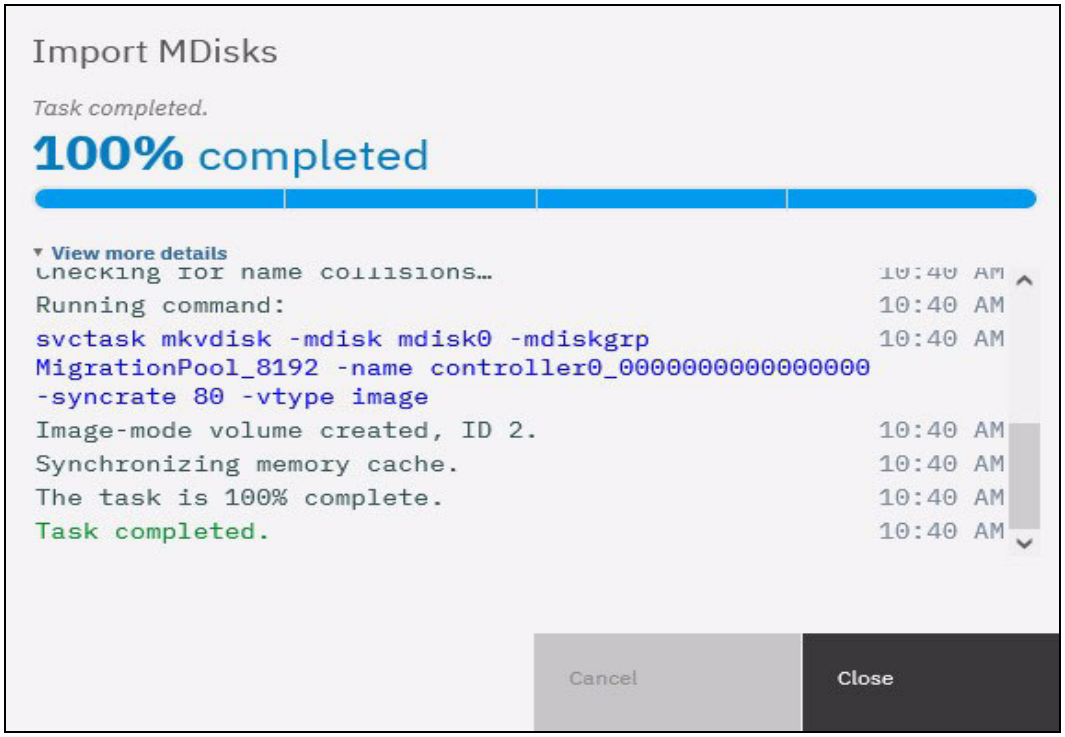


Figure 8-9 Storage Migration image mode volume creation detail

9. The next window lists the host that is configured on the IBM FlashSystem to which you can assign the volumes or configure new hosts. This step is optional and can be bypassed by clicking **Next**. In this example, the host `Host_A` is configured, as shown in Figure 8-10 on page 649. If no host is selected, you can create a host after the migration completes and map the imported volumes to it.
10. If the host that needs access to the migrated data is not configured, select **Add Host** to begin the Add Host wizard. Enter the host connection type, name, and connection details. Optionally, click **Advanced** to modify the host type and I/O group assignment. Figure 8-11 on page 649 shows that Add Host wizard with the details completed.

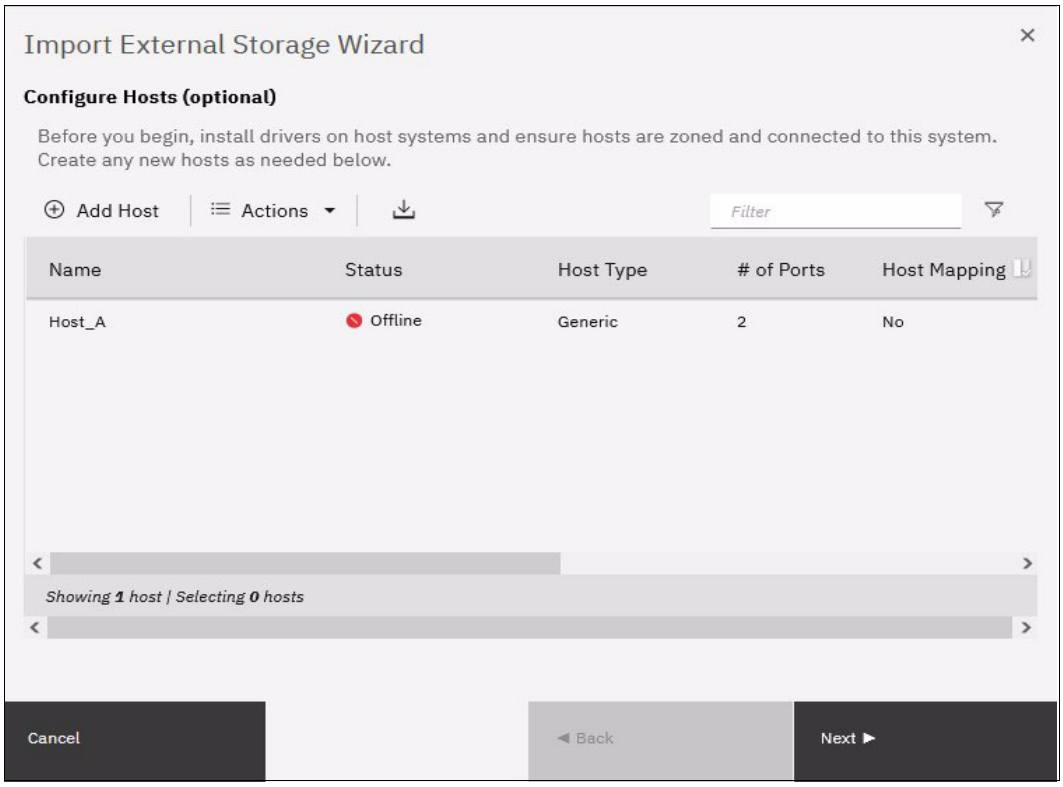


Figure 8-10 List of configured hosts to which to map the imported volume

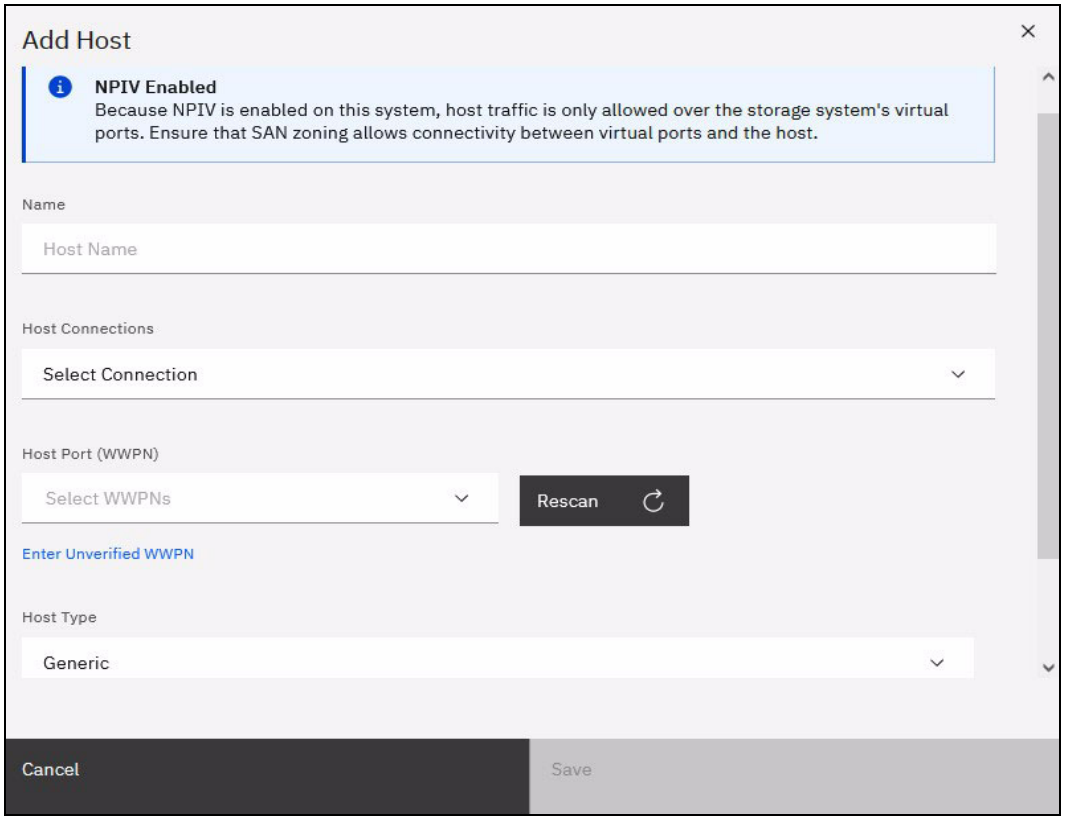


Figure 8-11 Creating a host during the migration process

For more information about the Add Host wizard, see Chapter 7, “Hosts” on page 517.

11. Click **Save** The host is created and now listed in the Configure Hosts window, as shown in Figure 8-10 on page 649. Click **Next** to proceed.

12. The next window lists the new volumes to where you can map them to hosts, as shown in Figure 8-12. The volumes are listed with names that were automatically assigned by the system. The names can be changed to reflect something more meaningful to the user by selecting the volume and clicking **Rename** in the **Actions** menu.

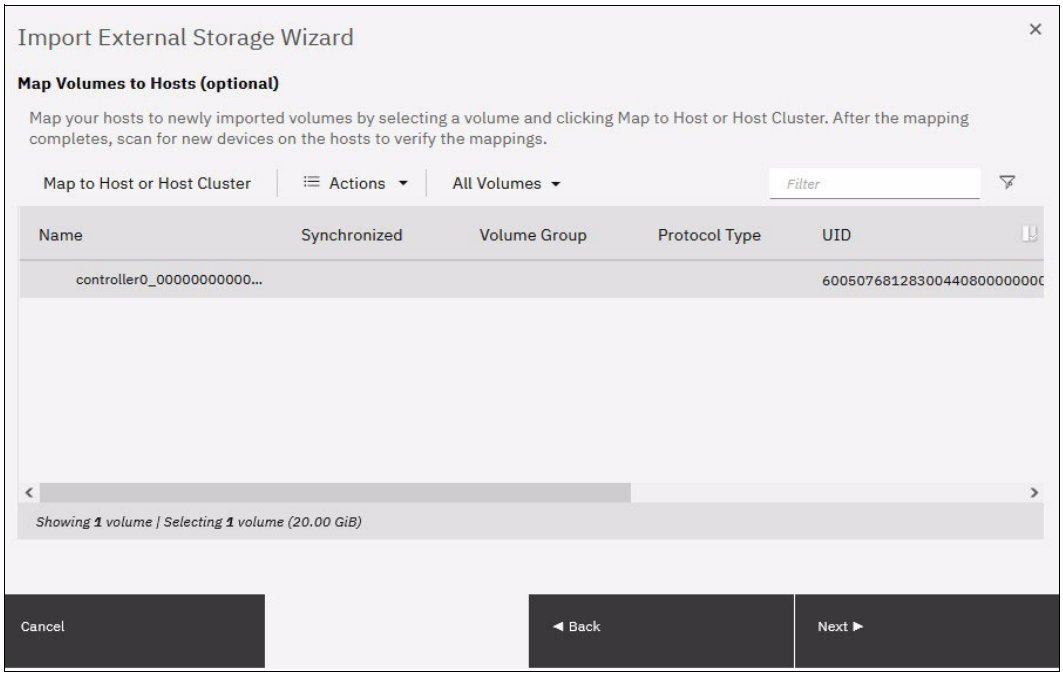


Figure 8-12 Mapping volumes to hosts

13. (This step is optional and can be bypassed by clicking **Next**.) Map the volumes to the hosts by selecting the volumes and clicking **Map to Host or Host Cluster**, as shown in Figure 8-13 on page 651.

    You can manually assign a SCSI ID to the LUNs you are mapping. This technique is useful when the host must have the same LUN ID for a LUN before and after it is migrated. To assign the SCSI ID manually, select the **Self Assign** option and follow the instructions as shown in Figure 8-14 on page 651.

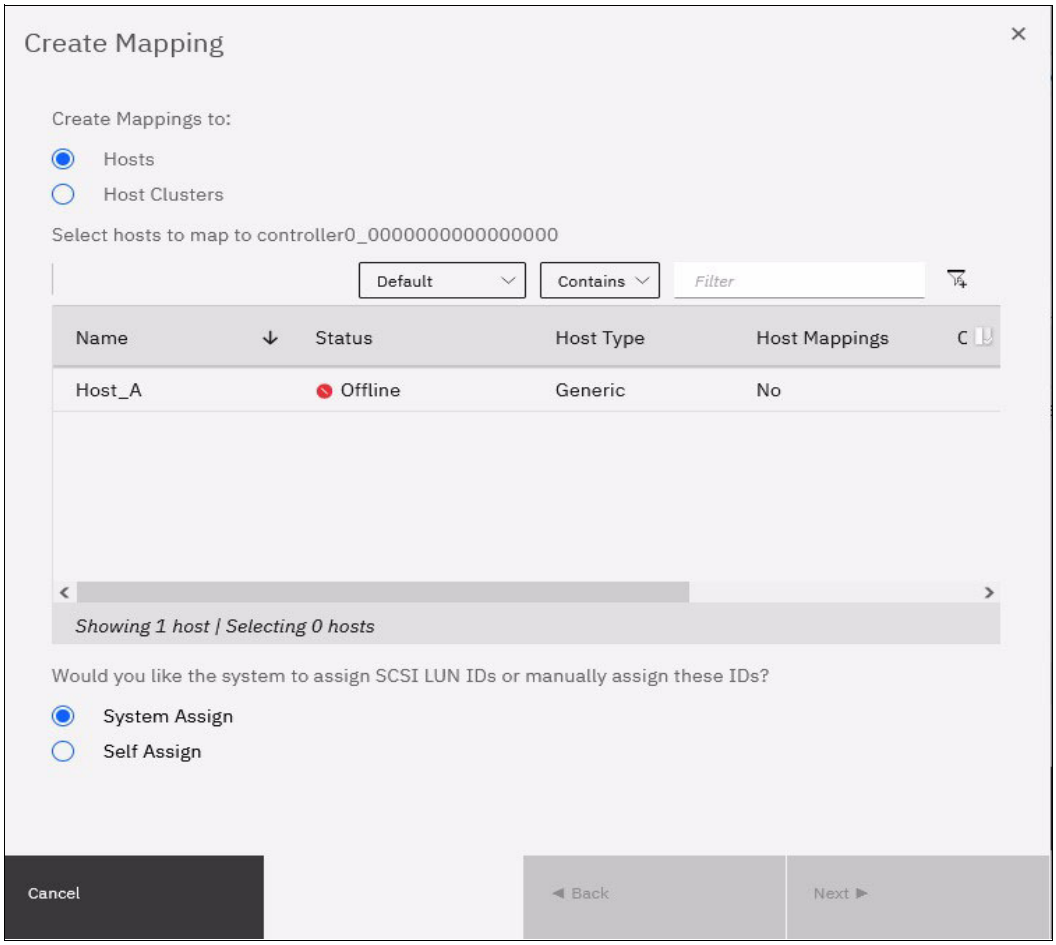


Figure 8-13 Selecting the host to which to map the new volume

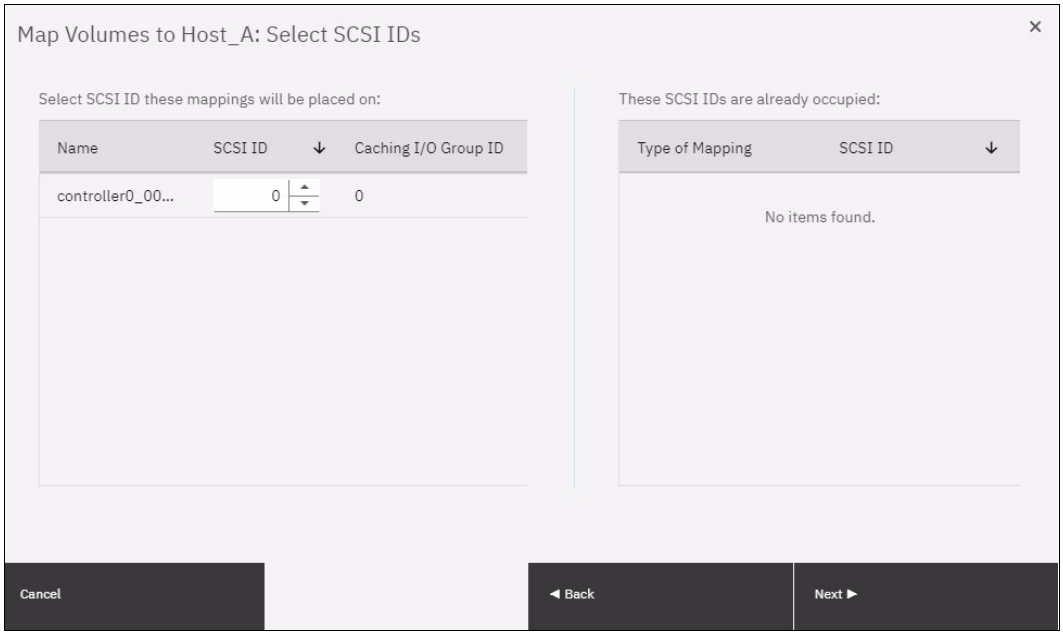


Figure 8-14 Manually assign a LUN SCSI ID to a mapped volume

When your LUN mapping is ready, click **Next**. A new dialog box opens with a summary of the new and existing mappings, as shown in Figure 8-15.

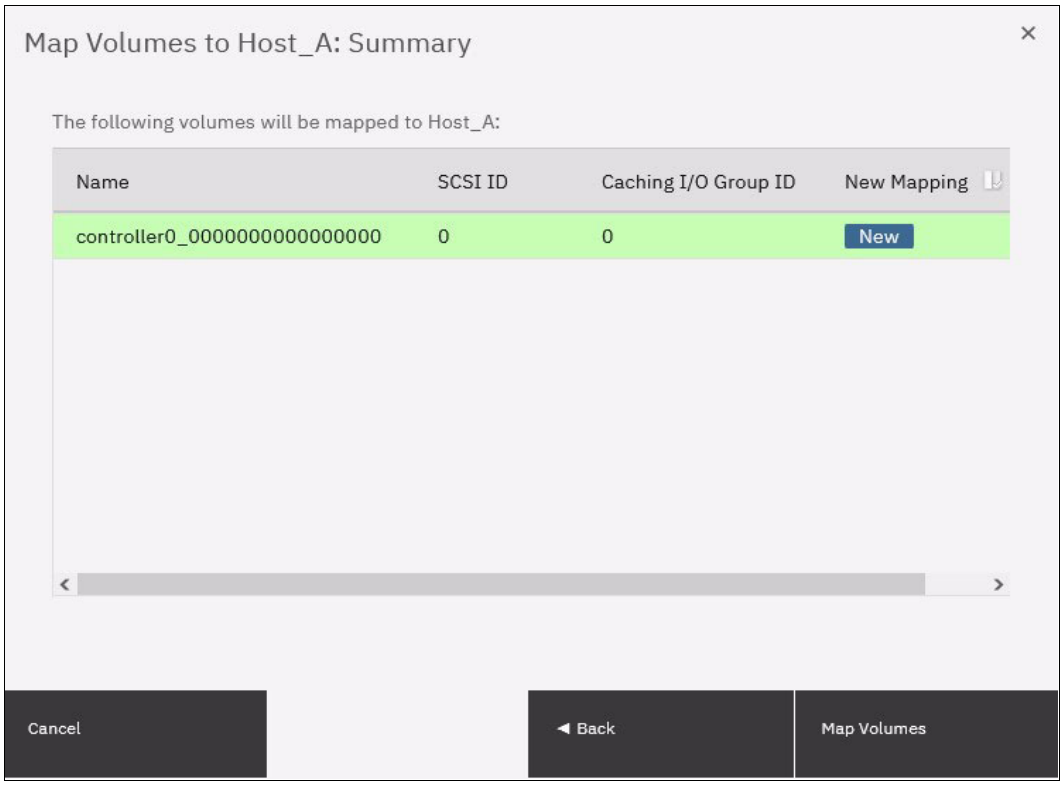


Figure 8-15 Volumes mapping summary before migration

Click **Map Volumes** and wait for the mappings to be created. Continue to map volumes to hosts until all mappings are created. Click **Next** to continue with the next migration step.

14. This step is optional.

    Select the storage pool into which you want to migrate the imported volumes. Ensure that the selected storage pool has enough space to accommodate the migrated volumes before you continue.

    You can decide not to migrate to a storage pool and to leave the imported MDisk as an image mode volume.

    However, this technique is not recommended because no volume mirroring is created. Therefore, no protection is available for the imported MDisk, and no data transfer occurs from the controller to be migrated to the system. So, although it is acceptable to delay the mirroring at some point, it should be done.

    Click **Next**, as shown in Figure 8-16 on page 653.

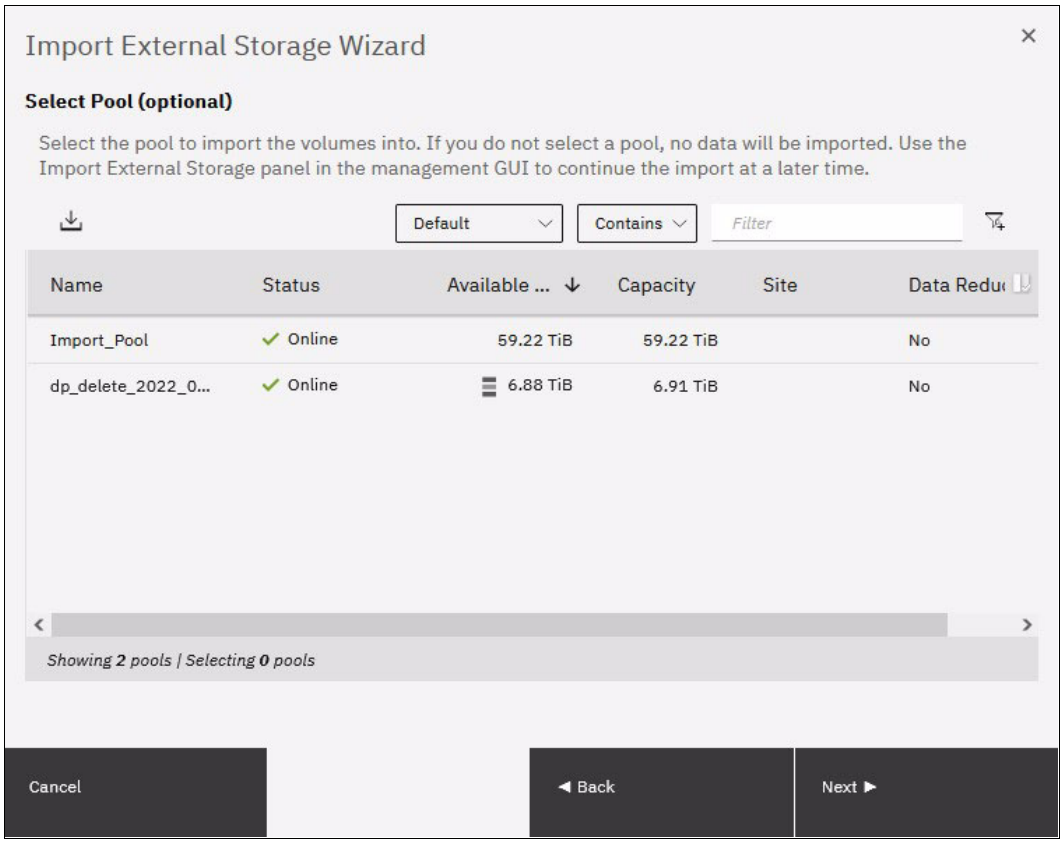


Figure 8-16 Selecting the target pool for the migration of the image mode MDisk

The migration starts. This task continues running in the background and uses the volume mirroring function to place a generic copy of the image mode volumes in the selected storage pool.

Note: With volume mirroring, the system creates two copies (`Copy0` and `Copy1`) of a volume. Typically, `Copy0` is in the migration pool, and `Copy1` is created in the target pool of the migration. When the host generates a write I/O on the volume, data is written concurrently on both copies. Read I/Os are performed on the primary copy only.

In the background, a mirror synchronization of the two copies is performed and runs until the two copies are synchronized. The speed of this background synchronization can be changed in the volume properties.

15. Click **Finish** to end the storage migration wizard, as shown in Figure 8-17.

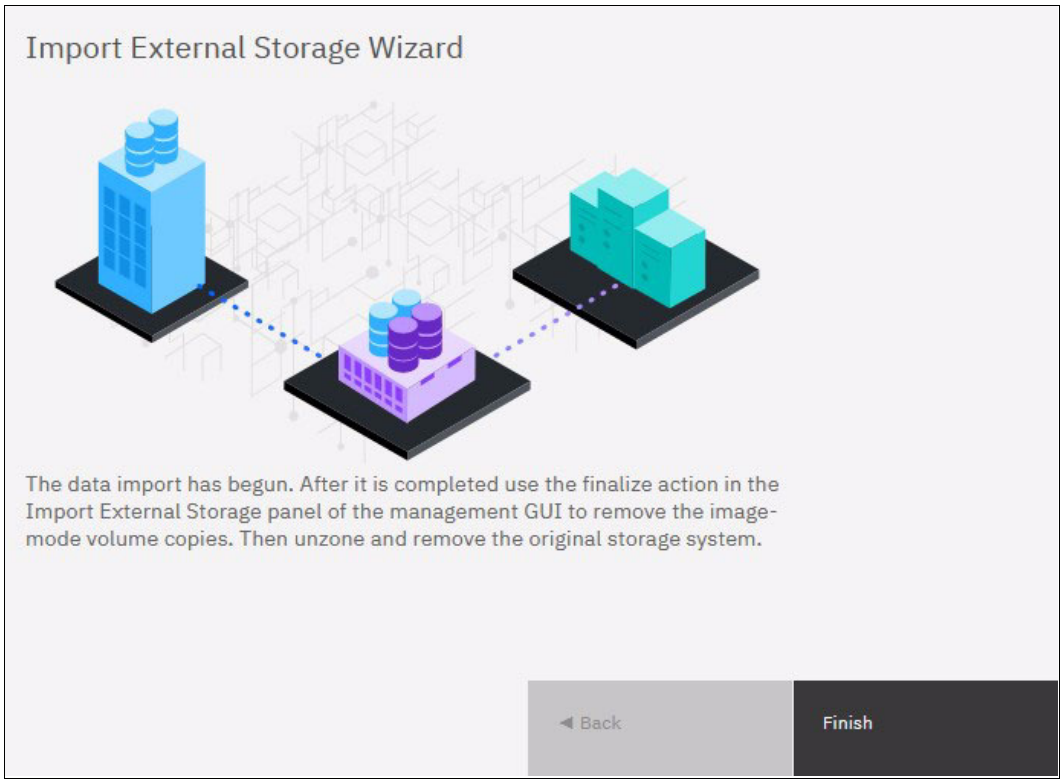


Figure 8-17 Migration is started

The end of the wizard is not the end of the migration task. You can find the progress of the migration in the Storage Migration window, as shown in Figure 8-18. The target storage pool and the progress of the volume copy synchronization is also displayed there.

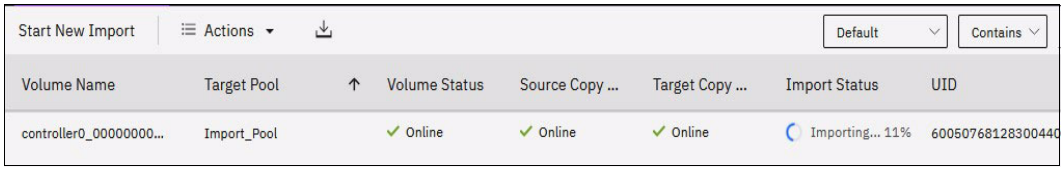


Figure 8-18 The ongoing migration is listed in the Storage Migration window

16. If you want to check the progress by using the command line interface (CLI), run the **`lsvdisksyncprogress`** command because the process is essentially a volume copy, as shown in Example 8-1.

Example 8-1 Migration progress on the command line interface

```
IBM_FlashSystem:FS5200_PFE:superuser>lsvdisksyncprogress
vdisk_id vdisk_name                    copy_id progress estimated_completion_time
2        controller0_0000000000000000 1       18       220328161517
```

17. When the migration completes, select all of the migrations that you want to finalize, right-click the selection and then click **Finalize**, as shown in Figure 8-19.

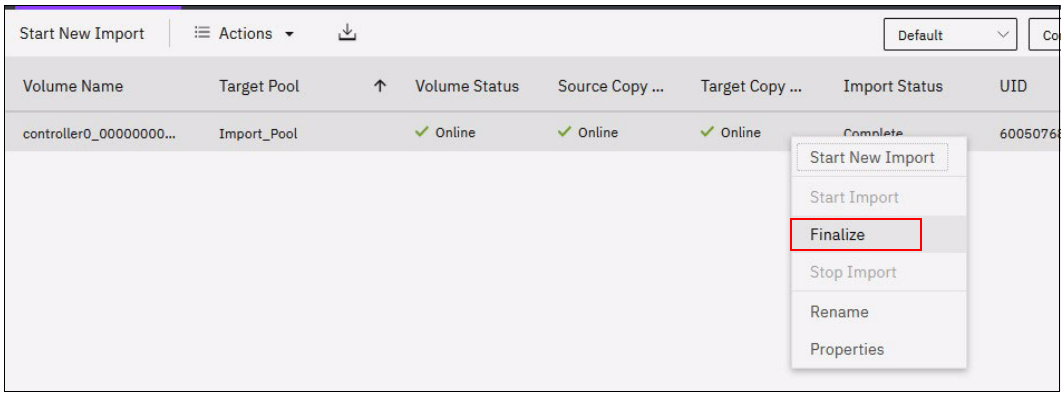


Figure 8-19 Finalizing a migration

You are asked to confirm the Finalize action because this process removes the MDisk from the Migration Pool and deletes the primary copy of the mirrored volume. The secondary copy remains in the destination pool and becomes the primary. Figure 8-20 shows the confirmation message.

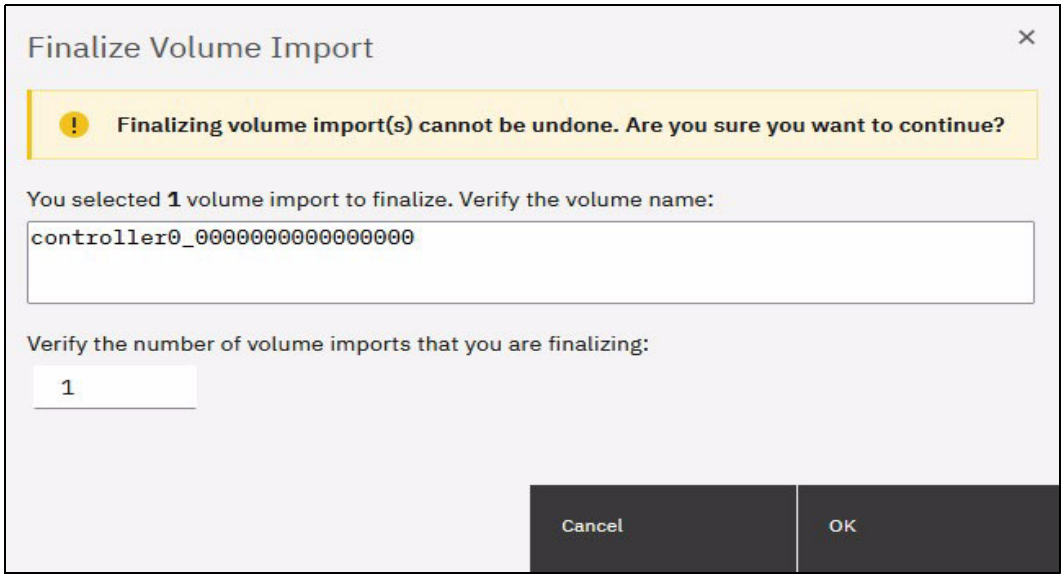


Figure 8-20 Migration finalization confirmation

18. When finalized, the image mode copies of the volumes are deleted and the associated MDisks are removed from the migration pool. The status of those MDisks returns to `unmanaged`. You can verify the status of the MDisks by selecting **Pools** → **External Storage**, as shown in Figure 8-21. In the example, `mdisk8` was migrated and finalized. It appears as `unmanaged` in the external storage window.

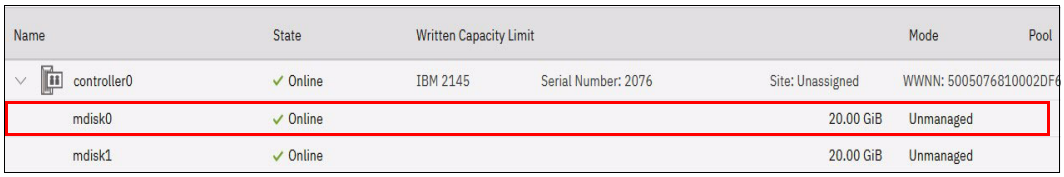


Figure 8-21 External Storage MDisks window

All the steps that are described in the Storage Migration wizard can be performed manually with the GUI and the CLI, but you should use the wizard as a guide.

8.4 Enclosure upgrade migration

The IBM FlashSystem enclosures can also be clustered like the IBM Storwize enclosures, and extra options for migrating the data are available. This action assumes that the IBM Storwize enclosure is a generation that can support the code that is required for the new hardware. For example, an IBM Storwize V7000 system must be a Gen2, Gen2+ or Gen3 to support the Version 8.3.1 code that is needed to cluster with an IBM FlashSystem 7200 or IBM FlashSystem 9200.

Note: The entry storage models 5010 and 5015 do not support clustering.

With the clustering capability, you can concurrently migrate the access to volumes from the IBM Storwize enclosure to the IBM FlashSystem enclosure and migrate the data from the IBM Storwize internal storage pool to the IBM FlashSystem internal storage pool.

The I/O group access change can be performed at any time, but ideally is done during a period of low production activity. It must also be coordinated with the operating system administrator to ensure that path discovery occurs.

Note: The NDVM process includes a limitation that prevents you from changing I/O groups if a volume is in a FlashCopy map or replication relationship. In those instances, the maps and relationships must be deleted and re-created. If an outage can be tolerated, use the `-sync` flag for relationship re-creation to avoid a resync. Otherwise, if no downtime is tolerable and a resync is acceptable, the process can be concurrent and transparent to the host.

For more information about volume mirroring, see 6.6, “Operations on volumes” on page 447. Volume mirroring can be performed by using the CLI or GUI and be moderated to lessen or eliminate the effect on performance by using the sync rate volume property.

The access change can be done before mirroring and vice versa. However, you should complete the second process without too much delay. Also, consider performing the mirroring first to minimize the added effect of accessing volumes through the IBM FlashSystem enclosure while the data is still on the IBM Storwize system, which affects performance.

8.5 Migrating data between systems nondisruptively

By using the volume mobility feature in IBM Storage Virtualize 8.4.2 and later, you can migrate volume data between systems nondisruptively.

By using non-disruptive system migration, we have the ability to move volumes from one IBM Storage Virtualize system to another without experiencing any application downtime. This feature caters to various use cases, such as load balancing across multiple systems, hardware updates, and hardware decommissioning. It also facilitates the migration of data between node-based and enclosure-based systems.

It's important to note that, unlike replication remote-copy types, non-disruptive system migration does not mandate a Remote Mirroring license before configuring a remote-copy relationship for migration purposes. This streamlines the migration process and provides a greater convenience in managing the storage environment.

A subtle difference exists between the nondisruptive volume move between I/O groups. In this case, we can migrate volumes between systems that do not need to be clustered (IBM SAN Volume Controller into an enclosure-based system, such as IBM FlashSystem or vice-versa if required).

Similarly, if clusters are reaching maximum limits from a cluster perspective, such as maximum VDisks, this method moves volumes nondisruptively while maintaining I/O to the host application through the entire duration.

The use of the remote copy (Metro Mirror) functions and enhancements to the Asymmetric Logical Unit Access (ALUNA) path state of features of SCSI made this new migration method available. This nondisruptive volume mobility feature is an alternative, but not a replacement for the image mode and clustering-based methods.

The following specific restrictions exist by using this method:

- No 3-site support

  Migration relationship cannot be converted into a 3-site relationship.
- Not a DR or HA solution:
  - Migration relationship cannot be converted into another type of remote-copy relationship.
  - Stop-with-access is prohibited. The volumes cannot act independently if they use the same Universally UID (UUID).
- No support for consistency groups, change volumes or expanding volumes:
  - Migration relationships cannot be added to a consistency group.
  - Associating change volumes to a migration relationship is prohibited.
  - Volumes cannot be resized until the migration is completed or canceled.
- Partnership requirements are equivalent to Metro Mirror.
- Performance considerations apply as with Metro Mirror.
- Both systems must be running 8.4.2 release or later with FC or IP partnerships between them. For more information about validating whether the systems are suitable to be upgraded to 8.4.2, see Concurrent Compatibility and Code Cross Reference for IBM Storge Virtualize.
- Current host support and restrictions:
  - IBM AIX, RHEL, SLES, VMware ESXi, Solaris, HP-UX:

- IBM AIX 7.3 TL1 or later: iSCSI and FC connectivity
- Solaris 10, 11 and HP-UX 11iV3: FC connectivity
- RHEL 7.x, 8.x and SLES 12.x, 15.x: iSCSI and FC connectivity
- VMware ESXi 7.x: iSCSI, iSER, and FC connectivity
- VMware ESXi 6.x: iSCSI and FC connectivity

– Not supported:

- IBM i, AIX 7.1, AIX 7.2, Hyper-V, and Windows
- SAN-boot volumes
- Volumes that are mapped to NVMe-attached hosts

– No SCSI persistent reservations or Offload Data Transfer (ODX):

- Migration relationships cannot be created while ODX is enabled.
- ODX must be disabled on both systems while migrating volumes.

Note: If ODX is disabled, it cannot be reenabled on the 8.4.2 software. As of this writing, ODX support is *not* available on v8.4.2.

For more information, see Migrating data between systems non-disruptively.

8.5.1 Nondisruptive volume migration procedure

In this section, we present an example of a host with two volumes that are migrated nondisruptively from a source IBM FlashSystem 5200 to the target IBM FlashSystem 9200 by using the graphical user interface (GUI).

For more information about the GUI and CLI instructions, see Data migration.

Figure 8-22 shows two host-attached volumes from the source FlashSystem 5200 with the following specific UUIDs:

- 600507680B8107FF10000000000000E2
- 600507680B8107FF10000000000000E3

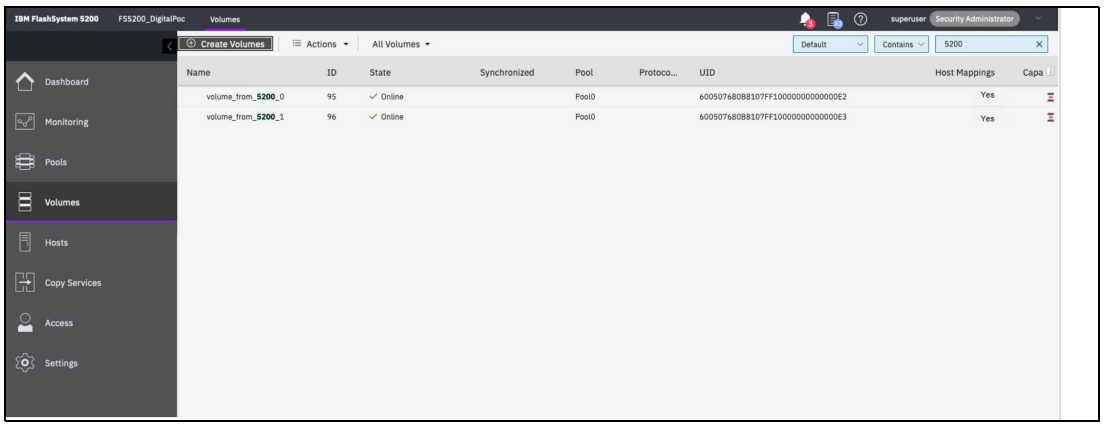


Figure 8-22 Existing volume on source system

Note: With the current restrictions, consistency groups are not supported. Therefore, each volume is migrated individually.

A partnership must exist or be created (as shown in Figure 8-23) between the source and target system in scope. This partnership is required so that the volume mobility feature can use the remote copy replication. It supports IP or FC partnership, which can be created by using the usual CLI and GUI methods.

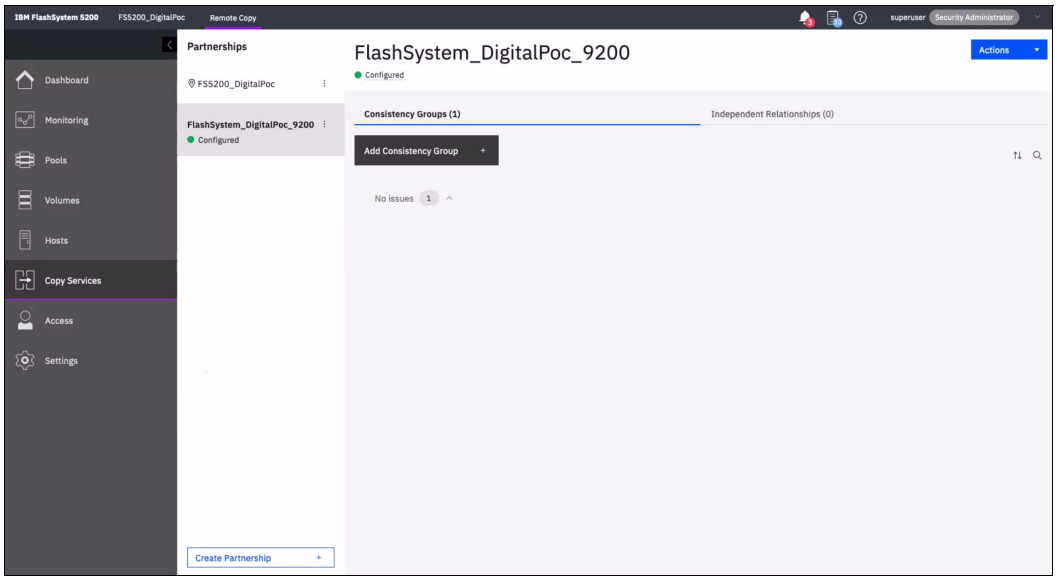


Figure 8-23 Establish partnership between source and target systems

Complete the following steps:

1. Ensure to create volumes of matching sizes on the target system, but with a different UUIDs 60050768108482D750000000000003C4 and 60050768108482D750000000000003C3 than the source system, as shown in Figure 8-24.

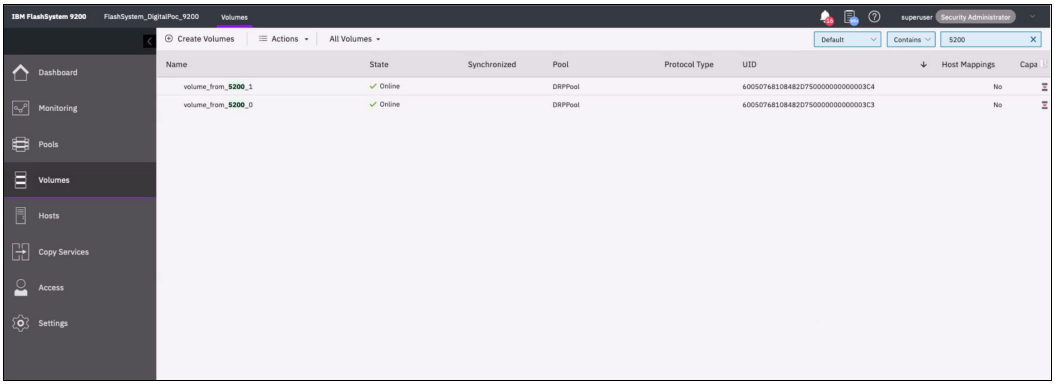


Figure 8-24 Creating two volumes on the target system

2. The Metro Mirror relationship is created manually by using a dedicated **`migration`** switch parameter through the CLI. This step can also be done through the GUI, as shown in Figure 8-25. Select **Migration** → **Non-disruptive System Migration** while creating the relationship.

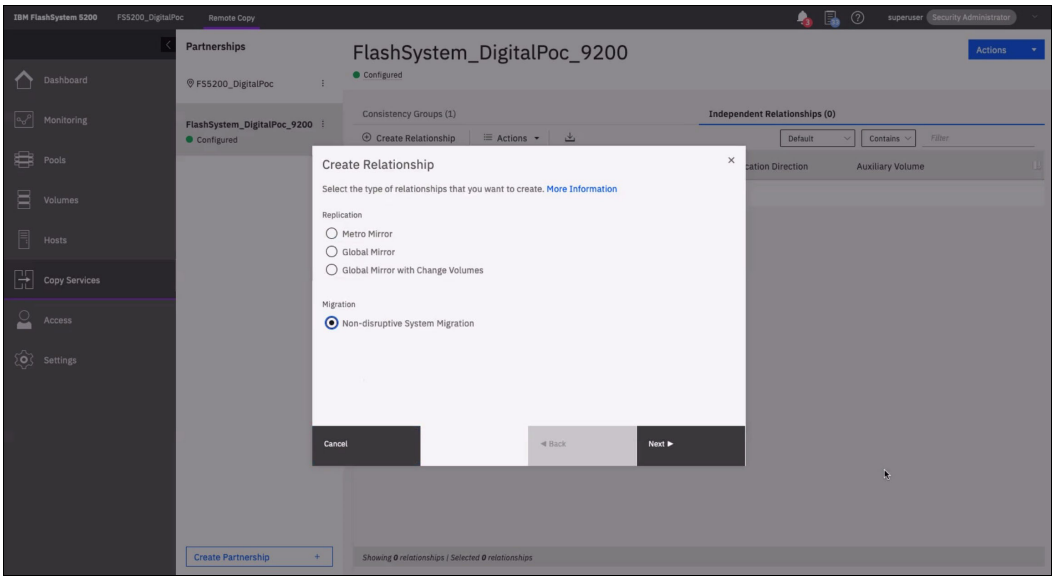


Figure 8-25 Creating a nondisruptive system migration relationship

3. Select the target system on which the auxiliary volumes are stored, as shown in Figure 8-26.

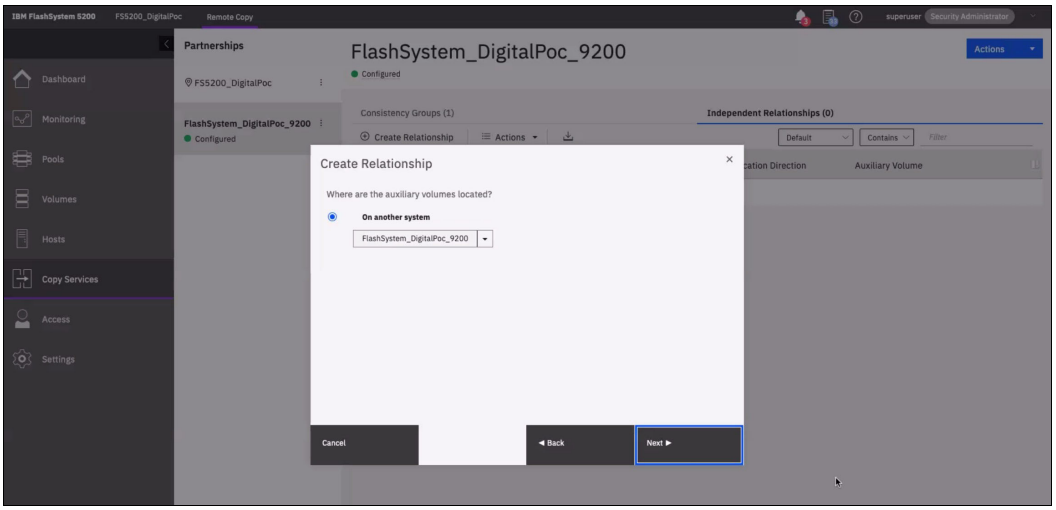


Figure 8-26 Choosing a target system

4. Select the master and auxiliary volumes in scope for the migration, as shown in Figure 8-27.

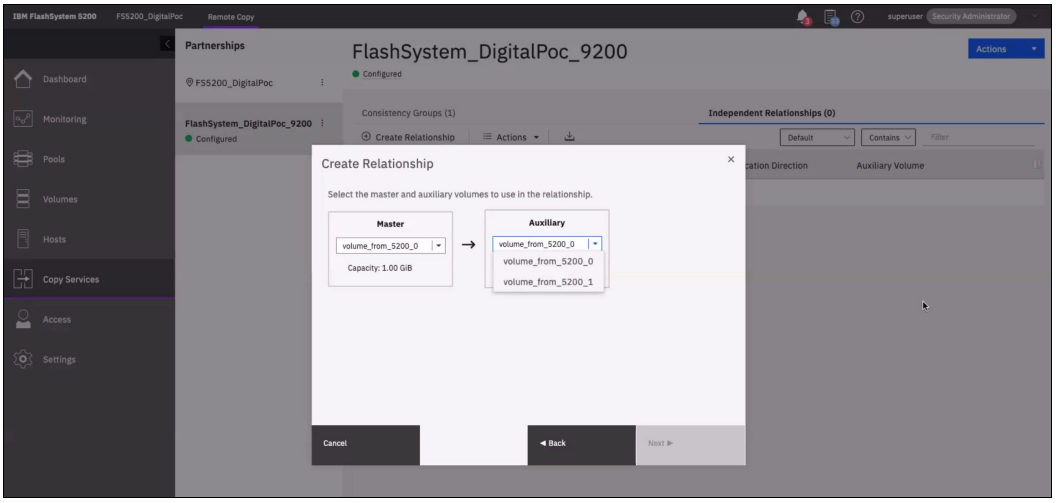


Figure 8-27 Choosing master and auxiliary volumes

5. Follow the same process to add the remaining volumes by clicking **Add**, as shown in Figure 8-28 and Figure 8-29 on page 662.

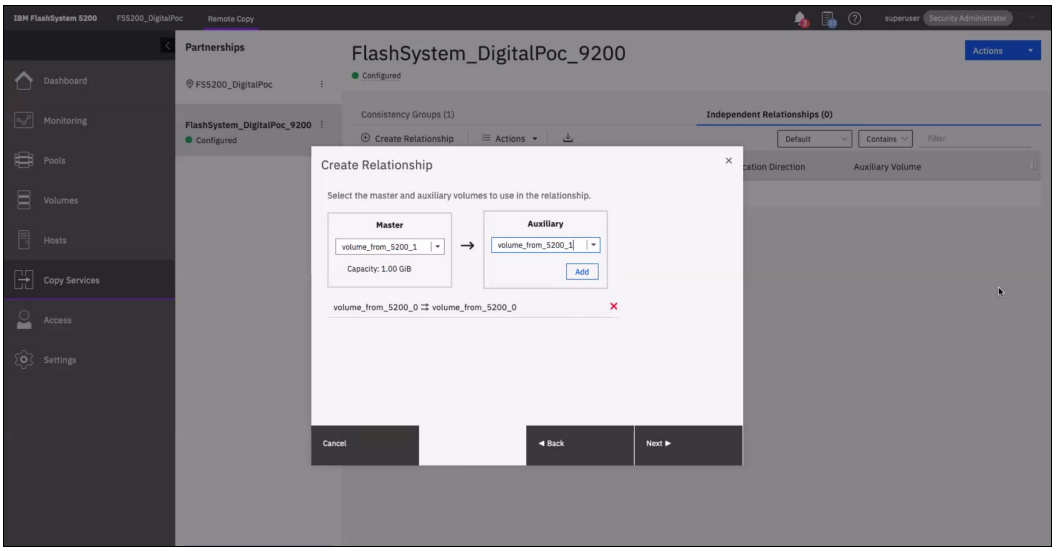


Figure 8-28 Adding master and auxiliary volumes

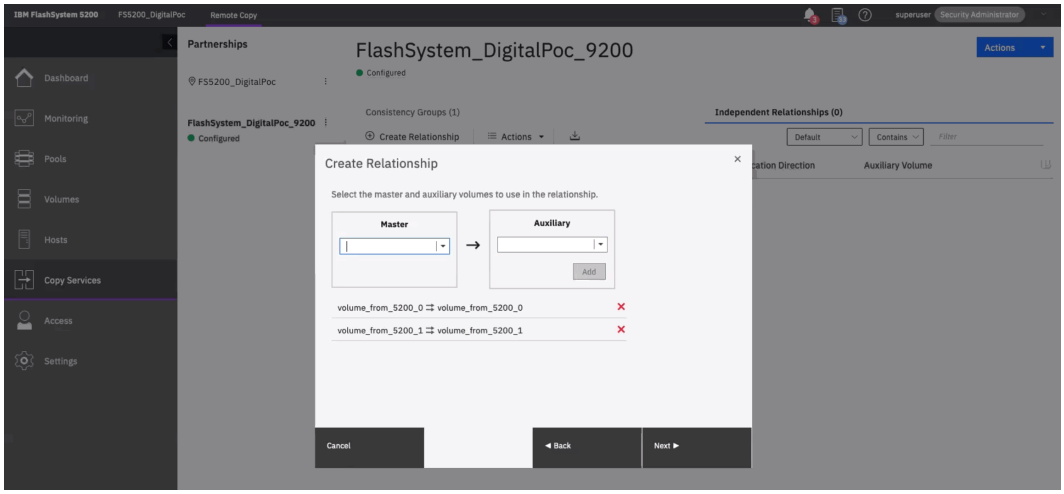


Figure 8-29 Final list of volumes to migrate

The remote copy relationship is created, as shown in Figure 8-30 and Figure 8-31 on page 663.

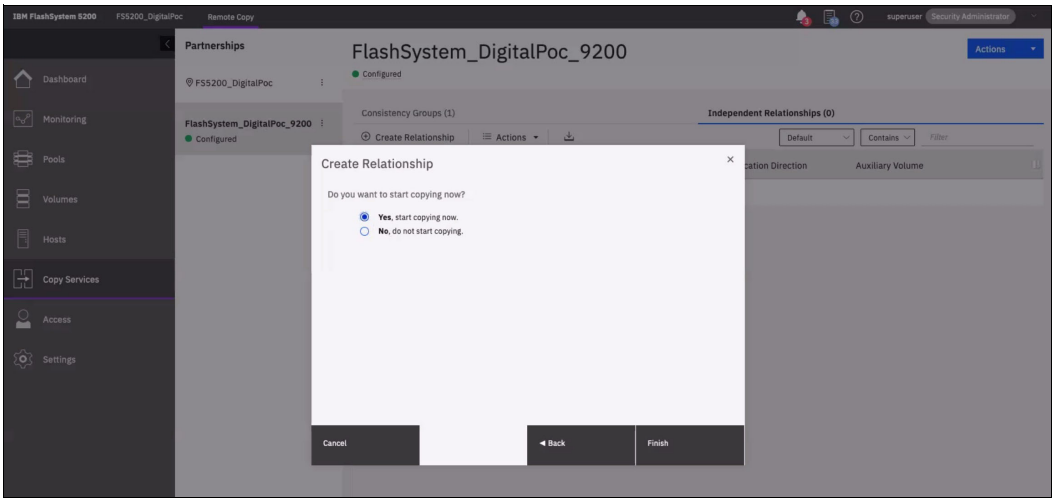


Figure 8-30 Creating a relationship

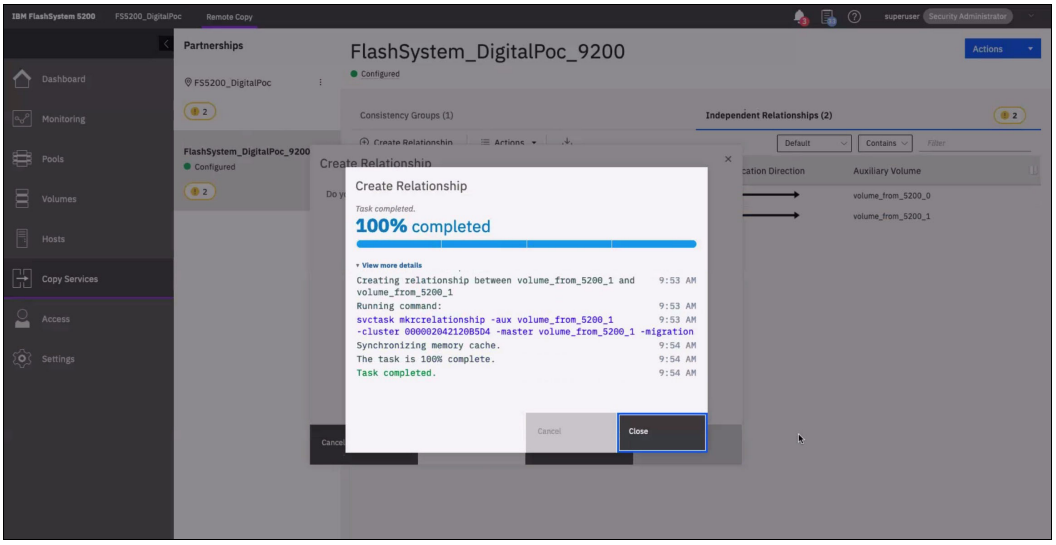


Figure 8-31 Creating a relationship successful completion

Note: You can select the **No, do not start copying** option (see Figure 8-30 on page 662) to create the relationship, which results in volume relationships in an `Inconsistent Stopped` state.

With the current restrictions, consistency groups are not supported. Therefore, each volume must be individually selected and manually started.

To enable the copying process, select one volume at a time, right-click and then select **Start.** If multiple volumes are selected, then the **Start** is unavailable.

Follow the same process for all volumes until they are all enabled to start the copy. The state changes from `Inconsistent Copying` to `Consistent Synchronized` after the copy is complete.

Do not proceed until all the volumes are `Consistent Synchronized`.

The UUID of the source volume is migrated over to the target from the source system. After the remote copy relationship completes the synchronization, the migration is in a consistent state, as shown in Figure 8-32. Therefore, it is fully replicated and is now similar to a traditional Metro Mirror relationship in which I/O from the source is replicated to the target before notification of completion to the host.

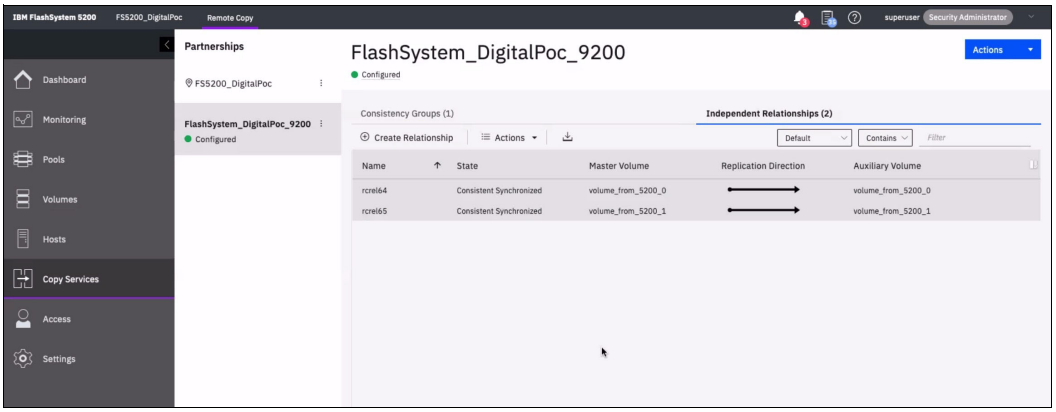


Figure 8-32 Consistent synchronized state

6. Create VDisk host maps or volume host cluster maps on the target system. Ensure that the host that is connected to the source system can also access the new imported UUID volumes on the target system.

   By using this dedicated remote copy partnership and remote copy migration relationship, the paths that are being used and presented are in standby state. Therefore, this new SCSI feature is available in IBM Storage Virtualize 8.4.2 and ongoing versions.

   At this point, the host is still performing active I/O through the primary system and no I/O is sent through the secondary. The active I/O paths are still being used as the preferred node and the partner node at the source system.

   Note: Rescan the HBAs on the host to ensure that all the paths are detected to the volumes on the target system. Record the current path states on the host and identify the WWPNs used for the active and standby (ghost) paths.

   Do not proceed if the extra standby paths are not visible on the host.

   Standby paths might be listed under a different name on the host, such as `ghost` paths. Data access can be affected if all standby paths are not visible to the host when the direction on the relationship is switched over.

7. Issue the **`switchrelationship`** command through the CLI, which changes the direction of the relationship so that the target system becomes the master and the source system becomes the auxiliary in this relationship. This step can also be done by using the GUI, as shown in Figure 8-33. Select the individual volume and right-click to choose the **`Switch`** option.

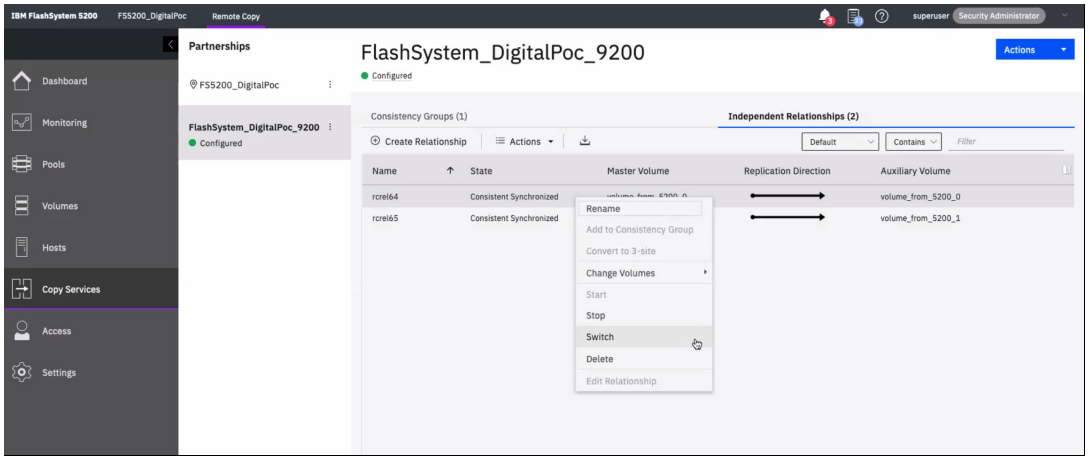


Figure 8-33 Approval state to switch relationship

The next window shows a warning message (see Figure 8-34).

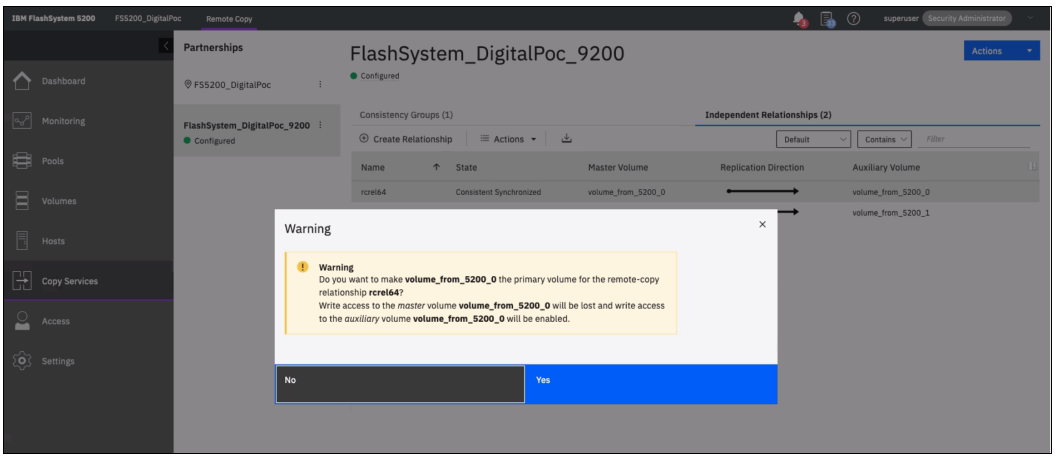


Figure 8-34 Approval state to switch relationship

8. Click **Yes** to proceed to the successful completion of the role reversal, as shown in Figure 8-35.

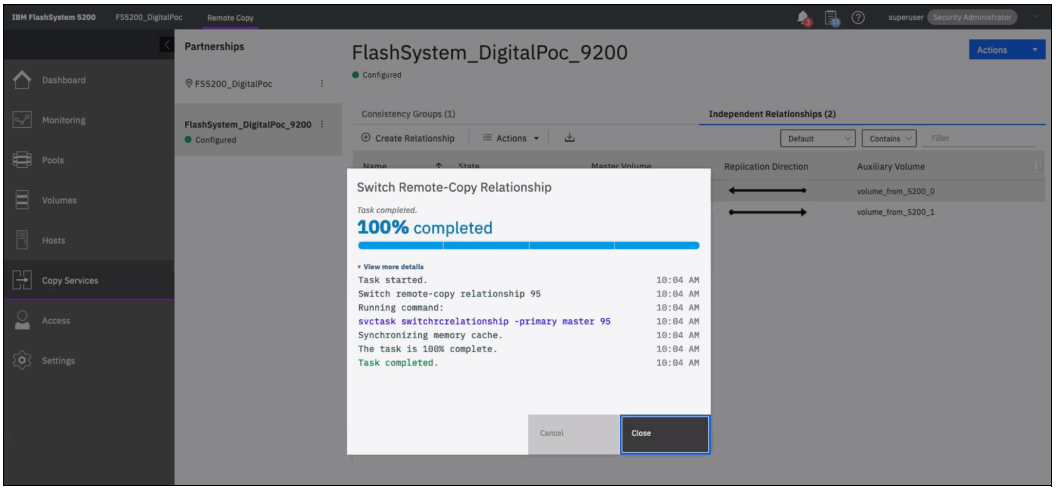


Figure 8-35 Successful completion of switch relationship for one volume

9. After one volume is completed (see Figure 8-36), follow the same steps as described in Step 7 on page 665 for the next volume.

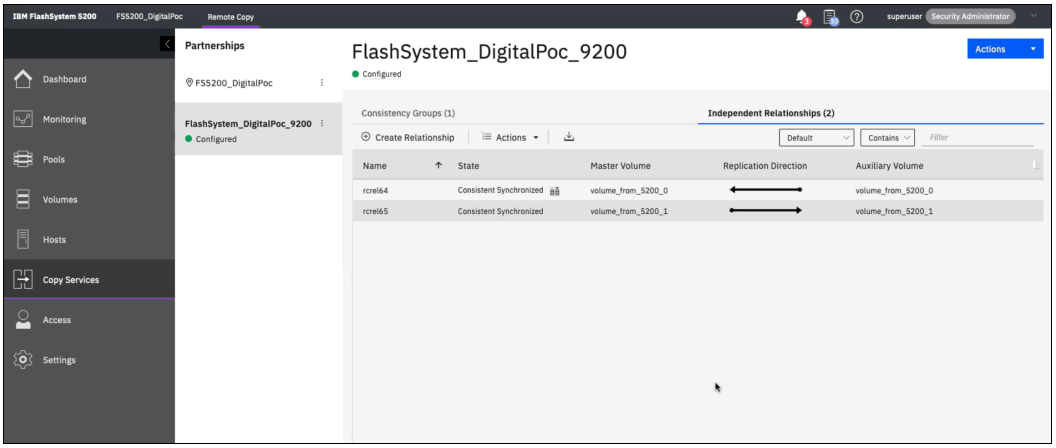


Figure 8-36 Completion of one volume

10. The standby (or ghost) paths that were presented from the target system are now converted to active paths. Also, the previously active paths that are presented from the original source system are now converted to standby (or ghost) paths. The GUI shows the direction of replication, as shown in Figure 8-37.

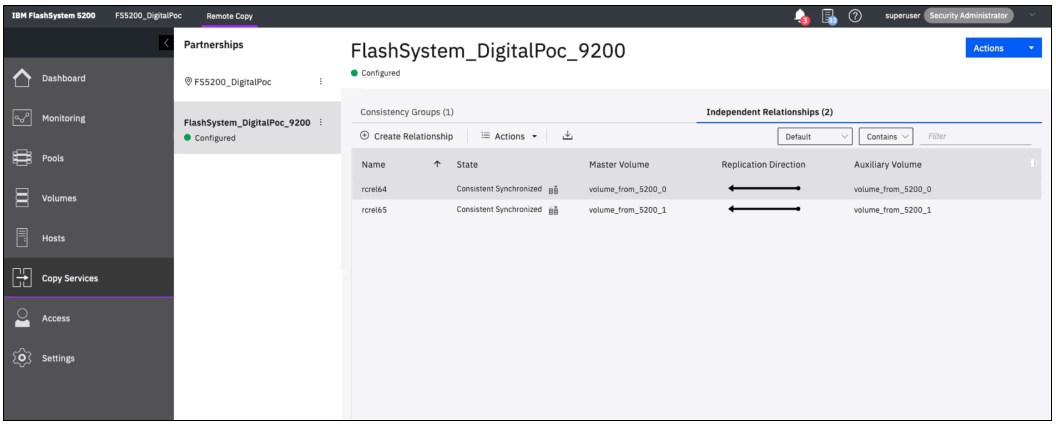


Figure 8-37 Switch relationship completion for all volumes

Note: Validate that the host uses the volumes from the target system for I/O. It is recommended to restart the host to ensure that it comes up clean and no issues exist before the volumes from the source system are unmapped and deleted.

11. The host is now performing I/O to the target system, which is the new system and the standby paths are to the originating source system. Unmap the hosts from the volumes on the source system.
12. Delete the volumes that were migrated from the source system.

Note: A warning message appears that indicates that the volume is in a remote-copy relationship. This warning can be ignored if the previous steps are completed correctly.

At the end of this stage, we now have volumes that were migrated to the target system and retain the same UUID `60050768 0B8107FF10000000000000E2` and `60050768 0B8107FF10000000000000E3` from the source system, as shown in Figure 8-38 on page 668. This entire process is not apparent to the host that is accessing the volumes.

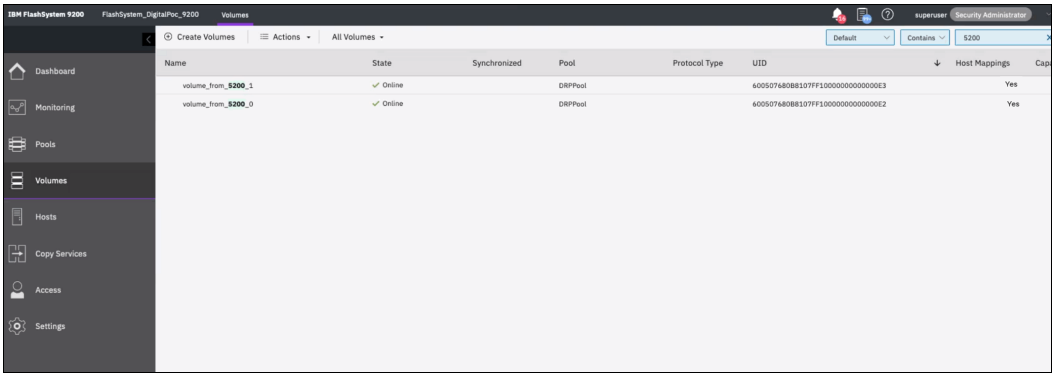


Figure 8-38 Volume migration completed with source UUID on target system

13. Remove the partnership between the two systems, as shown in Figure 8-23 on page 659. This process can be done after all the volumes are migrated.

9

Advanced features for storage efficiency

IBM Storage Virtualize running inside a storage system offers several functions for storage optimization and efficiency. This chapter introduces the basic concepts of those functions, and also provides a short technical overview and implementation recommendations.

For more information about the planning and configuration of storage efficiency features, see the following publications:

- *Performance and Best Practices Guide for IBM Storage FlashSystem and IBM SAN Volume Controller: Updated for IBM Storage Virtualize Version 8.6*, SG24-8543
- *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430

This chapter includes the following topics:

- 9.1, “IBM Easy Tier” on page 670
- 9.2, “Thin-provisioned volumes” on page 690
- 9.3, “UNMAP” on page 692
- 9.4, “Data reduction pools” on page 695
- 9.5, “Saving estimations for compression and deduplication” on page 707
- 9.6, “Overprovisioning and data reduction on external storage” on page 709
- 9.7, “Safeguarded child pool capability: Protection from logical data corruption” on page 714

© Copyright IBM Corp. 2024. All rights reserved.

9.1 IBM Easy Tier

IBM Storage Virtualize includes the IBM System Storage Easy Tier function, which enables automated subvolume data placement throughout different storage tiers. It also automatically moves extents within the same storage tier to intelligently align the system with workload requirements. Easy Tier works with all available storage tiers and drive modules:

- Storage-class memory (SCM)
- Flash drives
- Hard disk drives (HDDs)

Many applications manifest a significant skew in the distribution of I/O workload. That is, a small fraction of the storage is responsible for a disproportionately large fraction of the total I/O workload of an environment.

Easy Tier acts to identify this skew and automatically place data to take advantage of it. By moving the “hottest” data onto the fastest tier of storage, the workload on the remainder of the storage is reduced. By servicing most of the application workload from the fastest storage, Easy Tier accelerates application performance and increases overall server usage, which can reduce costs regarding servers and application licenses.

Easy Tier also reduces storage cost because the system always places the data with the highest I/O workload on the fastest tier of storage. Depending on the workload pattern, a large portion of the capacity can be provided by a lower and less expensive tier without impacting application performance.

Note: Easy Tier is a licensed function. On IBM FlashSystem 9200, IBM FlashSystem 7200, and IBM FlashSystem 5200, it is included in the base code. No actions are required to activate the Easy Tier license on these systems. No actions are required to activate the Easy Tier license on the IBM SAN Volume Controller.

On IBM FlashSystem 5100, you must have the suitable number of licenses to run Easy Tier.

The IBM FlashSystem 5000 entry systems also require a license for Easy Tier, which is a one-time charge per system.

Without a license, Easy Tier balances I/O workload only between managed disks (MDisks) in the same tier.

In HyperSwap environments, all member controllers must be licensed with Easy Tier to enable this function. For example, you need two licenses when two IBM FlashSystem 5035 systems are clustered.

9.1.1 Easy Tier concepts

Easy Tier is a performance optimization function that automatically migrates extents that belong to a volume between different storage tiers based on their I/O load. Movement of the extents is concurrent with I/O and transparent from the host point of view.

As a result of extent movement, the volume no longer has all its data in one tier, but rather in two or more tiers. Each tier provides optimal performance for the extent, as shown in Figure 9-1 on page 671.

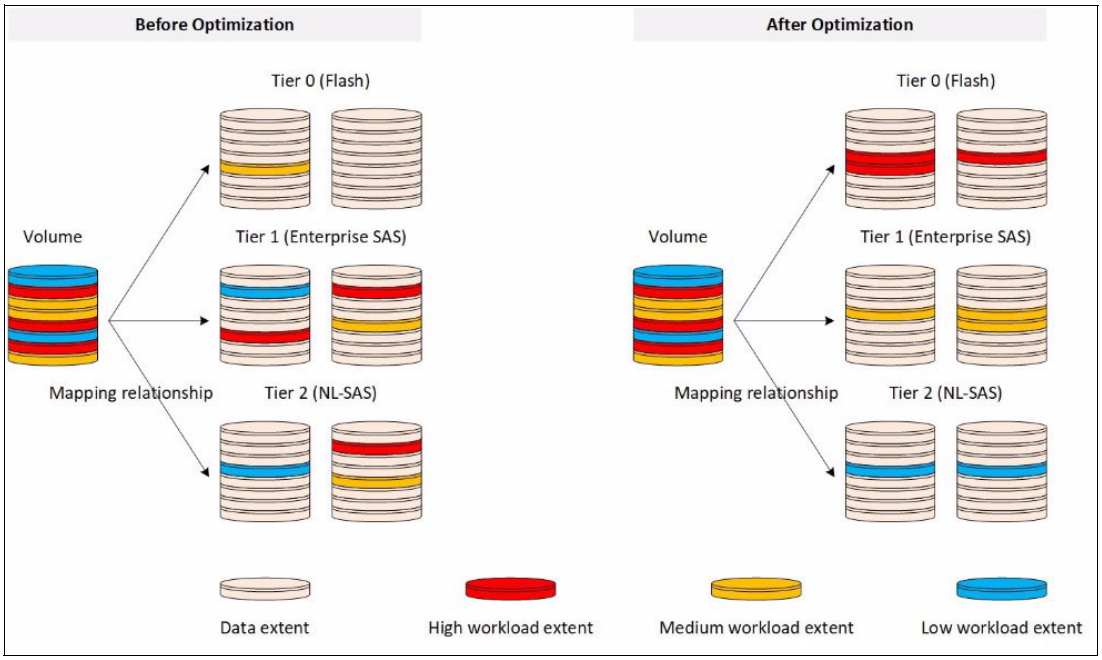


Figure 9-1 Easy Tier

Easy Tier monitors the I/O activity and latency of the extents on all Easy Tier enabled storage pools. Based on the performance log, it creates an extent migration plan and *promotes* (moves) high activity or hot extents to a higher disk tier within the same storage pool. It also *demotes* extents whose activity dropped off, or cooled, by moving them from a higher disk tier MDisk back to a lower tier MDisk.

If a pool contains only MDisks of a single tier, Easy Tier operates only in balancing mode. Extents are moved between MDisks in the same tier to balance the I/O workload within that tier.

Tiers of storage

The MDisks (external logical units [LUs] or redundant array of independent disks [RAID] arrays) that are presented to the system might have different performance attributes because of their technology type, such as flash drives, SCM, HDDs, and other characteristics.

The system divides available storage into the following tiers:

- SCM

  The SCM tier is used when the pool contains drives that use persistent memory technologies that improve the endurance and speed of current flash storage device technologies. SCM drives are available only in Non-Volatile Memory Express (NVMe)-based controller systems.

- Tier 0 flash

  Tier 0 flash drives are high-performance flash drives that use enterprise flash technology.

- Tier 1 flash

  Tier 1 flash drives represent the read-intensive flash drive technology. Tier 1 flash drives are lower-cost flash drives that typically offer capacities larger than enterprise-class flash, but have lower performance and write endurance characteristics.

- Enterprise tier

  The enterprise tier is used when the pool contains MDisks on enterprise-class HDDs, which are disk drives that are optimized for performance.

- Nearline (NL) tier

  The NL tier is used when the pool has MDisks on NL-class disks drives that are optimized for capacity.

The system automatically sets the tier for internal array mode MDisks because it knows the capabilities of array members, physical drives, and modules. External MDisks need manual tier assignment when they are added to a storage pool.

Note: The tier of MDisks that is mapped from certain types of IBM System Storage Enterprise Flash is fixed to tier0_flash, and cannot be changed.

Although the system can distinguish between five tiers, Easy Tier manages only a three-tier storage architecture within each storage pool. MDisk tiers are mapped to Easy Tier tiers depending on the pool configuration, as listed in Table 9-1.

Note: The table represents all the possible pool configurations. Some entries in the table contain *optional* tiers (shown in *italic* font), but the configurations without the optional tiers are also valid.

Table 9-1 Storage tier to Easy Tier mapping

Configuration	Easy Tier top tier	Easy Tier middle tier	Easy Tier bottom tier
SCM (*+ Tier0_Flash*)	SCM	(*Tier0_Flash*)	
SCM + Tier0_Flash (*+ Tier1_Flash*)	SCM	Tier0_Flash	(*Tier1_Flash*)
SCM + Tier0_Flash (*+ Tier1_Flash*) + Enterprise + NL (unsupported)	SCM	Tier0_Flash (*+ Tier1_Flash*)	Enterprise + NL
SCM + Tier0_Flash + Enterprise/NL	SCM	Tier0_Flash	Enterprise/NL
SCM + Tier0_Flash + Tier1_Flash + Enterprise/NL (unsupported)	SCM	Tier0_Flash + Tier1_Flash	Enterprise/NL
SCM + Tier1_Flash (*+ Enterprise/NL*)	SCM	Tier1_Flash	(*Enterprise/NL*)
SCM + Tier1_Flash + Enterprise + NL	SCM	Tier1_Flash + Enterprise	NL
SCM + Enterprise/NL	SCM	Enterprise/NL	
SCM + Enterprise + NL	SCM	Enterprise	NL
Tier0_Flash (*+ Tier1_Flash*)	Tier0_Flash	(*Tier1_Flash*)	

Configuration	Easy Tier top tier	Easy Tier middle tier	Easy Tier bottom tier
Tier0_Flash + Tier1_Flash + Enterprise/NL	Tier0_Flash	Tier1_Flash	Enterprise/NL
Tier0_Flash + Tier1_Flash + Enterprise + NL	Tier0_Flash	Tier1_Flash + Enterprise	NL
Tier0_Flash + Enterprise (*+ NL*)	Tier0_Flash	Enterprise	(*NL*)
Tier0_Flash + NL	Tier0_Flash	NL	
Tier1_Flash (+ *Enterprise/NL*)		Tier1_Flash	(*Enterprise/NL*)
Tier1_Flash + Enterprise + NL	Tier1_Flash	Enterprise	NL
Enterprise (*+ NL*)		Enterprise	(*NL*)
NL			NL

Sometimes, a single Easy Tier tier contains MDisks from more than one storage tier. For example, consider a pool with SCM, Tier1_Flash, Enterprise, and NL. SCM is the top tier, and Tier1_Flash and Enterprise share the middle tier. NL is represented by the bottom tier.

Note: Some storage pool configurations with four or more different tiers are not supported. If such a configuration is detected, an error is logged and Easy Tier enters measure mode, which means no extent migrations are performed.

For more information about planning and configuration considerations or best practices, see *Performance and Best Practices Guide for IBM Storage FlashSystem and IBM SAN Volume Controller: Updated for IBM Storage Virtualize Version 8.6*, SG24-8543.

Easy Tier automatic data placement

Easy Tier continuously monitors volumes for host I/O activity. It collects performance statistics for each extent, and derives exponential moving averages for a rolling 24-hour period of I/O activity. Random and sequential I/O rate, I/O block size and bandwidth for reads and writes, and I/O response time are collected.

A set of algorithms is used to decide where the extents should be and whether extent relocation is required:

- Once per day, Easy Tier analyzes the statistics to determine which data should be sent to a higher performing tier or a lower tier.
- Four times per day, it analyzes the statistics to identify whether any data must be rebalanced between MDisks in the same tier.
- Once every 5 minutes, Easy Tier checks the statistics to identify whether any of the MDisks are overloaded.

Based on this information, Easy Tier generates a migration plan that must be run for optimal data placement. The system spends the necessary time running the migration plan. The migration rate is limited to make sure host I/O performance is not affected while data is relocated.

The migration plan can consist of the data movement actions on volume extents, as shown in Figure 9-2. Although each action is shown once, all movement actions can be performed between any pair of adjacent tiers.

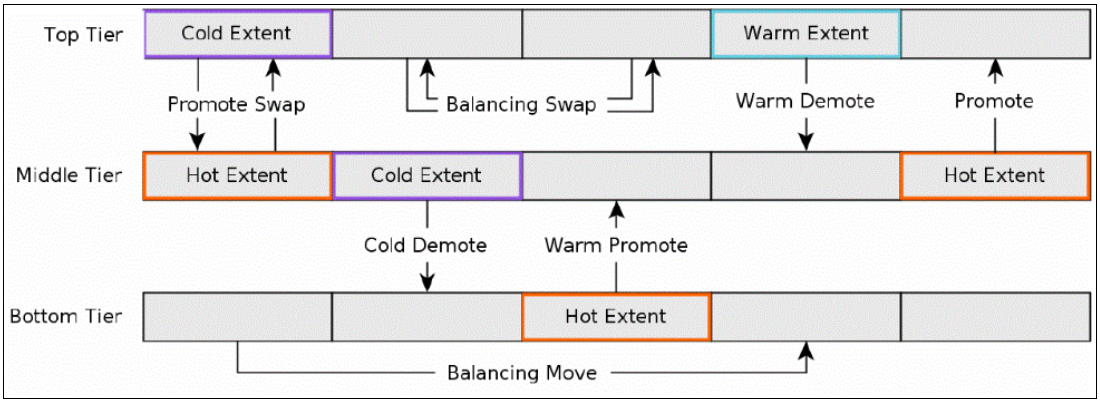


Figure 9-2 Actions on extents

The following possible actions are available:

- Promote

  Active data is moved from a lower tier of storage to a higher tier to improve the overall system performance.

- Promote Swap

  Active data is moved from a lower tier of storage to a higher tier to improve overall system performance. Less active data is moved first from the higher tier to the lower tier to make space.

- Warm Promote

  When an MDisk becomes overloaded, active data is moved from a lower tier to a higher tier to reduce the workload on the MDisk, which addresses the situation where a lower tier suddenly becomes active. Instead of waiting for the next migration plan, Easy Tier can react immediately. Warm promote acts in a similar way to warm demote. If the 5-minute average performance shows that a layer is overloaded, Easy Tier immediately starts to promote extents until the condition is relieved.

- Cold Demote

  Inactive or less active data is moved from a higher tier of storage to a lower tier to free space on the higher tier. Easy Tier automatically frees extents on the higher storage tier before the extents on the lower tier become hot, which helps the system to be more responsive to new hot data.

- Warm Demote

  When an MDisk becomes overloaded, active data is moved from a higher tier to a lower tier to reduce the workload on the MDisk. Easy Tier continuously ensures that the higher performance tier does not suffer from saturation or overload conditions that might affect the overall performance in the pool. This action is triggered when bandwidth or input/output operations per second (IOPS) exceeds a predefined threshold of an MDisk and causes the movement of selected extents from the higher-performance tier to the lower-performance tier to prevent MDisk overload.

- Balancing Move

  Data is moved within the same tier from an MDisk with a higher workload to one with a lower workload to balance the workload within the tier, which automatically populates new MDisks that were added to the pool.

- Balancing Swap

  Data is moved within the same tier from an MDisk with higher workload to one with a lower workload to balance the workload within the tier. Other less active data is moved first to make space.

Extent migration occurs at a maximum rate of 12 GB every 5 minutes for the entire system. It prioritizes the following actions:

- Promote and rebalance get equal priority.
- Demote is 1 GB every 5 minutes, and then gets whatever is left.

Note: Extent promotion or demotion occurs only between adjacent tiers. In a three-tier storage pool, Easy Tier does not move extents from the top directly to the bottom tier or vice versa without moving to the middle tier first.

The Easy Tier overload protection is designed to avoid overloading any type of MDisk with too much work. To achieve this task, Easy Tier must have an indication of the maximum capability of a MDisk.

For an array made of locally attached drives, the system can calculate the performance of the MDisk because it is pre-programmed with performance characteristics for different drives and array configurations. For a storage area network (SAN)-attached MDisk, the system *cannot* calculate the performance capabilities. Therefore, follow the best practice guidelines when configuring external storage, particularly the ratio between physical disks and MDisks that is presented to the system.

Each MDisk has an Easy Tier load parameter (low, medium, high, or very_high) that can be fine-tuned manually. If you analyze the statistics and find that the system does not appear to be sending enough IOPS to your external MDisk, you can increase the load parameter.

Easy Tier operating modes

Easy Tier includes the following main operating modes:

- Off

  When off, no statistics are recorded and no cross-tier extent migration occurs. Also, with Easy Tier turned off, no storage pool balancing across MDisks in the same tier is performed, even in single-tier pools.

- Evaluation or measurement only

  When in this mode, Easy Tier collects only usage statistics for each extent in a storage pool if it is enabled on the volume *and* the pool. No extents are moved. This collection is typically done for a single-tier pool that contains only HDDs so that the benefits of adding flash drives to the pool can be evaluated before any major hardware acquisition.

- Automatic data placement and storage pool balancing

  In this mode, usage statistics are collected and extent migration is performed between tiers (if more than one tier in a pool exists). Also, auto-balancing among MDisks in each tier is performed.

The default operation mode is `Enabled`. Therefore, the system balances storage pools. If the required licenses are installed, they also optimize performance.

Note: The auto-balance process automatically balances data when MDisks are added to a pool. However, the process does not migrate extents from existing MDisks to achieve even extent distribution among all old and new MDisks in the storage pool. The Easy Tier migration plan is based on performance. It is *not* based on the capacity of the underlying MDisks or on the number of extents on them.

Implementation considerations

Consider the following implementation and operational rules when you use the IBM System Storage Easy Tier function on the storage system:

- If the system contains self-compressing drives (IBM FlashCore Module [FCM] drives) in the top tier of storage in a pool with multiple tiers and Easy Tier is in use, consider setting an overallocation limit within these pools, as described in “Overallocation limit” on page 681.
- Volumes that are added to storage pools will use extents from the middle tier of a three-tier model, if available. Easy Tier then collects usage statistics to determine which extents to move to “faster” or “slower” tiers. If no free extents exist in the middle tier, extents from the other tiers are used (bottom tier if possible, otherwise top tier).
- When an MDisk with allocated extents is deleted from a storage pool, extents in use are migrated to MDisks in the same tier as the MDisk that is being removed, if possible. If insufficient extents exist in that tier, extents from another tier are used.
- Easy Tier monitors the extent I/O activity of each copy of a mirrored volume. Easy Tier works with each copy independently of the other copy. This situation applies to volume mirroring and IBM HyperSwap and Remote Copy (RC).

  Note: Volume mirroring can have different workload characteristics on each copy of the data because reads are normally directed to the primary copy and writes occur to both copies. Therefore, the number of extents that Easy Tier migrates between the tiers might differ for each copy.

- Easy Tier automatic data placement is not supported on image mode or sequential volumes. However, it supports evaluation mode for such volumes. I/O monitoring is supported and statistics are accumulated.
- When a volume is migrated out of a storage pool that is managed with Easy Tier, Easy Tier automatic data placement mode is no longer active on that volume. Automatic data placement is also turned off while a volume is being migrated, even when it is between pools that both have Easy Tier automatic data placement enabled. Automatic data placement for the volume is reenabled when the migration is complete.

  When the system migrates a volume from one storage pool to another, it attempts to migrate each extent to an extent in the new storage pool from the same tier as the original extent, if possible.
- When Easy Tier automatic data placement is enabled for a volume, you cannot use the **`svctask migrateexts`** CLI command on that volume.

9.1.2 Implementing and tuning Easy Tier

The Easy Tier function is enabled by default. It starts monitoring I/O activity immediately after the storage pool and volumes are created.

Without the proper licenses installed, the system rebalances only storage pools.

A few parameters can be adjusted. Also, Easy Tier can be turned off on selected volumes in storage pools.

MDisk settings

The tier for internal (array) MDisks is detected automatically and depends on the type of drives, which are its members. No adjustments are needed.

For an external MDisk, the tier is assigned when it is added to a storage pool. To assign the MDisk, select **Pools** → **External Storage**, select the MDisk (or MDisks) to add, and click **Assign** in **Actions** (on top) or by right-clicking the menu.

Note: The tier of MDisks that is mapped from specific types of IBM System Storage Enterprise Flash is fixed to tier0_flash and cannot be changed.

You can choose the target storage pool and storage tier that is assigned, as shown in Figure 9-3.

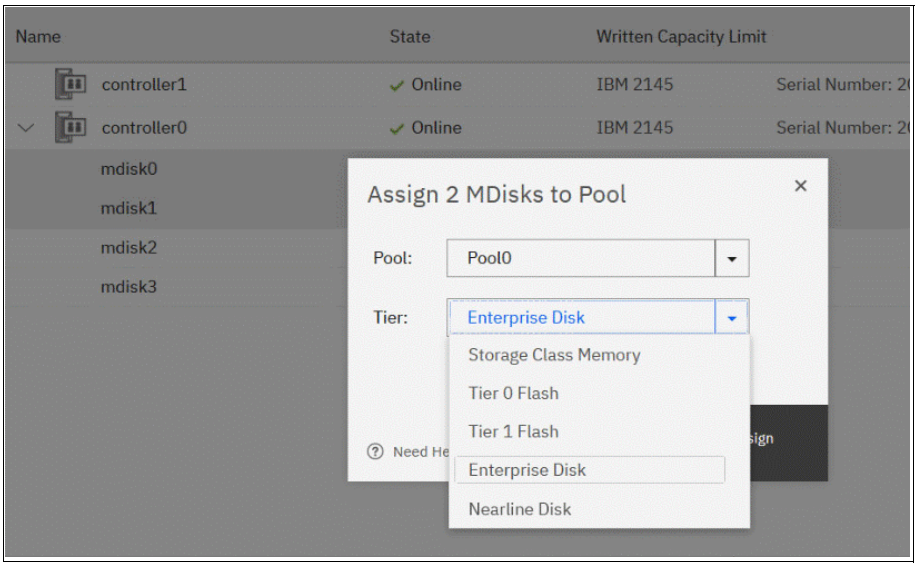


Figure 9-3 Choosing the tier when assigning MDisks

To change the storage tier for an MDisk that is assigned, select **Pools** → **External Storage**, right-click one or more selected MDisks, and choose **Modify Tier**, as shown in Figure 9-4.

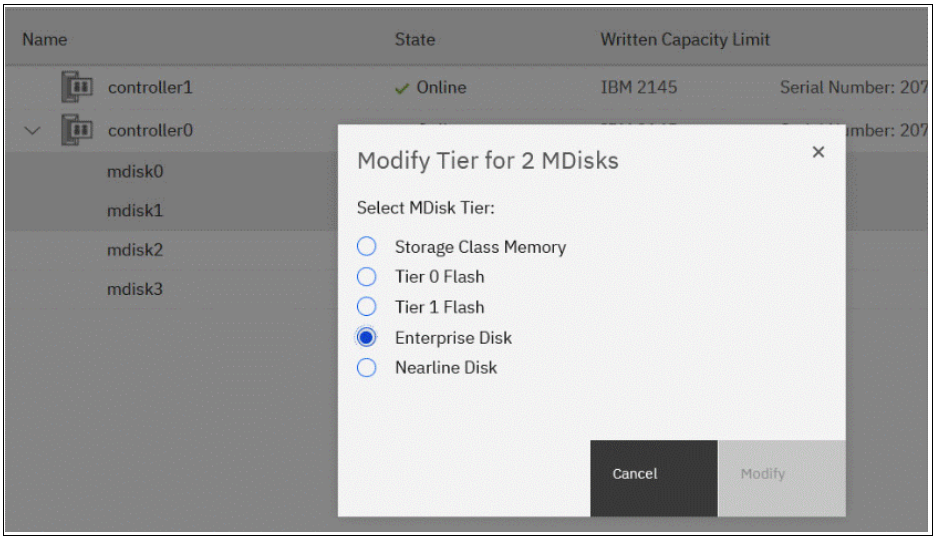


Figure 9-4 Changing the MDisk tier

Note: Assigning a tier to an external MDisk that does not match the physical back-end storage type is not supported by IBM and can lead to unpredictable consequences.

To determine what tier is assigned to an MDisk, select **Pools** → **External Storage**, select **Actions** → **Customize columns** and then select **Tier**. This action includes the current tier setting into a list of MDisk parameters that are shown in the External Storage window.

You can also find this information in MDisk properties.

To show this information, right-click **MDisk**, select **Properties** and then click **View more details**, as shown in Figure 9-5.

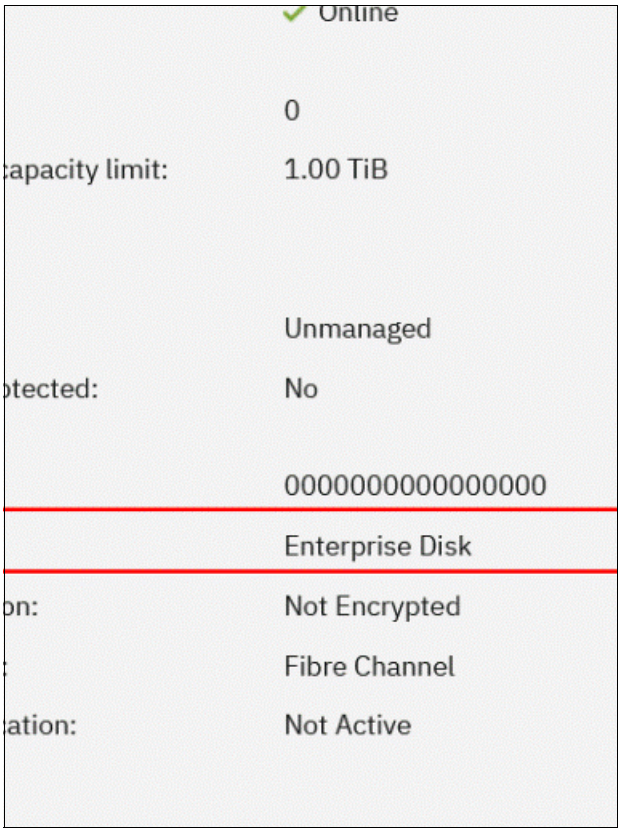


Figure 9-5 MDisk properties

To list MDisk parameters with the command line interface (CLI), run the **lsmdisk** command. The current tier for each MDisk is shown. To change the external MDisk tier, run the **chmdisk** command with the **-tier** parameter, as shown in Example 9-1.

Example 9-1 Listing and changing tiers for MDisks (partially shown)

```
IBM FlashSystem 7200:ITSOFS7K:superuser>lsmdisk
id name   status mode      mdisk_grp_id ... tier            encrypt
1  mdisk1 online unmanaged              ... tier0_flash     no
2  mdisk2 online managed   0            ... tier_enterprise no
3  mdisk3 online managed   0            ... tier_enterprise no
<...>
IBM FlashSystem 7200:ITSOFS7K:superuser>chmdisk -tier tier1_flash mdisk2
IBM FlashSystem 7200:ITSOFS7K:superuser>
```

For an external MDisk, because the system cannot calculate its exact performance capabilities, it has several predefined levels. In rare cases, statistics analysis might show that Easy Tier is overusing or underusing an MDisk. If so, levels can be adjusted only by using the CLI. Run **chmdisk** with the **-easytierload** parameter.

To reset the Easy Tier load to the system default for the chosen MDisk, use **-easytier default**, as shown in Example 9-2 on page 680.

Example 9-2 Changing the Easy Tier load

```
IBM FlashSystem 7200:ITSOFS7K:superuser>chmdisk -easytierload default mdisk2
IBM FlashSystem 7200:ITSOFS7K:superuser>
IBM FlashSystem 7200:ITSOFS7K:superuser>lsmdisk mdisk2 | grep tier
tier tier_enterprise
easy_tier_load high
IBM FlashSystem 7200:ITSOFS7K:superuser>
```

Note: Adjust the Easy Tier load settings only if instructed to do so by IBM Technical Support or your solution architect.

To list the current Easy Tier load setting of an MDisk, run **`lsmdisk`** with the MDisk name or ID as a parameter.

Storage pool settings

When a storage pool (standard pool or Data Reduction Pool [DRP]) is created, Easy Tier is turned on by default. The system automatically enables Easy Tier functions when the storage pool contains an MDisk from more than one tier. It also enables automatic rebalancing when the storage pool contains an MDisk from only one tier.

You can disable Easy Tier or switch it to measure-only mode when creating a pool or at any other time. This task cannot be done by using the GUI, but can be done by using the CLI.

To check the current Easy Tier function state on a pool, select **Pools** → **Pools**, right-click the selected pool and then click **Properties**, as shown in Figure 9-6. This window also shows the amount of data that is stored on each tier.

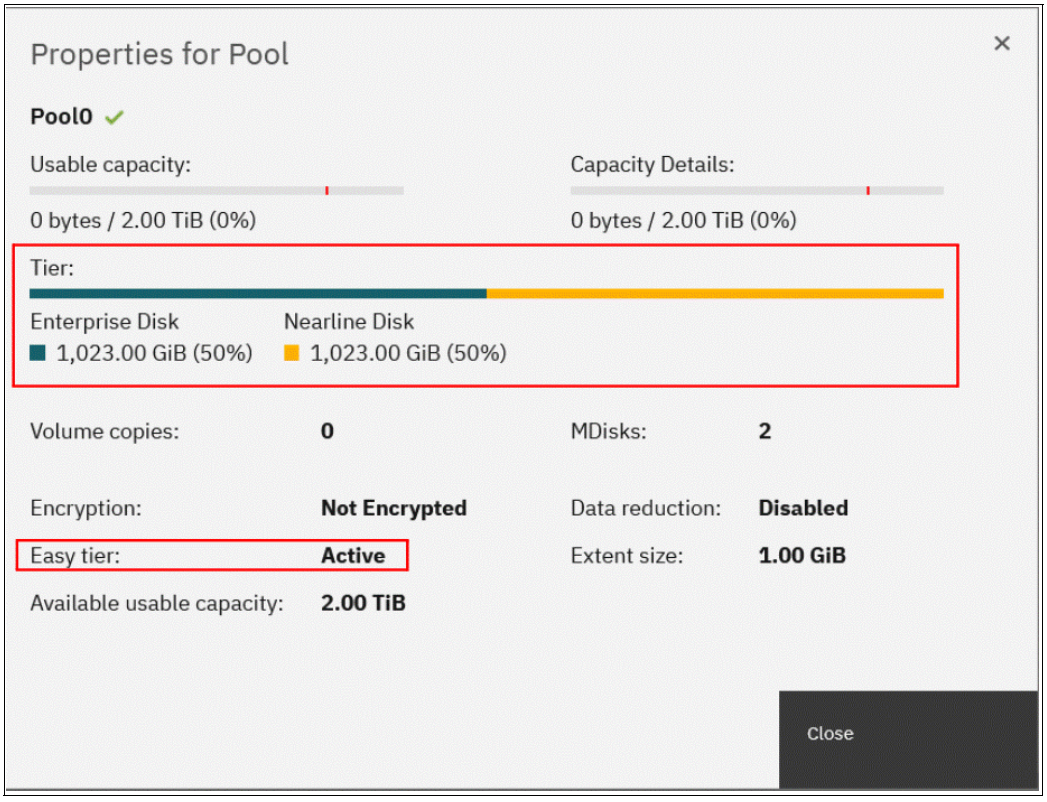


Figure 9-6 Pool properties

Easy Tier can be in one of the following states:

- Active

  Indicates that a pool is being managed by Easy Tier, and extent migrations between tiers can be performed. Performance-based pool balancing of MDisks in the same tier is also enabled. This state is the expected one for a pool with two or more tiers of storage.

- Balanced

  Indicates that a pool is being managed by Easy Tier to provide performance-based pool balancing of MDisks in the same tier. This state is the expected one for a pool with a single tier of storage.

- Inactive

  Indicates that Easy Tier is inactive (disabled).

- Measured

  Shows that Easy Tier statistics are being collected but no extent movement can be performed.

To find the state of the Easy Tier function on the pools by using the CLI, run the **lsmdiskgrp** command without any parameters. To turn off or on Easy Tier, run the **chmdiskgrp** command, as shown in Example 9-3. By running **lsmdiskgrp** with pool name/ID as a parameter, you can also determine how much storage of each tier is available within the pool.

Example 9-3 Listing and changing the Easy Tier status on pools

```
IBM FlashSystem 7200:ITSOFS7K:superuser>lsmdiskgrp
id name       status mdisk_count ... easy_tier easy_tier_status
0  TieredPool online 1           ... auto      balanced
IBM FlashSystem 7200:ITSOFS7K:superuser>chmdiskgrp -easytier measure TieredPool
IBM FlashSystem 7200:ITSOFS7K:superuser>chmdiskgrp -easytier auto TieredPool
IBM FlashSystem 7200:ITSOFS7K:superuser>
```

Overallocation limit

If the system contains self-compressing drives (FCM drives) in the top tier of storage in a pool with multiple tiers and Easy Tier is in use, consider setting an *overallocation limit* within these pools. The overallocation limit has no effect in pools with a different configuration.

Arrays that are created from self-compressing drives have a written capacity limit (virtual capacity before compression) that is higher than the array's usable capacity (physical capacity). Writing highly compressible data to the array means that the written capacity limit can be reached without running out of usable capacity. However, if data is not compressible or the compression ratio is low, it is possible to run out of usable capacity before reaching the written capacity limit of the array, which means the amount of data that is written to a self-compressing array must be controlled to prevent the array from running out of space.

Without a maximum overallocation limit, Easy Tier scales the usable capacity of the array based on the actual compression ratio of the data that is stored on the array at a point in time (PiT). Easy Tier migrates data to the array and might use a large percentage of the usable capacity in doing so, but it stops migrating to the array when the array is close to using all of its usable capacity. then it might start migrating data away from the array again to free up space.

However, Easy Tier migrates storage only at a slow rate, which might not keep up with changes to the compression ratio within the tier. When Easy Tier swaps extents or data is overwritten by hosts, compressible data might be replaced with data that is less compressible, which increases the amount of usable capacity that is consumed by extents and might result

in self-compressing arrays running out of space, which can cause a loss of access to data until the condition is resolved.

So, the user might specify the maximum overallocation ratio for pools that contain self-compressing arrays to prevent out-of-space scenarios. The value acts as a multiplier of the physically available space in self-compressing arrays. The allowed values are a percentage in the range of 100% (default) to 400% or off. The default setting allows no overallocation on new pools. Setting the value to off disables this feature.

When enabled, Easy Tier scales the available usable capacity of self-compressing arrays by using the specified overallocation limit and adjusts the migration plan to make sure the fullness of these arrays stays below the maximum overallocation. Specify the maximum overallocation limit based on the estimated lowest compression ratio of the data that is written to the pool.

For example, for an estimated compression ratio of 1.2:1, specify an overallocation limit of 120% to put a limit on the overallocation. Easy Tier stops migrating data to self-compressing arrays in the pool after the written capacity reaches 120% of the physical (usable) capacity of the array, which is the case even if the written capacity limit of the array is not reached yet or the current compression ratio of the data that is stored on the array is higher than 1.2:1 (and thus more usable capacity is available). This setting prevents changes to the compression ratio within the specified limits from causing the array to run out of space.

To modify the maximum overallocation limit of a pool by using the GUI, select **Pools** → **Pools**, right-click a pool, and select **Easy Tier Overallocation Limit**, as shown in Figure 9-7.

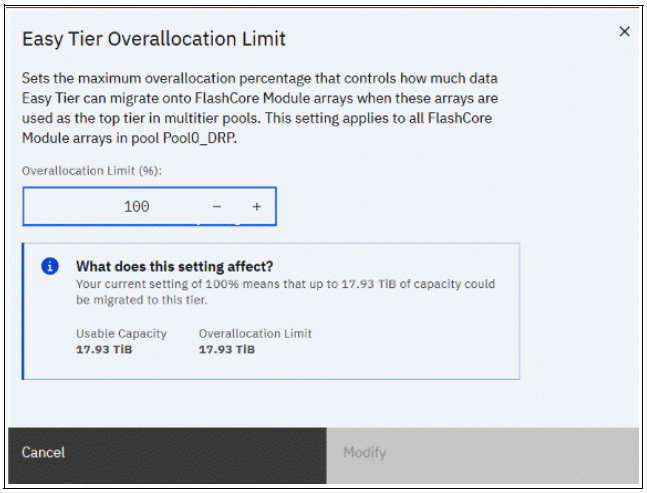


Figure 9-7 Modifying the pool overallocation limit

On the CLI, run the **chmdiskgrp** command with the **-etfcmoverallocationmax** parameter to set a percentage or use **off** to disable the limit.

Volume settings

By default, each striped-type volume enables Easy Tier to manage its extents. If you need to fix the volume extent location (for example, to prevent extent demotes and to keep the volume in the higher-performing tier), you can turn off Easy Tier management for a particular volume copy.

Note: Thin-provisioned and compressed volumes in a DRP cannot have Easy Tier turned off. You can turn off Easy Tier only at a pool level.

You can do this task only by using the CLI. Run the **lsvdisk** command to check and the **chvdisk** command to modify the Easy Tier function status on a volume copy, as shown in Example 9-4.

Example 9-4 Checking and modifying the Easy Tier settings on a volume

```
IBM_FlashSystem 7200:ITSO-V7k:superuser>lsvdisk  vdisk0 |grep easy_tier
easy_tier on
easy_tier_status balanced
IBM_FlashSystem 7200:ITSO-V7k:superuser>chvdisk -easytier off vdisk0
IBM FlashSystem 7200:ITSOFS7K:superuser>
```

System-wide settings

There is a system-wide setting that is called *Easy Tier acceleration* that is disabled by default. Turning it on makes Easy Tier move extents up to four times faster than the default setting. In acceleration mode, Easy Tier can move up to 48 GiB per 5 minutes, but in normal mode it moves up to 12 GiB. The following use cases are the most probable use cases for acceleration:

- When adding capacity to the pool either by adding to an existing tier or by adding a tier to the pool, accelerating Easy Tier can quickly spread volumes onto the new MDisks.
- Migrating the volumes between the storage pools when the target storage pool has more tiers than the source storage pool, so Easy Tier can quickly promote or demote extents in the target pool.

Note: Enabling Easy Tier acceleration is advised only during periods of low system activity only after migrations or storage reconfiguration occurred. It is a best practice to keep off the Easy Tier acceleration mode during normal system operation to avoid the performance impacts that are caused by accelerated data migrations.

This setting can be changed non-disruptively, but only by using the CLI. To turn on or off Easy Tier acceleration mode, run the **chsystem** command. Run the **lssystem** command to check its current state, as shown in Example 9-5.

Example 9-5 The chsystem command

```
IBM FlashSystem 7200:ITSOFS7K:superuser>lssystem |grep easy_tier
easy_tier_acceleration off
IBM FlashSystem 7200:ITSOFS7K:superuser>chsystem -easytieracceleration on
IBM FlashSystem 7200:ITSOFS7K:superuser>
```

9.1.3 Monitoring Easy Tier activity

When Easy Tier is active, it constantly monitors and records I/O activity, collecting extent heat data. Heat data files are produced approximately once a day and summarize the activity per volume since the last heat data file was produced. Easy Tier activity can be monitored by using the GUI or the external IBM Storage Tier Advisor Tool (IBM STAT) application.

Monitoring Easy Tier by using the GUI

To view the most recent Easy Tier statistics, select **Monitoring** → **Easy Tier Reports**. Select the storage pool that you want to see reports for in the filter section on the left, as shown in Figure 9-8 on page 684. It takes approximately 24 hours for reports to be available after turning on Easy Tier or after a configuration node failover occurred. If no reports are available,

the error message in the figure is shown. In this case, wait until new reports were generated and then revisit the GUI.

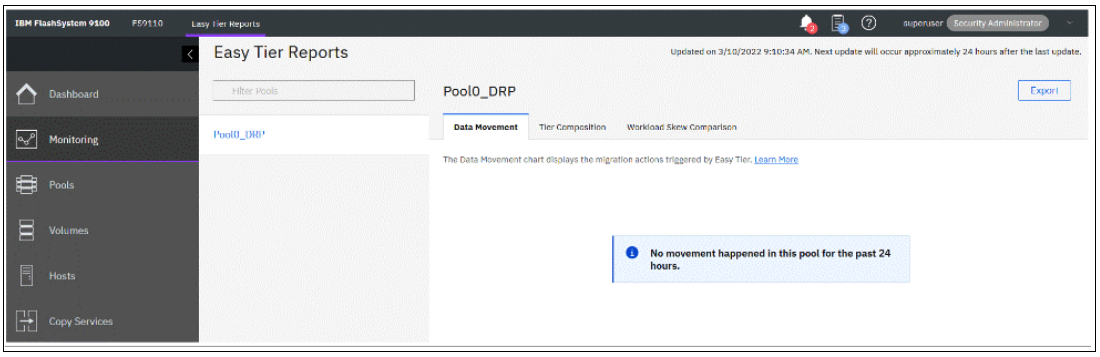


Figure 9-8 Easy Tier reports are not available

Three types of reports are available per storage pool: Data Movement, Tier Composition, and Workload Skew Comparison. Select the corresponding tabs in the GUI to view the charts. Alternatively, click **Export** or **Export All** to download the reports in comma-separated value (CSV) format.

Data Movement report

The Data Movement report shows the extent migrations that Easy Tier performed to relocate data between different tiers of storage and within the same tier for optimal performance, as shown in Figure 9-9. The chart displays the data for the previous 24-hour period, in one-hour increments. You can change the time span and the increments for a more detailed view.

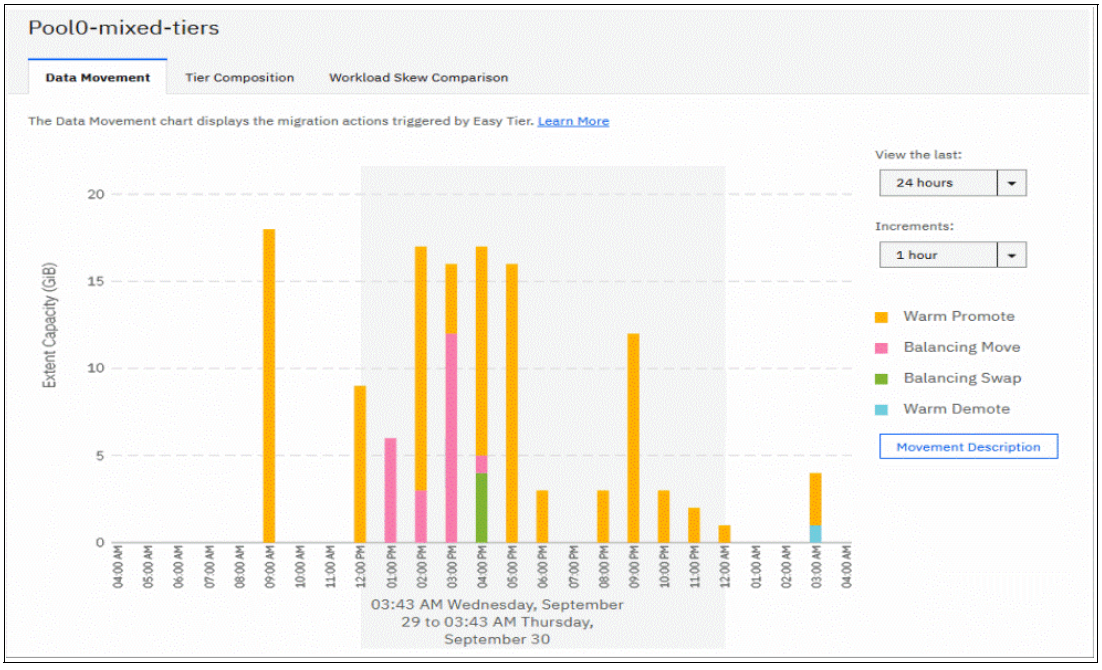


Figure 9-9 Easy Tier Data Movement report

The X-axis shows a timeline for the selected period by using the selected increments. The Y-axis indicates the amount of extent capacity that is moved. For each time increment, a color-coded bar displays the amount of data that is moved by each Easy Tier data movement action, such as promote or cold demote.

For more information about the different movement actions, see “Easy Tier automatic data placement” on page 673 or click **Movement Description** that is next to the chart to see an explanation in the GUI, as shown in Figure 9-10.

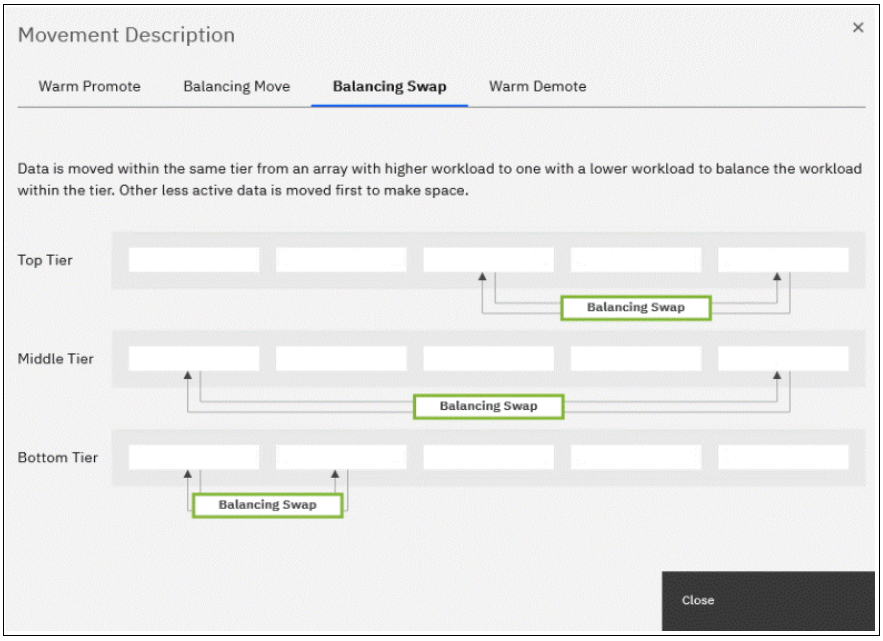


Figure 9-10 Easy Tier movement description

Tier Composition chart

The Tier Composition chart shows how different types of workloads are distributed between top, middle, and bottom tiers of storage in the selected pool, as shown in Figure 9-11.

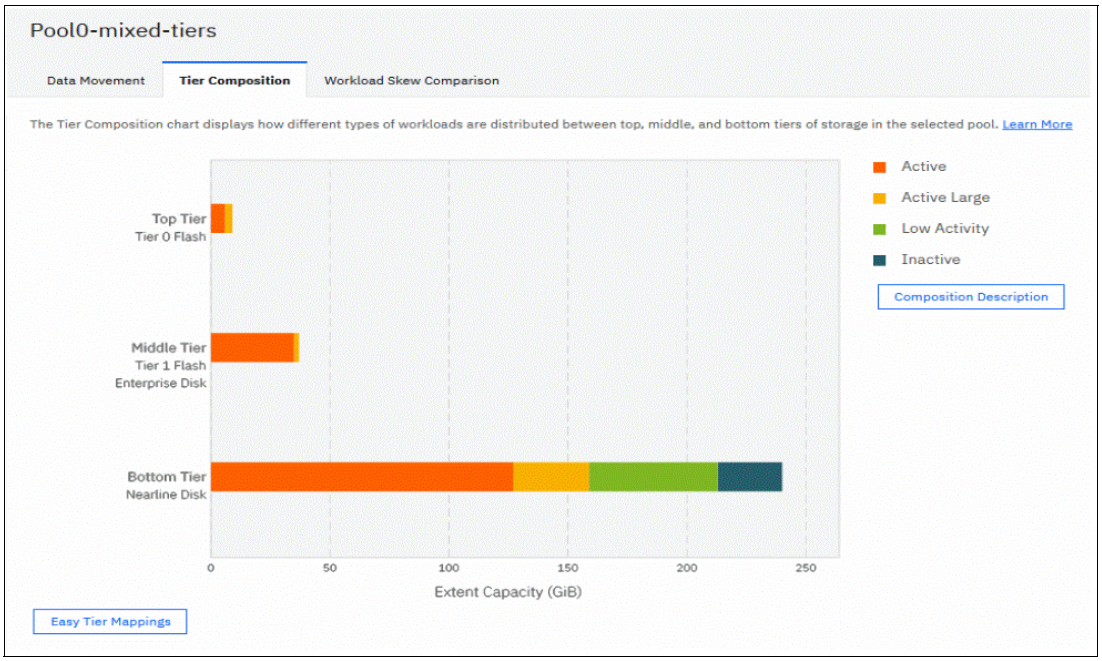


Figure 9-11 Tier Composition chart

A color-coded bar for each tier shows which workload types are present in that tier and how much of the extent capacity in that tier to which they can be attributed. Easy Tier distinguishes between the following workload types. Click **Composition Description** to show a short explanation for each workload type in the GUI as shown in Figure 9-12.

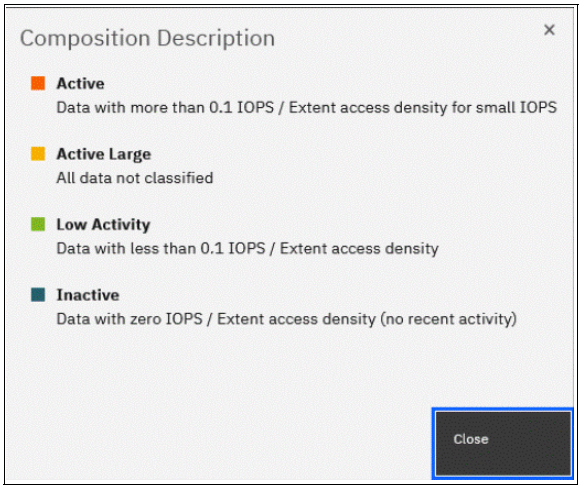
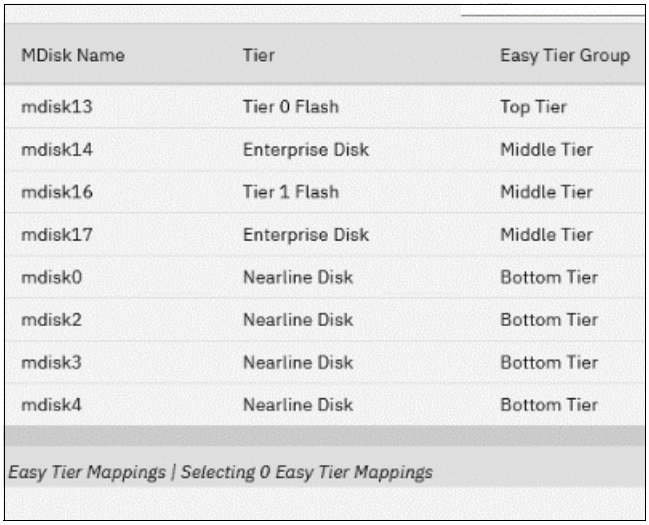


Figure 9-12 Composition description

Click the **Easy Tier Mappings** button to show which MDisks were assigned to which of the three tiers of Easy Tier, as shown in Figure 9-13.



MDisk Name	Tier	Easy Tier Group
mdisk13	Tier 0 Flash	Top Tier
mdisk14	Enterprise Disk	Middle Tier
mdisk16	Tier 1 Flash	Middle Tier
mdisk17	Enterprise Disk	Middle Tier
mdisk0	Nearline Disk	Bottom Tier
mdisk2	Nearline Disk	Bottom Tier
mdisk3	Nearline Disk	Bottom Tier
mdisk4	Nearline Disk	Bottom Tier

Easy Tier Mappings | Selecting 0 Easy Tier Mappings

Figure 9-13 Easy Tier Mappings

How storage tiers in the system are mapped to Easy Tier tiers depends on the available storage tiers in the pool. For a list of all possible mappings, see Table 9-1 on page 672.

Workload Skew Comparison

The Workload Skew Comparison chart displays the percentage of I/O workload that is attributed to a percentage of the total capacity, as shown in Figure 9-14.

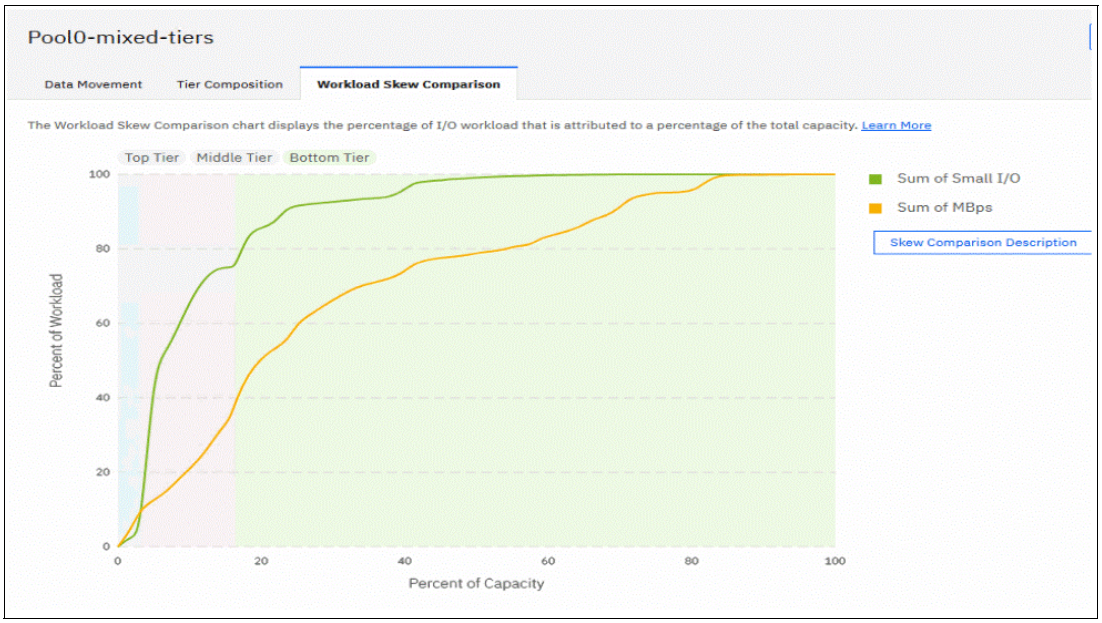


Figure 9-14 Easy Tier Workload Skew Comparison

The X-axis shows the percentage of capacity and the Y-axis shows the corresponding percentage of workload on that capacity. Workload is classified in small I/O (sum of small reads and writes) and megabytes per second (MBps) (sum of small and large bandwidth). The portion of capacity and workload that is attributed to a tier is color-coded in the chart with a legend above the chart.

Figure 9-14 shows that the top tier (Tier1 Flash) contributes only a tiny percentage of capacity to the pool, but handles around 85% of the IOPS and more than 40% of the bandwidth in that pool. The middle tier (enterprise disk) handles almost all the remaining IOPS and an extra 20% of the bandwidth. The bottom tier (NL disk) provides most of the capacity to the pool but does almost no small I/O workload.

Use this chart to estimate how much storage capacity in the high tiers must be available to handle most of the workload.

To view a description of skew comparison click **Skew Comparison Description** as shown in Figure 9-15.



Figure 9-15 Skew Comparison Description

Monitoring Easy Tier by using the IBM Storage Tier Advisor Tool

The IBM STAT is a Windows console application that can analyze heat data files that are generated by Easy Tier and produce a graphical display of the amount of “hot” data per volume and predictions of the performance benefits of adding more capacity to a tier in a storage pool.

Using this method of monitoring, Easy Tier can provide more insights on top of the information that is available in the GUI.

IBM STAT can be downloaded from IBM Support IBM Storage Tier Advisor Tool.

You can download the IBM STAT and install it on your Windows-based computer. The tool is packaged as an ISO file that must be extracted to a temporary location.

The tool installer is at `temporary_location\IMAGES\STAT\Disk1\InstData\NoVM\`. By default, the IBM STAT is installed in `C:\Program Files\IBM\STAT\`.

On the system, the heat data files are found in the `/dumps/easytier` directory on the configuration node and are named `dpa_heat.node_panel_name.time_stamp.data`. Any heat data file is erased when it exists for longer than 7 days.

Heat files must be offloaded and IBM STAT started from a Windows command prompt console with the file specified as a parameter, as shown in Example 9-6.

Example 9-6 Running IBM STAT by using the Windows command prompt

```
C:\Program Files (x86)\IBM\STAT>stat dpa_heat.78DXRY0.191021.075420.data
```

The IBM STAT creates a set of `.html` and `.csv` files that can be used for Easy Tier analysis.

To download a heat data file, select **Settings** → **Support** → **Support Package** → **Download Support Package** → **Download Existing Package**, as shown in Figure 9-16.

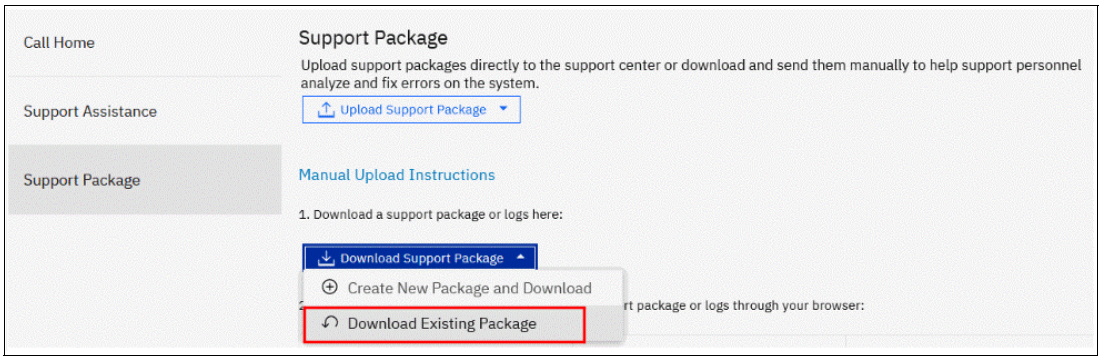


Figure 9-16 Downloading an Easy Tier heat file: Download Support Package

A download window opens that shows all the files in `/dumps` and its subfolders on a configuration node. You can filter the list by using the “easytier” keyword, select the **dpa_heat** file or files that are analyzed, and then click **Download**, as shown in Figure 9-17. Save them in a convenient location (for example, to a subfolder that holds the IBM STAT executable file).

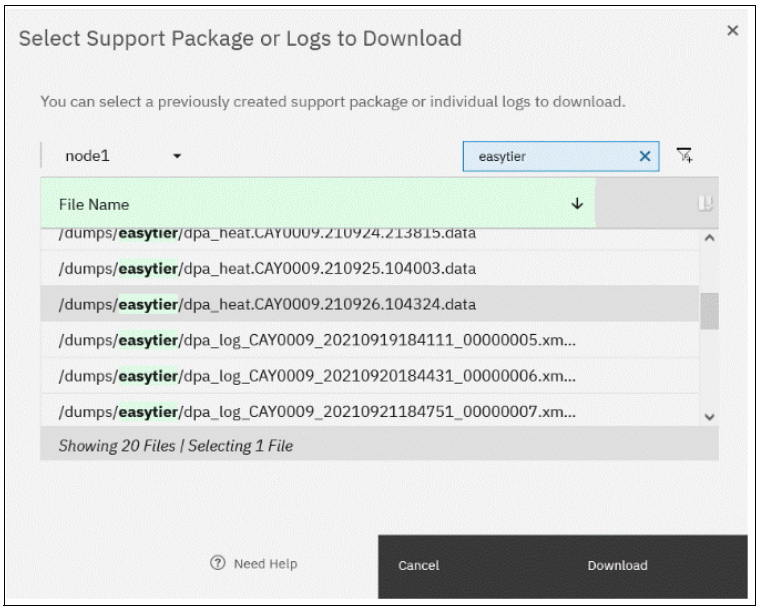


Figure 9-17 Downloading Easy Tier heat data file: dpa_heat files

You can also specify the output directory. IBM STAT creates a set of HTML files, and the user can then open the `index.html` file in a browser to view the results. Also, the following CSV files are created and placed in the `Data_files` directory:

- `<panel_name>_data_movement.csv`
- `<panel_name>_skew_curve.csv`
- `<panel_name>_workload_ctg.csv`

These files can be used as input data for other utilities, such as the IBM STAT Charting Utility.

For more information about how to interpret IBM STAT tool output and CSV files analysis, see *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521. XXX Vasfi to also update this

9.2 Thin-provisioned volumes

In a shared storage environment, *thin provisioning* is a method for optimizing the usage of available storage. It relies on the allocation of capacity on demand instead of the traditional method of allocating all the capacity at the time of initial provisioning. Using this principle means that storage environments can achieve higher utilization of physical storage resources by eliminating the unused allocated capacity.

Traditional storage allocation methods often provision large amounts of storage to individual hosts, but some of it remains unused (not written to), which might result in poor usage rates (often as low as 10%) of the underlying physical storage resources. Thin provisioning avoids this issue by presenting more storage capacity to the hosts than it uses from the storage pool. Physical storage resources can be expanded over time to respond to growth.

9.2.1 Concepts

The system supports thin-provisioned volumes in standard pools and in DRPs.

Each volume has a *provisioned capacity* and a *real capacity*. Provisioned capacity is the volume storage capacity that is available to a host. It is the capacity that is detected by host operating systems and applications and can be used when creating a file system. Real capacity is the storage capacity that is reserved to a volume copy from a pool.

In a standard-provisioned volume, the provisioned capacity and real capacity are the same. However, in a thin-provisioned volume, the provisioned capacity can be much larger than the real capacity.

The provisioned capacity of a thin-provisioned volume is larger than its real capacity. As more information is written by the host to the volume, more of the real capacity is used. The system identifies read operations to unwritten parts of the provisioned capacity and returns zeros to the server without using any real capacity.

The autoexpand feature prevents a thin-provisioned volume from using up its capacity and going offline. As a thin-provisioned volume uses capacity, the autoexpand feature maintains a fixed amount of unused real capacity that is called the *contingency capacity*. For thin-provisioned volumes in standard pools, the autoexpand feature can be turned on and off. For thin-provisioned volumes in DRPs, the autoexpand feature is always enabled.

The capacity of a thin-provisioned volume is split into chunks that are called *grains*. Write I/O to grains that have not previously been written to causes real capacity to be used to store data and metadata. The grain size of thin-provisioned volumes in standard pools can be 32 KB, 64 KB, 128 KB, or 256 KB. Generally, smaller grain sizes save space but require more metadata access, which can adversely impact performance. When you use thin provisioning with IBM FlashCopy, specify the same grain size for the thin-provisioned volume and FlashCopy. The grain size of thin-provisioned volumes in DRPs cannot be changed from the default of 8 KB.

A thin-provisioned volume can be converted non-disruptively to a fully allocated volume or vice versa by using the volume mirroring function. For example, you can add a thin-provisioned copy to a fully allocated volume and then remove the fully allocated copy from the volume after it is synchronized.

The fully allocated to thin-provisioned migration procedure uses a zero-detection algorithm so that grains that contain all zeros do not cause any real capacity to be used. Usually, if the system is to detect zeros on the volume, you must use software on the host side to write zeros to all unused space on the disk or file system.

9.2.2 Implementation

For more information about creating thin-provisioned volumes, see Chapter 6, “Volumes” on page 419.

Metadata

In a standard pool, the system uses real capacity to store data that is written to the volume and metadata that describes the thin-provisioned configuration of the volume. The metadata that is required for a thin-provisioned volume is usually less than 0.1% of the provisioned capacity.

If the host uses 100% of the provisioned capacity, some extra space is required on your storage pool to store thin-provisioned metadata. In the worst case, the real size of a thin-provisioned volume can be 100.1% of its virtual capacity.

In a DRP, metadata for a thin-provisioned volume is stored separately from user data and not reflected in the volume’s real capacity. Capacity reporting is handled at the pool level.

Volume parameters

When creating a thin-provisioned volume in a standard pool, some of its parameters can be modified in Custom mode, as shown in Figure 9-18.

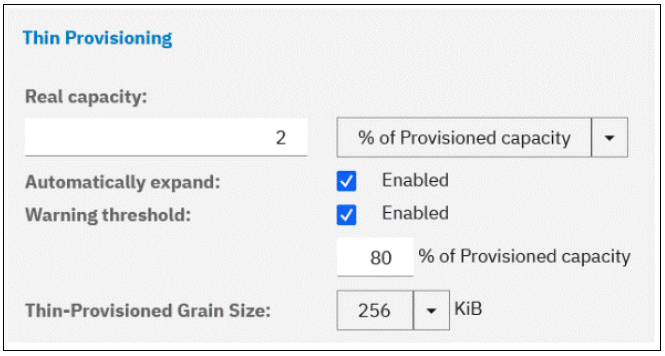


Figure 9-18 Volume parameters for thin provisioning

Real capacity defines both the initial volume real capacity and the amount of contingency capacity. When autoexpand is enabled, the system tries to maintain the contingency capacity always by allocating extra real capacity when hosts write to the volume.

The warning threshold can be used to send a notification when the volume is about to run out of space.

In a DRP, fine-tuning of these parameters is not required. The real capacity and warning threshold are handled at the pool level. The grain size is always 8 KB, and autoexpand is always on.

Host considerations: Do not use defragmentation applications on thin-provisioned volumes. The defragmentation process can write data to different areas of a volume, which can cause a thin-provisioned volume to grow up to its provisioned size.

9.3 UNMAP

IBM Storage Virtualize systems running Version 8.1.0 and later support the Small Computer System Interface (SCSI) UNMAP command. This command enables hosts to notify the storage controller of capacity that is no longer required, which can improve capacity savings and performance of flash storage.

9.3.1 The SCSI UNMAP command

UNMAP is a set of SCSI primitives that enable hosts to indicate to a storage system that space that is allocated to a range of blocks on a storage volume is no longer required. This command enables the storage system to take measures and optimize the system so that the space can be reused for other purposes.

When a host writes to a volume, storage is allocated from the storage pool. To free allocated space back to the pool, human intervention is needed on the storage system. The SCSI UNMAP feature is used to allow the host operating system to unprovision storage on the storage system, which means that the resources can automatically be released in the storage pools and used for other purposes.

One of the most common use cases is a host application, such as VMware, freeing storage within a file system. then the storage system can reorganize the space, such as optimizing the data on the volume or the pool so that space can be reclaimed.

A SCSI unmappable volume is a volume that can have storage unprovision and space reclamation that is triggered by the host operating system. The system can pass the SCSI **`UNMAP`** command through to back-end flash storage and external storage controllers that support the function.

9.3.2 Back-end SCSI UNMAP

The system can generate and send SCSI UNMAP commands to specific back-end storage controllers and internal flash storage. Support for SCSI UNMAP was introduced with Version 8.1.1.

This process occurs when volumes are formatted or deleted, extents are migrated, or an UNMAP command is received from the host. SCSI UNMAP commands are sent only to the following back-end controllers:

- IBM FlashSystem A9000
- IBM Storwize V5000 family, V5100, and V7000 (Version 8.1.0 or later)
- IBM FlashSystem 5000 family, FlashSystem 5200, 7200, 91x0, 9200, and FlashSystem V9000 (Version 8.1.0 or later)
- Beginning with 8.3.0.1: HPE Nimble storage systems
- Infinidat InfiniBox controllers (when the IBM system is running Version 8.4.0.2 or later)
- Pure storage systems

Back-end SCSI UNMAP commands help prevent an overprovisioned storage controller from running out of free capacity for write I/O requests, which means that when you use supported overprovisioned back-end storage, back-end SCSI UNMAP should be enabled.

Flash storage typically requires empty blocks to serve write I/O requests, which means UNMAP can improve flash performance by erasing blocks in advance.

This feature is turned on by default. It is a best practice to keep back-end UNMAP enabled, especially if a system is virtualizing an overprovisioned storage controller or uses FCM drives.

To verify that sending UNMAP commands to a back end is enabled, run the lssystem command, as shown in Example 9-7.

Example 9-7 Verifying the back-end UNMAP support status

```
IBM FlashSystem 7200:ITSOFS7K:superuser>lssystem | grep backend_unmap
backend_unmap on
```

9.3.3 Host SCSI UNMAP

The IBM Storage Virtualize system can advertise support for SCSI UNMAP to hosts. Hosts can then use the set of SCSI UNMAP commands to indicate that formerly used capacity is no longer required on a volume.

When these volumes are in DRPs, that capacity becomes reclaimable capacity and is monitored and collected, and eventually redistributed back to the pool for use by the system. Volumes in standard pools do not support automatic space reclamation after data is unmapped, and SCSI UNMAP commands are handled as though they were writes with zero data.

The system also sends SCSI UNMAP commands to back-end controllers that support them if host unmaps for corresponding blocks are received (and backend UNMAP is enabled).

With host SCSI UNMAP enabled, some host types (for example, Windows, Linux, or VMware) change their behavior when creating a file system on a volume, issuing SCSI UNMAP commands to the whole capacity of the volume. The format completes only after all of these UNMAP commands complete. Some host types run a background process (for example, fstrim on Linux), which periodically issues SCSI UNMAP commands for regions of a file system that are no longer required. Hosts might also send UNMAP commands when files are deleted in a file system.

Host SCSI UNMAP commands drive more I/O workload to back-end storage. In some circumstances (for example, volumes on a heavily loaded NL-serial-attached SCSI (SAS) array), this situation can cause an increase in response times on volumes that use the same storage. Also, host formatting time is likely to increase compared to a system that does not support the SCSI UNMAP command.

If you use DRPs, an overprovisioned back end that supports UNMAP, or FCM drives, it is a best practice to turn on SCSI UNMAP support. Host UNMAP support is enabled by default.

If only standard pools are configured and the back end is traditional (fully provisioned), consider keeping host UNMAP support turned off because it does not provide any benefit.

To check and modify the current settings for host SCSI UNMAP support, run the **lssystem** and **chsystem** CLI commands, as shown in Example 9-8.

Example 9-8 Turning on host UNMAP support

```
IBM FlashSystem 7200:ITSOFS7K:superuser>lssystem | grep host_unmap
host_unmap off
IBM FlashSystem 7200:ITSOFS7K:superuser>chsystem -hostunmap on
IBM FlashSystem 7200:ITSOFS7K:superuser>
```

Note: You can switch host UNMAP support on and off nondisruptively on the system side. However, hosts might need to rediscover storage, or (in the worst case) be restarted for them to stop sending UNMAP commands.

9.3.4 Offloading I/O throttle

Throttles are a mechanism to control the amount of resources that are used when the system is processing I/Os on supported objects. If a throttle limit is defined, the system processes the I/O for that object or delays the processing of the I/O to free resources for more critical I/O operations.

Offload commands, such as UNMAP and XCOPY, free hosts and speed the copy process by offloading the operations of certain types of hosts to a storage system. These commands are used by hosts to format new file systems, or copy volumes without the host needing to read and then write data.

Offload commands can sometimes create I/O-intensive workloads, potentially taking bandwidth from production volumes and affecting performance, especially if the underlying storage cannot handle the amount of I/O that is generated.

Throttles can be used to delay processing for offloads to free bandwidth for other more critical operations, which can improve performance but limits the rate at which host features, such as VMware VMotion, can copy data. It can also increase the time that it takes to format file systems on a host.

Note: For systems that are managing any NL storage, it might be a best practice to set the offload throttle to 100 MBps.

To implement an offload throttle, run the **mkthrottle** command with the **-type offload** parameter. Alternatively, in the GUI, select **Monitoring** → **System Hardware**, and then click **System Actions** → **Edit System Offload Throttle**, as shown in Figure 9-19.

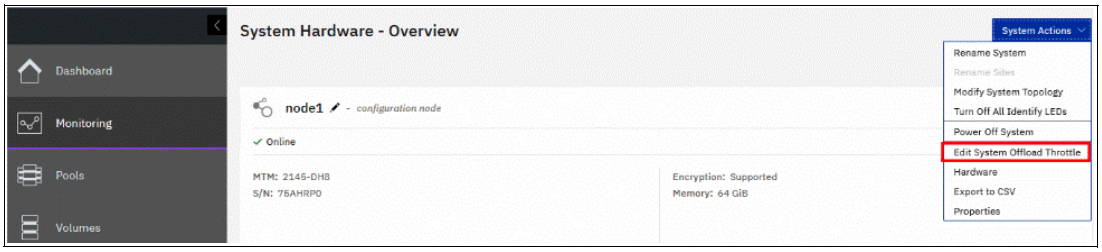


Figure 9-19 Setting an offload throttle

Enter the required settings (Bandwidth limit and IOPS limit) as necessary. Throttle limit is a per node limit, as shown in Figure 9-20 on page 695.

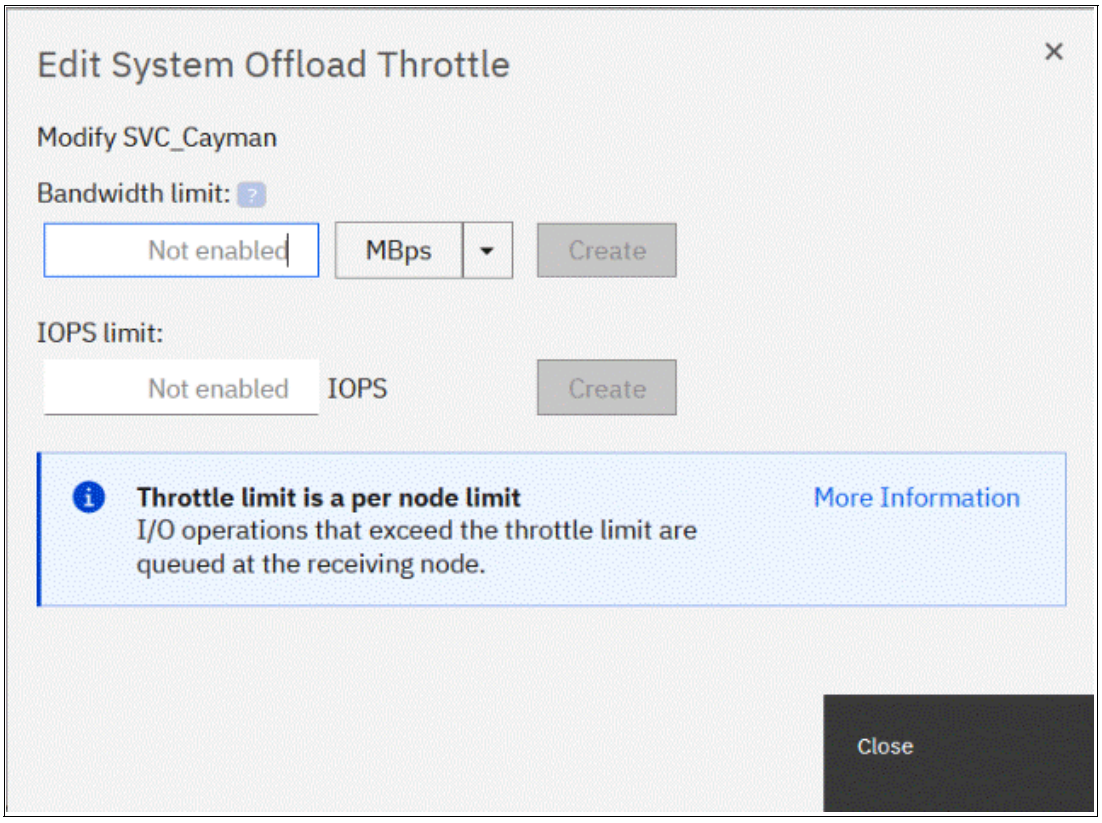


Figure 9-20 System Offload Throttle settings

9.4 Data reduction pools

Data reduction pools (DRP) provide a set of techniques that can be used to reduce the amount of usable capacity that is required to store data, which helps increase storage efficiency and reduce storage costs. Available techniques include thin provisioning, compression, and deduplication.

DRPs automatically reclaim used capacity that is no longer needed by host systems and return it back to the pool as available capacity for future reuse.

The data reduction in DRPs is embedded in this pool type and no separate license is necessary. This situation does not apply to real-time compression (RtC), where a specific capacity-based license is needed.

Note: This book provides only an overview of DRP. For more information, see *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430.

9.4.1 Introduction to DRP

The system can use different data reduction methods simultaneously, which increases the capacity savings across the entire storage pool.

DRPs support five types of volumes:

- Fully allocated

This type provides no data reduction.

- Thin provisioned

  This type provides data reduction by allocating storage on demand when writing to the volume.

- Thin and compressed

  In addition to being thin provisioned, data is compressed before being written to storage.

- Thin and deduplicated

  In addition to being thin provisioned, duplicates of data blocks are detected and replaced with references to the first copy.

- Thin, compressed, and deduplicated

  This type achieves maximum data reduction by combining thin provisioning, compression, and deduplication.

Volumes in a DRP track when capacity is freed from hosts and possible unused capacity that can be collected and reused within the storage pool. When a host no longer needs the data that is stored on a volume, the host system uses SCSI UNMAP commands to release that capacity from the volume. When these volumes are in DRPs, that capacity becomes reclaimable capacity, and is monitored, collected, and eventually redistributed back to the pool for use by the system.

Note: If the usable capacity usage of a DRP exceeds more than 85%, I/O performance can be affected. The system needs 15% of usable capacity available in DRPs to ensure that capacity reclamation can be performed efficiently.

At its core, a DRP uses a Log Structured Array (LSA) to allocate capacity. An LSA enables a tree-like directory to define the physical placement of data blocks independent of size and logical location.

Each volume has a range of logical block addresses (LBAs), starting from 0 and ending with the block address that fills the capacity. The LSA enables the system to allocate data sequentially when written to volumes (in any order) and provides a directory that provides a lookup to match volume LBA with physical addresses within the array. A volume in a DRP contains directory metadata to store the mapping from logical address on the volume to physical location on the back-end storage.

This directory is too large to store in memory, so it must be read from storage as required. The lookup and maintenance of this metadata results in I/O amplification. I/O amplification occurs when a single host-generated read or write I/O results in more than one back-end storage I/O request. For example, a read I/O request might need to read some directory metadata in addition to the actual data. A write I/O request might need to read directory metadata write updated directory metadata, journal metadata, and the actual data.

Conversely, data reduction reduces the size of data that uses compression and deduplication, so less data is written to the back-end storage.

IBM Storage Virtualize V8.4 introduced child pools for DRPs. A child pool is a folder-like object within a parent DRP that contains volumes. The child pool for DRP is quota-less and its capacity is the sum of all volumes within the child pool. A child pool can be assigned to an ownership group and to segment administrative domains. The parent pool and associated child pools share MDisk, deduplication hash table, and encryption keys. Therefore, it seems advisable to use this technology to separate departments of a single client, but not different clients.

Because of the nature of the newly introduced child pools, a new type of volume migration is available to move volumes within a single DRP and its affiliated child pools. With this migration, you can move volumes between all pools within one DRP entity.

9.4.2 DRP benefits

DRPs are a new type of storage pool that implement techniques such as thin provisioning, compression, and deduplication to reduce the amount of physical capacity that is required to store data. Savings in storage capacity requirements translate into a reduction in the cost of storing the data.

The cost reductions that are achieved through software can facilitate the transition to all flash storage. Flash storage has lower operating costs, lower power consumption, higher density, and is cheaper to cool than disk storage. However, the cost of flash storage is still higher. Data reduction can reduce the total cost of ownership (TCO) of an all-flash system to be competitive with HDDs.

One benefit of DRP is in the form of capacity savings that are achieved by deduplication and compression. Real-time deduplication identifies duplicate data blocks during write I/O operations and stores a reference to the first copy of the data instead of writing the data to the storage pool a second time. It does this task by maintaining a fingerprint database containing hashes of data blocks already written to the pool. If new data that is written by hosts matches an entry in this database, then a reference is generated in the directory metadata instead of writing the new data.

Compression reduces the size of the host data that is written to the storage pool. DRP uses the Lempel-Ziv based RtC and decompression algorithm. It offers a new implementation of data compression that is fully integrated into the IBM Storage Virtualize I/O stack. It makes optimal use of node resources such as memory and CPU cores, and uses hardware acceleration on supported platforms efficiently. DRP compression operates on small block sizes, which results in consistent and predictable performance.

Deduplication and compression can be combined, so the data is first deduplicated and then compressed. Therefore, deduplication references are created on the compressed data that is stored on the physical domain.

DRP supports end-to-end SCSI UNMAP functions. Hosts use the set of SCSI UNMAP commands to indicate that the formerly used capacity is no longer required on a target volume. Reclaimable capacity is unused capacity that is created when data is overwritten, volumes are deleted, or when data is marked as unneeded by a host by using the SCSI UNMAP command. That capacity can be collected and reused on the system.

DRPs, the directory, and the actual reduction techniques are designed around optimizing for flash and future solid-state storage technologies. All metadata operations are 4 KB, which is ideal for flash storage to maintain low and consistent latency. All data read operations are 8 KB (before reduction) and designed to minimize latency because flash storage is suitable for small block workloads with high IOPS. All write operations are coalesced into 256 KB sequential writes to simplify the garbage collection on flash devices and gain full stride writes from RAID arrays.

DRP works well with Easy Tier. The directory metadata of DRPs does not fit in memory, so it is stored on disk by using dedicated metadata volumes that are separate from the actual data. The metadata volumes are small but frequently accessed by small block I/O requests. Performance gains are expected because they are optimal candidates for promotion to the fastest tier of storage through Easy Tier. In contrast, data volumes with large but frequently rewritten data is grouped to consolidate “heat”. Easy Tier can accurately identify active data.

RAID Reconstruct Read (3R) is a technology to increase the reliability and availability of data that is stored in DRPs. 3R is introduced in IBM Storage Virtualize V8.4.

All reads are evaluated, and if there is a mismatch, the data is reconstructed by using the parity information. To eliminate rereading of corrupted data, the affiliate cache block is marked invalid. This process works for internal and external back-end devices.

9.4.3 Planning for DRP

Before configuring and using DRPs in production environments, you must plan the capacity and performance. DRP has different performance characteristics than standard pools, so existing sizing models cannot be used directly without modifications.

For more information about how to estimate the capacity savings that are achieved by compression and deduplication, see 9.5, “Saving estimations for compression and deduplication” on page 707.

The following software and hardware requirements must be met for DRP compression and deduplication:

- The system is running Version 8.1.3.2 or higher.
- All supported platforms need at least 32 GB of cache.

Consider the following points:

- Nodes that have more than 64 GB memory can use a bigger deduplication fingerprint database, which might lead to better deduplication.
- IBM FlashSystem 5015 is not supported.
- Avoid Global Mirror with Change Volumes to or from a deduplicated volume.

For a list of recommended firmware versions, see IBM Storage Virtualize Family of Products Upgrade Planning.

Note: DRPs can be created on FlashSystem 5015, but compression and deduplication are not supported. By enabling DRP on FlashSystem 5015, SCSI UNMAP is available and thin provisioned volumes can be created.

In most cases, it is a best practice to enable compression for all thin-provisioned and deduplicated volumes. Overhead in DRPs is caused by metadata handling, which is the same for compressed volumes and thin-provisioned volumes without compression.

In the IBM FlashSystem 5035 system, the limitation in CPU power and the lack of a hardware accelerator might lead to a performance impact.

If the system contains self-compressing drives, DRPs provide a major benefit only if deduplication is used and the estimated deduplication savings are significant. If you are not planning to use deduplication or the expected deduplication ratio is low, consider the use of fully allocated volumes instead and use drive compression for capacity savings.

For more information about how to estimate deduplication savings, see 9.5.2, “Evaluating compression and deduplication” on page 709.

In systems with self-compressing drives, certain system configurations make determining accurate physical capacity on the system difficult. If the system contains self-compressing drives and DRPs with thin-provisioned volumes without compression, the system cannot determine the accurate amount of physical capacity that is used on the system. In this case,

overcommitting and losing access to write operations is possible. To prevent this situation from occurring, use compressed volumes (with or without deduplication) or fully allocated volumes. Separate compressed volumes and fully allocated volumes by using separate pools.

Similar considerations apply to configurations with compressing back-end storage controllers, as described in 9.6, “Overprovisioning and data reduction on external storage” on page 709.

A maximum number of four DRPs can be used in a system. When this limit is reached, only standard pools can be created.

A DRP uses a customer data volume per I/O group to store volume data. There is a limit on the maximum size of a customer data volume of 524,288 extents per I/O group, which places a limit on the maximum physical capacity in a pool after data reduction that depends on the extent size, number of DRPs, and number of I/O groups, as listed in Table 9-2. DRPs have a minimum extent size of 1024 MB.

Table 9-2 Maximum physical capacity after data reduction

Extent size	1 DRP - 1 I/O group	1 DRP - 4 I/O groups	4 DRP - 4 I/O groups
1024 MB	128 TiB	512 TiB	2 PiB
2048 MB	256 TiB	1 PiB	4 PiB
4096 MB	512 TiB	2 PiB	8 PiB
8192 MB	1 PiB	4 PiB	16 PiB

Overwriting data, unmapping data, and deleting volumes cause reclaimable capacity in the pool to increase. Garbage collection is performed in the background to convert reclaimable capacity to available capacity. This operation requires free capacity in the pool to operate efficiently without impacting I/O performance. A best practice is to ensure that the provisioned capacity with the DRP does not exceed 85% of the total usable capacity of the DRP.

To ensure that garbage collection is working properly, there is a minimum capacity limit in a single DRP depending on extent size and number of I/O groups, as listed in Table 9-3. Even when no volumes are in the pool, some of the space is used to store metadata. The required metadata capacity depends on the total capacity of the storage pool and on the extent size, which should be considered when planning capacity.

Table 9-3 Minimum capacity in a single DRP

Extent size	1 I/O group	4 I/O groups
1024 MB	255 GiB	1 TiB
2048 MB	0.5 TiB	2 TiB
4096 MB	1 TiB	4 TiB
8192 MB	2 TiB	8 TiB

Note: The default extent size in a DRP is 4 GB. If the estimated total capacity in the pool exceeds the documented limits, choose a larger extent size. If the estimated total capacity is relatively small, you can use a smaller extent size for a smaller metadata impact and lower minimum capacity limit.

For more information about the considerations of using data reduction on the system and the back-end storage, see 9.6, “Overprovisioning and data reduction on external storage” on page 709.

9.4.4 Implementing DRP with compression and deduplication

To use all data reduction technologies on the system, you must create a DRP, create volumes within the DRP, and map these volumes to hosts that support SCSI UNMAP commands. The implementation process for DRP is like standard pools, but has its own specifics.

Creating pools and volumes

To create a DRP, select **Pools** → **Pools**, and select **Data Reduction** in the Create Pool dialog. For more information about how to create a storage pool and populate it with MDisks, see Chapter 5, “Using storage pools” on page 365.

To create a volume within a DRP, select **Volumes** → **Volumes** and then click **Create Volumes**.

Figure 9-21 shows the Create Volumes dialog. In the **Capacity Savings** menu, the following selections are available: **None**, **thin provisioned**, and **Compressed**. If **Compressed** or **thin provisioned** is selected, the **Deduplicated** option also becomes available and can be selected.

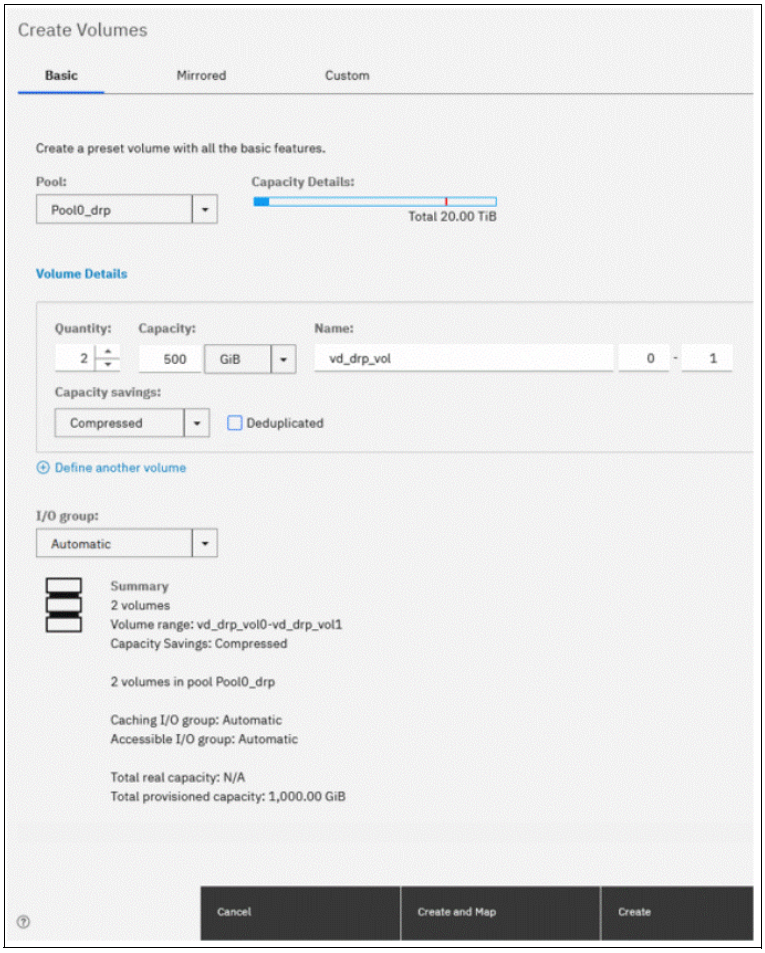


Figure 9-21 Creating compressed volumes

Capacity monitoring

Capacity monitoring in DRPs is mainly done on the system and storage pool levels. Use the Dashboard in the GUI to view a summary of the capacity usage and capacity savings of the entire system.

The Pools page in the management GUI is used for reporting on the storage pool level and displays Usable Capacity and Capacity Details. Usable Capacity indicates the amount of capacity that is available for storing data on a pool after formatting and RAID techniques are applied. Capacity Details is the capacity that is available for volumes before any capacity savings methods are applied.

To monitor this capacity, select **Pools** → **Pools**, as shown in Figure 9-22.

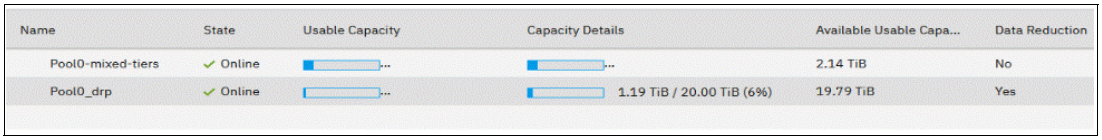


Figure 9-22 DRP capacity overview

To see more detailed capacity reporting including the warning threshold and capacity savings, open the pool properties dialog by right-clicking a pool and selecting **Properties**. This dialog shows the savings that are achieved by thin provisioning, compression, and deduplication, and the total data reduction savings in the pool, as shown in Figure 9-23.

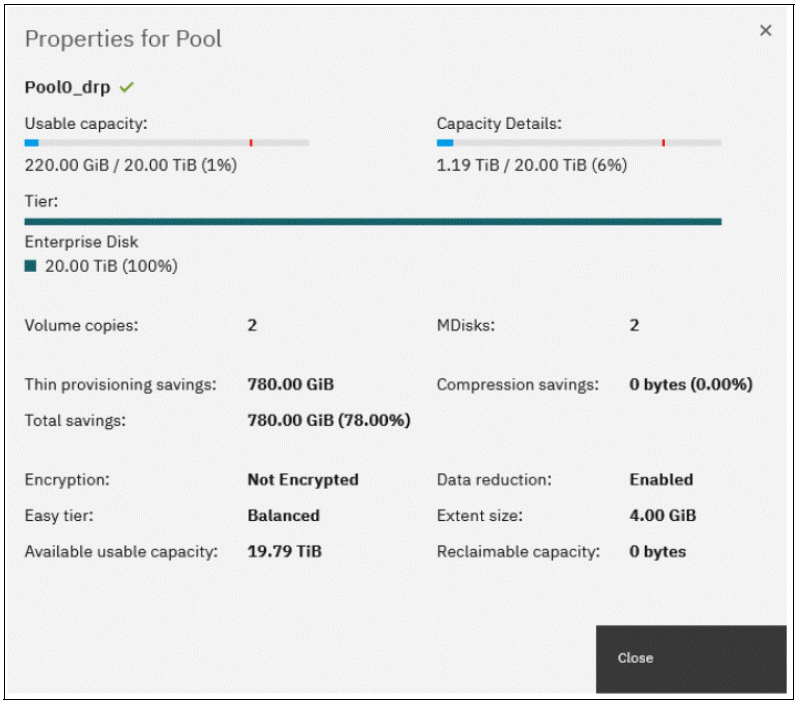


Figure 9-23 Capacity reporting in a DRP

In addition, the Reclaimable capacity is shown, which is unused capacity that is created when data is overwritten, volumes are deleted, or when data is marked as unneeded by a host by using the SCSI **UNMAP** command. This capacity is converted to available capacity by the garbage collection background process.

The capacity reporting shows 1.19 TiB capacity usage. The DRP reserves space for deduplication and compression. In this case, the DRP reserves approximately 1 TiB before

creating any volumes in the DRP. Some extents are marked used, and only some bytes are written.

With increasing usage of the pool, the effect of this reservation decreases. Creating two compressed volumes of each 500 GiB initially takes only another 0.19 TiB from the storage pool.

Thin-provisioned, compressed, and deduplicated volumes do not provide detailed per-volume capacity reporting, as shown in the Volumes by Pool window in Figure 9-24. Only the Capacity (which is the provisioned capacity that is available to hosts) is shown.

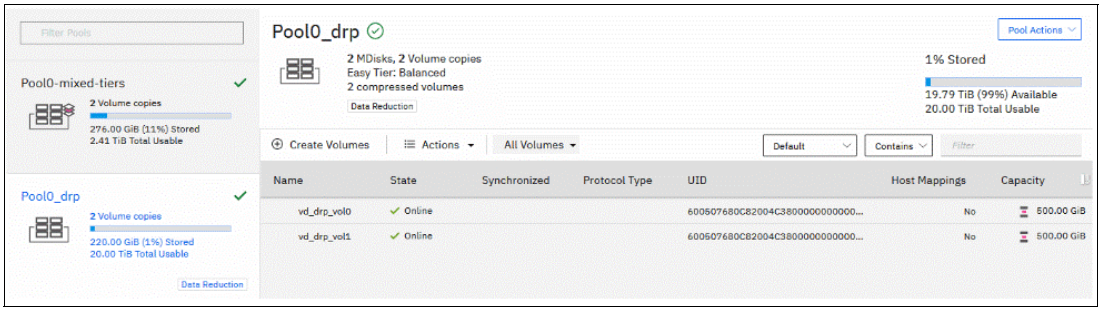


Figure 9-24 Volumes in a DRP

Real capacity, Used capacity, and Compression savings are not applicable for volumes with capacity savings. Only fully allocated volumes display those parameters.

Per-volume compression savings are not visible directly, but they can be accurately estimated by using the IBM Comprestimator, which is described in 9.5.1, “Evaluating compression savings by using IBM Comprestimator” on page 707. The IBM Comprestimator can be used on compressed volumes to analyze the volume level compression savings.

The CLI can be used for limited capacity reporting on the volume level. The `used_capacity_before_reduction` entry indicates the total amount of data that is written to a thin-provisioned or compressed volume copy in a data reduction storage pool before data reduction occurs. This field is empty for fully allocated volume copies and volume copies not in a DRP.

To find this value, run the **`lsvdisk`** command with a volume name or ID as a parameter, as shown in Example 9-9. It shows a thin-provisioned volume without compression and deduplication with a virtual size of 1 TiB that is provisioned to the host. A 53 GB file was written from the host.

Example 9-9 Data Reduction Pool volume capacity reporting on the CLI

```
IBM FlashSystem 7200:ITSOFS7K:superuser>lsvdisk thin_provisioned
id 34
name vdisk1

capacity 1.00TB

used_capacity
real_capacity
free_capacity

tier tier_scm
tier_capacity 0.00MB
tier tier0_flash
```

```
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 0.00MB

compressed_copy no
uncompressed_used_capacity
deduplicated_copy no
used_capacity_before_reduction 53.04GB
```

The used, real, and free capacity, and the capacity that is stored on each storage tier, is not shown for volumes (except fully allocated volumes) in DRPs.

Capacity reporting on the pool level is available by running the **lsmdiskgrp** command with the pool ID or name as a parameter, as shown in Example 9-10.

Example 9-10 Data Reduction Pool capacity reporting on the CLI

```
IBM FlashSystem 7200:ITSOFS7K:superuser>lsmdiskgrp 1 | grep -E
"capacity|compression|tier tier"
capacity 5.00TB
free_capacity 2.87TB
virtual_capacity 4.00TB
used_capacity 1.14TB
real_capacity 1.14TB
tier tier_scm
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier0_flash
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier1_flash
tier_capacity 5.00TB
tier_free_capacity 3.85TB
tier tier_enterprise
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier_nearline
tier_capacity 0.00MB
tier_free_capacity 0.00MB
compression_active no
compression_virtual_capacity 0.00MB
compression_compressed_capacity 0.00MB
compression_uncompressed_capacity 0.00MB
child_mdisk_grp_capacity 0.00MB
used_capacity_before_reduction 143.68GB
used_capacity_after_reduction 94.64GB
overhead_capacity 52.00GB
deduplication_capacity_saving 36.20GB
reclaimable_capacity 0.00MB
physical_capacity 5.00TB
physical_free_capacity 3.85TB
```

Compression-related properties are not valid for DRPs.

For more information about every reported value, see IBM FlashSystem 9200 Documentation and expand **Command line interface** → **Storage pool commands** → **lsmdiskgrp**.

Migrating to and from a DRP

Data migration from or to a DRP is done by using volume mirroring. A second copy in the target pool is added to the source volume, and the original copy is optionally removed after the synchronization process completes. If the volume already has two copies, one of the copies must be removed or a more complex migration scheme that uses FlashCopy, RC, host mirroring, or similar must be used.

To migrate in one step when scenario allows, complete the following steps:

1. To create a second copy, right-click the source volume and choose **Add Volume Copy**, as shown in Figure 9-25 on page 705. Choose the target pool of the migration for the second copy and select the capacity savings.
2. After you click **Add**, synchronization starts. The time that synchronization takes to complete depends on the size of the volume, system performance, and the configured migration rate. You can increase the synchronization rate by right-clicking the volume and selecting **Modify Mirror Sync Rate**.

   When both copies are synchronized, `Yes` is displayed for both copies in the Synchronized column in the Volumes window. You can track the synchronization process by using the Running tasks window, as shown in Figure 9-26 on page 705. After it reaches 100% and the copies are in-sync, you can complete migration by deleting the source copy.

Note: You can increase the synchronization rate by right-clicking the volume and selecting **Modify Mirror Sync Rate**.

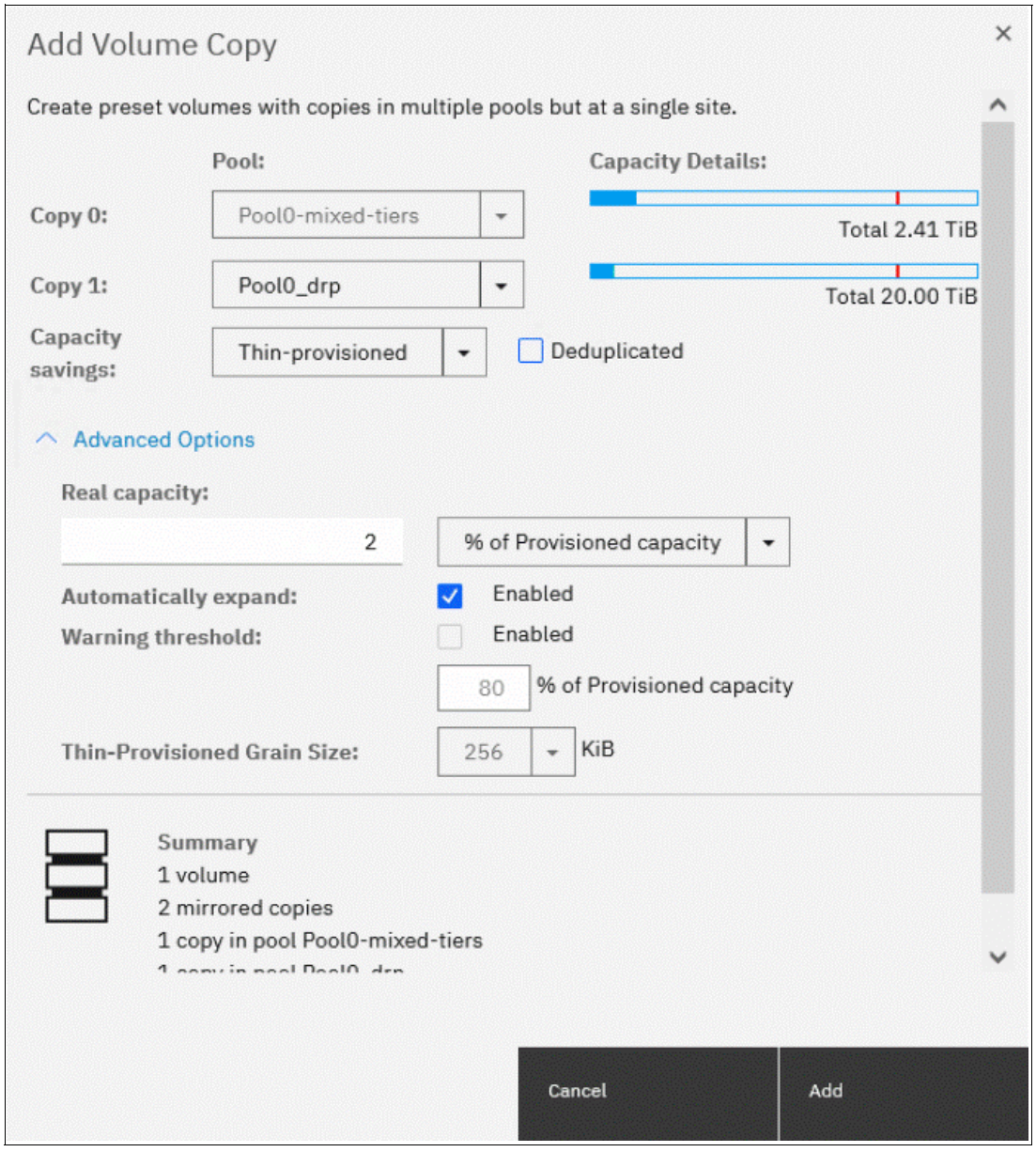


Figure 9-25 Add Volume Copy dialog

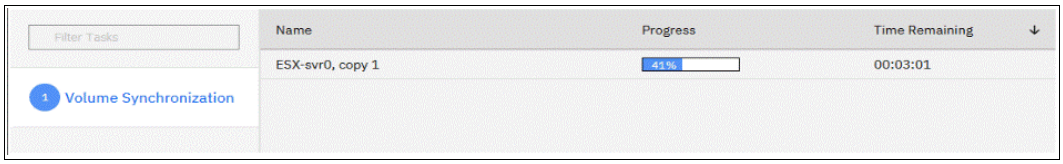


Figure 9-26 Synchronization progress

Two-step migration process

Some volume types in DRPs cannot coexist in the same I/O group as compressed volumes in standard pools. The following two-step migration process is used to work around these limitations:

1. Add a second copy to each source volume in the target pool (as described in the one-step migration process) but select capacity savings options that can coexist with the volumes in the current system configuration. For example, do not enable deduplication for the new volume copy if compressed volumes in standard pools exist in the same I/O group.

   Wait for synchronization to complete and then delete the source copy.

   Complete this process for all volumes that must be migrated and then verify that no volumes that cannot coexist with the requested volume type are left in the same I/O group. For example, verify that no more compressed volumes in standard pools are left in the same I/O group when migrating to deduplicated volumes in a DRP.
2. For all of the volumes that are handled in the previous step, add a second copy in the target pool, but select the requested capacity savings options. Wait for synchronization to complete and then delete the source copy to complete migration.

Alternatively, right-click one of the volumes, click **Modify Capacity Savings**, and select the wanted options. A second copy is created and the source copy is deleted automatically after synchronization completes.

Garbage collection and volume deletion

DRP includes built-in capabilities to enable garbage collection of unused storage capacity. *Garbage collection* is a DRP process that reduces the amount of data that is stored on external storage systems and internal drives by reclaiming previously used storage resources that are no longer needed by host systems.

When a DRP is created, the system monitors the pool for reclaimable capacity from host **UNMAP** operations. When space is freed from a host operating system, it is a process called *unmapping*. Hosts indicate that the allocated capacity is no longer required on a target volume. The freed space is collected and reused by the system automatically without having to reallocate the capacity manually.

Removing thin-provisioned or compressed volume copies in a DRP is an asynchronous operation. Volume copies that are removed from the system enter the *deleting* state, during which the used capacity of the copies is converted to reclaimable capacity in the pool by using a background deletion process. The removal process of deduplicated volume copies searches and moves deduplication references that other volumes might have to the deleting volume copies. This task is done to ensure that deduplicated data that was on deleted copies continues to be available for other volumes in the system.

After this process completes, the volume copies are deleted and disappear from the system configuration. In a second step, garbage collection can give the reclaimable capacity that is generated in the first step back to the pool as available capacity, which means that the used capacity of a removed volume is not available for reuse immediately after the removal.

The time that it takes to delete a thin-provisioned or compressed volume copy depends on the size of the volume, the system configuration, and the workload. For deduplicated copies, the duration also depends on the amount and size of other deduplicated copies in the pool, which means that it might take a long time to delete a small deduplicated copy if there are many other deduplicated volumes in the same pool. The deletion process is a background process that might impact system overall performance.

The deleting state of a volume or volume copy can be seen by running the **lsvdisk** command. The GUI hides volumes in this state, but it shows deleting volume copies if the volume contains another copy.

Note: Removing thin-provisioned or compressed volume copies in a DRP might take a long time to complete. Used capacity is not immediately given back to the pool as available capacity.

When one copy of a mirrored volume is in the deleting state, it is not possible to add a copy to the volume before the deletion finishes. If a new copy must be added without waiting for the deletion to complete, first split the copy that must be deleted into a new volume, and then delete the new volume and add a new second copy to the original volume. To split a copy into a new volume, right-click the copy and select **Split into New Volume** in the GUI or run the **splitvdiskcopy** command on the CLI.

9.5 Saving estimations for compression and deduplication

This section provides information about the tools that are used for sizing the environment for compression and deduplication.

9.5.1 Evaluating compression savings by using IBM Comprestimator

IBM Comprestimator is a utility that estimates the capacity savings that can be achieved when compression is used for storage volumes. The utility is integrated into the system by using the GUI and the CLI. It can also be installed and used on host systems.

Starting with IBM Storage Virtualize V8.4, the integrated Comprestimator is always enabled and running continuously, thus providing up-to-date compression estimation over the entire cluster, both in GUI and IBM Storage Insights.

IBM Comprestimator provides a quick and accurate estimation of compression and thin-provisioning benefits. The utility performs read-only operations, so it does not affect the data that is stored on the volume.

If the compression savings prove to be beneficial in your environment, volume mirroring can be used to convert volumes to compressed volumes.

To see the results and the date of the latest estimation cycle (see Figure 9-27 on page 708), go to the **Volumes** window, right-click any volume, and select **Space Savings** → **Estimate Compression Savings**. If no analysis was done, the system suggests running it. A new estimation of all volumes can be started from this dialog. To run or rerun analysis on a single volume, select **Analyze** in the **Space Savings** submenu.

To analyze all the volumes on the system from the CLI, run the **analyzevdiskbysystem** command.

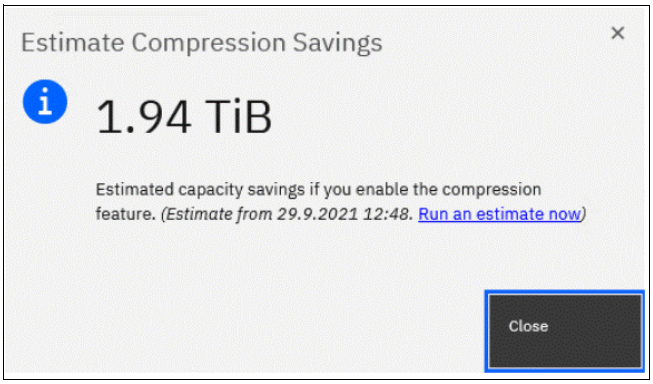


Figure 9-27 Estimate Compression Savings

The command analyzes all the current volumes that are created on the system. Volumes that are created during or after the analysis are not included and can be analyzed individually. The time that it takes to analyze all the volumes on the system depends on the number of volumes that are being analyzed, and results can be expected at about a minute per volume. For example, if a system has 50 volumes, compression savings analysis takes approximately 50 minutes.

You can run an analysis on a single volume by specifying its name or ID as a parameter for the **`analyzevdisk`** CLI command.

To check the progress of the analysis, run the **`lsvdiskanalysisprogress`** command. This command displays the total number of volumes on the system, total number of volumes that are remaining to be analyzed, and the estimated time of completion.

To display information for the thin provisioning and compression estimation analysis report for all volumes, run the **`lsvdiskanalysis`** command.

If you use a version of IBM Storage Virtualize that is older than Version 7.6 or if you want to estimate the compression savings of another IBM or non-IBM storage system, the separate IBM Compretimator Utility can be installed on a host that is connected to the device that needs to be analyzed. For more information and the latest version of this utility, see IBM FlashSystem Comprestimator.

Consider the following best practices for using IBM Comprestimator:

- Run the IBM Comprestimator Utility before implementing an IBM Storage Virtualize solution and DRPs.
- Download the latest version of the IBM Comprestimator Utility if you are not using one that is included in your IBM Storage Virtualize solution.
- Use IBM Comprestimator to analyze volumes that contain as much active data as possible rather than volumes that are nearly empty or newly created to ensure more accuracy when sizing your environment for compression and DRPs.

Note: IBM Comprestimator can run for a long period (a few hours) when it is scanning a relatively empty device. The utility randomly selects and reads 256 KB samples from the device. If the sample is empty (that is, full of null values), it is skipped. A minimum number of samples with data is required to provide an accurate estimation. When a device is mostly empty, many random samples are empty. As a result, the utility runs for a longer time as it tries to gather enough non-empty samples that are required for an accurate estimate. The scan is stopped if the number of empty samples is over 95%.

9.5.2 Evaluating compression and deduplication

To help with the profiling and analysis of user workloads that are to be migrated to the new system, IBM provides a highly accurate Data Reduction Estimation Tool (DRET) that supports both deduplication and compression. The tool operates by scanning target workloads on any established array (from IBM or a third party) and then merging all scan results to provide an integrated system-level data reduction estimate. It provides a report of what it expects the deduplication and compression savings to be from data that is written to a disk.

The DRET utility uses advanced mathematical and statistical algorithms to perform an analysis with a low memory footprint. The utility runs on a host that has access to the devices to be analyzed. It performs only read operations, so it has no effect on the data that is stored on the device.

The following sections provide information about installing DRET on a host and using it to analyze devices on it. Depending on the environment configuration, in many cases DRET is used on more than one host to analyze more data types.

When DRET is used to analyze a block device that is used by a file system, all underlying data in the device is analyzed regardless of whether this data belongs to files that were already deleted from the file system. For example, you can fill a 100 GB file system and make it 100% used, and then delete all the files in the file system to make it 0% used. When scanning the block device that is used for storing the file system in this example, DRET accesses the data that belongs to the files that are deleted.

Important: The preferred method of using DRET is to analyze volumes that contain as much active data as possible rather than volumes that are mostly empty of data, which increases the accuracy level and reduces the risk of analyzing old data that is deleted but might still have traces on the device.

For more information and the latest version of this utility, see the Data Reduction Estimator Tool web page.

9.6 Overprovisioning and data reduction on external storage

Starting with IBM Storage Virtualize V8.1.x, overprovisioning on selected back-end controllers is supported, which means that if back-end storage performs data deduplication or data compression on LUs that are provisioned from it, they still can be used as external MDisks on the system. However, more configuration and monitoring considerations must be accounted for.

Overprovisioned MDisks from controllers that are supported by this feature can be used as managed mode MDisks in the system and can be added to storage pools (including DRPs).

Implementation steps for overprovisioned MDisks are the same as for fully allocated storage controllers. The system detects whether the MDisk is overprovisioned, its total physical capacity, and used and remaining physical capacity. It detects whether SCSI UNMAP commands are supported by the back end. By sending SCSI UNMAP commands to overprovisioned MDisks, the system marks data that is no longer in use. then the garbage collection processes on the back end can free unused capacity and convert it to free space.

At the time of this writing, the following back-end controllers are supported by overprovisioned MDisks:

- IBM FlashSystem A9000 V12.1.0 and later
- IBM FlashSystem 900 V1.4 and later
- IBM FlashSystem V9000 AE2 and AE3 expansions
- IBM Storwize or IBM FlashSystem family systems Version 8.1.0 and later
- IBM PureSystems storage
- HPE Nimble

Extra caution is required when planning and monitoring capacity for such configurations. Table 9-4 lists configuration guidelines when overprovisioned external storage controllers are used.

Table 9-4 Using data reduction at two levels

System	Back end	Comments
DRP	Fully allocated	Recommended. Use DRP on the system to plan for compression and deduplication. DRP at the top level provides the best application capacity reporting.
Fully allocated	Overprovisioned, single tier of storage	Recommended with appropriate precautions. Track physical capacity use carefully to avoid out-of-space conditions. The system can report physical use but does not manage to avoid out-of-space conditions. There is no visibility of each application's use at the system layer. If the back end runs out of space, there is a limited ability to recover. Consider creating a sacrificial emergency space volume.
DRP with compression	Overprovisioned	Recommended with appropriate precautions. Assume 1:1 compression in back-end storage and do not overcommit capacity in the back end. Small extra savings are achieved from compressing DRP metadata.
Fully allocated	Overprovisioned, multiple tiers of storage	Use with great care. Easy Tier is unaware of physical capacity in tiers of a hybrid pool, so it tends to fill the top tier with the hottest data. Changes in the compressibility of data in the top tier can overcommit the storage, which leads to out-of-space conditions.
DRP with thin-provisioned or fully allocated volumes	Overprovisioned	Avoid. Difficult to understand the physical capacity usage of the uncompressed volumes. High risk of overcommitting the back end. If a mix of DRP and fully allocated volumes is required, use separate pools.
DRP	DRP	Avoid. Creates two levels of I/O amplification and capacity impact. DRP at the bottom layer provides no benefit.

When DRPs are used with a compressing back-end controller, use compression in DRP and avoid overcommitting the back end by assuming a 1:1 compression ratio in back-end storage. Small extra savings are realized from compressing metadata.

If the back-end controller uses FCM drives that are always compressing with hardware acceleration, the same methodology should be used. The eventual capacity savings should be used by creating more MDisks to be implemented in the DRP.

Note: Fully allocated volumes that are above overprovisioned MDisks configurations must be avoided or used with extreme caution because it can lead to overcommitting back-end storage.

The concept of provisioning groups is used for capacity reporting and monitoring of overprovisioned external storage controllers. A provisioning group is an object that represents a set of MDisks that share physical resources. Each overprovisioned MDisk is part of a provisioning group that defines the physical storage resources that are available to a set of MDisks.

Storage controllers report the usable capacity of an overprovisioned MDisk based on its provisioning group. If multiple MDisks are part of the same provisioning group, then all these MDisks share the physical storage resources and report the same usable capacity. However, this usable capacity is not available to each MDisk individually because it is shared between all these MDisks.

Provisioning groups are used differently depending on the back-end storage, as shown in the following examples:

- IBM FlashSystem A9000 and IBM FlashSystem 900: The entire subsystem forms one provisioning group.
- Storwize and IBM FlashSystem family systems: The storage pool forms a provisioning group, which enables more than one independent provisioning group in a system.
- RAID with compressing drives: An array is a provisioning group that presents the physical storage that is in use much like an external array.

Capacity usage should be monitored primarily on the overprovisioned back-end storage controller.

From the system, capacity usage can be monitored on overprovisioned MDisks by using one of the following methods:

- The GUI dashboard, as shown in Figure 9-28 on page 712.

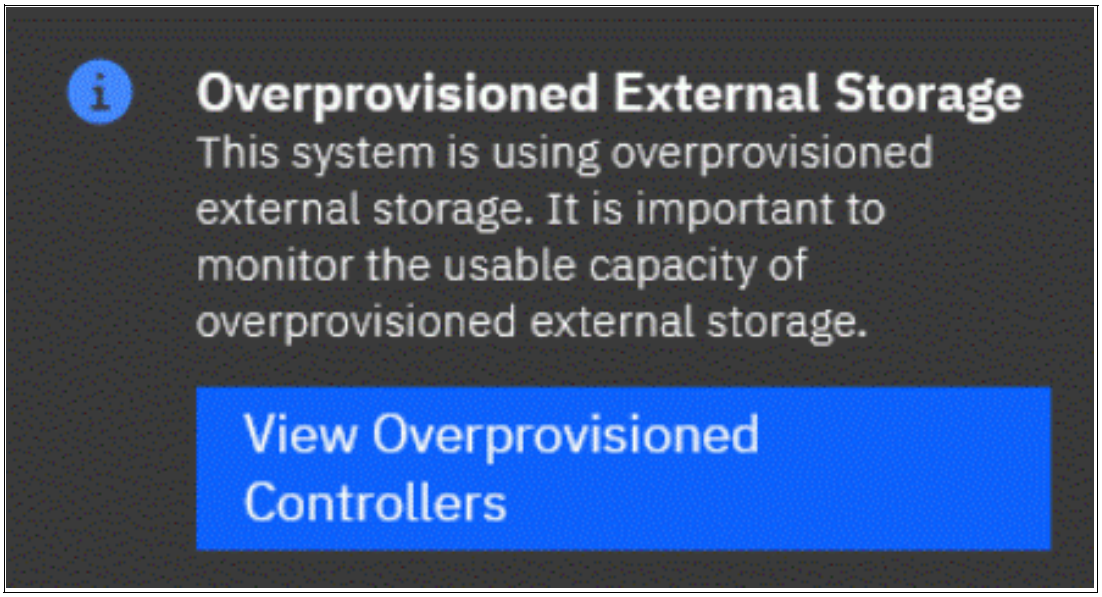


Figure 9-28 Dashboard button for overprovisioned storage monitoring

Click **View Overprovisioned Controllers** to show an overview of overprovisioned MDisks used by the system, as shown in Figure 9-29.

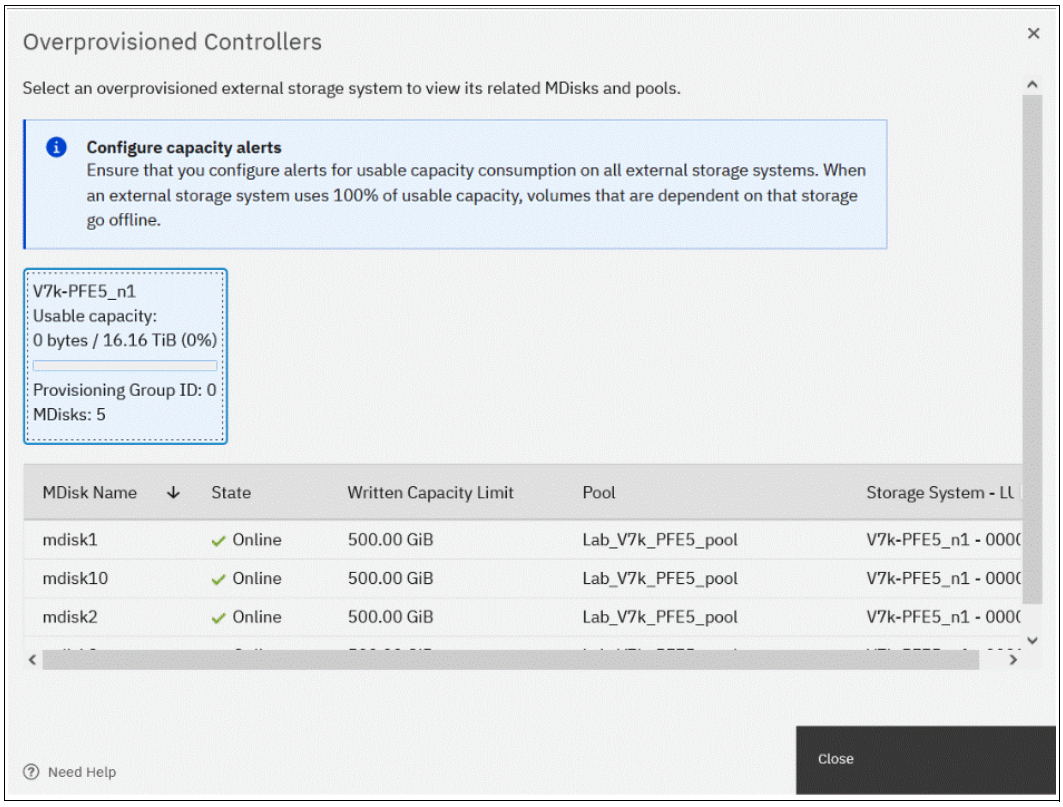


Figure 9-29 View overprovisioned controllers and mdisks

- The MDisk properties window, which opens by selecting **Pools** → **MDisks by Pools**, right-clicking an MDisk, and then selecting the **Properties** option, as shown in Figure 9-30.

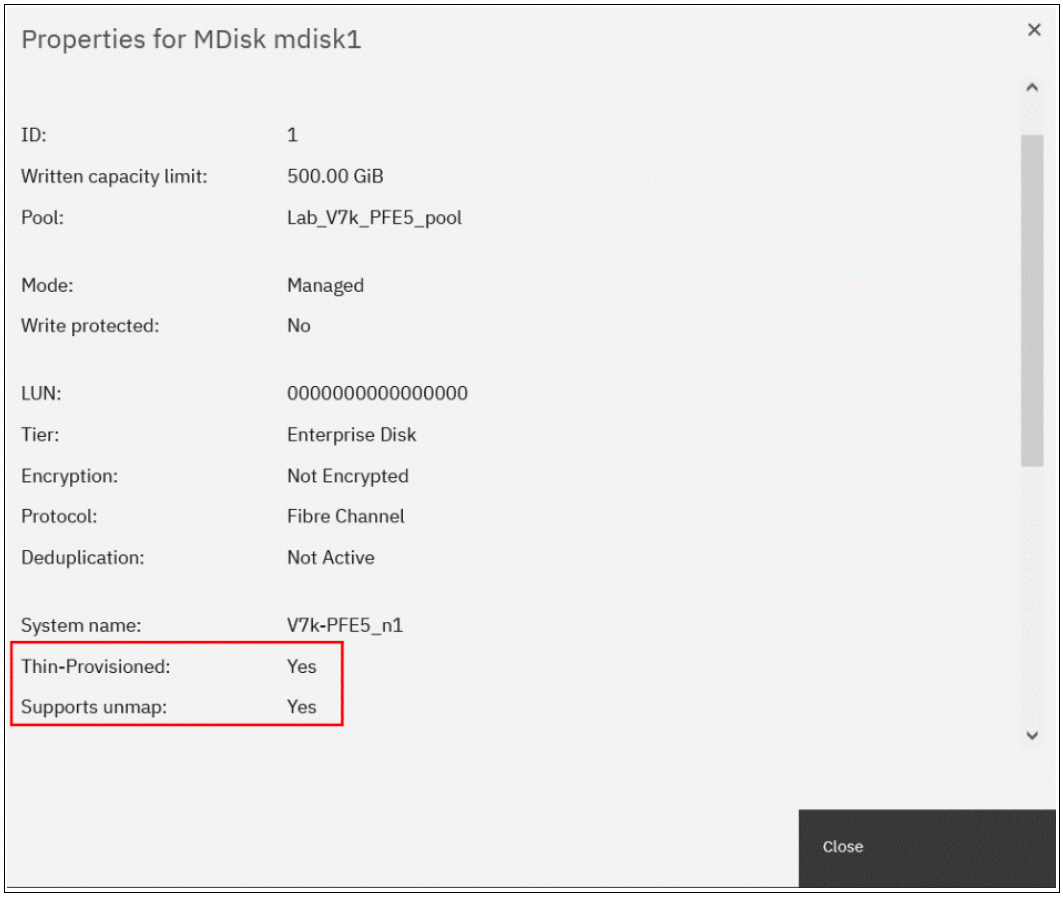


Figure 9-30 Thin-provisioned MDisk properties

- Running **lsmdisk** with an MDisk name or ID as a parameter displays the full properties of a thin-provisioned volume, as shown in Example 9-11.

Example 9-11 The lsmdisk parameters for thin-provisioned MDisks

```
IBM FlashSystem 7200:ITSOFS7K:superuser>lsmdisk mdisk2
id 2
name mdisk2
status online
mode managed
<...>
dedupe no
<...>
over_provisioned yes
supports_unmap yes
provisioning_group_id
physical_capacity 299.00GB
physical_free_capacity 288.00GB
write_protected no
allocated_capacity 11.00GB
effective_used_capacity 300.00GB
```

The overprovisioning status and SCSI UNMAP support for the selected MDisk are displayed.

The **physical_capacity** and **physical_free_capacity** parameters belong to the MDisk's provisioning group. They indicate the total physical storage capacity and formatted available physical space in the provisioning group that contains this MDisk.

Note: It is not a best practice to create multiple storage pools from MDisks in a single provisioning group.

9.7 Safeguarded child pool capability: Protection from logical data corruption

The Safeguarded child pool capability was introduced to the IBM Storage Virtualize products family (IBM FlashSystem, IBM SAN Volume Controller, and IBM Storage Virtualize for Public Cloud) in version 8.4.2.0.

It was introduced as a way to protect logical data from being corrupted by taking the periodic snapshots of the data and to store it in a non-mutable state, without access from hosts or servers, administrators, and applications. In this case, Safeguarded copies can serve as a point of recovery for the data if a logical corruption of the original data occurs. The data can be restored to a pre-corruption state from these copies.

Key components and concept of Safeguarded copies

Safeguarded copies of the volumes that are identified for protection are created and exist in a Safeguarded child pool. This Safeguarded child pool must be created in the parent pool of the to-be protected volumes.

The volume group must be created for those volumes. then the Safeguarded policy is created or selected from available options and should be assigned or associated to or with the volume group of the to-be protected volumes.

IBM Storage Virtualize creates Safeguard copies, which is the main role it plays in IBM Copy Services Manager.

IBM Copy Services Manager's role is to periodically scan the managed storage systems for the existence of volume groups that are associated with a Safeguarded policy. By performing this scan, it orchestrates the creation of the Safeguarded volume copies according to the backup interval that is specified in the Safeguard policy.

IBM Storage Virtualize keeps up with the retention policy that is specified in the Safeguarded policy and deletes the IBM Copy Services Manager-orchestrated Safeguarded Copy volumes accordingly.

Safeguarded copies can be used for recovery or restoration of the data by way of the IBM Copy Services Manager-mediated Safeguarded Copy function.

Safeguarded copies include the following components:

- Safeguarded child pool: Stores Safeguarded copies.
- Volume groups: Implements Safeguarded policies to volumes and for orchestration.
- Safeguarded policies - implements necessary policies for backup intervals and retention
- IBM Copy Services Manager integration: Orchestrates the creation of copies according to Safeguarded policies

For more information, see *IBM FlashSystem Safeguarded Copy Implementation Guide*, REDP-5654.

10

Advanced Copy Services

This chapter describes Advanced Copy Services, which is a group of functions that provide different methods of data copy. It also describes the storage software capabilities to support the interaction with hybrid clouds. These functions are enabled by IBM Storage Virtualize software.

This chapter includes the following topics:

- 10.1, “Maximum configuration updates” on page 718
- 10.2, “Safeguarded Copy” on page 718
- 10.3, “IBM FlashCopy” on page 719
- 10.4, “Managing FlashCopy by using the GUI” on page 752
- 10.5, “Transparent Cloud Tiering” on page 789
- 10.6, “Implementing Transparent Cloud Tiering” on page 792
- 10.7, “Volume mirroring and migration options” on page 804
- 10.8, “Remote Copy” on page 806
- 10.9, “Native IP replication” on page 841
- 10.10, “Managing Remote Copy by using the GUI” on page 862
- 10.11, “Remote Copy memory allocation” on page 885
- 10.12, “A real-life implementation of IP replication: Anadolu Sigorta” on page 886
- 10.13, “Troubleshooting Remote Copy” on page 888
- 10.14, “3-Site Replication” on page 891
- 10.15, “IBM Storage volume group and volume group snapshots” on page 894
- 10.16, “IBM Storage policy-based replication” on page 905
- 10.17, “Converting Global Mirror to policy-based replication” on page 938
- 10.18, “Monitoring options for policy-based replication” on page 946
- 10.19, “Troubleshooting policy-based replication” on page 955

© Copyright IBM Corp. 2024. All rights reserved.

10.1 Maximum configuration updates

Storage Virtualize 8.6.0 features several updates to the maximum configuration limits that affect FlashCopy and other Advanced Copy Services Features. Table 10-1 lists the maximum configuration limits updates.

Table 10-1 Maximum Configuration updates

	8.3.1	**8.4.0**	**8.4.2**	**8.5.4[a]**	**8.6.0[b]**
Number of VDisks	10,000	10,000	15,864[2]	15,864[2]	15,864[2]
Host-Mappable Volumes	10,000	10,000	15,864[2]	15,864[2]	15,864[2]
FlashCopy Mappings	5,000	10,000	15,864[2] 8,192[3]	15,864[2] 8,192[3]	15,864[2] 8,192[3]
Number of Asynchronous Volumes	5,000	5,000	7932[2]	7932[2]	7932[2]
Maximum Asynchronous Capacity	720 TiB	720 TiB	2 PiB[1]	2 PiB[2]	2 PiB[2]
Number of HA Volumes	1250	2000	2000	2000	2000
Maximum HA Capacity	720 TiB	720 TiB	2 PiB[1]	2 PiB[2]	2 PiB[2]
HA + 3-Site Volumes	1250	2000	2000	2000	2000

a. IBM Storage Virtualize 8.5 does not support IBM Storwize V5000 and V5100
b. IBM Storage Virtualize 8.6 does not support IBM Storwize V7000

1. Only applies to FS9200, FS7200, SA2, and SV2.
2. Only applies to V7000, FS7200, FS9100, FS9200, FS9500, and IBM SAN Volume Controller.
3. Only applies to FS5x00.

10.2 Safeguarded Copy

Traditional disaster recovery is designed around recovering from or preventing data loss. Data loss can occur from catastrophic or accidental events, and strategies to prevent data loss include the following examples:

- Backup or archive of data to various media including disk and tape.
- Off-site transport of backup media (or directly backing up data off-site).
- Replication of data to another site. This approach ensures that data remains active and can be accomplished by the storage layer, a dedicated application, or middleware.
- Various methods for continuous operation, including the HyperSwap feature.

All of these approaches are designed to prevent data loss or enable recovery if data loss occurs. However, they do not increase cyber resiliency, which refers to the ability of an organization to deliver a service continuously, despite any adverse cyber events. This concept brings together information security, organizational resilience, and business continuity.

Cyber events include the following examples:

- Cyberattacks:
  - Phishing, web attacks, and denial of service

  - Malware, such as ransomware or Trojans
  - Man-in-the-middle (eavesdropping)
  - Insider threats
  - Password (brute force) hacks
- Data breaches:
  - User error
  - Physical theft or intrusion
  - Unauthorized access

These threats make it apparent that securing information and systems is more than simply configuring firewalls or enforcing secure passwords. Threats now include mobile devices or even trusted employees going rogue. Also, many of these attacks are frequently carried out by state or state-sponsored threat actors.

The IBM Safeguarded Copy feature in IBM Storage Virtualize 8.4.2 and above, prevents point-in-time data from being deleted or changed. It prevents user error, malicious destruction, or ransomware attacks from modifying this data.

Safeguarded Copy provides the following benefits:

- Separation of Duties means that privileged users cannot compromise production data. Bad local actors cannot delete or corrupt copies of data.
- Automation capabilities to create regular immutable copies of data. These copies cannot be mounted, read from, or written to without the use of a separate function of software outside the storage administrator's domain of control.
- Data reduced (thin) copies to limit the growth requirements of the immutable copies.
- Restore data instantly to any point in time before any attack.

For more information, including supported storage systems and implementation, see *Implementation Guide for FlashSystem Safeguarded Copy*, REDP-5654.

10.3 IBM FlashCopy

Through the IBM FlashCopy function of IBM Storage Virtualize, you can perform a point-in-time (PiT) copy of one or more volumes. This section describes the inner workings of FlashCopy and provides more information about its configuration and use.

You can use FlashCopy to help you solve critical and challenging business needs that require duplication of data of your source volume. Volumes can remain online and active while you create consistent copies of the data sets. Because the copy is performed at the block level, it operates below the host operating system and its cache. Therefore, the copy is not apparent to the host unless it is mapped.

While the FlashCopy operation is performed, the source volume is frozen briefly to initialize the FlashCopy bitmap after which I/O can resume. Several FlashCopy options require the data to be copied from the source to the target in the background, which can take time to complete. However, the resulting data on the target volume is presented so that the copy appears to complete immediately. This feature means that the copy can immediately be mapped to a host and is directly accessible for read and write operations.

10.3.1 Business requirements for FlashCopy

When you are deciding whether FlashCopy addresses your needs, you must adopt a combined business and technical view of the problems that you want to solve. First, determine the needs from a business perspective. Then determine whether FlashCopy can address the technical needs of those business requirements.

The business applications for FlashCopy are wide-ranging. Common use cases for FlashCopy include, but are not limited to, the following examples of rapidly creating:

- Consistent backups of dynamically changing data
- Consistent copies of production data to facilitate data movement or migration between hosts
- Copies of production data sets for application development and testing, auditing, data mining, and for quality assurance purposes

Regardless of your business needs, FlashCopy within IBM Storage Virtualize is flexible and offers a broad feature set, which makes it applicable to numerous scenarios.

Back up improvements with FlashCopy

FlashCopy does not reduce the time that it takes to perform a backup to traditional backup infrastructure. However, it can be used to minimize and under certain conditions, eliminate application downtime that is associated with performing backups. FlashCopy can also transfer the resource usage of performing intensive backups from production systems.

After the FlashCopy is performed, the resulting image of the data can be backed up to tape, as though it were the source system. After the copy to tape is completed, the image data is redundant, and the target volumes can be discarded. For time-limited applications, no copy" or incremental FlashCopy is used most often. The use of these methods puts less load on your servers infrastructure.

When FlashCopy is used for backup purposes, the target data is often managed as read-only at the operating system level. This approach provides extra security by ensuring that your target data was not modified and remains true to the source.

Restore with FlashCopy

FlashCopy can perform a restore from any FlashCopy mapping. Therefore, you can restore (or copy) from the target to the source of your regular FlashCopy relationships. When restoring data from FlashCopy, this method can be qualified as reversing the direction of the FlashCopy mappings.

This capability features the following benefits:

- FlashCopy creation mistakes are not a concern. You trigger a restore.
- The process appears instantaneous.
- You can maintain a pristine image of your data while you are restoring what was the primary data.

This approach can be used for various applications, such as recovering your production database application after an errant batch process that caused extensive damage.

Preferred practices: Although restoring from a FlashCopy is quicker than a traditional tape media restore, you must not use restoring from a FlashCopy as a substitute for good backup and archiving practices. Instead, keep one to several iterations of your FlashCopies so that you can near-instantly recover your data from the most recent history, and keep your long-term backup and archive as suitable for your business.

In addition to the restore option that copies the original blocks from the target volume to modified blocks on the source volume, the target can be used to perform a restore of individual files. To restore files, you make the target available on a different host because the use of the original host can cause confusion for the host operating system if the same disk is seen twice. You can then copy the files to the source by using normal host data copy methods for your environment.

For more information about how to use reverse FlashCopy, see 10.3.13, “Reverse FlashCopy” on page 743.

Moving and migrating data with FlashCopy

FlashCopy can be used to facilitate the movement or migration of data between hosts while minimizing downtime for applications. By using FlashCopy, application data can be copied from source volumes to new target volumes while applications remain online.

After the volumes are fully copied and synchronized, the application can be brought down and a final incremental synchronization that is performed to capture any updates. The application can then be immediately brought back up on the new server that is accessing the new FlashCopy target volumes.

This method differs from the other migration methods, which are described later in this chapter. Common uses for this capability are host and back-end storage hardware refreshes.

Application testing with FlashCopy

It is often important to test a new version of an application or operating system by using a copy of the production data. This test ensures the highest quality possible for your environment. FlashCopy makes this type of testing easy to accomplish without putting the production data at risk.

Note: Although a FlashCopy can be initiated while the application or server is operational, it is considered to be a *crash consistent* copy. Therefore, it is preferable to initiate a FlashCopy when the application is quiesced or in read-only mode (during backup) to achieve a consistent copy of the data.

You can create a FlashCopy of your source volumes and use that for your testing. This copy is a duplicate of your production data down to the block level so that even physical disk identifiers are copied. Therefore, it is impossible for your applications to tell the difference between the original source volumes or the FlashCopy volumes.

You can also use the FlashCopy feature to create restart points for long-running batch jobs. This option means that if a batch job fails several days into its run, it might be possible to restart the job from a saved copy of its data rather than rerunning the entire multiday job.

However, because all systems and applications require patching to keep them up to date with security and bug fixes, it is a good idea to see what effect those updates have on an operational environment before committing them.

To reduce risk and accelerate roll-out, FlashCopy can be used to create clones upon which the patches can be applied to first understand their impact. This process also enables quick roll-back onto non-patched volumes if the updates negatively affect the operating environment.

10.3.2 FlashCopy principles and terminology

The FlashCopy function creates a point in time (PiT) or time-zero (T0) copy of data that is stored on a source volume to a target volume by using Copy on Write (CoW) and copy on-demand, or a Redirect on Write (RoW) mechanism.

The system automatically activates RoW when the following conditions are met for both the source and volume targets:

- Both are in the same data reduction pool with deduplication enabled.
- FlashCopy mappings are in the same I/O group.
- Neither are mirrored.

Unless all of these conditions are met, the FlashCopy function uses CoW to complete write operations. The RoW function uses lighter-weight metadata references than CoW, which attempts to improve (when possible) the data reduction ratio and system performance.

When a FlashCopy operation starts, a checkpoint creates a bitmap table that indicates that no part of the source volume was copied. Each bit in the bitmap table represents one region of the source volume and its corresponding region on the target volume. Each region is called a grain.

The relationship between two volumes defines the way data is copied and is called a FlashCopy mapping.

FlashCopy mappings between multiple volumes can be grouped in a Consistency group to ensure their PiT (or T0) is identical for all of them. A simple one-to-one FlashCopy mapping does not need to belong to a consistency group.

Figure 10-1 shows the basic terms that are used with FlashCopy. All elements are explained in this chapter.

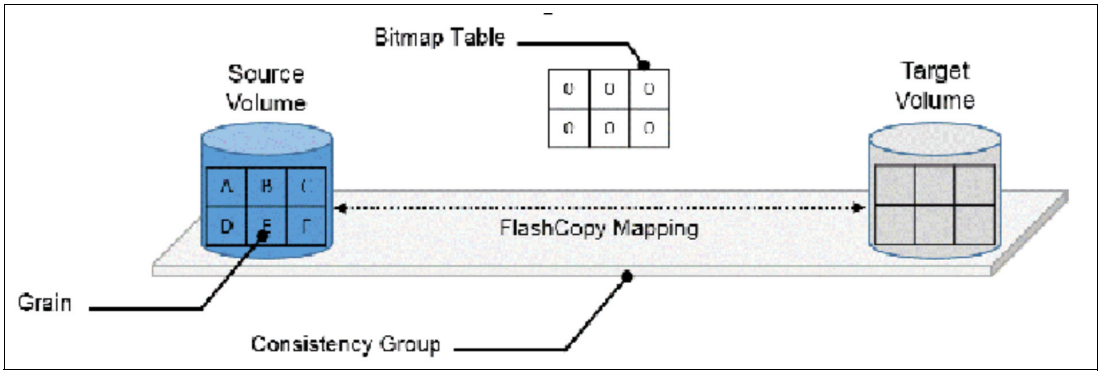


Figure 10-1 FlashCopy terminology

10.3.3 FlashCopy mapping

The relationship between the source volume and the target volume is defined by a FlashCopy mapping. The FlashCopy mapping can have three different types, four attributes, and seven different states.

The FlashCopy mapping can be one of the following types:

- Snapshot. Sometimes referred to as *nocopy*, a snapshot is a PiT copy of a volume without a background copy of the data from the source volume to the target. Only the changed blocks on the source volume are copied. The target copy cannot be used without an active link to the source.
- Clone. Sometimes referred to as *full copy*, a clone is a PiT copy of a volume with background copy of the data from the source volume to the target. All blocks from the source volume are copied to the target volume. The target copy becomes a usable independent volume.
- Backup. Sometimes referred to as *incremental*, a backup FlashCopy mapping consists of a PiT full copy of a source volume, plus periodic increments or “deltas” of data that changed between two points in time.

The FlashCopy mapping has four property attributes (clean rate, copy rate, autodelete, incremental) and seven different states that are described later in this chapter. Users can perform the following actions on a FlashCopy mapping:

- Create: Define a source and target, and set the properties of the mapping.
- Prepare: The system must be prepared before a FlashCopy copy starts. It flushes the cache and makes it “transparent” for a short time, so no data is lost.
- Start: The FlashCopy mapping is started and the copy begins immediately. The target volume is immediately accessible.
- Stop: The FlashCopy mapping is stopped (by the system or by the user). Depending on the state of the mapping, the target volume is usable or not usable.
- Modify: Some properties of the FlashCopy mapping can be modified after creation.
- Delete: Delete the FlashCopy mapping. This action does not delete volumes (source or target) from the mapping.

In addition, FlashCopy has a few other restrictions, which must be understood including:

- The source and target volumes must be the same size.
- The minimum granularity that IBM Storage Virtualize supports for FlashCopy is an entire volume. It is not possible to use FlashCopy to copy only part of a volume.

Important: As with any PiT copy technology, you are bound by operating system and application requirements for interdependent data and the restriction to an entire volume.

The source and target volumes must belong to the same IBM Storage Virtualize based system. They do not have to be in the same I/O group or storage pool, although it is recommended that they have the same preferred node for best performance.

Volumes that are members of a FlashCopy mapping cannot have their size increased or decreased while they are members of the FlashCopy mapping.

All FlashCopy operations occur on FlashCopy mappings. FlashCopy does not alter the volumes. However, multiple operations can occur at the same time on multiple FlashCopy mappings because of the use of consistency groups.

In a multi-tenanted environment, access to FlashCopy mappings can also be controlled by using Ownership groups. Ownership group contains a subset of users and objects within a system. An ownership group can be used to further restrict access to specific resources to

that which is defined within it. Only users with Security Administrator roles can configure and manage ownership groups.

10.3.4 Consistency groups

To overcome the issue of dependent writes across volumes and to create a consistent image of the client data, a FlashCopy operation must be performed on multiple volumes as an atomic operation. To accomplish this, IBM Storage Virtualize supports the concept of consistency groups.

Consistency groups address the requirement to preserve PiT data consistency across multiple volumes for applications that include related data spanning multiple volumes. For these volumes, consistency groups maintain the integrity of the FlashCopy by ensuring that “dependent writes” are run in the application’s intended sequence. Also, consistency groups provide an easy way to manage several mappings.

FlashCopy mappings can be part of a consistency group, even if only one mapping exists in the consistency group. If a FlashCopy mapping is not part of any consistency group, it is referred as stand-alone.

Dependent writes

It is crucial to use consistency groups when a data set spans multiple volumes. Consider the following typical sequence of writes for a database update transaction:

1. A write is run to update the database log, which indicates that a database update is about to be performed.
2. A second write is run to perform the update to the database.
3. A third write is run to update the database log, which indicates that the database update completed successfully.

The database ensures the correct ordering of these writes by waiting for each step to complete before the next step is started. However, if the database log (updates 1 and 3) and the database (update 2) are on separate volumes, it is possible for the FlashCopy of the database volume to occur before the FlashCopy of the database log. This sequence can result in the target volumes seeing writes 1 and 3 but not 2 because the FlashCopy of the database volume occurred before the write was completed.

In this case, if the database was restarted by using the backup that was made from the FlashCopy target volumes, the database log indicates that the transaction was successful. In fact, it was not successful because the FlashCopy of the volume with the database file was started (the bitmap was created) before the write completed to the volume. Therefore, the transaction is lost and the integrity of the database is in question.

Most of the commands that the user can perform on FlashCopy mappings are the same for consistency groups, such as **`prestartfcconsistgrp`** and **`startfcconsistgrp`**, instead of **`prestartfcmap`** and **`startfcmap`**.

10.3.5 Crash consistent copy and hosts considerations

FlashCopy consistency groups do not provide application consistency. It ensures only that volume points-in-time are consistent between them.

Because FlashCopy is at the block level, it is necessary to understand the interaction between your application and the host operating system. From a logical standpoint, it is

easiest to think of these objects as “layers” that sit on top of one another. The application is the topmost layer, and beneath it is the operating system layer.

Both of these layers feature various levels and methods of caching data to provide better speed. Therefore, because IBM Storage Virtualize and FlashCopy sit below these layers, they are unaware of the cache at the application or operating system layers.

To ensure the integrity of the copy that is made, it is necessary to flush the host operating system and application cache for any outstanding reads or writes before the FlashCopy operation is performed. Failing to flush the host operating system and application cache produces what is referred to as a *crash consistent* copy.

The resulting copy requires the same type of recovery procedure, such as log replay and file system checks that are required following a host crash. FlashCopies that are crash consistent can often be used after file system and application recovery procedures.

Various operating systems and applications provide facilities to stop I/O operations and ensure that all data is flushed from host cache. If these facilities are available, they can be used to prepare for a FlashCopy operation. When this type of facility is unavailable, the host cache must be flushed manually by quiescing the application and unmounting the file system or drives.

Before the FlashCopy mappings are started, any data that is held on the host operating system or application caches for the target volumes must be discarded. The easiest way to ensure that no data is held in these caches is to unmount the target volumes before the FlashCopy operation starts.

Preferred practice: From a practical standpoint, when you have an application that is backed by a database and you want to make a FlashCopy of that application’s data, you can typically use the write-suspend method that is available in most modern databases. This is possible because the database maintains strict control over I/O.

This method is an alternative to flushing data from the application and backing database, which is always the suggested method because it is safer. However, this alternative method can be used when facilities do not exist or your environment includes time sensitivity.

IBM FlashCopy application integrated solutions

IBM FlashCopy is not application aware and a third-party tool is needed to link the application to the FlashCopy operations.

IBM Spectrum Protect Snapshot protects data with integrated, application-aware snapshot backup and restore capabilities that use FlashCopy technologies in the IBM Storage Virtualize.

You can protect data that is stored by IBM Db2, SAP, Oracle, Microsoft Exchange, and Microsoft SQL Server applications. You can create and manage volume-level snapshots for file systems and custom applications.

In addition, it enables you to manage frequent, near-instant, nondisruptive, application-aware backups and restores that use integrated application and VMware snapshot technologies. IBM Spectrum Protect Snapshot can be widely used in IBM and non-IBM storage systems.

Other IBM products are also available for application aware backup and restore capabilities, such as IBM Spectrum Protect Plus (SPP) and IBM Copy Data Manager (CDM). Consult with your IBM representative for more information about these offerings.

Note: To see how IBM Spectrum Protect Snapshot, IBM Spectrum Protect Plus and IBM Copy Data Manager can help your business, see the following IBM Documentation web pages:

- IBM Storage Protect Snapshot documentation
- IBM Storage Protect Plus documentation
- IBM Storage Copy Data Management documentation

10.3.6 Volume group snapshot

A volume group serves as a container for managing a collection of associated volumes as a unified entity. The primary purpose of a volume group is to ensure consistency across all volumes within the group.

Volume groups have various applications:

- Safeguarded copy function. You can group volumes for configuration as safeguarded copies. The safeguarded copy function is a cybersecurity resilience feature that creates unalterable copies of data, providing protection against manipulation or changes. For more information, refer to 10.2, “Safeguarded Copy” on page 718.
- Policy-based replication. Volume groups can be used for policy-based replication. When you assign a replication policy to a volume group, policy-based replication is configured for all volumes within that group. For more information, refer to 10.16, “IBM Storage policy-based replication” on page 905.
- Snapshot function. Snapshots represent read-only, point-in-time copies of a volume group that are not directly accessible from hosts. To access the contents of a snapshot, a clone or thin clone of the volume group snapshot can be created.

Volume groups are designed to function as a cohesive unit, and they are effective when considered as a whole. When a group of thin clones or clones is populated, the snapshot function ensures that the images remain mutually consistent. If volumes are added or removed from a group, the host applications must ensure mutual consistency among the volume groups. For more information, refer to 10.15, “IBM Storage volume group and volume group snapshots” on page 894.

10.3.7 Grains and bitmap: I/O indirection

When a FlashCopy operation starts, a checkpoint is made of the source volume. No data is copied at the time that a start operation occurs. Instead, the checkpoint creates a bitmap that indicates that no part of the source volume was copied. Each bit in the bitmap represents one region of the source volume. Each region is called a *grain*.

You can think of the bitmap as a simple table of ones or zeros. The table tracks the difference between a source volume grains and a target volume grains. At the creation of the FlashCopy mapping, the table is filled with zeros, which indicates that no grain is copied yet.

When a grain is copied from source to target, the region of the bitmap that refers to that grain is updated (for example, from “0” to “1”), as shown in Figure 10-2.

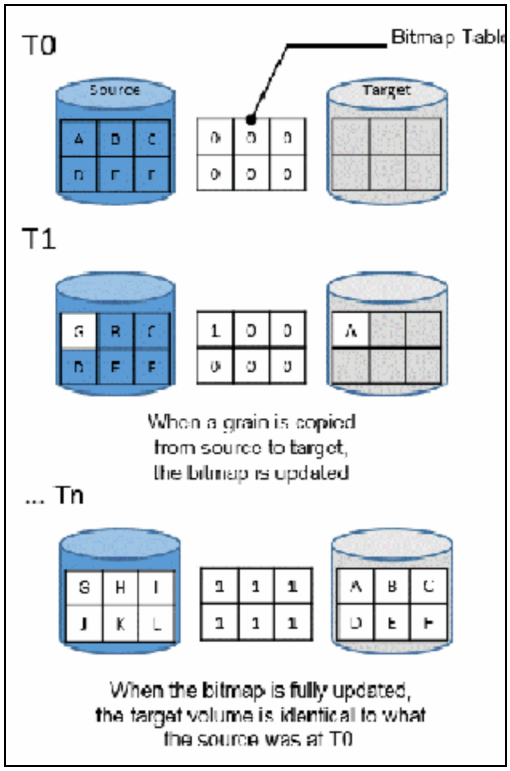


Figure 10-2 A simplified representation of grains and bitmap

The grain size can be 64 KB or 256 KB (the default is 256 KB). The grain size cannot be selected by the user when a FlashCopy mapping is created from the graphical user interface (GUI). The FlashCopy bitmap contains 1 bit for each grain. The bit records whether the associated grain is split by copying the grain from the source to the target.

After a FlashCopy mapping is created, the grain size for that FlashCopy mapping cannot be changed. When a FlashCopy mapping is created, the grain size of that mapping is used if the grain size parameter is not specified and one of the volumes is part of a FlashCopy mapping.

If neither volume in the new mapping is part of another FlashCopy mapping and at least one of the volumes in the mapping is a compressed volume, the default grain size is 64 KB for performance considerations. Other than in this situation, the default grain size is 256 KB.

Copy on Write, Redirect on Write, and Copy on Demand

With IBM Storage Virtualize version 8.4 or later, FlashCopy uses a Copy on Write (CoW) mechanism to copy data from a source volume to a target volume in standard pools (non-DRP), and RoW mechanism in data reduction pools (DRPs).

With CoW, as shown in Figure 10-3 on page 728, when data is written on a source volume, the grain where the blocks that are to be changed are stored is first copied to the target volume and then modified on the source volume. The bitmap is updated to track the copy.

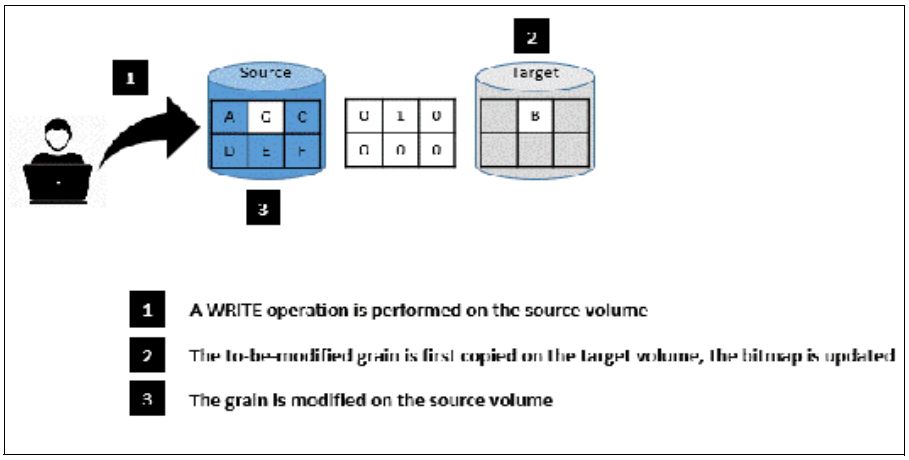


Figure 10-3 CoW steps

With Redirect on Write (RoW), when the source volume is modified, the updated grain is written directly to a new block in the DRP customer data volume. The source volume metadata and FlashCopy bitmap are then updated to reflect this update.

RoW was introduced with IBM Storage Virtualize release 8.4 for DRPs only. When compared to CoW, RoW reduces the back-end activity by removing the copy operation, which improves the overall performance of FlashCopy operations.

Note: At the time of this writing, RoW is used only for volumes with supported deduplication, without a mirroring relationship, and when the source and target volumes are within the same pool and I/O group. Whether to use CoW or RoW is automatically done by the FlashCopy software-based in these conditions.

With IBM FlashCopy, the target volume is immediately accessible for read *and* write operations. Therefore, a target volume can be modified, even if it is part of a FlashCopy mapping.

In standard pools, as shown in Figure 10-4, when a write operation is performed on the *target* volume, the grain that contains the blocks to be changed is first copied from the source. This is called Copy on-Demand. It is then modified with the new value. The bitmap is modified so the grain from the source is *not* copied again, even if it is changed or if a background copy is enabled.

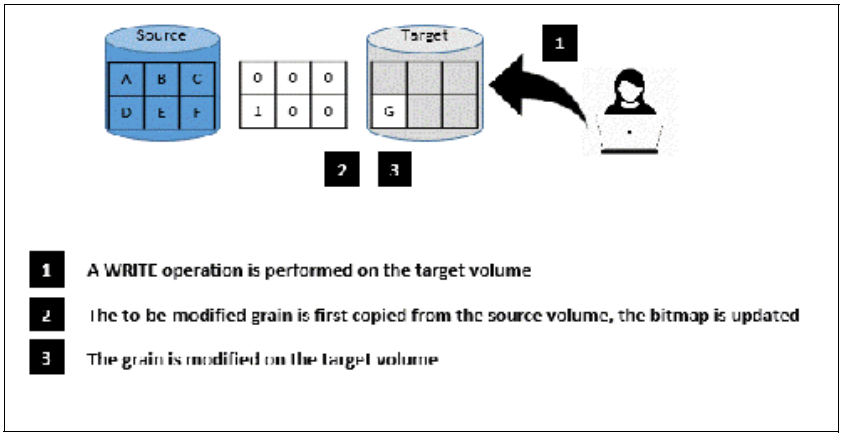


Figure 10-4 Copy on-Demand steps

For IBM Storage Virtualize version 8.4 or later, this behavior is slightly different in DRPs. It reads the grain to be updated from the source volume, modifies it with the new value in the cache, writes the modified grain to the DRP customer data volume, and then updates the FlashCopy bitmap.

Note: If all the blocks of the grain to be modified are changed, the source grain does not need to be read or copied first. No copy on demand is available and it is directly modified at the target volume.

FlashCopy indirection layer

The FlashCopy indirection layer governs the I/O to the source and target volumes when a FlashCopy mapping is started, which is done by using the FlashCopy bitmap. The purpose of the FlashCopy indirection layer is to enable the source and target volumes for read and write I/O immediately after the FlashCopy is started.

The indirection Layer intercepts any I/O coming from a host (read or write operation) and addressed to a FlashCopy volume (source or target). It determines whether the addressed volume is a source or a target, its direction (read or write), and the state of the bitmap table for the FlashCopy mapping that the addressed volume is in. It then decides what operation to perform. The different I/O indirections are described next.

Read from the source Volume

When a user performs a read operation on the source volume, there is no redirection. The operation is similar to what is done with a volume that is not part of a FlashCopy mapping.

Write on the source volume

Performing a write operation on the source volume modifies a block or a set of blocks, which modifies a grain on the source. It generates one of the following actions, depending on the state of the grain to be modified.

Consider the following points:

- If the bitmap indicates that the grain was copied, the source grain is changed and the target volume and the bitmap table remain unchanged, as shown in Figure 10-5.

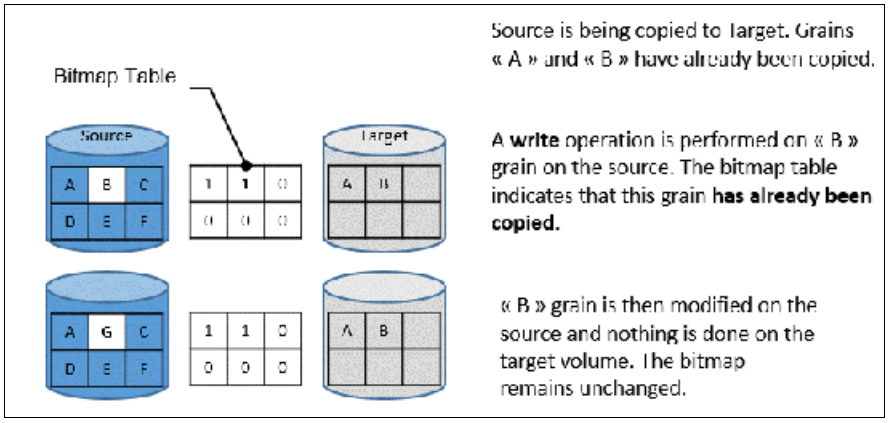


Figure 10-5 Modifying a copied grain on the source

- If the bitmap indicates that the grain is not yet copied, the grain is first copied on the target (CoW), the bitmap table is updated, and the grain is modified on the source, as shown in Figure 10-6. This behavior is true for standard pools in IBM Storage Virtualize release 8.4 or later, or any pool type if the code version is earlier than 8.4.

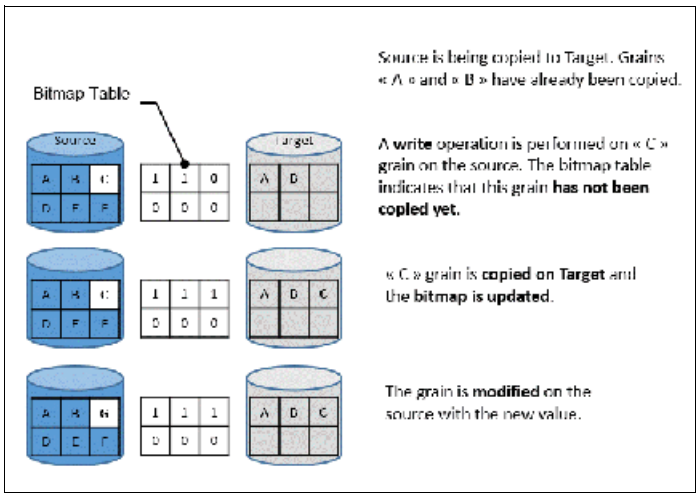


Figure 10-6 Modifying a non-copied grain on the source

- If the target of the write operation is a DRP in a system that is running IBM Storage Virtualize release 8.4 or later, the system performs a RoW operation, as described in “Copy on Write, Redirect on Write, and Copy on Demand” on page 727.

Write on a target volume

Because FlashCopy target volumes are immediately accessible in Read and Write mode, it is possible to perform write operations on the target volume when the FlashCopy mapping is started. Performing a write operation on the target generates one of the following actions, depending on the bitmap:

- If the bitmap indicates the grain to be modified on the target was not yet copied, it is first copied from the source (copy on demand). The bitmap is updated, and the grain is modified on the target with the new value, as shown in Figure 10-7 on page 731. The source volume remains unchanged.

> **Note:** If the entire grain is to be modified and not only some of the blocks, the copy on-demand is bypassed. The bitmap is updated, and the grain on the target is modified but not copied first.

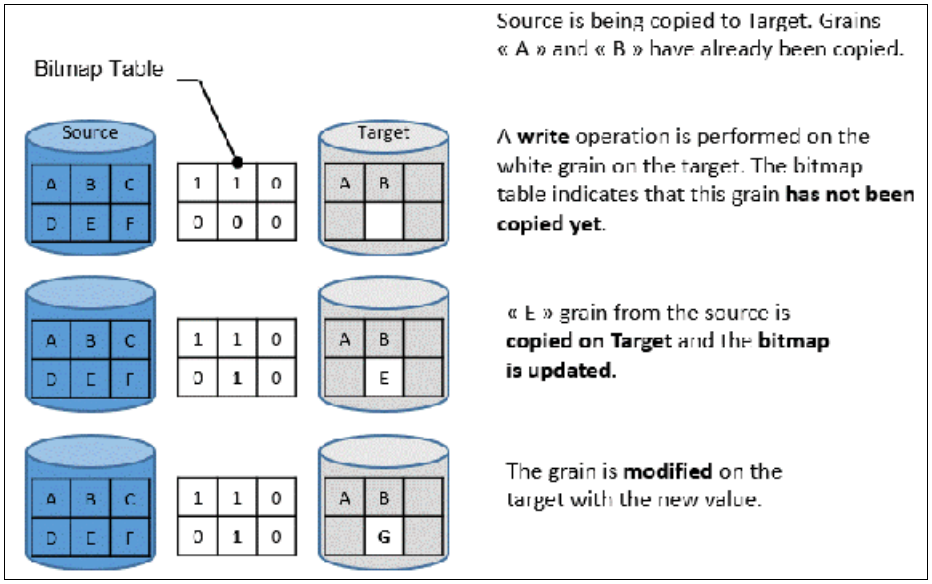


Figure 10-7 Modifying a non-copied grain on the target

- If the bitmap indicates the grain to be modified on the target was copied, it is directly changed. The bitmap is *not* updated, and the grain is modified on the target with the new value, as shown in Figure 10-8.

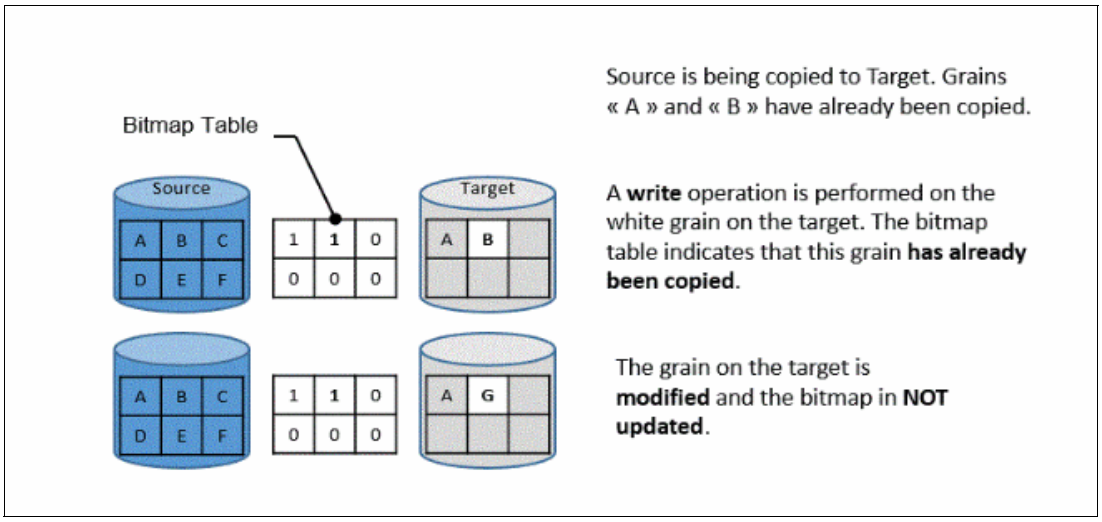


Figure 10-8 Modifying an already copied grain on the target

Note: The bitmap is not updated in that case. Otherwise, it might be copied from the source late if a background copy is ongoing or if write operations are made on the source. That process over-writes the changed grain on the target.

Read from a target volume

Performing a read operation on the target volume returns the value in the grain on the source or on the target, depending on the bitmap. Consider the following points:

- If the bitmap indicates that the grain was copied from the source or that the grain was modified on the target, the grain on the target is read, as shown in Figure 10-9.
- If the bitmap indicates that the grain was not yet copied from the source or was not modified on the target, the grain on the source is read, as shown in Figure 10-9.

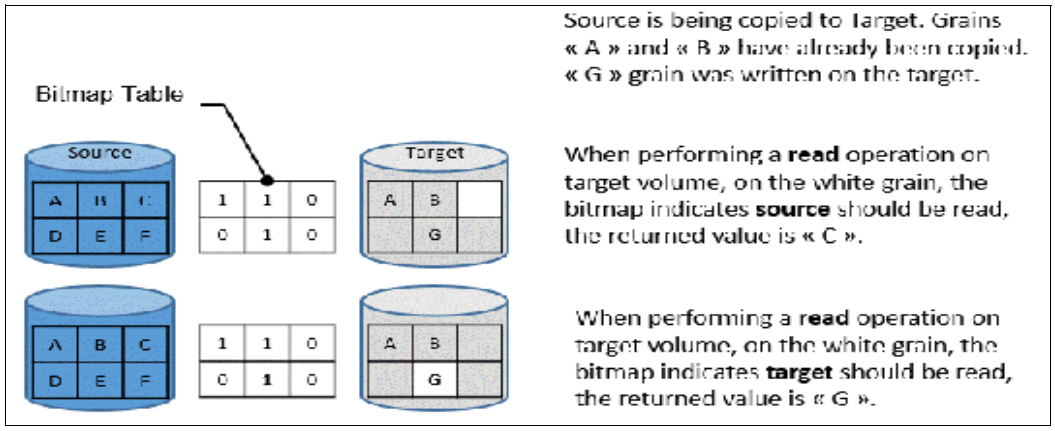


Figure 10-9 Reading a grain on target

If the source features multiple targets, the Indirection layer algorithm behaves differently on Target I/Os. For more information about multi-target operations, see 10.3.12, “Multiple target FlashCopy” on page 738.

10.3.8 Interaction with cache

The cache of IBM Storage Virtualize-based systems is divided into upper and lower cache. Upper cache serves mostly as write cache and hides the write latency from the hosts and application. Lower cache is a read/write cache and optimizes I/O to and from disks. Figure 10-10 on page 733 shows the IBM Storage Virtualize software stack including cache architecture.

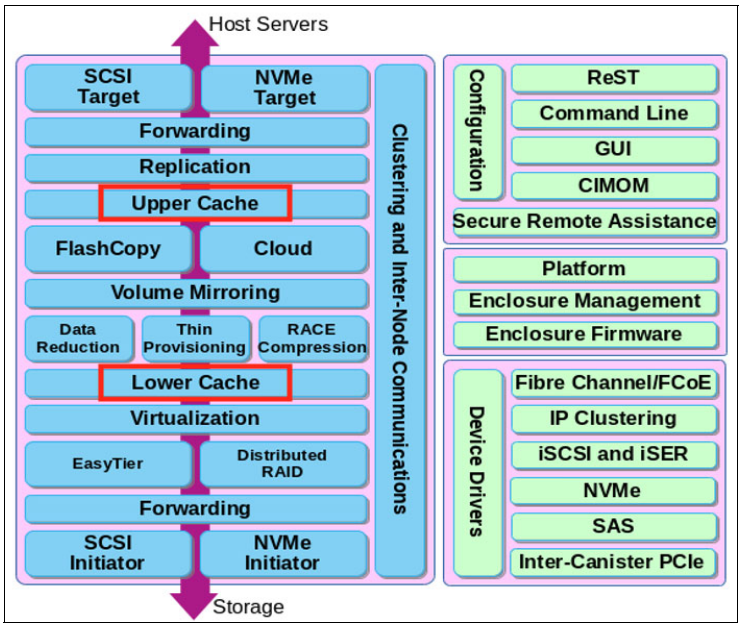


Figure 10-10 IBM Storage Virtualize software architecture

The copy-on-write process introduces significant latency into write operations. To isolate the active application from this extra latency, the FlashCopy indirection layer is placed logically between upper and lower cache.

Therefore, the extra latency that is introduced by the copy-on-write process is encountered only by the internal cache operations and not by the application. With IBM Storage Virtualize release 8.4 or later, the RoW mechanism that is used for DRPs aims to reduce the overhead that is introduced by CoW.

Also, the two-level cache that is described here provides more performance improvements to the FlashCopy mechanism. Because the FlashCopy layer is above the lower cache in the IBM Storage Virtualize software stack, it can benefit from read prefetching and coalescing writes to backend storage.

Preparing FlashCopy benefits from the two-level cache because upper cache write data does not have to go directly to backend storage, but to the lower cache layer instead.

10.3.9 Background copy rate

The background copy rate is a property of a FlashCopy mapping. A grain copy from the source to the target can occur when triggered by a write operation on the source or target volume, or when background copy is enabled. With background copy enabled, the target volume eventually becomes a clone of the source volume at the time the mapping was started (T0). When the copy is completed, the mapping can be removed between the two volumes, and you can end up with two independent volumes.

The background copy rate property determines the speed at which grains are copied as a background operation, immediately after the FlashCopy mapping is started. That speed is defined by the user when the FlashCopy mapping is created, and can be changed dynamically for each individual mapping, whatever its state. Mapping copy rate values can be 0–150, with the corresponding speeds that are listed in Table 10-2.

Table 10-2 Copy rate values

User-specified copy rate attribute value	Data copied per second	256 KB grains per second	64 KB grains per second
1–10	128 kibibytes (KiB)	0.5	2
11–20	256 KiB	1	4
21–30	512 KiB	2	8
31–40	1 MiB	4	16
41–50	2 MiB	8	32
51–60	4 MiB	16	64
61–70	8 MiB	32	128
71–80	16 MiB	64	256
81–90	32 MiB	128	512
91–100	64 MiB	256	1024
101–110	128 MiB	512	2048
111–120	256 MiB	1024	4096
121–130	512 MiB	2048	8192
131–140	1 GiB	4096	16384
141–150	2 GiB	8192	32768

When the background copy function is not performed (copy rate = 0), the target volume remains a valid copy of the source data only while the FlashCopy mapping remains in place.

The grains per second numbers represent the maximum number of grains that the IBM Storage Virtualize copies per second. This amount assumes that the bandwidth to the managed disks (MDisks) can accommodate this rate.

If the IBM Storage Virtualize cannot achieve these copy rates because of insufficient bandwidth from the nodes to the MDisks, the background copy I/O contends for resources on an equal basis with the I/O that is arriving from the hosts.

The FlashCopy background copy process might affect the incoming host I/O because of resource contention. Therefore, this issue affects the host I/O response times and the throughput, which affects the processing time of the host jobs.

FlashCopy background copy and foreground host I/O coexist and continue to progress, and do not stop, hang, nor cause the node to fail.

The FlashCopy background copy is performed by the node in the I/O group in which the source volumes are found. This responsibility is moved to the partner node in the I/O group if the original node fails.

10.3.10 Incremental FlashCopy

When a FlashCopy mapping is stopped (because the entire source volume was copied onto the target volume or a user manually stopped it), the bitmap table is reset. Therefore, when the same FlashCopy is started again, the copy process is restarted from the beginning.

Running the **`-incremental`** option when creating the FlashCopy mapping allows the system to keep the bitmap as it is when the mapping is stopped. Therefore, when the mapping is started again (at another PiT), the bitmap is reused and only changes between the two copies are applied to the target.

A system that provides Incremental FlashCopy capability allows the system administrator to refresh a target volume without having to wait for a full copy of the source volume to be complete. At the point of refreshing the target volume, if the data was changed on the source or target volumes for a particular grain, the grain from the source volume is copied to the target.

The advantages of Incremental FlashCopy are useful only if a previous full copy of the source volume was obtained. Incremental FlashCopy helps with only further recovery time objectives (RTOs). It does *not* help with the initial RTO. An RTO is the time that is needed to recover data from a previous state.

For example, as shown in Figure 10-11, a FlashCopy mapping was defined between a source volume and a target volume by using the **`-incremental`** option.

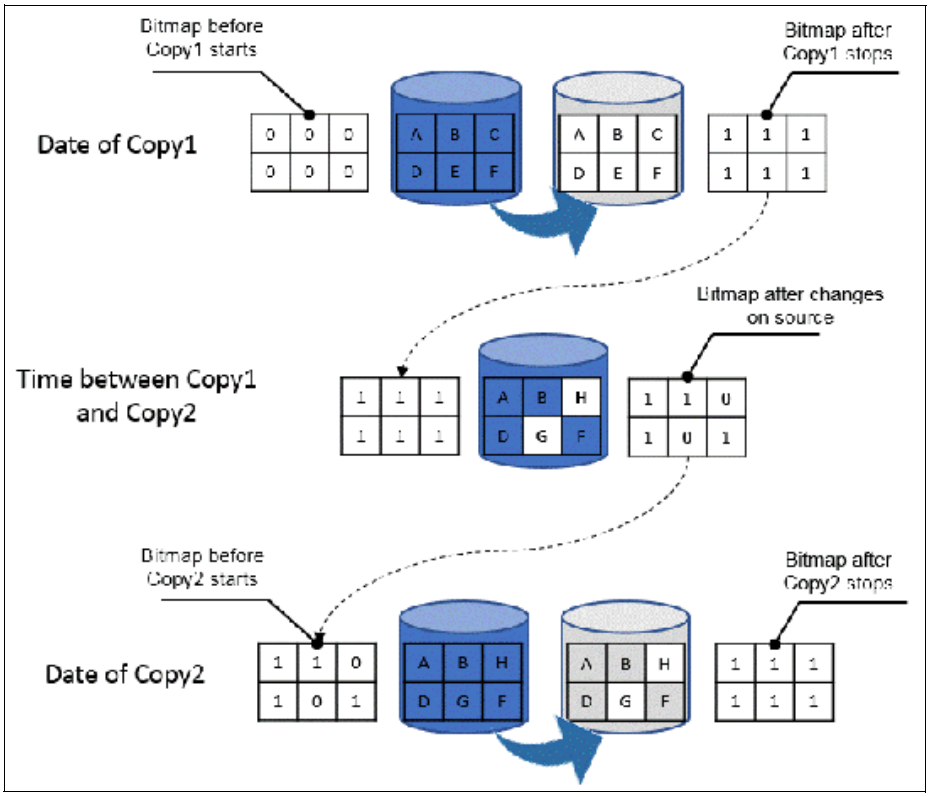


Figure 10-11 Incremental FlashCopy example

Consider the following points:

- The mapping is started on Copy1 date. A *full copy* of the source volume is made, and the bitmap is updated every time that a grain is copied. At the end of Copy1, all grains are

copied and the target volume is an exact replica of the source volume at the beginning of Copy1. Although the mapping is stopped, the bitmap is maintained because of the **-incremental** option.

- Changes are made on the source volume and the bitmap is updated, although the FlashCopy mapping is not active. For example, grains E and C on the source are changed in G and H, their corresponding bits are changed in the bitmap. The target volume is untouched.
- The mapping is started again on Copy2 date. The bitmap indicates that only grains E and C were changed; therefore, only G and H are copied on the target volume. The other grains do not need to be copied because they were copied the first time. The copy time is much quicker than for the first copy as only a fraction of the source volume is copied.

10.3.11 Starting FlashCopy mappings and consistency groups

You can prepare, start, or stop FlashCopy on a stand-alone mapping or a consistency group.

When the CLI is used to perform FlashCopy on volumes, run a **prestartfcmap** or **prestartfcconsistgrp** command *before* you start a FlashCopy (regardless of the type and options specified). These commands put the cache into write-through mode and provides a flushing of the I/O that is bound for your volume. After FlashCopy is started, an effective copy of a source volume to a target volume is created.

The content of the source volume is presented immediately on the target volume and the original content of the target volume is lost.

FlashCopy commands can then be run to the FlashCopy consistency group and therefore, simultaneously for all of the FlashCopy mappings that are defined in the consistency group. For example, when a FlashCopy **start** command is run to the consistency group, all of the FlashCopy mappings in the consistency group are started at the same time. This simultaneous start results in a PiT copy that is consistent across all of the FlashCopy mappings that are contained in the consistency group.

Rather than running **prestartfcmap** or **prestartfcconsistgrp**, you can also use the **-prep** parameter in the **startfcmap** or **startfcconsistgrp** command to prepare and start FlashCopy in one step.

Important: After an individual FlashCopy mapping is added to a consistency group, it can be managed as part of the group only. Operations, such as `prepare`, `start`, and `stop`, are no longer allowed on the individual mapping.

FlashCopy mapping states

At any point, a mapping is in one of the following states:

- Idle or copied

  The source and target volumes act as independent volumes, even if a mapping exists between the two. Read and write caching is enabled for the source and the target volumes. If the mapping is incremental and the background copy is complete, the mapping records only the differences between the source and target volumes. If the connection to both nodes in the I/O group that the mapping is assigned to is lost, the source and target volumes are offline.

- Copying

  The copy is in progress. Read and write caching is enabled on the source and the target volumes.

- Prepared

  The mapping is ready to start. The target volume is online, but is not accessible. The target volume cannot perform read or write caching. Read and write caching is failed by the Small Computer System Interface (SCSI) front end as a hardware error.

  If the mapping is incremental and a previous mapping that is completed, the mapping records only the differences between the source and target volumes. If the connection to both nodes in the I/O group that the mapping is assigned to is lost, the source and target volumes go offline.

- Preparing

  The target volume is online, but not accessible. The target volume cannot perform read or write caching. Read and write caching is failed by the SCSI front end as a hardware error. Any changed write data for the source volume is flushed from the cache. Any read or write data for the target volume is discarded from the cache.

  If the mapping is incremental and a previous mapping that is completed, the mapping records only the differences between the source and target volumes. If the connection to both nodes in the I/O group that the mapping is assigned to is lost, the source and target volumes go offline.

- Stopped

  The mapping is stopped because you issued a stop command or an I/O error occurred. The target volume is offline and its data is lost. To access the target volume, you must restart or delete the mapping. The source volume is accessible and the read and write cache is enabled. If the mapping is incremental, the mapping is recording write operations to the source volume. If the connection to both nodes in the I/O group that the mapping is assigned to is lost, the source and target volumes go offline.

- Stopping

  The mapping is copying data to another mapping. If the background copy process is complete, the target volume is online while the stopping copy process completes. If the background copy process is incomplete, data is discarded from the target volume cache. The target volume is offline while the stopping copy process runs. The source volume is accessible for I/O operations.

- Suspended

  The mapping started, but it did not complete. Access to the metadata is lost, which causes the source and target volume to go offline. When access to the metadata is restored, the mapping returns to the copying or stopping state and the source and target volumes return online. The background copy process resumes. If the data was not flushed and was written to the source or target volume before the suspension, it is in the cache until the mapping leaves the suspended state.

Summary of FlashCopy mapping states

Table 10-3 on page 738 lists the various FlashCopy mapping states and the corresponding states of the source and target volumes.

Table 10-3 FlashCopy mapping state summary

State	Source		Target	
	Online/Offline	**Cache state**	**Online/Offline**	**Cache state**
Idling/Copied	Online	Write-back	Online	Write-back
Copying	Online	Write-back	Online	Write-back
Stopped	Online	Write-back	Offline	N/A
Stopping	Online	Write-back	► Online if copy complete ► Offline if copy incomplete	N/A
Suspended	Offline	Write-back	Offline	N/A
Preparing	Online	Write-through	Online but not accessible	N/A
Prepared	Online	Write-through	Online but not accessible	N/A

10.3.12 Multiple target FlashCopy

Various methods and configurations are available to enable one-t- many FlashCopy operations, including the following examples:

- A volume can be the source of multiple target volumes.
- A target volume can also be the source of another target volume.
- A target volume can have only one source volume.
- A source volume can have multiple target volumes in one or multiple consistency groups.
- A consistency group can contain multiple FlashCopy mappings (source-target relations).
- A source volume can belong to multiple consistency groups.

Figure 10-12 on page 739 shows these different possibilities.

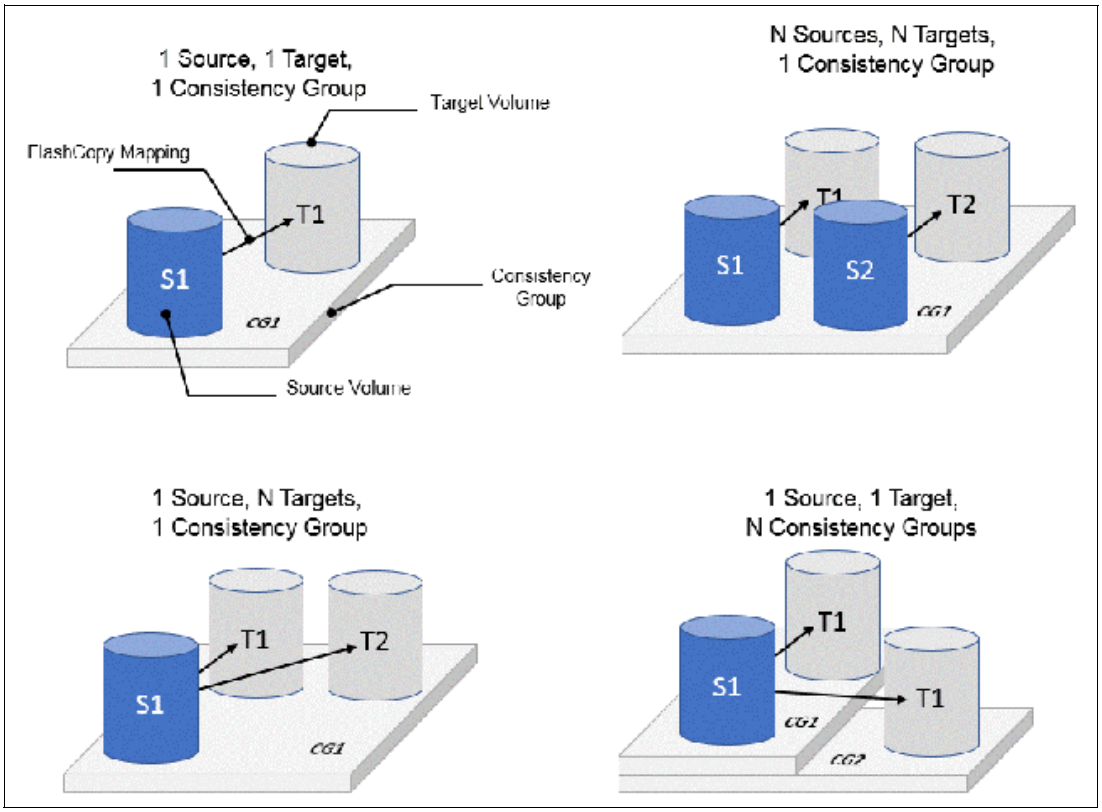


Figure 10-12 Consistency groups and mappings combinations

Every source-target relation is a FlashCopy mapping and is maintained with its own bitmap table. No consistency group bitmap table exists.

When a source volume is in a FlashCopy mapping with multiple targets, in multiple consistency groups, it allows the copy of a single source at multiple points in time and therefore, keeps multiple versions of a single volume.

Consistency group with multiple target FlashCopy

A consistency group aggregates FlashCopy mappings, not volumes. Therefore, where a source volume has multiple FlashCopy mappings, they can be in the same or separate consistency groups.

If a particular volume is the source volume for multiple FlashCopy mappings, you might want to create separate consistency groups to separate each mapping of the same source volume. Regardless of whether the source volume with multiple target volumes are in the same consistency group or in separate consistency groups, the resulting FlashCopy produces multiple identical copies of the source data.

Dependencies

When a source volume has multiple target volumes, a mapping is created for each source-target relationship. When data is changed on the source volume, it is first copied to the target volume because of the copy-on-write mechanism that is used by FlashCopy.

If IBM Storage Virtualize release 8.4 or later is run, it uses RoW mechanism instead for DRP pools, as described in “Copy on Write, Redirect on Write, and Copy on Demand” on page 727.

You can create up to 256 targets for a single source volume. Therefore, a single write operation on the source volume might result in 256 write operations (one per target volume) when CoW is used. This configuration generates a large workload that the system cannot be able to handle, which can lead to a heavy performance impact on front-end operations.

To avoid any significant effect on performance because of multiple targets, FlashCopy creates dependencies between the targets. Dependencies can be considered as “hidden” FlashCopy mappings that are not visible to and cannot be managed by the user. A dependency is created between the most recent target and the previous one (in order of start time). Figure 10-13 shows an example of a source volume with three targets.

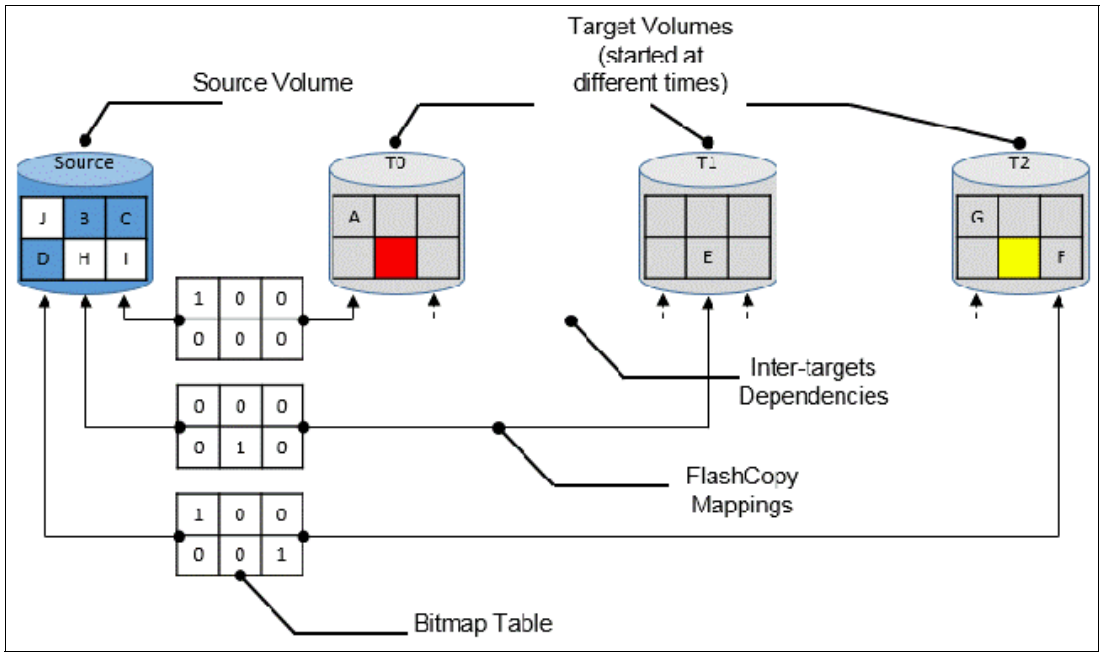


Figure 10-13 FlashCopy dependencies example

When the three targets are started, Target T0 was started first and considered the “oldest.” Target T1 was started next and is considered “next oldest,” and finally, Target T2 was started last and considered the “most recent” or “newest.” The “next oldest” target for T2 is T1. The “next oldest” target for T1 is T0. T1 is newer than T2, and T0 is newer than T1.

Source read with multiple target FlashCopy

No specific behavior is shown for read operations on source volumes when multiple targets exist for that volume. The data is always read from the source.

Source write with multiple target FlashCopy (CoW)

A write to the source volume does not cause its data to be copied to all of the targets. Instead, it is copied to the most recent target volume only. For example, consider the sequence of events that are listed in Table 10-4 for a source volume and three targets that are started at different times. In this example, no background copy exists. The “most recent” target is indicated with an asterisk.

Table 10-4 Sequence example of write IOs on a source with multiple targets

	Source volume	Target T0	Target T1	Target T2
Time 0: mapping with T0 is started	A B C D E F	_ _ _* _ _ _	Not started	Not started

	Source volume	Target T0	Target T1	Target T2
Time 1: change of “A” is made on source (->“G”)	G B C D E F	A _ _* _ _ _	Not started	Not started
Time 2: mapping with T1 is started	G B C D E F	A _ _ _ _ _	_ _ _* _ _ _	Not started
Time 3: change of “E” is made on source (->“H”)	G B C D H F	A _ _ _ _ _	_ _ _* _ E_	Not started
Time 4: mapping with T2 is started	G B C D H F	A _ _ _ _ _	_ _ _ _ E_	_ _ _* _ _ _
Time 5: change of “F” is made on source (->“I”)	G B C D H I	A _ _ _ _ _	_ _ _ _ E_	_ _ _* _ _ F
Time 6: change of “G” is made on source (->“J”)	J B C D H I	A _ _ _ _ _	_ _ _ _ E_	G _ _* _ _ F
Time 7: stop of Source-T2 mapping	J B C D H I	A _ _ _ _ _	G _ _ * _ E F	Stopped
Time 8: stop of Source-T1 mapping	J B C D H I	A _ _ * _ E F	Stopped	Stopped
				* *“most recent” target*

An intermediate target disk, not the oldest or the newest, treats the set of newer target volumes and the true source volume as a type of composite source. It treats all older volumes as a target and behaves like a source to them.

Target read with multiple target FlashCopy

Target reading with multiple targets depends on whether the grain was copied. Consider the following points:

- If the grain that is read is copied from the source to the target, the read returns data from the target that is read.
- If the grain is not yet copied, each of the newer mappings is examined in turn. The read is performed from the first copy (the oldest) that is found. If none is found, the read is performed from the source.

  For example (Figure 10-13 on page 740), if the yellow grain on T2 is read, it returns “H” because no newer target than T2 exists. Therefore, the source is read.

  As another example (see Figure 10-13 on page 740), if the red grain on T0 is read, it returns “E” because two newer targets exist for T0, and T1 is the oldest of those targets.

Target write with multiple target FlashCopy (Copy on Demand)

A write to an intermediate or the newest target volume must consider the state of the grain within its own mapping and the state of the grain of the next oldest mapping. Consider the following points:

- If the grain in the target that is written is copied and if the grain of the next oldest mapping is not yet copied, the grain must be copied before the write can proceed to preserve the contents of the next oldest mapping.

  For example, in Figure 10-13 on page 740, if the grain “G” is changed on T2, it must be copied to T1 (next oldest not yet copied) first and then changed on T2.

- If the grain in the target that is being written is not yet copied, the grain is copied from the oldest copied grain in the mappings that are newer than the target, or from the source if none is copied. For example, in Figure 10-13 on page 740, if the red grain on T0 is written, it is first copied from T1 (data “E”). After this copy is done, the write can be applied to the target.

Table 10-5 lists the indirection layer algorithm in a multi-target FlashCopy.

Table 10-5 Summary table of the FlashCopy indirection layer algorithm t

Accessed volume	Was the grain copied?	Host I/O operation	
		Read	Write
Source	No	Read from the source volume.	Copy grain to most recently started target for this source, then write to the source.
	Yes	Read from the source volume.	Write to the source volume.
Target	No	If any newer targets exist for this source in which this grain was copied, read from the oldest of these targets. Otherwise, read from the source.	Hold the write. Check the dependency target volumes to see whether the grain was copied. If the grain is not copied to the next oldest target for this source, copy the grain to the next oldest target. Then write to the target.
	Yes	Read from the target volume.	Write to the target volume.

Stopping the process in a multiple target FlashCopy: Cleaning Mode

When a mapping that contains a target that includes dependent mappings is stopped, the mapping enters the `stopping` state. It then begins copying all grains that are uniquely held on the target volume of the mapping that is being stopped to the next oldest mapping that is in the copying state. The mapping remains in the `stopping` state until all grains are copied, and then enters the `stopped` state. This mode is referred to as the *Cleaning Mode*.

For example, if the mapping Source-T2 was stopped, the mapping enters the `stopping` state while the cleaning process copies the data of T2 to T1 (next oldest). After all of the data is copied, Source-T2 mapping enters the `stopped` state, and T1 is no longer dependent upon T2. However, T0 remains dependent upon T1.

For example, as shown in Table 10-4 on page 740, if you stop the Source-T2 mapping on “Time 7,” then the grains that are not yet copied on T1 are copied from T2 to T1. Reading T1 is then like reading the source at the time T1 was started (“Time 2”).

As another example, with Table 10-4 on page 740, if you stop the Source-T1 mapping on “Time 8,” the grains that are not yet copied on T0 are copied from T1 to T0. Reading T0 is then similar to reading the source at the time T0 was started (“Time 0”).

If you stop the Source-T1 mapping while Source-T0 mapping and Source-T2 are still in copying mode, the grains that are not yet copied on T0 are copied from T1 to T0 (next oldest). T0 now depends upon T2.

Your target volume is still accessible while the cleaning process is running. When the system is operating in this mode, it is possible that host I/O operations can prevent the cleaning process from reaching 100% if the I/O operations continue to copy new data to the target volumes.

Cleaning rate

The data rate at which data is copied from the target of the mapping being stopped to the next oldest target is determined by the *cleaning rate*. This property of FlashCopy mapping can be changed dynamically. It is measured as is the copyrate property, but both properties are independent. Table 10-6 lists the relationship of the cleaning rate values to the attempted number of grains to be split per second.

Table 10-6 Cleaning rate values

User-specified copy rate attribute value	Data copied/sec	256 KB grains/sec	64 KB grains/sec
1 - 10	128 KiB	0.5	2
11 - 20	256 KiB	1	4
21 - 30	512 KiB	2	8
31 - 40	1 MiB	4	16
41 - 50	2 MiB	8	32
51 - 60	4 MiB	16	64
61 - 70	8 MiB	32	128
71 - 80	16 MiB	64	256
81 - 90	32 MiB	128	512
91 - 100	64 MiB	256	1024
101 - 110	128 MiB	512	2048
111 - 120	256 MiB	1024	4096
121 - 130	512 MiB	2048	8192
131 - 140	1 GiB	4096	16384
141 - 150	2 GiB	8192	32768

10.3.13 Reverse FlashCopy

Reverse FlashCopy enables FlashCopy targets to become restore points for the source without breaking the FlashCopy mapping, and without having to wait for the original copy operation to complete. Therefore, a FlashCopy source supports multiple targets (up to 256), and multiple rollback points.

A key advantage of the IBM Storage Virtualize Multiple Target Reverse FlashCopy function is that the reverse FlashCopy does not destroy the original target. This feature enables processes that use the target, such as a tape backup or tests, to continue uninterrupted.

IBM Storage Virtualize can also create an optional copy of the source volume to be made before the reverse copy operation starts. This ability to restore back to the original source data can be useful for diagnostic purposes.

The production disk is instantly available with the backup data. Figure 10-14 shows an example of Reverse FlashCopy with a simple FlashCopy mapping (single target).

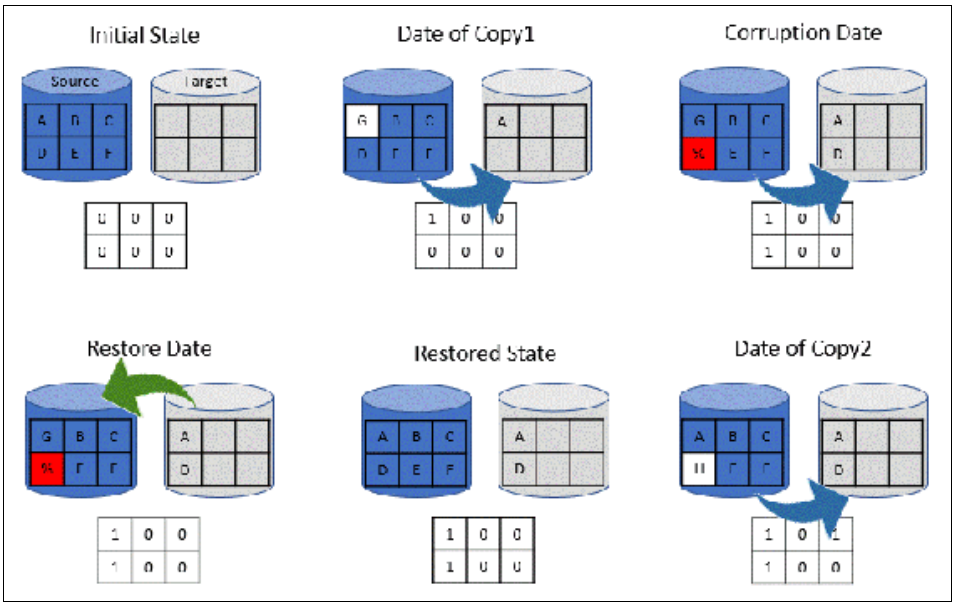


Figure 10-14 A reverse FlashCopy example for data restoration

This example assumes that a simple FlashCopy mapping was created between the “source” volume and “target” volume, and no background copy is set.

When the FlashCopy mapping starts (Date of Copy1), if source volume is changed (write operations on grain “A”), the modified grains are first copied to target, the bitmap table is updated, and the source grain is modified (from “A” to “G”).

At a specific time (“Corruption Date”), data is modified on another grain (grain “D” below), so it is first written on the target volume and the bitmap table is updated. Unfortunately, the new data is corrupted on the source volume.

The storage administrator can then use the Reverse FlashCopy feature by completing the following steps:

1. Create a mapping from target to source (if not already created). Because FlashCopy recognizes that the target volume of this new mapping is a source in another mapping, it does not create another bitmap table. It uses the existing bitmap table instead, with its updated bits.
2. Start the new mapping. Because of the existing bitmap table, only the *modified* grains are copied.

After the restoration is complete, at the “Restored State” time, source volume data is similar to what it was before the Corruption Date. The copy can resume with the restored data (Date of Copy2) and for example, data on the source volume can be modified. “D” grain is changed in “H” grain in the example above. In this last case, because the “D” grain was copied, it is not copied again on the target volume.

Consistency groups are reversed by creating a set of reverse FlashCopy mappings and adding them to a new reverse consistency group. Consistency groups cannot contain more than one FlashCopy mapping with the same target volume.

10.3.14 FlashCopy and image mode volumes

FlashCopy can be used with image mode volumes. Because the source and target volumes must be the same size, you must create a target volume with the same size as the image mode volume when you are creating a FlashCopy mapping. To accomplish this task by using the CLI, run the **svcinfo lsvdisk -bytes volumename** command. The size in bytes is then used to create the volume that is used in the FlashCopy mapping.

This method provides an exact number of bytes because image mode volumes might not line up one-to-one on other measurement unit boundaries. Example 10-1 shows the size of the ITSO-RS-TST volume. The ITSO-TST01 volume is then created, which specifies the same size.

Example 10-1 Listing the size of a volume in bytes and creating a volume of equal size

```
IBM_2145:ITSO-SV1:superuser>lsvdisk -bytes ITSO-RS-TST
id 42
name ITSO-RS-TST
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name Pool0
capacity 21474836480
type striped
formatted no
formatting yes
mdisk_id
mdisk_name
FC_id
......

IBM_2145:ITSO-SV1:superuser>mkvdisk -mdiskgrp Pool0 -iogrp 0 -size 21474836480
-unit b -name ITSO-TST01
Virtual Disk, id [43], successfully created
IBM_2145:ITSO-SV1:superuser>

IBM_2145:ITSO-SV1:superuser>lsvdisk -delim " "
42 ITSO-RS-TST 0 io_grp0 online 0 Pool0 20.00GB striped
600507680C9B80004800000000000002C 0 1 not_empty 0 no 0 0 Pool0 yes no 42
ITSO-RS-TST
43 ITSO-TST01 0 io_grp0 online 0 Pool0 20.00GB image
600507680C9B80004800000000000002D 0 1 not_empty 0 no 0 0 Pool0 yes no 43 ITSO-TST01
IBM_2145:ITSO-SV1:superuser>
```

Tip: Alternatively, you can run the **expandvdisksize** and **shrinkvdisksize** volume commands to modify the size of the volume.

These actions must be performed before a mapping is created.

10.3.15 FlashCopy mapping events

This section describes the events that modify the states of a FlashCopy. It also describes the mapping events that are listed in Table 10-7.

Overview of a FlashCopy sequence of events: The FlashCopy sequence includes the following tasks:

1. Associate the source data set with a target location (one or more source and target volumes).
2. Create a FlashCopy mapping for each source volume to the corresponding target volume. The target volume must be equal in size to the source volume.
3. Discontinue access to the target (application dependent).
4. Prepare (pre-trigger) the FlashCopy:
   a. Flush the cache for the source.
   b. Discard the cache for the target.
5. Start (trigger) the FlashCopy:
   a. Pause I/O (briefly) on the source.
   b. Resume I/O on the source.
   c. Start I/O on the target.

Table 10-7 Mapping events

Mapping event	Description
Create	A FlashCopy mapping is created between the specified source volume and the specified target volume. The operation fails if any one of the following conditions is true: ► The source volume is a member of 256 FlashCopy mappings. ► The node has insufficient bitmap memory. ► The source and target volumes are different sizes.
Prepare	The **prestartfcmap** or **prestartfcconsistgrp** command is directed to a consistency group for FlashCopy mappings that are members of a normal consistency group or to the mapping name for FlashCopy mappings that are stand-alone mappings. The **prestartfcmap** or **prestartfcconsistgrp** command places the FlashCopy mapping into the Preparing state. The **prestartfcmap** or **prestartfcconsistgrp** command can corrupt any data that was on the target volume because cached writes are discarded. Even if the FlashCopy mapping is never started, the data from the target might be changed logically during the act of preparing to start the FlashCopy mapping.
Flush done	The FlashCopy mapping automatically moves from the `preparing` state to the `prepared` state after all cached data for the source is flushed and all cached data for the target is no longer valid.

Mapping event	Description
Start	When all of the FlashCopy mappings in a consistency group are in the `prepared` state, the FlashCopy mappings can be started. To preserve the cross-volume consistency group, the start of all of the FlashCopy mappings in the consistency group must be synchronized correctly concerning I/Os that are directed at the volumes by running the **`startfcmap`** or **`startfcconsistgrp`** command. The following actions occur during the running of the **`startfcmap`** command or the **`startfcconsistgrp`** command: ► New reads and writes to all source volumes in the consistency group are paused in the cache layer until all ongoing reads and writes beneath the cache layer are completed. ► After all FlashCopy mappings in the consistency group are paused, the internal cluster state is set to enable FlashCopy operations. ► After the cluster state is set for all FlashCopy mappings in the consistency group, read and write operations continue on the source volumes. ► The target volumes are brought online. As part of the **`startfcmap`** or **`startfcconsistgrp`** command, read and write caching is enabled for the source and target volumes.
Modify	The following FlashCopy mapping properties can be modified: ► FlashCopy mapping name ► Clean rate ► Consistency group ► Copy rate (for background copy or stopping copy priority) ► Automatic deletion of the mapping when the background copy is complete
Stop	The following separate mechanisms can be used to stop a FlashCopy mapping: ► Issue a command ► An I/O error occurred
Delete	This command requests that the specified FlashCopy mapping is deleted. If the FlashCopy mapping is in the `copying` state, the **`force`** flag must be used.
Flush failed	If the flush of data from the cache cannot be completed, the FlashCopy mapping enters the `stopped` state.
Copy complete	After all of the source data is copied to the target and there are no dependent mappings, the state is set to `copied`. If the option to automatically delete the mapping after the background `copy complete` is specified, the FlashCopy mapping is deleted automatically. If this option is not specified, the FlashCopy mapping is not deleted automatically and can be reactivated by preparing and starting again.
Bitmap online/offline	The node failed.

10.3.16 Thin-provisioned FlashCopy

FlashCopy source and target volumes can be thin-provisioned.

Source or target thin-provisioned

The most common configuration is a fully allocated source and a thin-provisioned target. By using this configuration, the target uses a smaller amount of real storage than the source.

With this configuration, use a copyrate equal to 0 only. In this state, the virtual capacity of the target volume is identical to the capacity of the source volume, but the real capacity (the one used on the storage system) is lower, as shown on Figure 10-15.

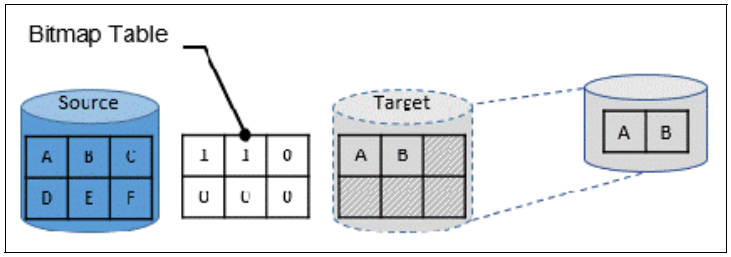


Figure 10-15 Thin-provisioned target volume

The real size of the target volume increases with writes that are performed on the source volume, on not already copied grains. Eventually, if the entire source volume is written (unlikely), the real capacity of the target volume is identical to the source's volume.

Source and target thin-provisioned

When the source and target volumes are thin-provisioned, only the data that is allocated to the source is copied to the target. In this configuration, the background copy option has no effect.

Performance: The best performance is obtained when the grain size of the thin-provisioned volume is the same as the grain size of the FlashCopy mapping.

Thin-provisioned incremental FlashCopy

The implementation of thin-provisioned volumes does not preclude the use of incremental FlashCopy on the same volumes. It does not make sense to have a fully allocated source volume and then use incremental FlashCopy (which is always a full copy the first time) to copy this fully allocated source volume to a thin-provisioned target volume. However, this action is not prohibited.

Consider the following optional configurations:

- A thin-provisioned source volume can be copied incrementally by using FlashCopy to a thin-provisioned target volume. Whenever the FlashCopy is performed, only data that was modified is recopied to the target. If space is allocated on the target because of I/O to the target volume, this space is not reclaimed with subsequent FlashCopy operations.
- A fully allocated source volume can be copied incrementally by using FlashCopy to another fully allocated volume at the same time as it is being copied to multiple thin-provisioned targets (taken at separate points in time). By using this combination, a single full backup can be kept for recovery purposes, and the backup workload is separated from the production workload. At the same time, older thin-provisioned backups can be retained.

10.3.17 Serialization of I/O by FlashCopy

In general, the FlashCopy function in the IBM Storage Virtualize introduces no explicit serialization into the I/O path. Therefore, many concurrent I/Os are allowed to the source and target volumes.

However, a lock exists for each grain and this lock can be in shared or exclusive mode. For multiple targets, a common lock is shared, and the mappings are derived from a particular source volume. The lock is used in the following modes under the following conditions:

- The lock is held in shared mode during a read from the target volume, which touches a grain that was not copied from the source.
- The lock is held in exclusive mode while a grain is being copied from the source to the target.

If the lock is held in shared mode and another process wants to use the lock in shared mode, this request is granted unless a process is waiting to use the lock in exclusive mode.

If the lock is held in shared mode and it is requested to be exclusive, the requesting process must wait until all holders of the shared lock free it.

Similarly, if the lock is held in exclusive mode, a process wanting to use the lock in shared or exclusive mode must wait for it to be freed.

10.3.18 Event handling

When a FlashCopy mapping is not copying or stopping, the FlashCopy function does not affect the handling or reporting of events for error conditions that are encountered in the I/O path. Event handling and reporting are affected only by FlashCopy when a FlashCopy mapping is copying or stopping that is actively moving data.

These scenarios are described next,

Node failure

Normally, two copies of the FlashCopy bitmap are maintained. One copy of the FlashCopy bitmap is on each of the two nodes that make up the I/O group of the source volume. When a node fails, one copy of the bitmap for all FlashCopy mappings whose source volume is a member of the failing node's I/O group becomes inaccessible.

FlashCopy continues with a single copy of the FlashCopy bitmap that is stored as non-volatile in the remaining node in the source I/O group. The system metadata is updated to indicate that the missing node no longer holds a current bitmap. When the failing node recovers or a replacement node is added to the I/O group, the bitmap redundancy is restored.

Path failure (Path Offline state)

In a fully functioning system, all of the nodes have a software representation of every volume in the system within their application hierarchy.

Because the storage area network (SAN) that links IBM Storage Virtualize nodes to each other and to the MDisks is made up of many independent links, it is possible for a subset of the nodes to be temporarily isolated from several of the MDisks. When this situation occurs, the MDisks are said to be *Path Offline* on certain nodes.

Other nodes: Other nodes might see the MDisks as online because their connection to the MDisks still exists.

Path Offline for the source volume

If a FlashCopy mapping is in the `copying` state and the source volume goes path offline, this path offline state is propagated to all target volumes up to, but not including, the target volume for the newest mapping that is 100% copied but remains in the `copying` state. If no mappings

are 100% copied, all of the target volumes are taken offline. `Path offline` is a state that exists on a per-node basis. Other nodes might not be affected. If the source volume comes online, the target and source volumes are brought back online.

Path Offline for the target volume

If a target volume goes path offline but the source volume is still online and if any dependent mappings exist, those target volumes also go path offline. The source volume remains online.

10.3.19 Asynchronous notifications

FlashCopy raises informational event log entries for certain mapping and consistency group state transitions. These state transitions occur as a result of configuration events that complete asynchronously. The informational events can be used to generate Simple Network Management Protocol (SNMP) traps to notify the user.

Other configuration events complete synchronously, and no informational events are logged as a result of the following events:

- `PREPARE_COMPLETED`

  This state transition is logged when the FlashCopy mapping or consistency group enters the `prepared` state as a result of a user request to prepare. The user can now start (or stop) the mapping or consistency group.

- `COPY_COMPLETED`

  This state transition is logged when the FlashCopy mapping or consistency group enters the `idle_or_copied` state when it was in the `copying` or `stopping` state. This state transition indicates that the target disk now contains a complete copy and no longer depends on the source.

- `STOP_COMPLETED`

  This state transition is logged when the FlashCopy mapping or consistency group enters the `stopped` state as a result of a user request to stop. It is logged after the automatic copy process completes. This state transition includes mappings where no copying needed to be performed. This state transition differs from the event that is logged when a mapping or group enters the `stopped` state as a result of an I/O error.

10.3.20 Interoperation with Metro Mirror and Global Mirror

A volume can be part of any copy relationship; that is, FlashCopy, Metro Mirror (MM]), or Remote Mirror. Therefore, FlashCopy can work with MM/Global Mirror (GM) to provide better protection of the data.

For example, you can perform an MM copy to duplicate data from Site_A to Site_B, and then perform a daily FlashCopy to back up the data to another location.

Note: A volume cannot be part of FlashCopy, MM, or Remote Mirror, if it is set to Transparent Cloud Tiering (TCT) function.

Table 10-8 on page 751 lists the supported combinations of FlashCopy and Remote Copy (RC). In the table, “RC” refers to MM and GM.

Table 10-8 FlashCopy and remote copy interaction

Component	RC primary site	RC secondary site
FlashCopy Source	Supported	Supported latency: When the FlashCopy relationship is in the `preparing` and `prepared` states, the cache at the RC secondary site operates in write-through mode. This process adds latency to the latent RC relationship.
FlashCopy Target	This is a supported combination and has the following restrictions: ► Running a **`stop -force`** might cause the RC relationship to be fully resynchronized. ► Code level must be 6.2.*x* or later. ► I/O group must be the same.	This is a supported combination with the major restriction that the FlashCopy mapping cannot be `copying`, `stopping`, or `suspended`. Otherwise, the restrictions are the same as at the RC primary site.

10.3.21 FlashCopy attributes and limitations

The FlashCopy function in IBM Storage Virtualize features the following attributes:

- The target is the T0 copy of the source, which is known as *FlashCopy mapping target*.
- FlashCopy produces an exact copy of the source volume, including any metadata that was written by the host operating system, Logical Volume Manager (LVM), and applications.
- The source volume and target volume are available (almost) immediately following the FlashCopy operation.
- The source and target volumes:
  - Must be the same “virtual” size
  - Must be on the same IBM Storage Virtualize system
  - Do not need to be in the same I/O group or storage pool, although it is recommended for them to have the same preferred node for best performance
- The storage pool extent sizes can differ between the source and target.
- The target volumes can be the source volumes for other FlashCopy mappings (*cascaded FlashCopy*). However, a target volume can have only one source copy.
- Consistency groups are supported to enable FlashCopy across multiple volumes at the same time.
- The target volume can be updated independently of the source volume.
- Bitmaps that are governing I/O redirection (I/O indirection layer) are maintained in both nodes of the IBM Storage Virtualize I/O group to prevent a single point of failure (SPOF).
- FlashCopy mapping and consistency groups can be automatically withdrawn after the completion of the background copy.
- Thin-provisioned FlashCopy (or Snapshot in the GUI) use disk space only when updates are made to the source or target data, and not for the entire capacity of a volume copy.
- FlashCopy licensing is based on the virtual capacity of the source volumes.

- Incremental FlashCopy copies all of the data when you first start FlashCopy, and then only the changes when you stop and start FlashCopy mapping again. Incremental FlashCopy can substantially reduce the time that is required to re-create an independent image.
- Reverse FlashCopy enables FlashCopy targets to become restore points for the source without breaking the FlashCopy relationship, and without having to wait for the original copy operation to complete.
- The size of the source and target volumes cannot be altered (increased or decreased) while a FlashCopy mapping is defined.

IBM FlashCopy limitations for IBM Storage Virtualize V8.6.0 are listed in Table 10-9.

Table 10-9 FlashCopy limitations in V8.6.0

Property	Maximum number
FlashCopy mappings per system	8192[1], 15864[2]
FlashCopy targets per source	256
FlashCopy mappings per consistency group	512
FlashCopy consistency groups per system	500
Total FlashCopy volume capacity per I/O group	4096 TiB
FlashCopy relationships per graph (backups per source)	256

1. Applies to IBM FlashSystem 5x00
2. Applies to IBM SAN Volume Controller DH8, SV2, SV3, SA2, FS7XXX, FS9XXX

10.3.22 Expanding Volumes In a FlashCopy Mapping

Volumes that are part of a FlashCopy mapping that is user defined can now be expanded. User-defined mappings are mappings that are manually created. Volume expansion features the following limitations:

- The source or the target volume in the mapping can be expanded at any time.
- For incremental FlashCopy mappings the target volume must be expanded before the source volume.
- Source and target volumes must be the same size when the mapping is prepared or started.
- The target volume cannot be shrunk after expansion.

10.4 Managing FlashCopy by using the GUI

It is often easier to work with the FlashCopy function from the GUI if you have a reasonably small number of host mappings. However, in enterprise data centers with many host mappings, it can be beneficial to use the CLI to run your FlashCopy commands.

10.4.1 FlashCopy presets

The IBM Storage Virtualize GUI interface provides three FlashCopy presets (Snapshot, Clone, and Backup) to simplify the more common FlashCopy operations.

Although these presets meet most FlashCopy requirements, they do not support all possible FlashCopy options. If more specialized options are required that are not supported by the presets, the options must be performed by using CLI commands.

This section describes the preset options and their use cases.

Snapshot

This preset creates a Point in Time copy that tracks only the changes that are made at the source or target volumes. The snapshot is not intended to be an independent copy. Instead, the copy is used to maintain a view of the production data at the time that the snapshot is created. Therefore, the snapshot holds only the data from regions of the production volume that changed since the snapshot was created. Because the snapshot preset uses thin provisioning, only the capacity that is required for the changes is used.

Snapshot uses the following preset parameters:

- Background copy: `None`
- Snapshot initial volume real-size: `Zero`
- Incremental: `No`
- Delete after completion: `No`
- Cleaning rate: `No`
- Snapshot pool: `Source volume pool`
- Snapshot preferred node: `Source volume node`

Use case

The user wants to produce a copy of a volume without affecting the availability of the volume. The user does not anticipate many changes to be made to the source or target volume; a significant proportion of the volumes remains unchanged.

By ensuring that only changes require a copy of data to be made, the total amount of disk space that is required for the copy is reduced. Therefore, many Snapshot copies can be used in the environment.

Snapshots are useful for providing protection against corruption or similar issues with the validity of the data, but they do not provide protection from physical controller failures. Snapshots can also provide a vehicle for performing repeatable testing (including “what-if” modeling that is based on production data) without requiring a full copy of the data to be provisioned.

For example, in Figure 10-16 on page 754, the source volume user can still work on the original data volume (as with a production volume) and the target volumes can be accessed instantly. Users of target volumes can modify the content and perform “what-if” tests; for example, versioning. Storage administrators do not need to perform full copies of a volume for temporary tests. However, the target volumes must remain linked to the source. When the link is broken (FlashCopy mapping that is stopped or deleted), the target volumes become unusable.

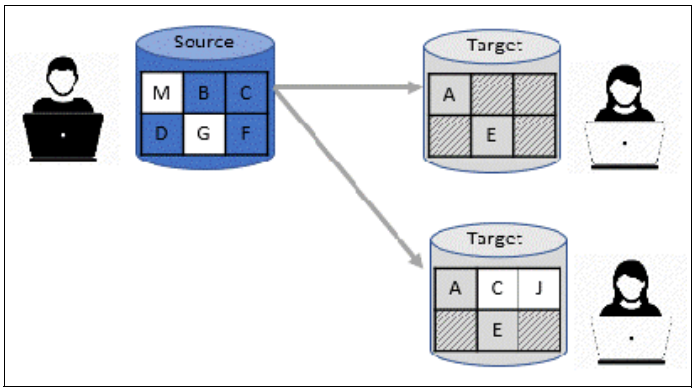


Figure 10-16 FlashCopy snapshot preset example

Clone

The clone preset creates a replica of the volume, which can be changed without affecting the original volume. After the copy completes, the mapping that was created by the preset is automatically deleted.

Clone uses the following preset parameters:

- Background copy rate: `50`
- Clone volume real-size: `Same as source volume`
- Incremental: `No`
- Delete after completion: `Yes`
- Cleaning rate: `50`
- Clone pool: `Source volume pool`
- Clone preferred node: `Source volume node`

Use case

Users want a copy of the volume that they can modify without affecting the original volume. After the clone is established, it is not expected that it is refreshed or that the original production data must be referenced again. If the source is thin-provisioned, the target is thin-provisioned for the auto-create target.

Backup

The backup preset creates an incremental PiT replica of the production data. After the copy completes, the backup view can be refreshed from the production data, with minimal copying of data from the production volume to the backup volume.

Backup uses the following preset parameters:

- Background Copy rate: `50`
- Backup volume real-size: `Same as source volume`
- Incremental: `Yes`
- Delete after completion: `No`
- Cleaning rate: `50`
- Backup pool: `Source volume pool`
- Backup preferred node: `Source volume node`

Use case

The user wants to create a copy of the volume that can be used as a backup if the source becomes unavailable, such as because of loss of the underlying physical controller. The user plans to periodically update the secondary copy, and does not want to suffer from the resource demands of creating a copy each time.

Incremental FlashCopy times are faster than full copy, which helps to reduce the window where the new backup is not yet fully effective. If the source is thin-provisioned, the target is also thin-provisioned in this option for the auto-create target.

Another use case, which is not supported by the name, is to create and maintain (periodically refresh) an independent image that can be subjected to intensive I/O (for example, data mining) without affecting the source volume's performance.

Note: IBM Storage Virtualize in general and FlashCopy in particular are not backup solutions on their own. For example, FlashCopy backup preset does not schedule a regular copy of your volumes. Instead, it overwrites the mapping target and does not make a copy of it before starting a new "backup" operation. It is the user's responsibility to handle the target volumes (for example, saving them to tapes) and the scheduling of the FlashCopy operations.

10.4.2 FlashCopy window

This section describes the tasks that you can perform at a FlashCopy level by using the IBM Storage Virtualize GUI.

When the IBM Storage Virtualize GUI is used, FlashCopy components can be seen in different windows. Three windows are related to FlashCopy and are available by using the **Copy Services** menu, as shown in Figure 10-17.

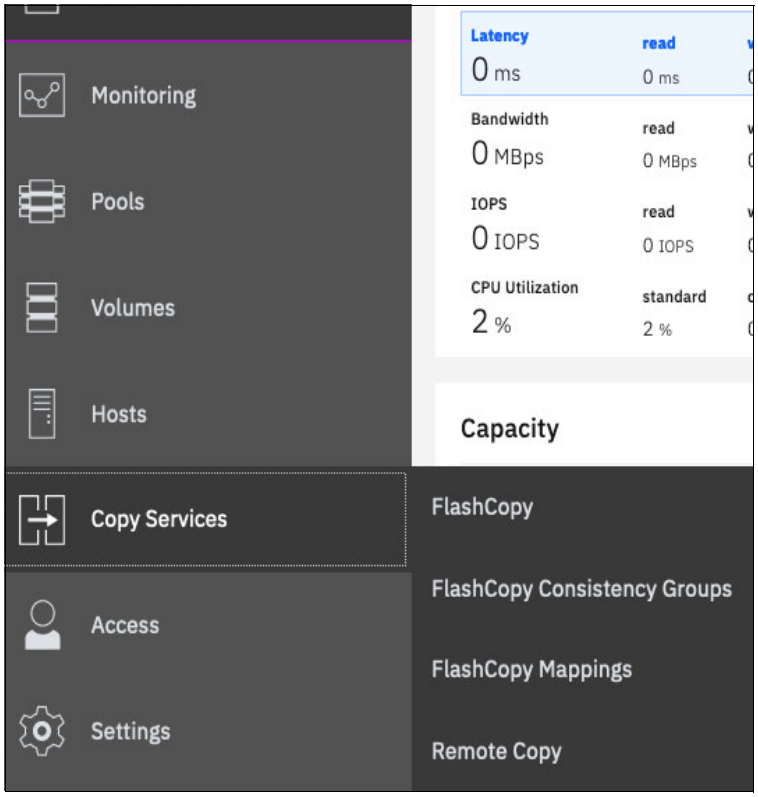


Figure 10-17 Copy Services menu

The FlashCopy window is accessible by clicking **Copy Services** → **FlashCopy**. It displays all of the volumes that are defined in the system. Volumes that are part of a FlashCopy mapping

appear, as shown in Figure 10-18. By clicking a source volume, you can display the list of its target volumes.

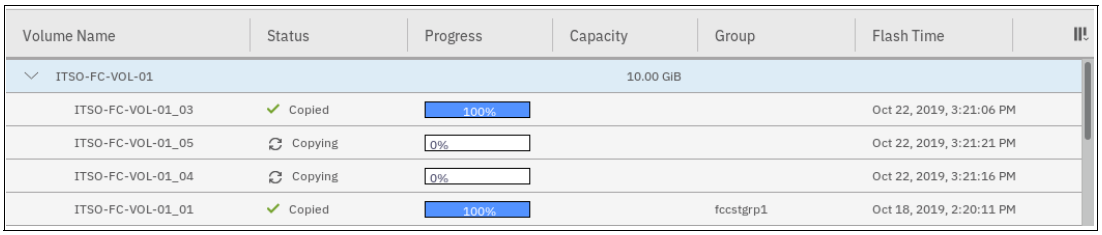


Figure 10-18 Source and target volumes that are displayed in the FlashCopy window

All volumes are listed in this window, and target volumes appear twice (as a regular volume and as a target volume in a FlashCopy mapping).

Consider the following points:

- The Consistency Group window is accessible by clicking **Copy Services** → **FlashCopy Consistency Groups**. Use the FlashCopy Consistency Groups window (as shown in Figure 10-19) to list the FlashCopy mappings that are part of consistency groups and part of no consistency groups.

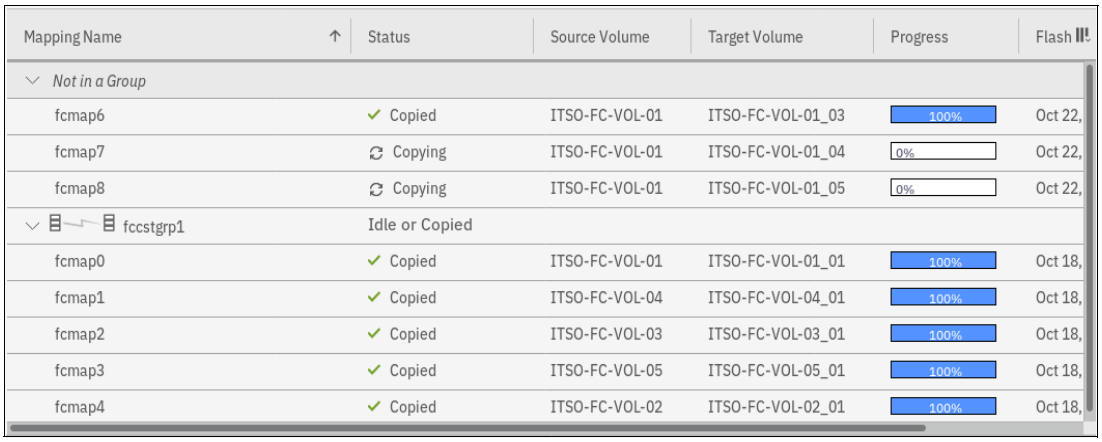


Figure 10-19 FlashCopy Consistency Groups window

- The FlashCopy Mappings window is accessible by clicking **Copy Services** → **FlashCopy Mappings**. Use the FlashCopy Mappings window (as shown in Figure 10-20 on page 757) to display the list of mappings between source volumes and target volumes.

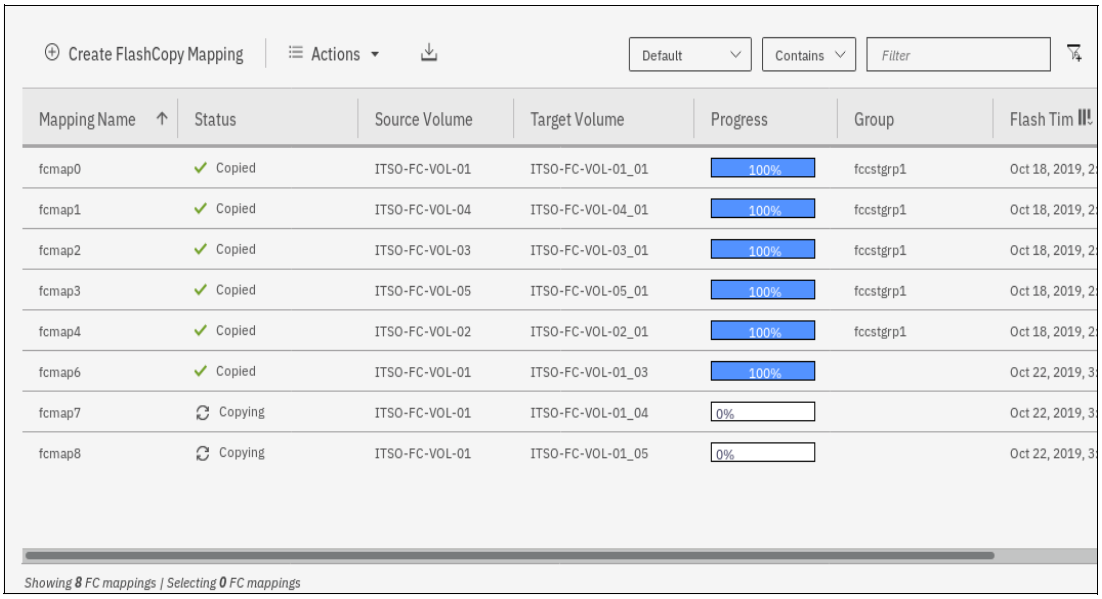


Figure 10-20 FlashCopy Mapping window

10.4.3 Creating a FlashCopy mapping

This section describes creating FlashCopy mappings for volumes and their targets.

Open the FlashCopy window from the **Copy Services** menu, as shown in Figure 10-21 on page 758. Select the volume for which you want to create the FlashCopy mapping. Right-click the volume or click the **Actions** menu.

Multiple FlashCopy mappings: To create multiple FlashCopy mappings at the same time, select multiple volumes by pressing and holding **Ctrl** and clicking the entries that you want.

Depending on whether you created the target volumes for your FlashCopy mappings or you want the system to create the target volumes for you, the following options are available:

- If you created the target volumes, see “Creating a FlashCopy mapping with existing target volumes” on page 758.
- If you want the system to create the target volumes for you, see “Creating a FlashCopy mapping and target volumes” on page 763.

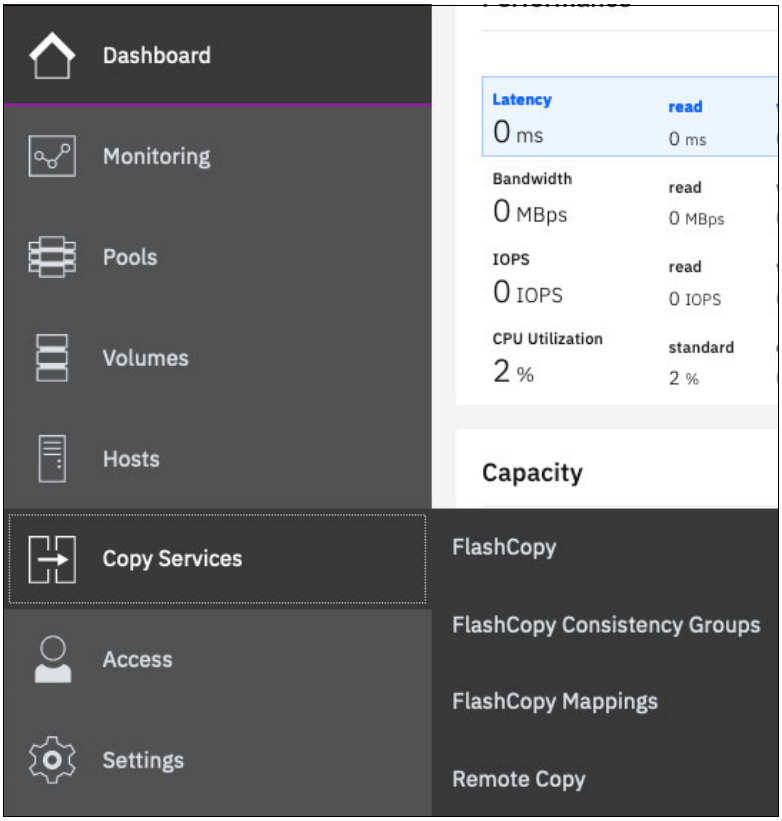


Figure 10-21 FlashCopy window

Creating a FlashCopy mapping with existing target volumes

Complete the following steps to use existing target volumes for the FlashCopy mappings:

Attention: When starting a FlashCopy mapping from a source volume to a target volume, data that is on the target is over-written. The system does not prevent you from selecting a target volume that is mapped to a host and contains data.

1. Right-click the volume that you want to create a FlashCopy mapping for, and select **Advanced FlashCopy** → **Use Existing Target Volumes**, as shown in Figure 10-22 on page 759.
2. When the Create FlashCopy Mapping window opens (Figure 10-23 on page 759,) create the mapping between the selected source volume and the target volume with which you want to create a mapping. Click **Add**.

Important: The source volume and the target volume must be of equal size. Therefore, only targets of the same size are shown in the list for a source volume.

Volumes that are a target in a FlashCopy mapping cannot be a target in a new mapping. Therefore, only volumes that are not targets can be selected.

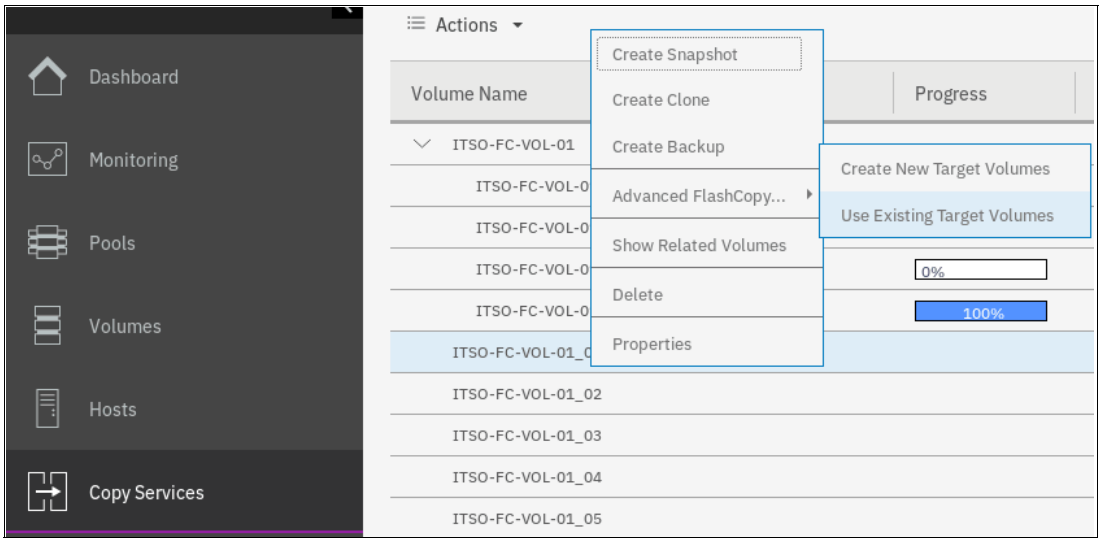


Figure 10-22 Creating a FlashCopy mapping with an existing target

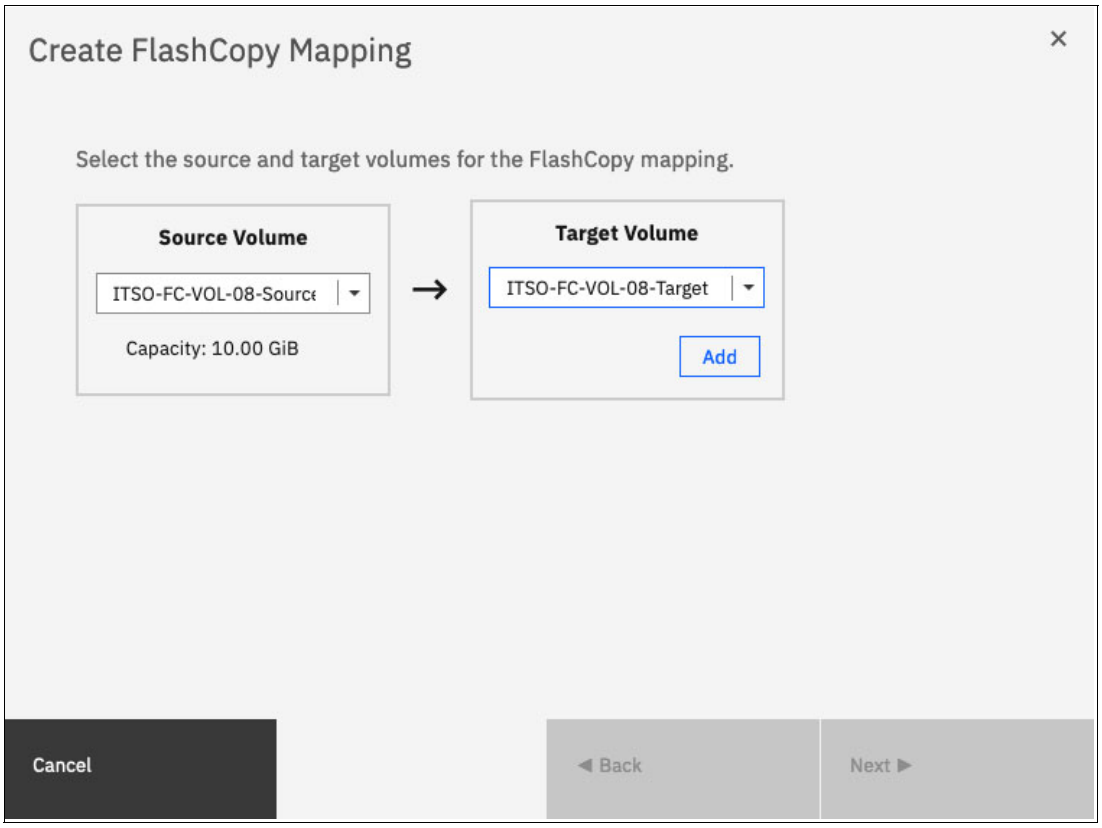


Figure 10-23 Selecting source and target for a FlashCopy mapping

3. To remove a mapping that was created, click ✖ (Figure 10-24 on page 760).

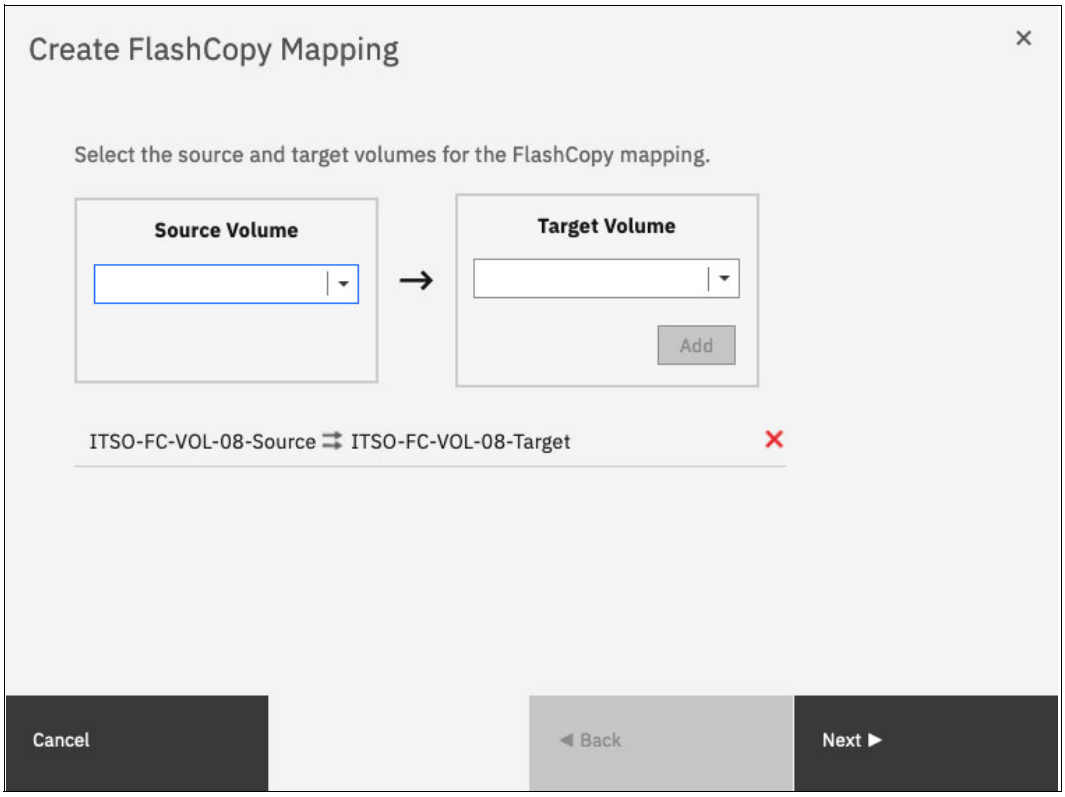


Figure 10-24 Viewing source and target at creation time

4. Click **Next** after you create all of the mappings that you need (see Figure 10-24).
5. In the next window, select one FlashCopy preset. The GUI provides the following presets to simplify common FlashCopy operations, as shown in Figure 10-25 on page 761. For more information about the presets, see 10.4.1, “FlashCopy presets” on page 752:
   - Snapshot: Creates a PiT snapshot copy of the source volume.
   - Clone: Creates a PiT replica of the source volume.
   - Backup: Creates an incremental FlashCopy mapping that can be used to recover data or objects if the system experiences data loss. These backups can be copied multiple times from source and target volumes.

Note: If you want to create a simple Snapshot of a volume, you likely want the target volume to be defined as thin-provisioned to save space on your system. If you use an existing target, ensure that it is thin-provisioned first. The use of the Snapshot preset does not make the system check whether the target volume is thin-provisioned.

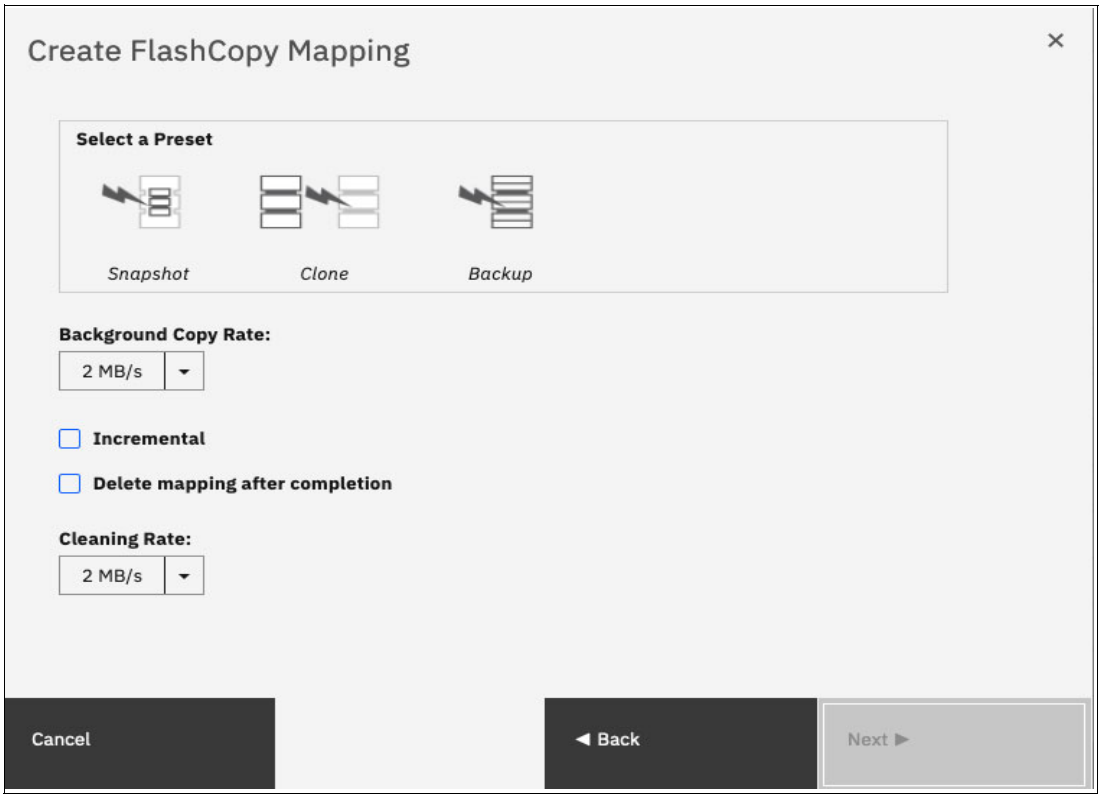


Figure 10-25 FlashCopy mapping preset selection

When selecting a preset, some options, such as Background Copy Rate, Incremental, and Delete mapping after completion, are automatically changed or selected. You can still change the automatic settings, but this is not recommended for the following reasons:

- If you select the **Backup** preset but then clear **Incremental** or select **Delete mapping after completion**, you lose the benefits of the incremental FlashCopy and must copy the entire source volume each time you start the mapping.
- If you select the **Snapshot** preset but then change the **Background Copy Rate**, you with a full copy of your source volume.

For more information about the Background Copy Rate and the Cleaning Rate, see Table 10-2 on page 734, or Table 10-6 on page 743.

When your FlashCopy mapping setup is ready, click **Next**.

6. You can choose whether to add the mappings to a consistency group, as shown in Figure 10-26.

   If you want to include this FlashCopy mapping in a consistency group, select **Yes, add the mappings to a consistency group** and select the consistency group from the drop-down menu.

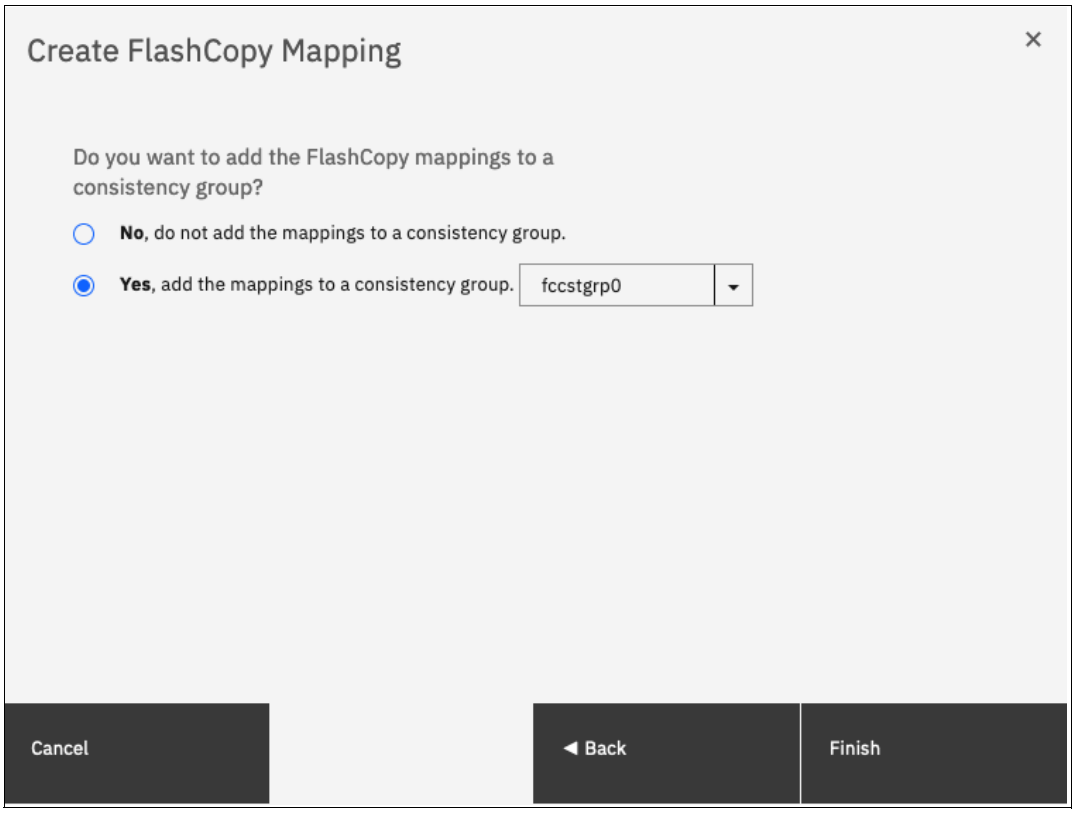


Figure 10-26 Select or not a consistency group for the FlashCopy mapping

7. It is possible to add a FlashCopy mapping to a consistency group or to remove a FlashCopy mapping from a consistency group after they are created. If you do not know at this stage what to do, you can change it later. Click **Finish**.

The FlashCopy mapping is now ready for use. It is visible in the three different windows: FlashCopy, FlashCopy mappings, and consistency groups.

Note: Creating a FlashCopy mapping does *not* automatically start any copy. You must manually start the mapping.

Creating a FlashCopy mapping and target volumes

Complete the following steps to create target volumes for FlashCopy mapping:

1. Right-click the volume that you want to create a FlashCopy mapping for and select **Advanced FlashCopy** → **Create New Target Volumes**, as shown in Figure 10-27.

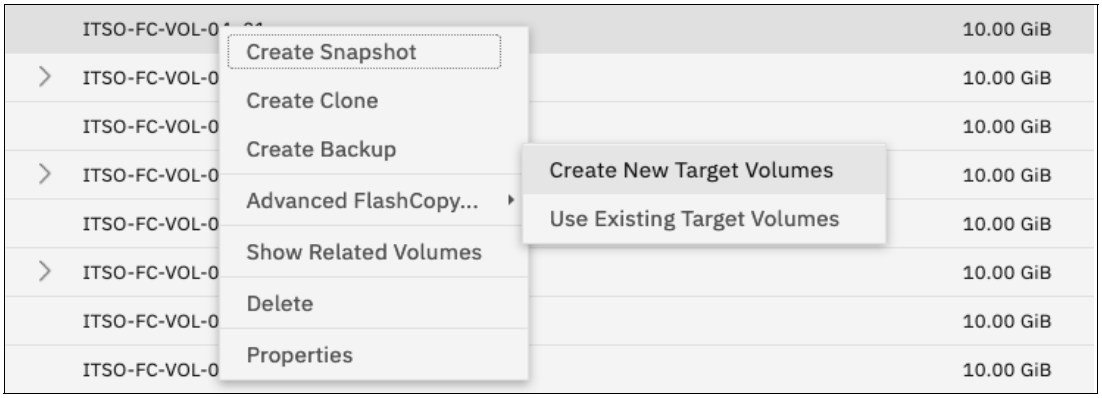


Figure 10-27 Creating a FlashCopy mapping and creating targets

2. In the next window, select one FlashCopy preset. The GUI provides the following presets to simplify common FlashCopy operations, as shown in Figure 10-28 on page 764. For more information about the presets, see 10.4.1, “FlashCopy presets” on page 752.
   - Snapshot: Creates a PiT snapshot copy of the source volume.
   - Clone: Creates a PiT replica of the source volume.
   - Backup: Creates an incremental FlashCopy mapping that can be used to recover data or objects if the system experiences data loss. These backups can be copied multiple times from source and target volumes.

Note: If you want to create a simple Snapshot of a volume, you likely want the target volume to be defined as thin-provisioned to save space on your system. If you use an existing target, ensure it is thin-provisioned first. The use of the Snapshot preset does not make the system check whether the target volume is thin-provisioned.

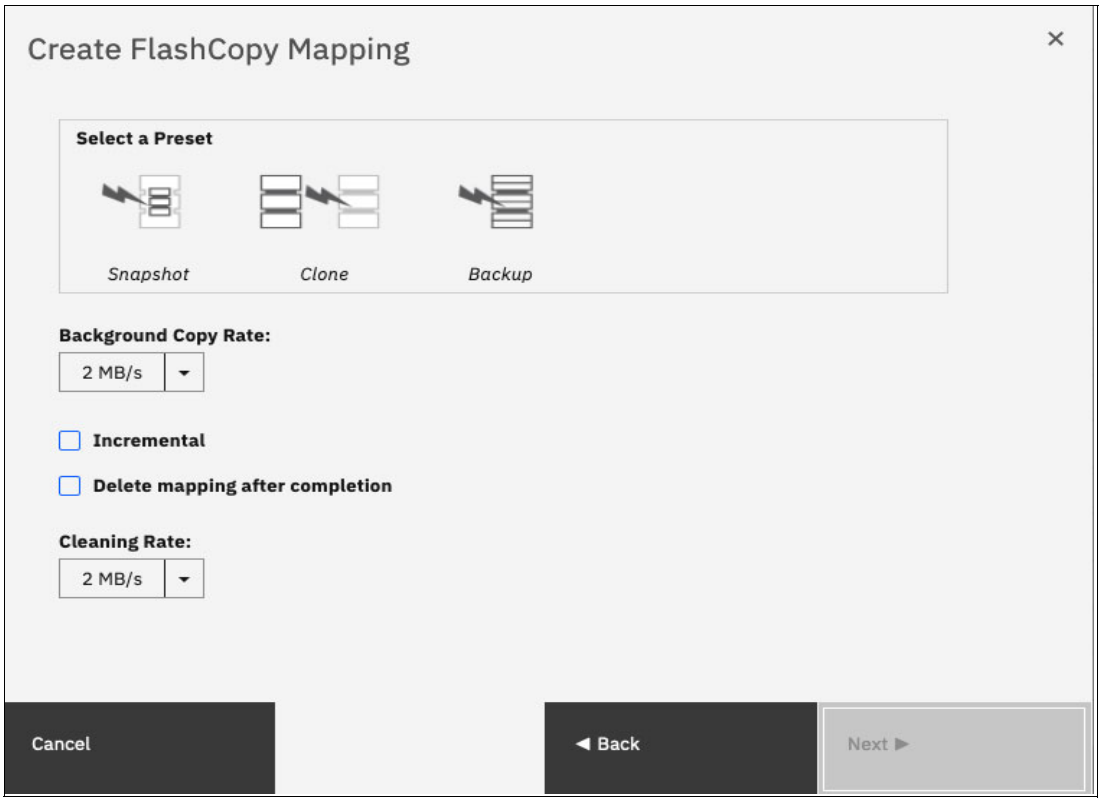


Figure 10-28 FlashCopy mapping preset selection

When selecting a preset, some options, such as Background Copy Rate, Incremental, and Delete mapping, after completion are automatically changed or selected. You can still change the automatic settings, but this is not recommended for the following reasons:

- If you select the **Backup** preset but then clear **Incremental** or select **Delete mapping after completion**, you lose the benefits of the incremental FlashCopy. You must copy the entire source volume each time you start the mapping.
- If you select the **Snapshot** preset but then change the **Background Copy Rate**, you have a full copy of your source volume.

For more information about the Background Copy Rate and the Cleaning Rate, see Table 10-2 on page 734, or Table 10-6 on page 743.

When your FlashCopy mapping setup is ready, click **Next**.

3. You can choose whether to add the mappings to a consistency group, as shown in Figure 10-29 on page 765.

   If you want to include this FlashCopy mapping in a consistency group, select **Yes, add the mappings to a consistency group**, and select the consistency group from the drop-down menu.

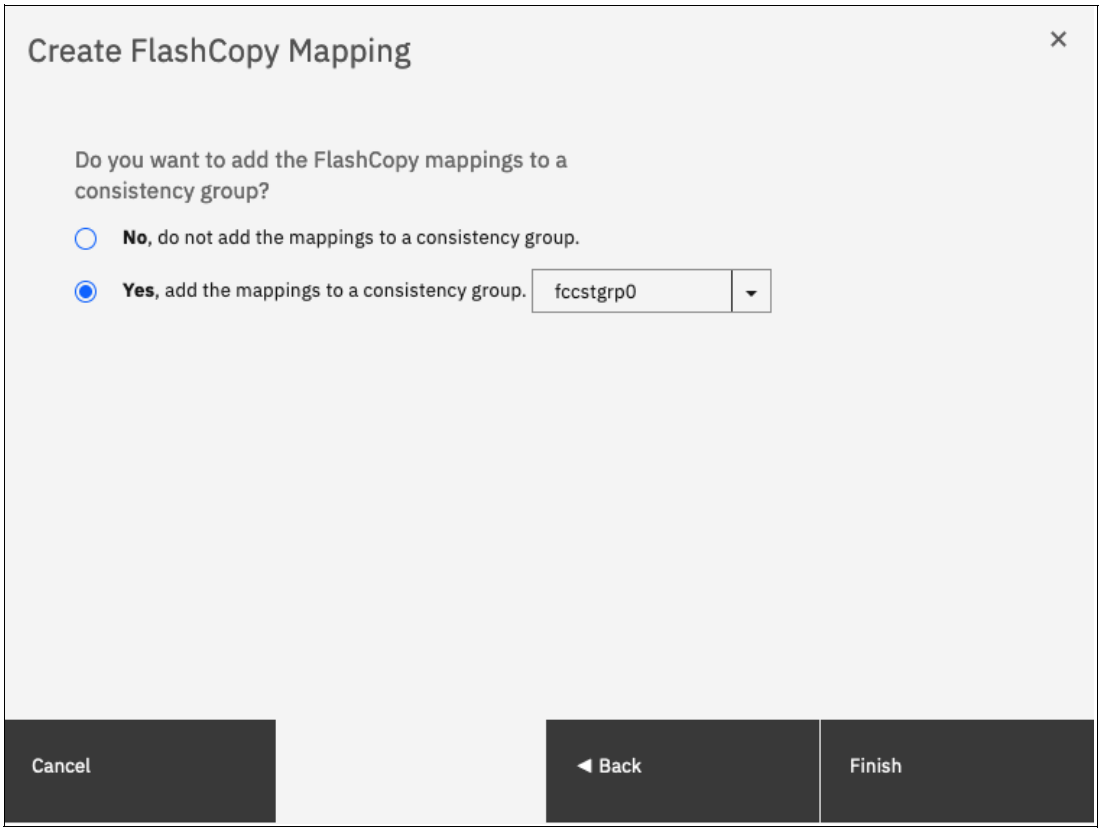


Figure 10-29 Select a consistency group for the FlashCopy mapping

4. It is possible to add a FlashCopy mapping to a consistency group or to remove a FlashCopy mapping from a consistency group after they are created. If you do not know at this stage what to do, you can change it later. Click **Next**.
5. The system prompts the user to select the pool that is used to automatically create targets, as shown in Figure 10-30 on page 766. Click **Next**.
6. The system prompts the user how to define the new volumes that are created, as shown in Figure 10-31 on page 766. It can be None, Thin-provisioned, or Inherit from source volume. If Inherit from source volume is selected, the system checks the type of the source volume, and then creates a target of the same type. Click **Finish**.

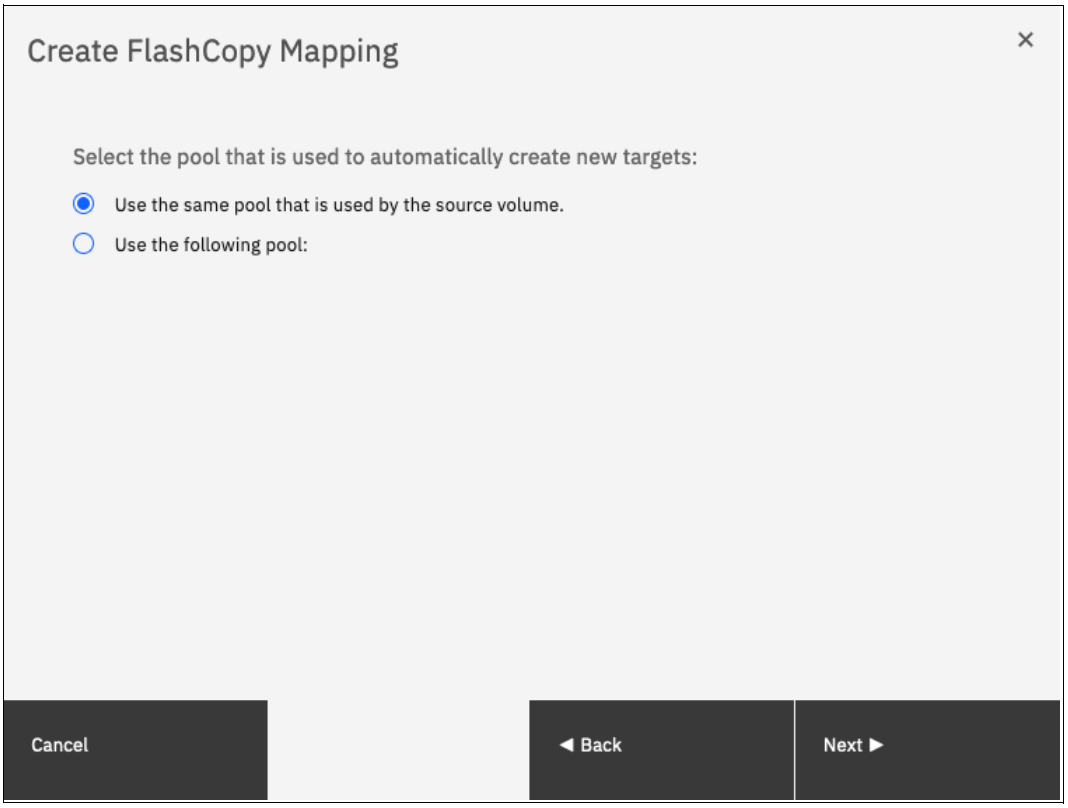


Figure 10-30 Select the pool

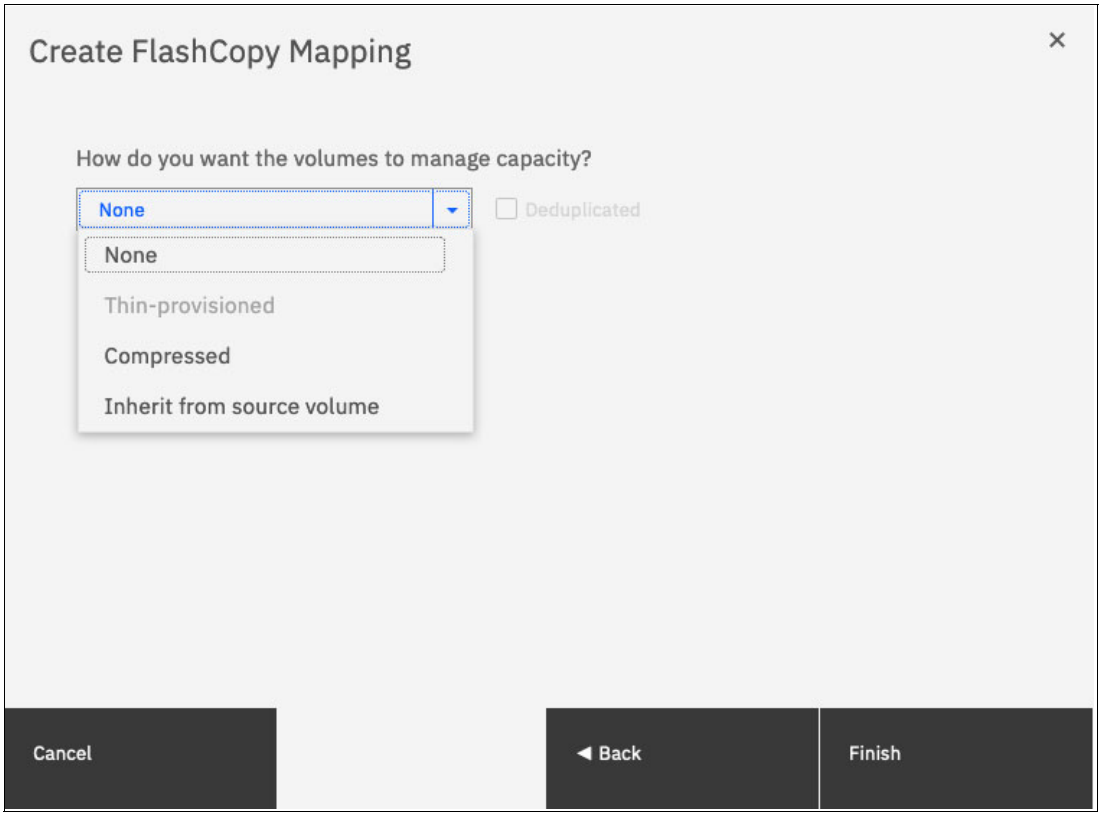


Figure 10-31 Select the type of volumes for the created targets

Note: If you selected multiple source volumes to create FlashCopy mappings, selecting **Inherit properties from source Volume** applies to each newly created target volume. For example, if you selected a compressed volume and a generic volume as sources for the new FlashCopy mappings, the system creates a compressed target and a generic target.

The FlashCopy mapping is now ready for use. It is visible in the three different windows: FlashCopy, FlashCopy mappings, and consistency groups.

10.4.4 Single-click snapshot

The snapshot creates a PiT backup of production data. The snapshot is not intended to be an independent copy. Instead, it is used to maintain a view of the production data at the time that the snapshot is created. Therefore, the snapshot holds only the data from regions of the production volume that changed since the snapshot was created. Because the snapshot preset uses thin provisioning, only the capacity that is required for the changes is used.

Snapshot uses the following preset parameters:

- Background copy: `No`
- Incremental: `No`
- Delete after completion: `No`
- Cleaning rate: `No`
- Primary copy source pool: `Target pool`
- Snapshot initial volume real-size: `Zero`

To create and start a snapshot, complete the following steps:

1. Open the FlashCopy window from the **Copy Services** → **FlashCopy** menu.
2. Select the volume that you want to create a snapshot of, and right-click it or click **Actions** → **Create Snapshot**, as shown in Figure 10-32.

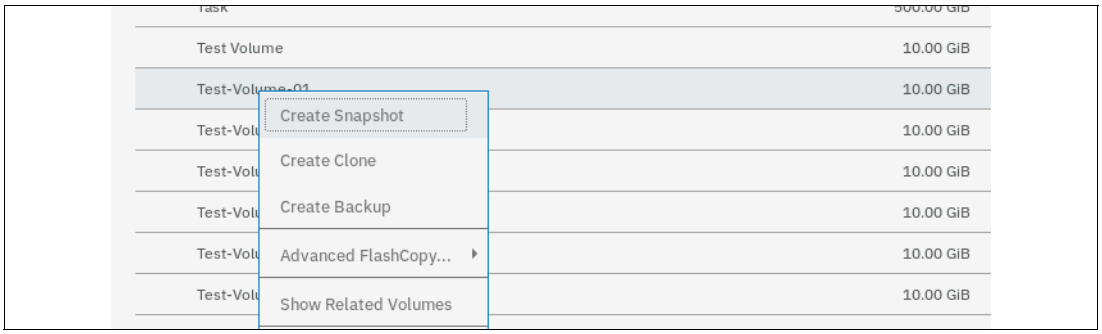


Figure 10-32 Single-click snapshot creation and start

3. You can select multiple volumes at a time, which creates as many snapshots automatically. The system then automatically groups the FlashCopy mappings in a new consistency group, as shown in Figure 10-33 on page 768.

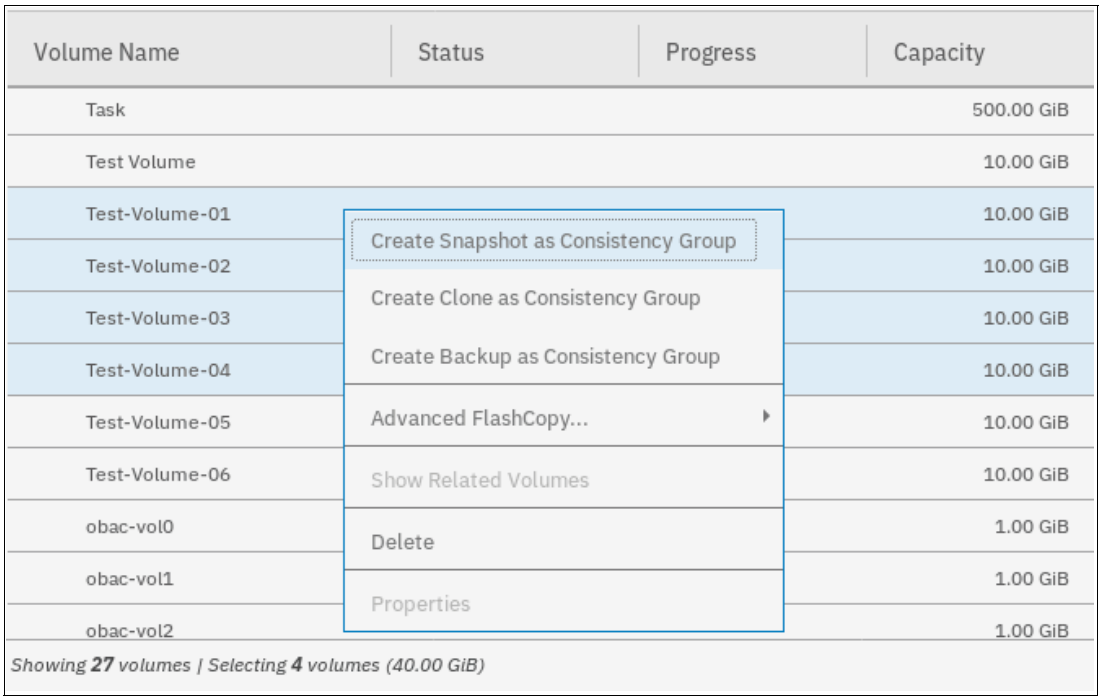


Figure 10-33 Selection single-click snapshot creation and start

For each selected source volume, the following actions occur:

- A FlashCopy mapping is automatically created. It is named by default `fcmapXX`.
- A target volume is created. By default the source name is appended with a `_XX` suffix.
- A consistency group is created for each mapping, unless multiple volumes were selected. Consistency groups are named by default `fccstgrpX`.

The newly created consistency group is automatically started.

10.4.5 Single-click clone

The *clone preset* creates a replica of the volume, which can be changed without affecting the original volume. After the copy completes, the mapping that was created by the preset is automatically deleted.

The clone preset uses the following parameters:

- Background copy rate: `50`
- Incremental: `No`
- Delete after completion: `Yes`
- Cleaning rate: `50`
- Primary copy source pool: `Target pool`
- Volume real-size: `Same as source`

To create and start a snapshot, complete the following steps:

1. Open the FlashCopy window from the **Copy Services** → **FlashCopy** menu.
2. Select the volume that you want to create a snapshot of, and right-click it or click **Actions** → **Create Clone**, as shown in Figure 10-34 on page 769.

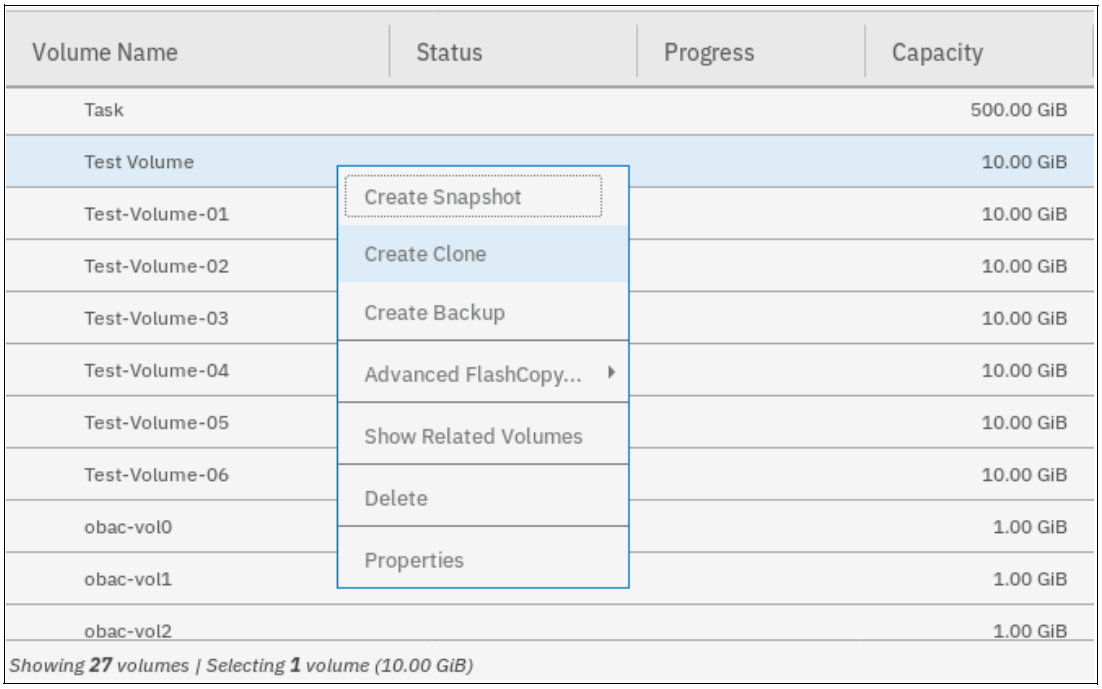


Figure 10-34 Single-click clone creation and start

3. You can select multiple volumes at a time, which creates as many snapshots automatically. The system then automatically groups the FlashCopy mappings in a new consistency group, as shown in Figure 10-35.

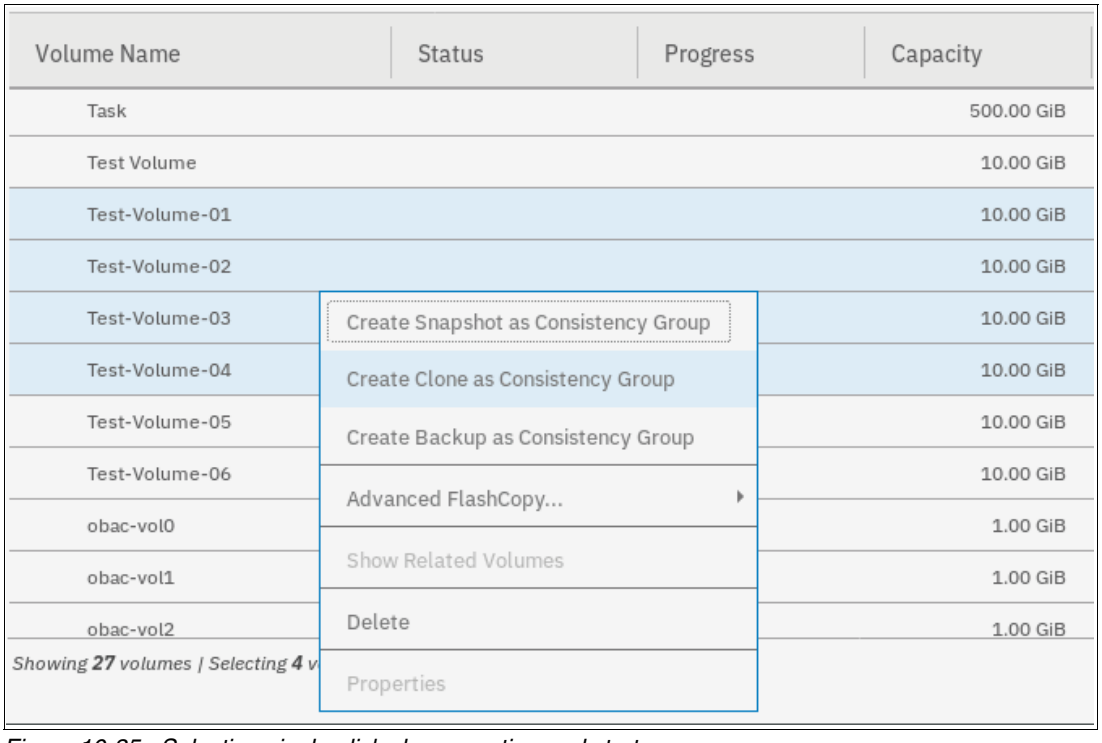


Figure 10-35 Selection single-click clone creation and start

For each selected source volume, the following actions occur:

- A FlashCopy mapping is automatically created. It is named by default `fcmapXX`.
- A target volume is created. The source name is appended with an `_XX` suffix.
- A consistency group is created for each mapping, unless multiple volumes were selected. Consistency groups are named by default `fccstgrpX`.
- The newly created consistency group is automatically started.

10.4.6 Single-click backup

The backup creates a PiT replica of the production data. After the copy completes, the backup view can be refreshed from the production data, with minimal copying of data from the production volume to the backup volume. The backup preset uses the following parameters:

- Background Copy rate: `50`
- Incremental: `Yes`
- Delete after completion: `No`
- Cleaning rate: `50`
- Primary copy source pool: `Target pool`
- Backup volume real-size: `Same as source`

To create and start a backup, complete the following steps:

1. Open the FlashCopy window from the **Copy Services** → **FlashCopy** menu.
2. Select the volume that you want to create a backup of, and right-click it or click **Actions** → **Create Backup**, as shown in Figure 10-36.

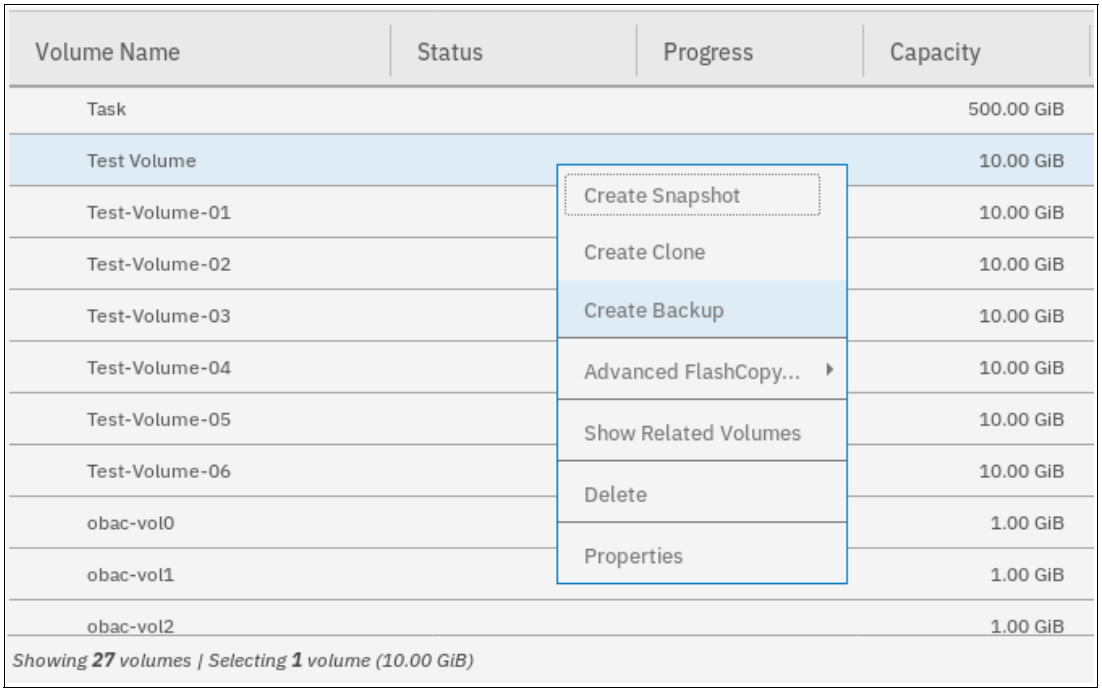


Figure 10-36 Single-click backup creation and start

3. You can select multiple volumes at a time, which creates as many snapshots automatically. The system then automatically groups the FlashCopy mappings in a new consistency group, as shown Figure 10-37 on page 771.

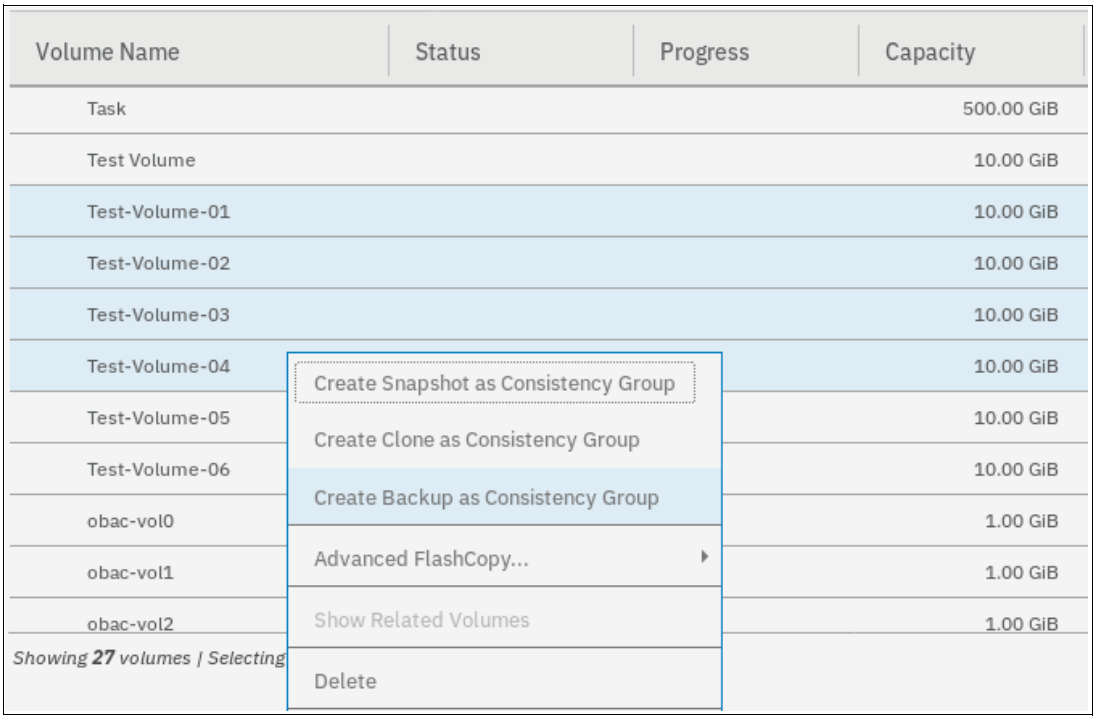


Figure 10-37 Selection single-click backup creation and start

For each selected source volume, the following actions occur:

- A FlashCopy mapping is automatically created. It is named by default `fcmapXX`.
- A target volume is created. It is named after the source name with a `_XX` suffix.
- A consistency group is created for each mapping, unless multiple volumes were selected. Consistency groups are named by default `fccstgrpX`.
- The newly created consistency group is automatically started.

10.4.7 Creating a FlashCopy consistency group

To create a FlashCopy consistency group in the GUI, complete the following steps:

1. Open the Consistency Groups window by clicking **Copy Services** → **FlashCopy Consistency Groups**. Click **Create Consistency Group**, as shown in Figure 10-38.

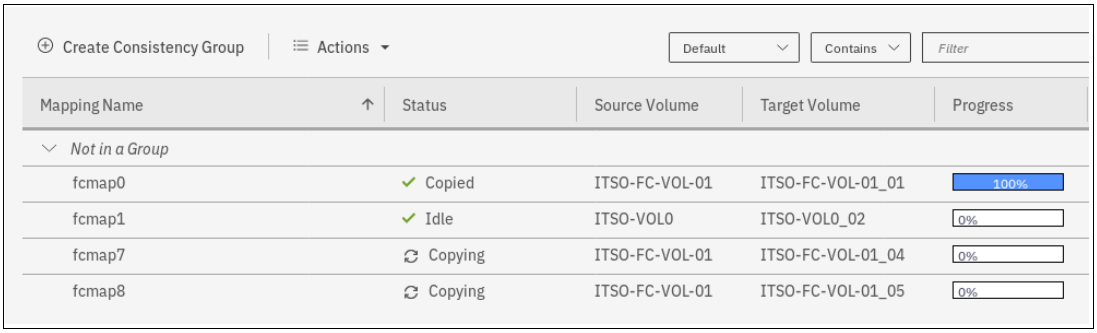


Figure 10-38 Creating a FlashCopy Consistency group

2. Enter the FlashCopy consistency group name that you want to use and the ownership group (optional) and then click **Create**, as shown in Figure 10-39.

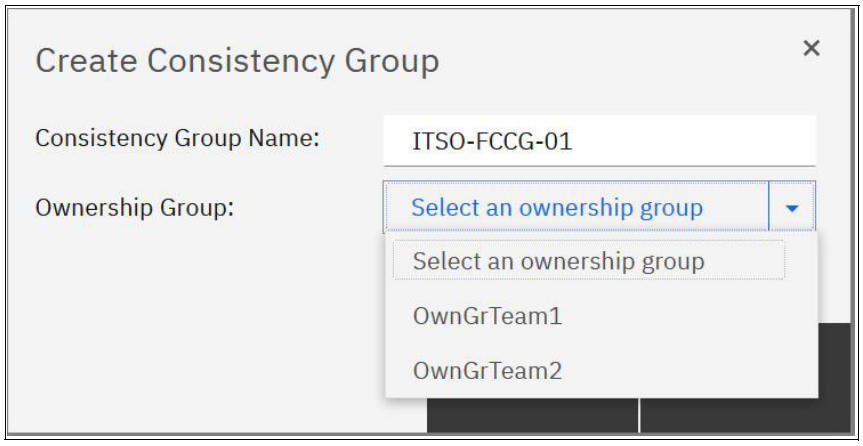


Figure 10-39 Enter the name and ownership group of the new consistency group

Consistency Group name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The volume name can be 1 - 63 characters.

10.4.8 Creating FlashCopy mappings in a consistency group

To create a FlashCopy Consistency Group in the GUI, complete the following steps:

1. Open the Consistency Groups window by clicking **Copy Services** → **FlashCopy Consistency Groups**. This example assumes that source and target volumes were previously created.
2. Select the consistency group in which you want to create the FlashCopy mapping. If you prefer not to create a FlashCopy mapping in a consistency group, select **Not in a Group**, and right-click the selected consistency group or click **Actions** → **Create FlashCopy Mapping**, as shown in Figure 10-40.

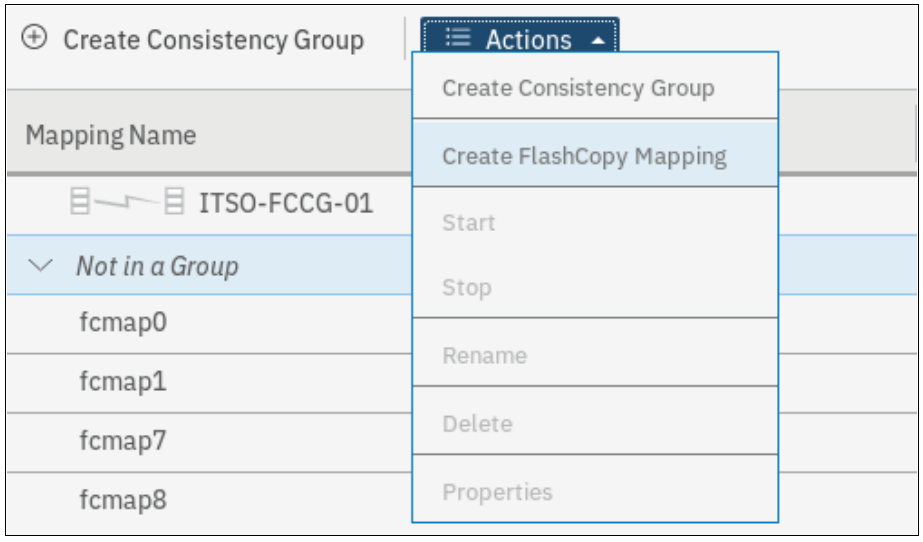


Figure 10-40 Creating a FlashCopy mapping

3. Select a volume in the source volume column by using the drop-down menu. Then select a volume in the target volume column by using the drop-down menu. Click **Add**, as shown in Figure 10-41.

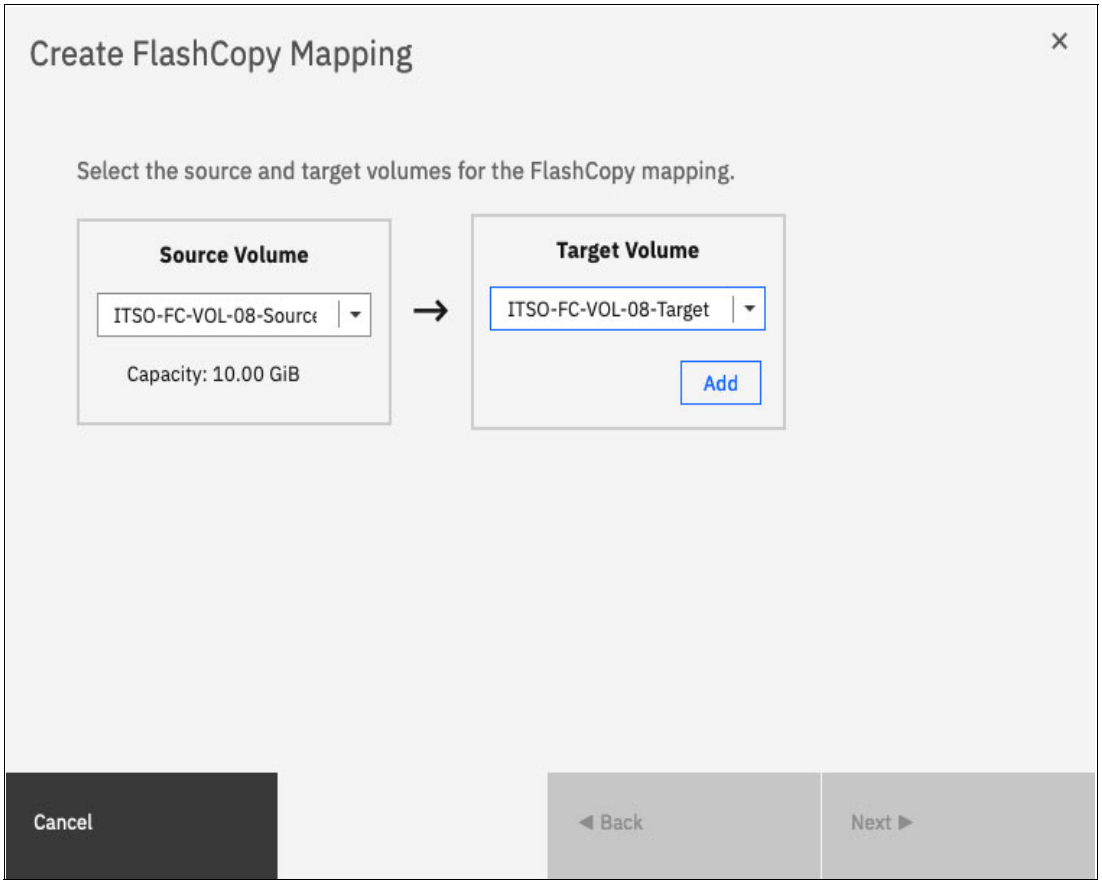


Figure 10-41 Select source and target volumes for the FlashCopy mapping

Repeat this step to create other mappings. To remove a mapping that was created, click ✖. Click **Next.**

> **Important:** The source and target volumes must be of equal size. Therefore, only the targets with the suitable size are shown for a source volume.
>
> Volumes that are target volumes in another FlashCopy mapping cannot be target of a new FlashCopy mapping. Therefore, they do not appear in the list.

4. In the next window, select one FlashCopy preset. The GUI provides the following presets to simplify common FlashCopy operations, as shown in Figure 10-42 on page 774. For more information about the presets, see 10.4.1, “FlashCopy presets” on page 752:
   - Snapshot: Creates a PiT snapshot copy of the source volume.
   - Clone: Creates a PiT replica of the source volume.
   - Backup: Creates an incremental FlashCopy mapping that can be used to recover data or objects if the system experiences data loss. These backups can be copied multiple times from source and target volumes.

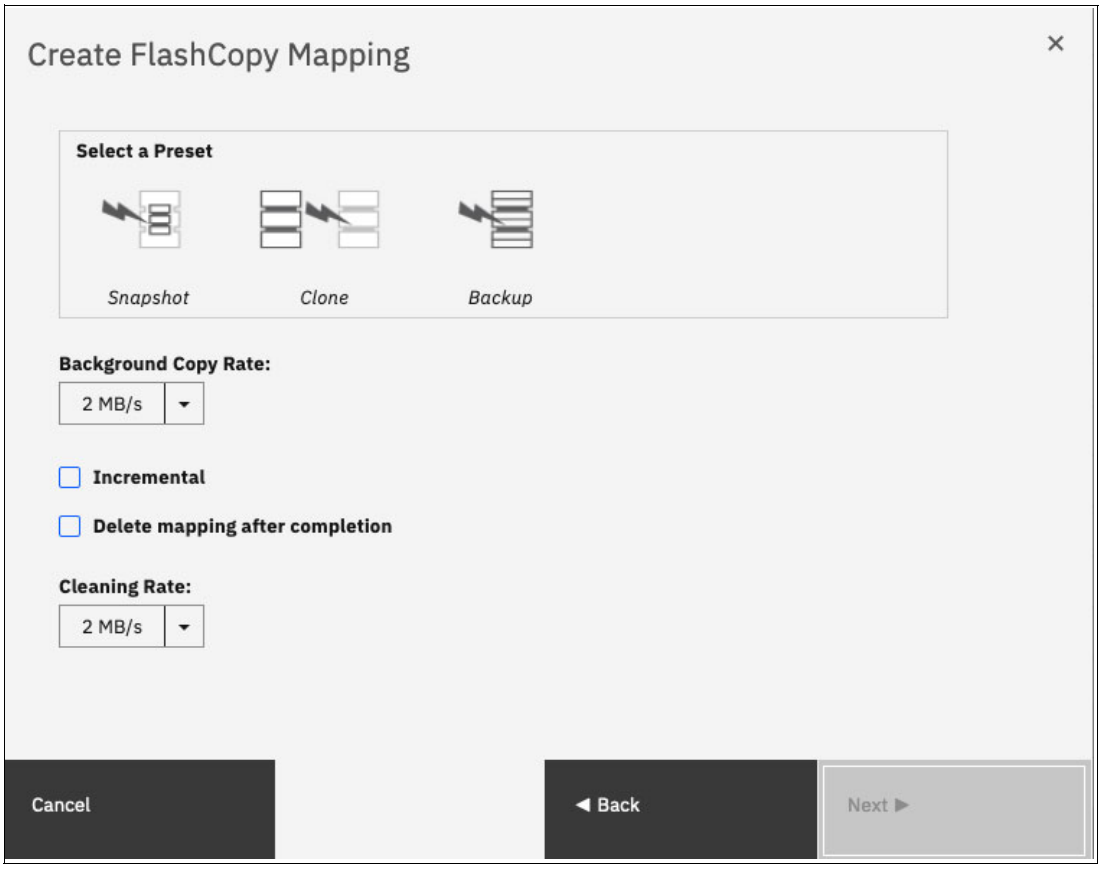


Figure 10-42 FlashCopy mapping preset selection

When selecting a preset, some options, such as Background Copy Rate, Incremental, and Delete mapping after completion, are automatically changed or selected. You can still change the automatic settings, but this is not recommended for the following reasons:

- If you select the **Backup** preset but then clear **Incremental** or select **Delete mapping after completion**, you lose the benefits of the incremental FlashCopy. Copy the entire source volume each time you start the mapping.
- If you select the **Snapshot** preset but then change the **Background Copy Rate**, you have a full copy of your source volume.

For more information about the Background Copy Rate and the Cleaning Rate, see Table 10-2 on page 734, or Table 10-6 on page 743.

5. Select **Yes**, **add the mappings to consistency group** and choose the consistency group from the drop-down menu (see Figure 10-43).

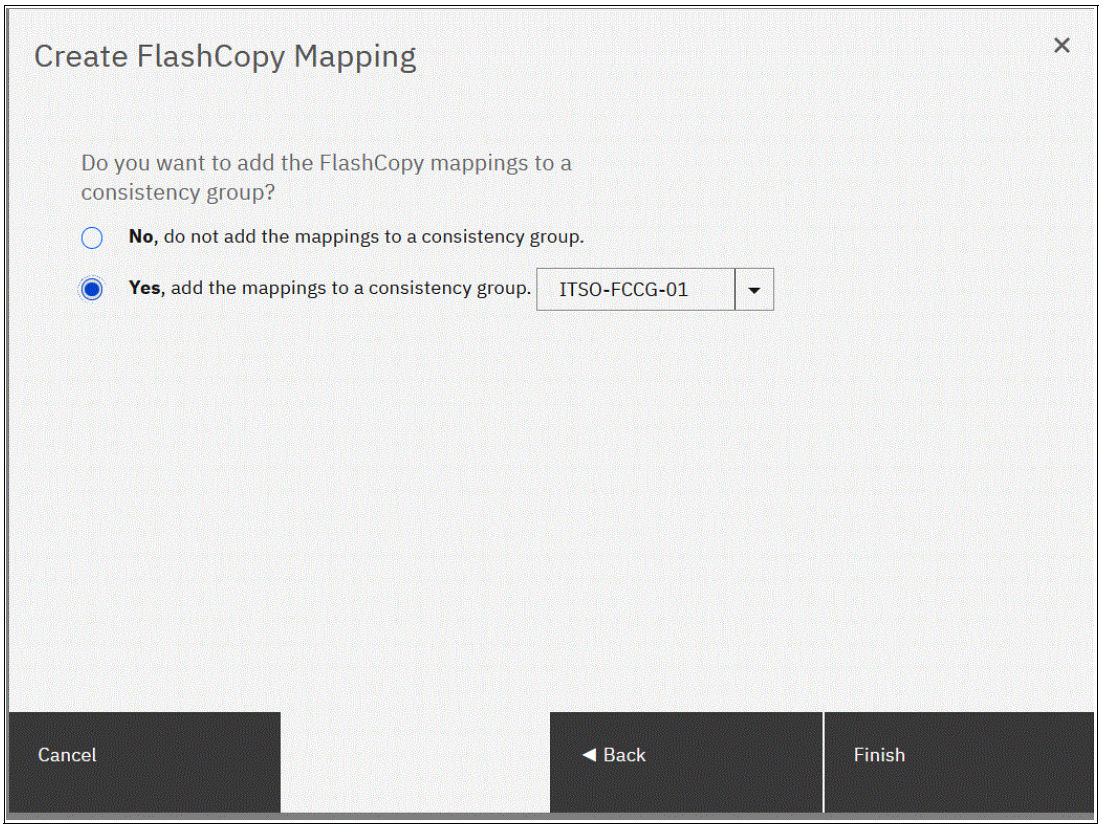


Figure 10-43 FlashCopy mapping Consistency Group selection

6. When your FlashCopy mapping setup is ready, click **Finish**.

10.4.9 Showing related volumes

To show related volumes for a specific FlashCopy mapping, complete the following steps:

1. Open the Copy Services FlashCopy Mappings window.
2. Right-click a FlashCopy mapping and select **Show Related Volumes**, as shown in Figure 10-44. Also, depending on which window you are inside Copy Services, you can right-click at mappings and select **Show Related Volumes**.

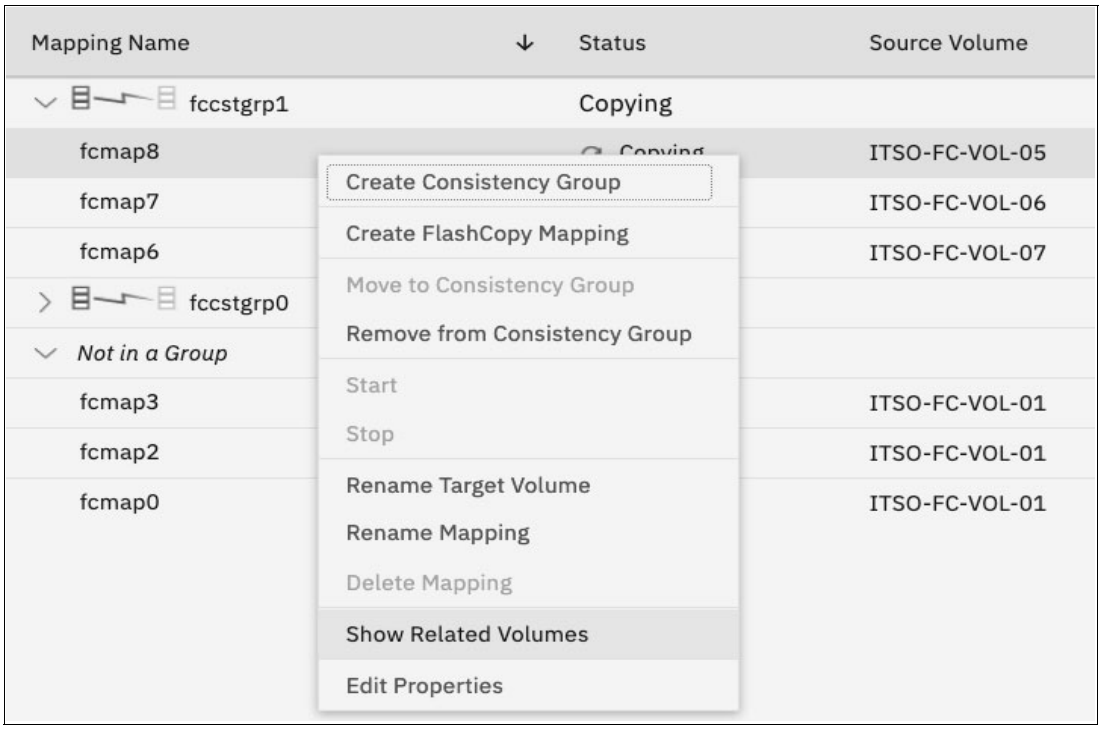


Figure 10-44 Showing related volumes for a mapping, a consistency group or another volume

3. In the related volumes window, you can see the related mapping for a volume, as shown in Figure 10-45. If you click one of these volumes, you can see its properties.



Figure 10-45 Showing related volumes list

10.4.10 Moving FlashCopy mappings across consistency groups

To move one or multiple FlashCopy mappings to a consistency group, complete the following steps:

1. Open the FlashCopy, Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to move and select **Move to Consistency Group**, as shown in Figure 10-46.

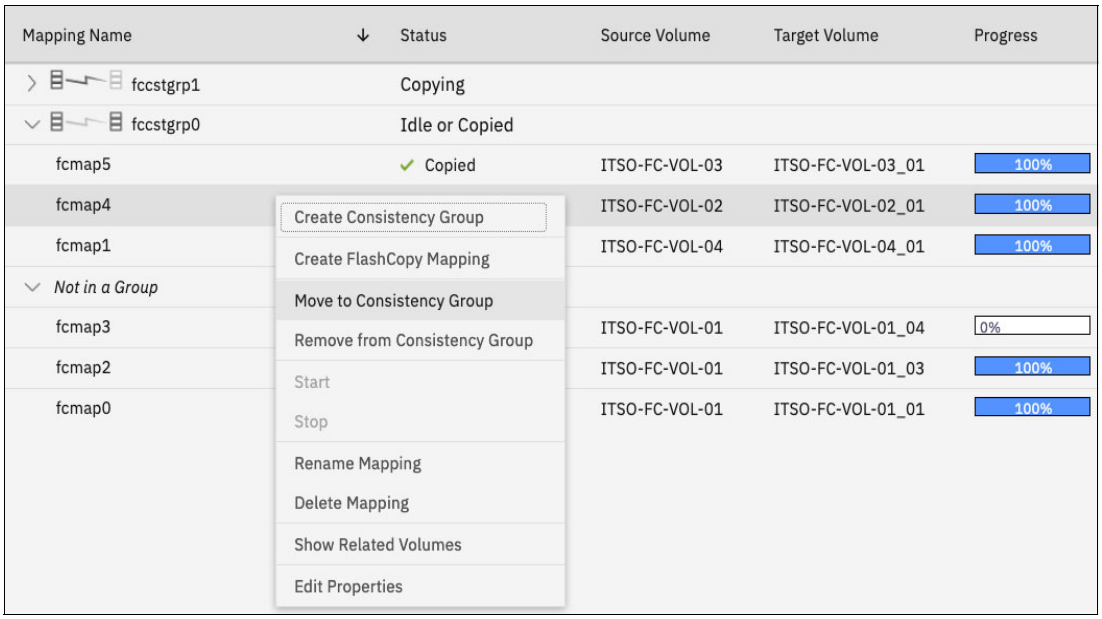


Figure 10-46 Moving a FlashCopy mapping to a consistency group

> **Note:** You cannot move a FlashCopy mapping that is in a `copying`, `stopping`, or `suspended` state. The mapping should be `idle-or-copied` or `stopped` to be moved.

3. In the Move FlashCopy Mapping to Consistency Group window, select the consistency group for the FlashCopy mappings selection by using the drop-down menu, as shown in Figure 10-47.

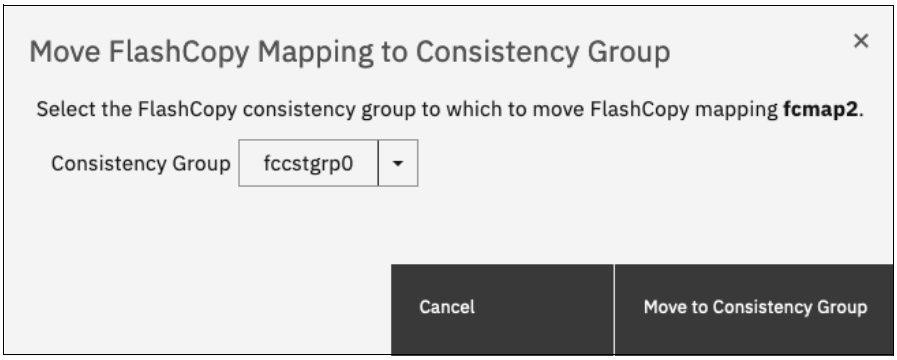


Figure 10-47 Selecting the consistency group where to move the FlashCopy mapping

4. Click **Move to Consistency Group** to confirm your changes.

10.4.11 Removing FlashCopy mappings from consistency groups

To remove one or multiple FlashCopy mappings from a consistency group, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to remove and select **Remove from Consistency Group**, as shown in Figure 10-48.

> **Note:** Only FlashCopy mappings that belong to a consistency group can be removed.

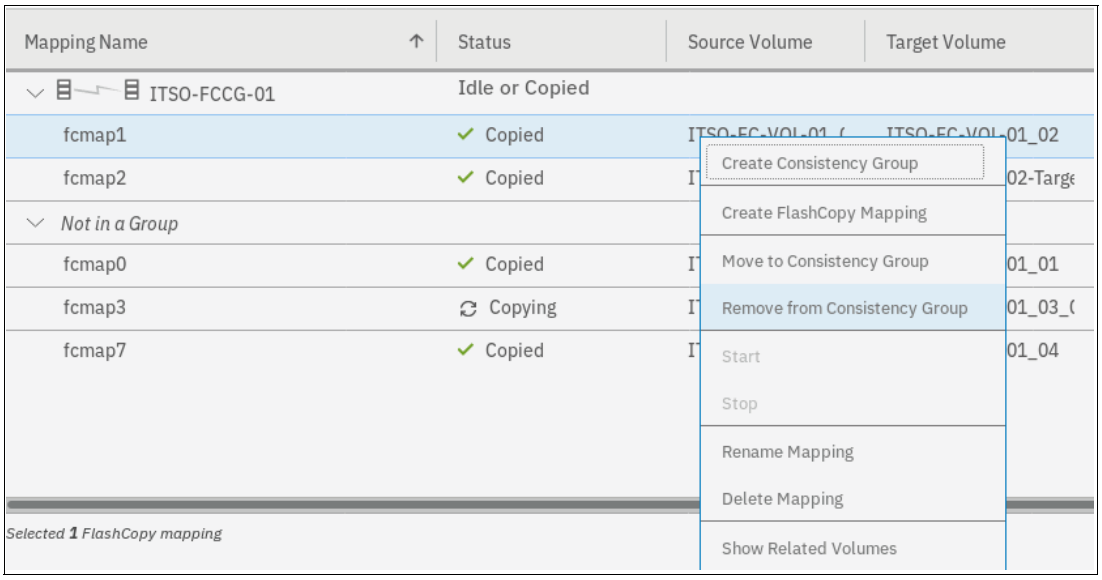


Figure 10-48 Removing FlashCopy mappings from a consistency group

3. In the Remove FlashCopy Mapping from Consistency Group window, click **Remove**, as shown in Figure 10-49 on page 779.

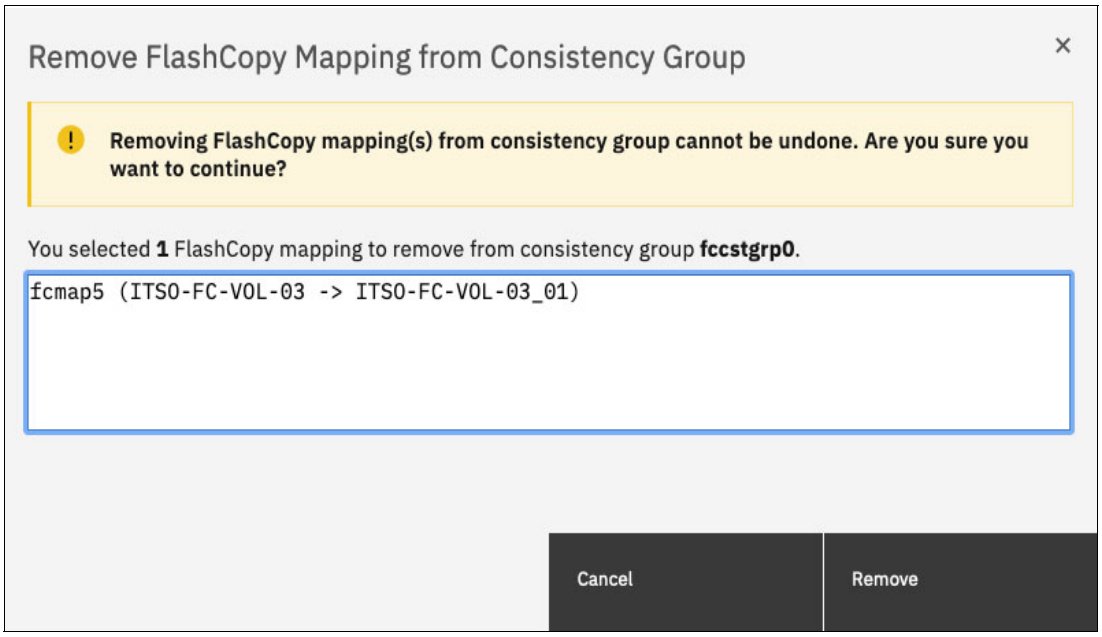


Figure 10-49 Confirm the selection of mappings to be removed

10.4.12 Modifying a FlashCopy mapping

To modify a FlashCopy mapping, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mapping that you want to edit and select **Edit Properties**, as shown in Figure 10-50.

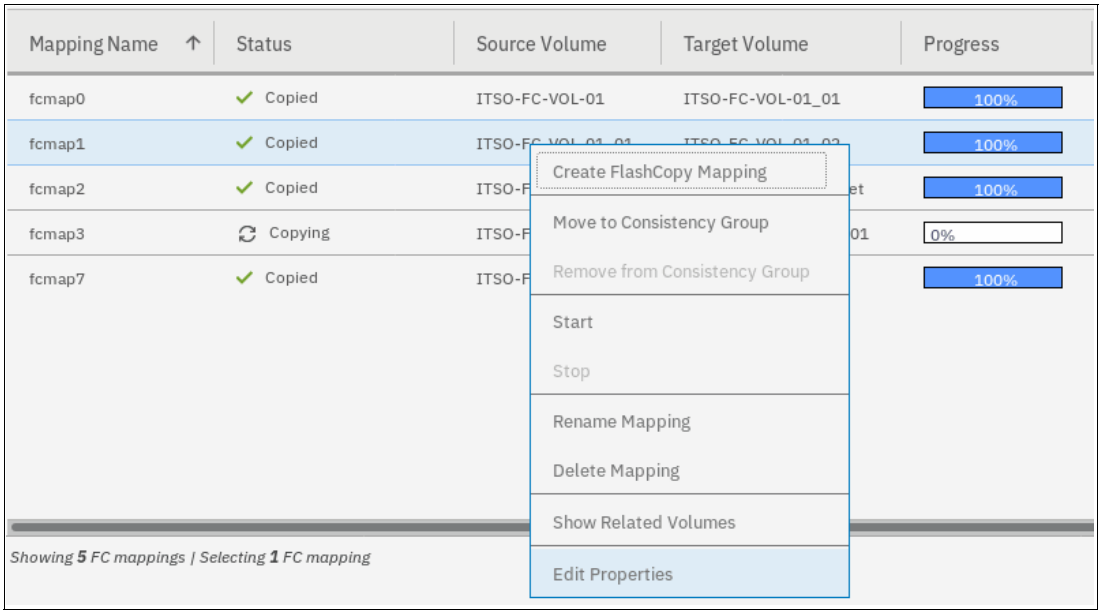


Figure 10-50 Editing a FlashCopy mapping properties

Note: It is not possible to select multiple FlashCopy mappings to edit their properties all at the same time.

3. In the Edit FlashCopy Mapping window, you can modify the background copy rate and the cleaning rate for a selected FlashCopy mapping, as shown in Figure 10-51.

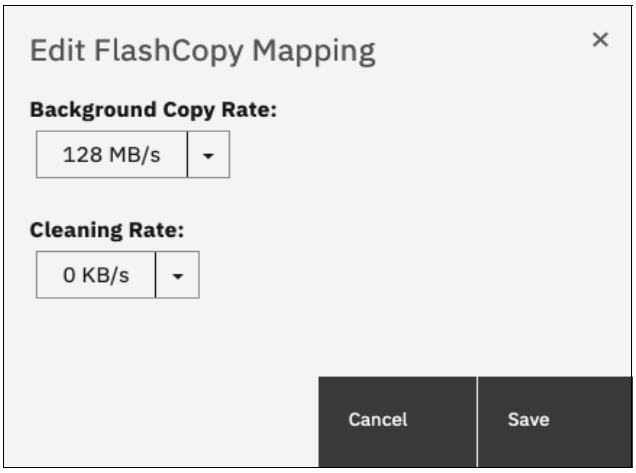


Figure 10-51 Editing copy and cleaning rates of a FlashCopy mapping

For more information about the Background Copy Rate and the Cleaning Rate, see Table 10-2 on page 734, or Table 10-6 on page 743.

4. Click **Save** to confirm your changes.

10.4.13 Renaming FlashCopy mappings

To rename one or multiple FlashCopy mappings, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to rename and select **Rename Mapping**, as shown in Figure 10-52.

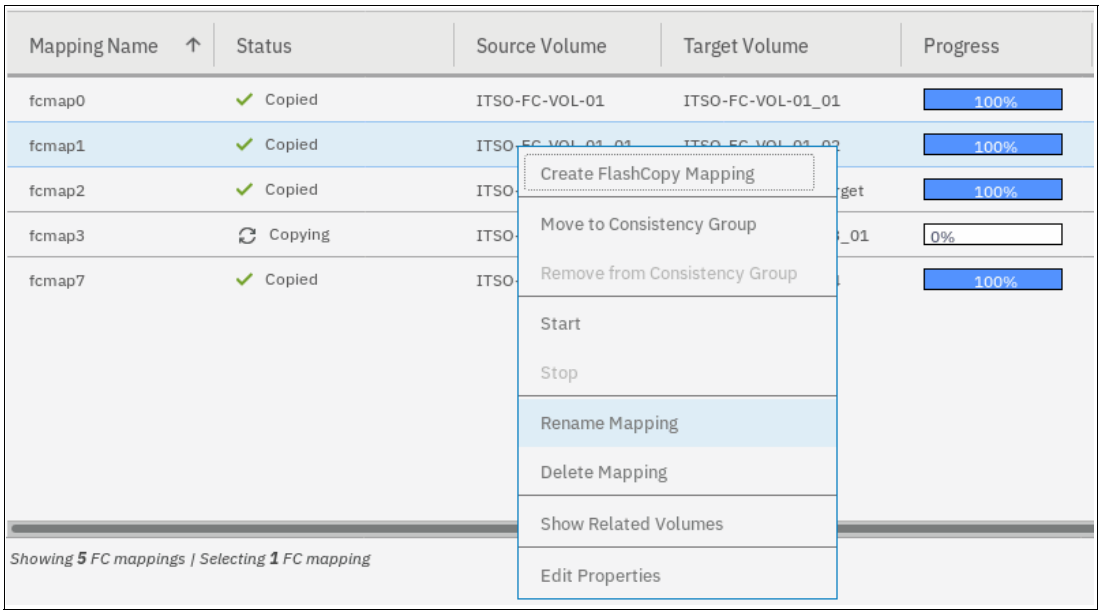


Figure 10-52 Renaming FlashCopy mappings

3. In the Rename FlashCopy Mapping window, enter the new name that you want to assign to each FlashCopy mapping and click **Rename**, as shown in Figure 10-53.

FlashCopy mapping name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The FlashCopy mapping name can be 1 - 63 characters.

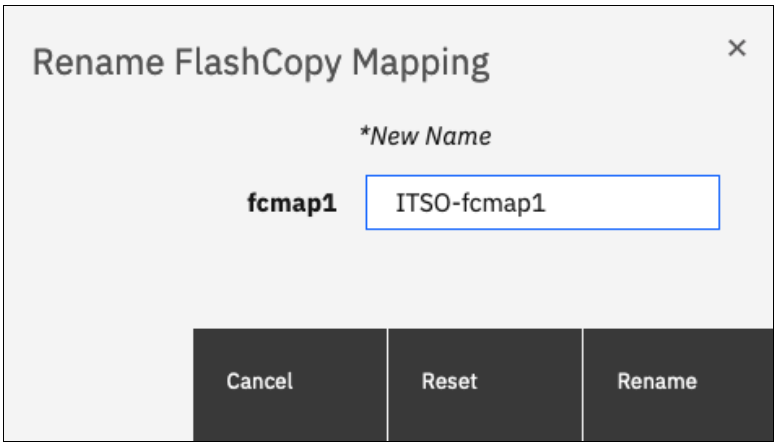


Figure 10-53 Renaming the selected FlashCopy mappings

Renaming a consistency group

To rename a consistency group, complete the following steps:

1. Open the Consistency Groups window.
2. Right-click the consistency group that you want to rename and select **Rename**, as shown in Figure 10-54.

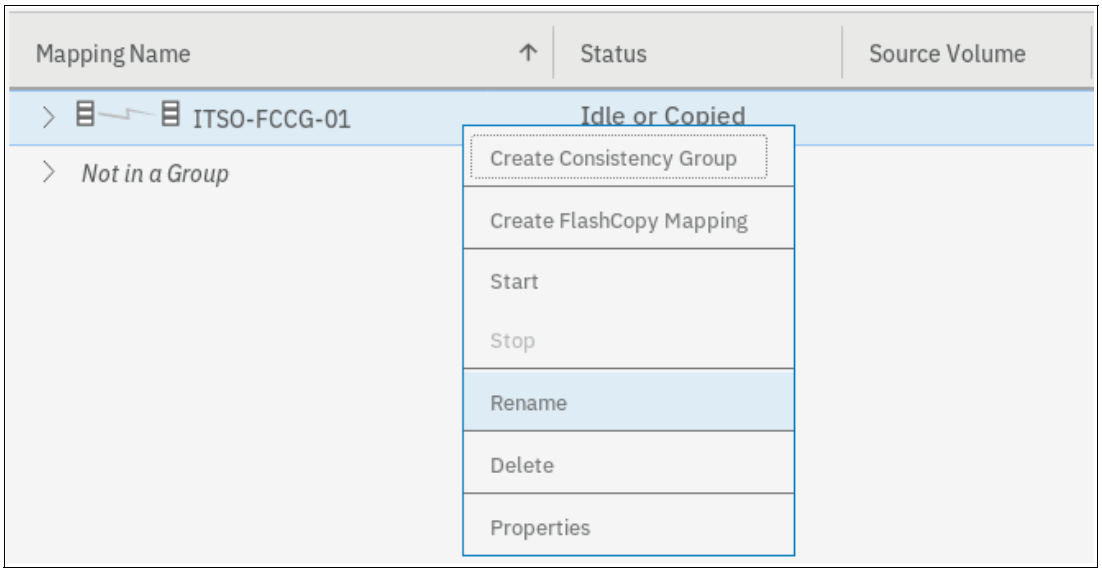


Figure 10-54 Renaming a consistency group

3. Enter the new name that you want to assign to the consistency group and click **Rename**, as shown in Figure 10-55.

> **Note:** It is not possible to select multiple consistency groups to edit their names all at the same time.

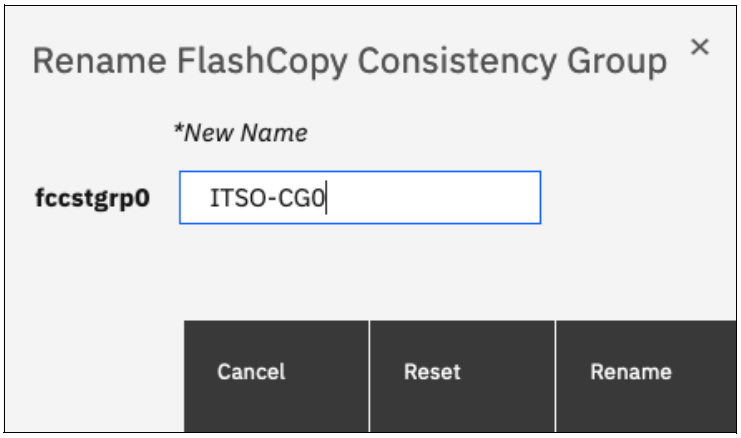


Figure 10-55 Renaming the selected consistency group

> **Consistency group name:** The name can consist of the letters A - Z and a - z, the numbers 0 - 9, the dash (-), and the underscore (_) character. The name can be 1 - 63 characters. However, the name cannot start with a number, a dash, or an underscore.

10.4.14 Deleting FlashCopy mappings

To delete one or multiple FlashCopy mappings, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to delete and select **Delete Mapping**, as shown in Figure 10-56 on page 783.

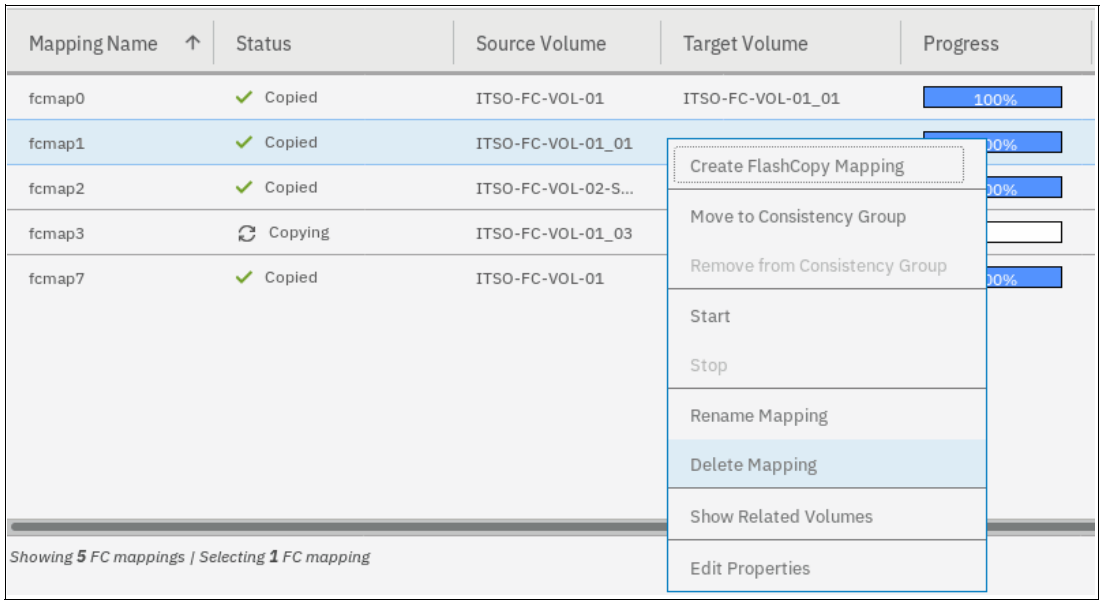


Figure 10-56 Deleting FlashCopy mappings

3. The Delete FlashCopy Mapping window opens, as shown in Figure 10-57. In the **Verify the number of FlashCopy mappings that you are deleting** field, enter the number of volumes that you want to remove. This verification was added to help avoid deleting the wrong mappings.

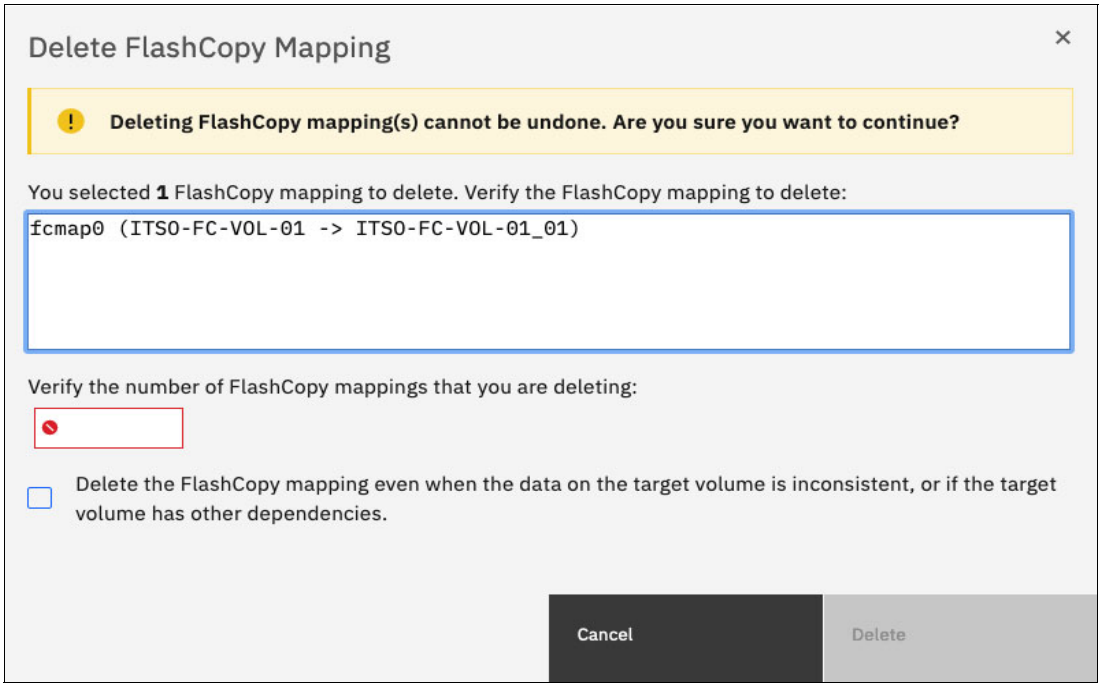


Figure 10-57 Confirming the selection of FlashCopy mappings to be deleted

4. If you still have target volumes that are inconsistent with the source volumes and you want to delete these FlashCopy mappings, select the **Delete the FlashCopy mapping even when the data on the target volume is inconsistent, or if the target volume has other dependencies** option. Click **Delete**.

10.4.15 Deleting a FlashCopy consistency group

Important: Deleting a consistency group does not delete the FlashCopy mappings that it contains.

To delete a consistency group, complete the following steps:

1. Open the Consistency Groups window.
2. Right-click the consistency group that you want to delete and select **Delete**, as shown in Figure 10-58.

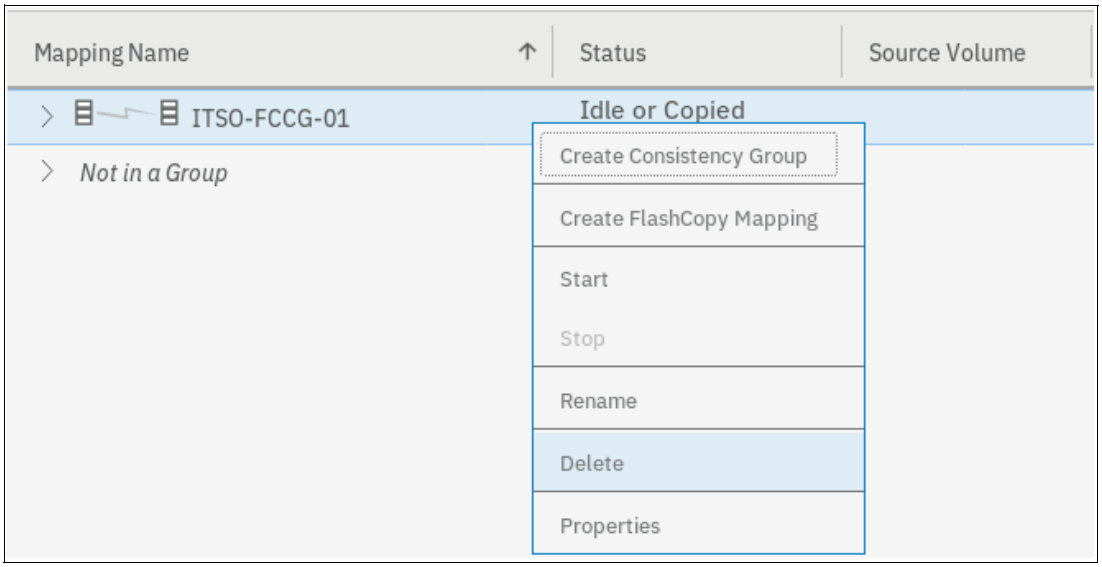


Figure 10-58 Deleting a consistency group

3. A warning message is displayed, as shown in Figure 10-59. Click **Yes**.

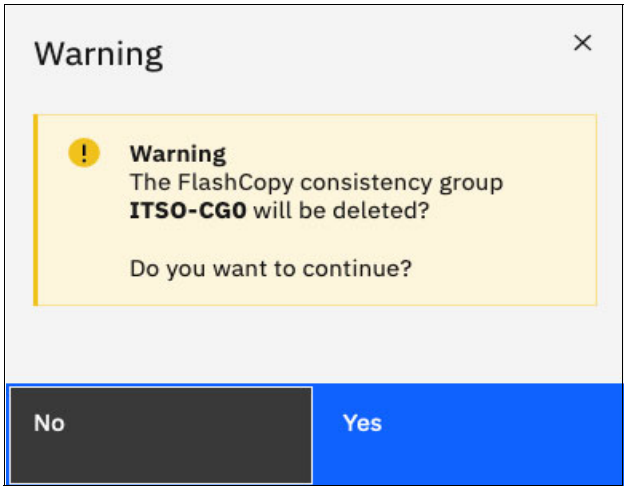


Figure 10-59 Confirming the consistency group deletion

10.4.16 Starting FlashCopy mappings

Important: Only FlashCopy mappings that do not belong to a consistency group can be started individually. If FlashCopy mappings are part of a consistency group, they can be started only all together by using the consistency group **`start`** command.

It is the **`start`** command that defines the “PiT”. It is the moment that is used as a reference (T0) for all subsequent operations on the source and the target volumes. To start one or multiple FlashCopy mappings that do not belong to a consistency group, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to start and select **Start**, as shown in Figure 10-60.

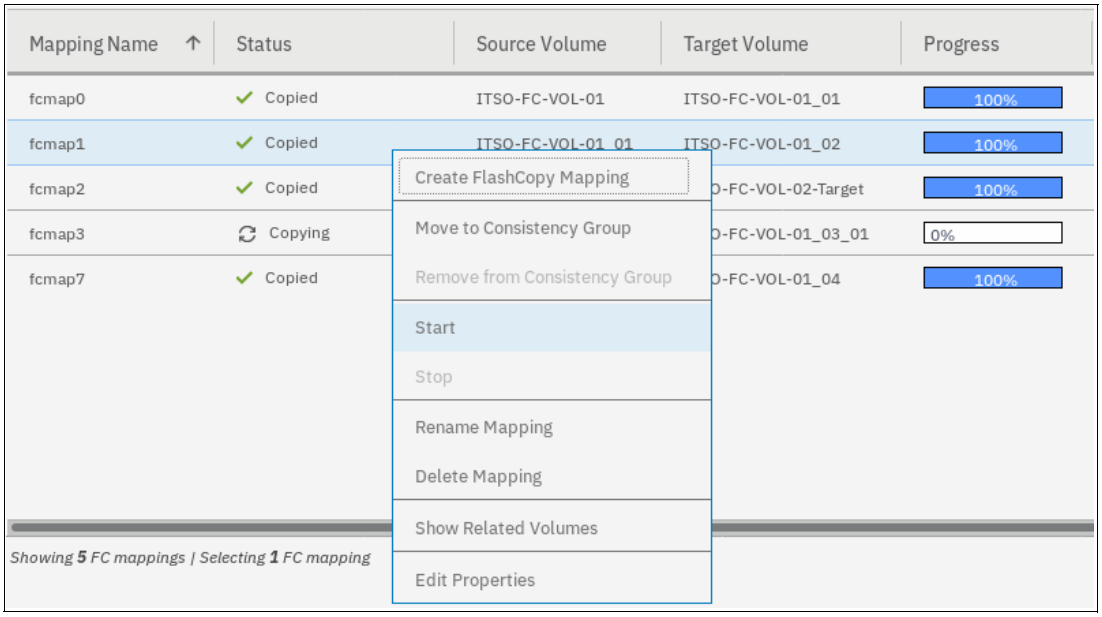


Figure 10-60 Starting FlashCopy mappings

You can check the FlashCopy state and the progress of the mappings in the Status and Progress columns of the table, as shown in Figure 10-61.

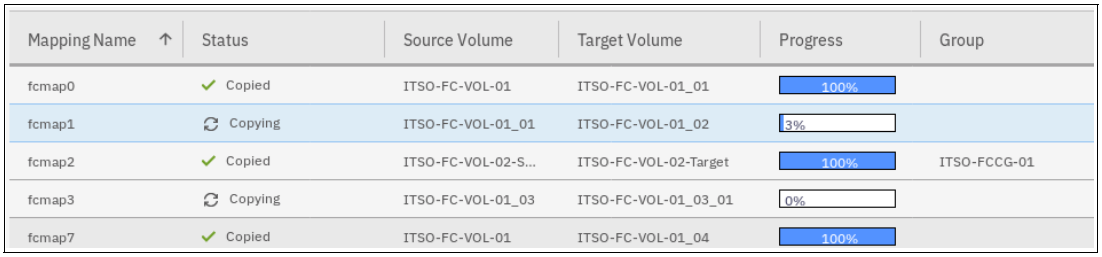


Figure 10-61 FlashCopy mappings status and progress examples

FlashCopy Snapshots depend on the source volume and should be in a “copying” state if the mapping is started.

FlashCopy clones and the first occurrence of FlashCopy backup can take some time to complete, depending on the copyrate value and the size of the source volume. The next occurrences of FlashCopy backups are faster because only the changes that were made during two occurrences are copied.

For more information about FlashCopy starting operations and states, see 10.3.11, “Starting FlashCopy mappings and consistency groups” on page 736.

10.4.17 Stopping FlashCopy mappings

Important: Only FlashCopy mappings that do not belong to a consistency group can be stopped individually. If FlashCopy mappings are part of a consistency group, they can be stopped all together only by using the consistency group **`stop`** command.

The only reason to stop a FlashCopy mapping is for incremental FlashCopy. When the first occurrence of an incremental FlashCopy is started, a full copy of the source volume is made. When 100% of the source volume is copied, the FlashCopy mapping does not stop automatically, and a manual stop can be performed. The target volume is available for read and write operations, during the copy, and after the mapping is stopped.

In any other case, stopping a FlashCopy mapping interrupts the copy and resets the bitmap table. Because only part of the data from the source volume was copied, the copied grains might be meaningless without the remaining grains. Therefore, the target volumes are placed offline and are unusable, as shown in Figure 10-62.

Mapping Name	Flash Time	Status	Source Volume	Target Volume	Progress	Group
fcmap0		Stopped	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	0%	
fcmap2	11/3/2020 5:43:07 PM	Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_03	100%	
fcmap3	11/3/2020 5:43:22 PM	Copying	ITSO-FC-VOL-01	ITSO-FC-VOL-01_04	0%	
fcmap4	11/3/2020 5:49:12 PM	Copied	ITSO-FC-VOL-02	ITSO-FC-VOL-02_01	100%	ITSO-CG0
fcmap5	11/3/2020 5:49:12 PM	Copied	ITSO-FC-VOL-03	ITSO-FC-VOL-03_01	100%	ITSO-CG0

Figure 10-62 Showing target volumes state and FlashCopy mappings status

To stop one or multiple FlashCopy mappings that do not belong to a consistency group, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to stop and select **Stop**, as shown in Figure 10-63 on page 787.

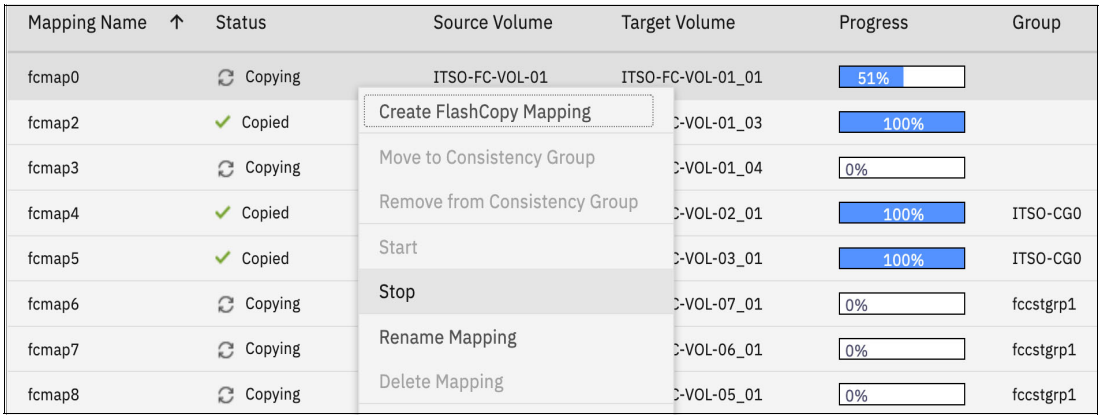


Figure 10-63 Stopping FlashCopy mappings

Note: FlashCopy mappings can be in a `stopping` state for some time if you created dependencies between several targets. It is in a cleaning mode. For more information about dependencies and stopping process, see “Stopping the process in a multiple target FlashCopy: Cleaning Mode” on page 742.

10.4.18 Memory allocation for FlashCopy

Copy Services features require that small amounts of volume cache be converted from cache memory into bitmap memory to allow the functions to operate at an I/O group level. If not enough bitmap space is allocated when you try to use one of the functions, you cannot complete the configuration. The total memory that can be dedicated to these functions is not defined by the physical memory in the system. The memory is constrained by the software functions that use the memory.

For every FlashCopy mapping that is created on an IBM Storage Virtualize system, a bitmap table is created to track the copied grains. By default, the system allocates 20 MiB of memory for a minimum of 10 TiB of FlashCopy source volume capacity and 5 TiB of incremental FlashCopy source volume capacity.

Depending on the grain size of the FlashCopy mapping, the memory capacity usage is different. One MiB of memory provides the following volume capacity for the specified I/O group:

- For clones and snapshots FlashCopy with 256 KiB grains size, 2 TiB of total FlashCopy source volume capacity
- For clones and snapshots FlashCopy with 64 KiB grains size, 512 GiB of total FlashCopy source volume capacity
- For incremental FlashCopy, with 256 KiB grains size, 1 TiB of total incremental FlashCopy source volume capacity
- For incremental FlashCopy, with 64 KiB grains size, 256 GiB of total incremental FlashCopy source volume capacity

Review Table 10-10 on page 788 to calculate the memory requirements and confirm that your system can accommodate the total installation size.

Table 10-10 Memory allocation for FlashCopy services

Minimum allocated bitmap space	Default allocated bitmap space	Maximum allocated bitmap space	Minimum[1] functionality when using the default values
0	20 MiB	2 GiB	10 TiB of FlashCopy source volume capacity 5 TiB of incremental FlashCopy source volume capacity
[1]The actual amount of functionality might increase based on settings, such as grain size and strip size.			

FlashCopy includes the FlashCopy function, Global Mirror with Change Volumes (GMCV), and active-active (HyperSwap) relationships.

For multiple FlashCopy targets, you must consider the number of mappings. For example, for a mapping with a grain size of 256 KiB, 8 KiB of memory allows one mapping between a 16 GiB source volume and a 16 GiB target volume. Alternatively, for a mapping with a 256 KiB grain size, 8 KiB of memory allows two mappings between one 8 GiB source volume and two 8 GiB target volumes.

When creating a FlashCopy mapping, if you specify an I/O group other than the I/O group of the source volume, the memory accounting goes toward the specified I/O group, not toward the I/O group of the source volume.

When creating FlashCopy relationships or mirrored volumes, more bitmap space is allocated automatically by the system, if required.

For FlashCopy mappings, only one I/O group uses bitmap space. By default, the I/O group of the source volume is used.

When you create a reverse mapping, such as when you run a restore operation from a snapshot to its source volume, a bitmap is created.

When you configure change volumes for use with GM, two internal FlashCopy mappings are created for each change volume.

You can modify the resource allocation for each I/O group of an IBM Storage Virtualize system by selecting **Settings** → **System** and clicking the **Resources** menu, as shown in Figure 10-64 on page 789. This value can also be adjusted using the **`chiogrp`** CLI command.

For more information about the syntax, see chiogrp.

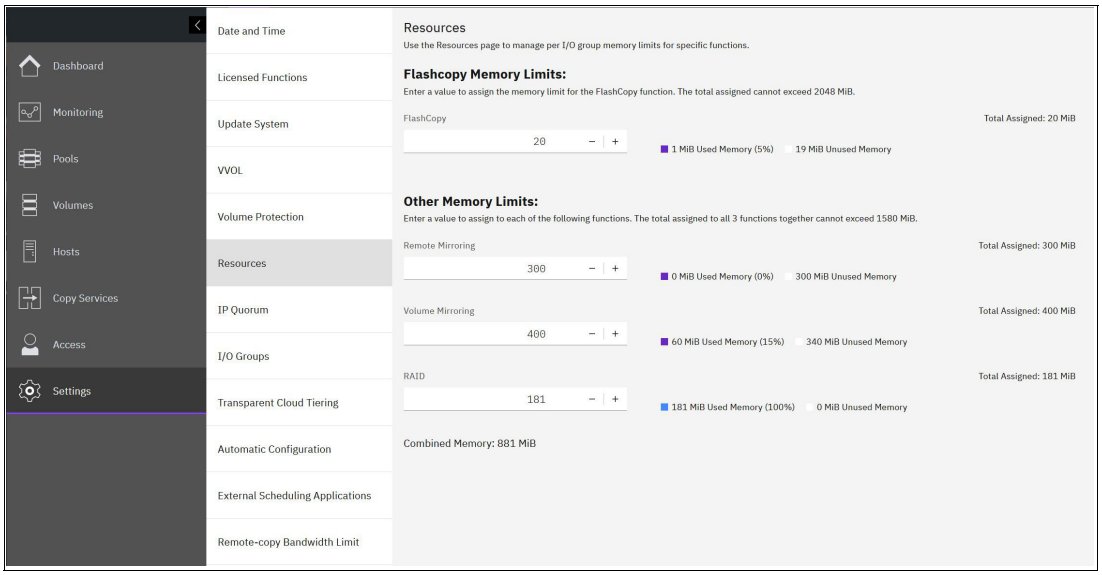


Figure 10-64 Modifying resources allocation per I/O group

10.5 Transparent Cloud Tiering

Introduced in V7.8, *Transparent Cloud Tiering* (TCT) is a function of IBM Storage Virtualize that uses IBM FlashCopy mechanisms to produce a PiT snapshot of the data. TCT helps to increase the flexibility to protect and transport data to public or private cloud infrastructure. This technology is built on top of IBM Storage Virtualize software capabilities. TCT uses the cloud to store snapshot targets and provides alternatives to restore snapshots from the private and public cloud of an entire volume or set of volumes.

Demonstration video: To view a demonstration video, see IBM Storage Virtualize V8.6: Tranparent Cloud Tiering.

TCT can help to solve business needs that require duplication of data of your source volume. Volumes can remain online and active while you create snapshot copies of the data sets. TCT operates below the host operating system and its cache. Therefore, the copy is not apparent to the host.

IBM Storage Virtualize features built-in software algorithms that allow the TCT function to securely interact; for example, with Information Dispersal Algorithms (IDA), which is essentially the interface to IBM Cloud Object Storage.

Object Storage is a general term that refers to the entity in which IBM Cloud Object Storage organizes, manages, and stores units of data. To transform these snapshots of traditional data into Object Storage, the storage nodes and the IDA import the data and transform it into several metadata and slices. The object can be read by using a subset of those slices. When an Object Storage entity is stored as IBM Cloud Object Storage, the objects must be manipulated or managed as a whole unit. Therefore, objects cannot be accessed or updated partially.

IBM Storage Virtualize uses internal software components to support HTTP-based REST application programming interface (API) to interact with an external cloud service provider (CSP) or private cloud.

For more information, see IBM Cloud Object Storage.

10.5.1 Considerations for using Transparent Cloud Tiering

TCT can help to address certain business needs. When considering whether to use TCT, adopt a combination of business and technical views of the challenges and determine whether TCT can solve both of those needs.

The use of TCT can help businesses to manipulate data as shown in the following examples:

- Creating a consistent snapshot of:
  - Dynamically changing data
  - Production data to facilitate data movement or migration between systems that are running at different locations
- Creating a snapshot of production data sets for:
  - Application development and testing
  - Quality assurance
- Using secure data tiering to off-premises cloud providers

From a technical standpoint, ensure that you evaluate the network capacity and bandwidth requirements to support your data migration to off-premises infrastructure. To maximize productivity, you must match your amount of data that must be transmitted off to the cloud plus your network capacity.

From a security standpoint, ensure that your on-premises or off-premises cloud infrastructure supports your requirements in terms of methods and level of encryption.

Regardless of your business needs, TCT within the IBM Storage Virtualize can provide opportunities to manage the exponential data growth and to manipulate data at low cost.

Today, many CSPs offers several storage-as-service solutions, such as content repository, backup, and archive. Combining all of these services, your IBM Storage Virtualize can help you solve many challenges that are related to rapid data growth, scalability, and manageability at attractive costs.

10.5.2 Transparent Cloud Tiering as backup solution and data migration

TCT can also be used as a backup and data migration solution. In certain conditions, this can be easily applied to eliminate the downtime that is associated with the need to import and export data.

When TCT is applied as your backup strategy, IBM Storage Virtualize uses the same FlashCopy functions to produce PiT snapshot of an entire volume or set of volumes.

To ensure the integrity of the snapshot, it might be necessary to flush the host operating system and application cache of any outstanding reads or writes before the snapshot is performed. Failing to flush the host operating system and application cache can produce inconsistent and useless data.

Many operating systems and applications provide a mechanism to stop I/O operations and ensure that all data is flushed from host cache. If these mechanisms are available, they can be used in combination with snapshot operations. When these mechanisms are not available, it might be necessary to flush the cache manually by quiescing the application and unmounting the file system or logical drives.

When choosing cloud Object Storage as a backup solution, be aware that the Object Storage must be managed as a whole. Backup and restore of individual files, folders, and partitions are not possible.

To interact with external cloud service providers (CSP) or a private cloud, IBM Storage Virtualize requires interaction with the correct architecture and specific properties. Conversely, CSPs offer attractive prices for Object Storage in the cloud and deliver an easy-to-use interface. Normally, cloud providers offer low-cost prices for Object Storage space, and charges are applied for the cloud outbound traffic only.

10.5.3 Restoring data by using Transparent Cloud Tiering

TCT can also be used to restore data from any snapshot that is stored in cloud providers. When the cloud accounts' technical and security requirements are met, the storage objects in the cloud can be used as a data recovery solution. The recovery method is similar to backup, except that the reverse direction is applied.

TCT running on IBM Storage Virtualize queries for Object Storage stored in a cloud infrastructure. It enables users to restore the objects into a new volume or set of volumes.

This approach can be used for various applications, such as recovering your production database application after an errant batch process that caused extensive damage.

Note: Always consider the bandwidth characteristics and network capabilities when choosing to use TCT.

Restoring individual files by using TCT is not possible. Object Storage is unlike a file or a block; therefore, Object Storage must be managed as a whole unit piece of storage, and not partially. Cloud Object Storage is accessible by using an HTTP-based REST API.

10.5.4 Transparent Cloud Tiering restrictions

The following restrictions must be considered before TCT is used:

- Because the Object Storage is normally accessed by using the HTTP protocol on top of a TPC/IP stack, all traffic that is associated with cloud service flows through the node management ports.
- The size of cloud-enabled volumes cannot change. If the size of the volume changes, a new snapshot must be created, so an Object Storage is constructed.
- TCT cannot be applied to volumes that are part of traditional copy services, such as FlashCopy, MM, GM, and HyperSwap.
- Volume containing two physical copies in two different storage pools cannot be part of TCT.
- Cloud Tiering snapshots cannot be taken from a volume that is part of migration activity across storage pools.

- Because VMware vSphere virtual volumes (vVOLs) are managed by a specific VMware application, these volumes are not candidates for TCT.
- File system volumes are not qualified for TCT.

10.6 Implementing Transparent Cloud Tiering

This section describes the steps and requirements to implement TCT by using your IBM Storage Virtualize.

10.6.1 Domain Name System configuration

Because most of IBM Cloud Object Storage is managed and accessible by using the HTTP protocol, the Domain Name System (DNS) setting is an important requirement to ensure consistent resolution of domain names to internet resources.

1. Using your IBM Storage Virtualize management GUI, click **Settings** → **Network** → **DNS** → **Add DNS server +** and enter your DNS Internet Protocol Version 4 (IPv4) address or Internet Protocol Version 6 (IPv6) address. The DNS name can be anything that you want and is used as a reference.

   Click **Save** after you complete the choices, as shown in Figure 10-65.

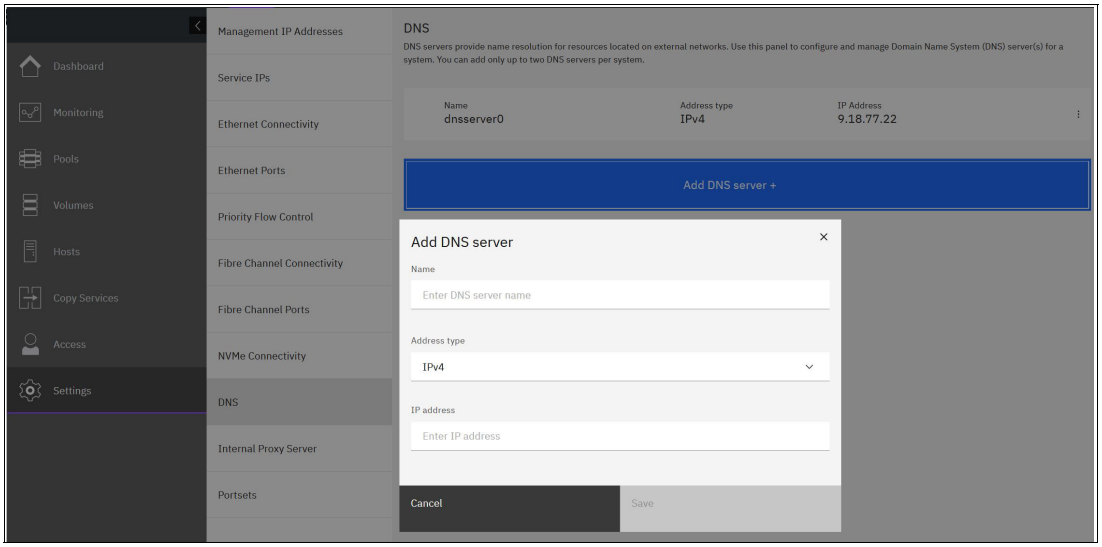


Figure 10-65 DNS settings

2. The command **lsdnsserver** can be used to retrieve information about DNS servers present within the system.

Example 10-2 lsdnsserver command

```
IBM_FlashSystem:FS9500_PRD:ysl>lsdnsserver
id name       type IP_address status
0  dnsserver0 ipv4 192.168.1.212  reachable
1  powermdns  ipv4 192.168.1.202  active


IBM_FlashSystem:FS9500_PRD:ysl>lsdnsserver  1
id 1
```

```
name powermdns
type ipv4
IP_address 192.168.1.202
status active

IBM_FlashSystem:FS9500_PRD:ysl>lsdnsserver  0
id 0
name dnsserver0
type ipv4
IP_address 192.168.1.212
status reachable

IBM_FlashSystem:FS9500_PRD:ysl>
```

Note: Starting with IBM Storage Virtualize 8.6, the DNS server status can have the following possible values:

- Active. This status indicates that the DNS server is currently operational and functioning as expected.
- Failed. The Failed status signifies that the DNS server has encountered an error or has stopped functioning correctly. It is unable to perform its intended DNS operations.
- Untried. When the DNS server is marked as Untried, it means that it has not been tested or attempted to establish a connection. This status typically occurs when the system has not yet initiated communication with the DNS server.
- Reachable (new). The Reachable status suggests that the DNS server is accessible and can be reached successfully. It implies that the system can establish a connection and communicate with the DNS server without any issues.
- Unresponsive (new). The Unresponsive status indicates that the DNS server is not responding to requests or queries. This status implies that there may be a network connectivity problem or an issue with the DNS server itself, which prevents it from responding to requests.

10.6.2 Enabling Transparent Cloud Tiering

After you complete the DNS settings, you can enable the TCT function in your IBM Storage Virtualize system by completing the following steps:

1. Using the IBM Storage Virtualize GUI, click **Settings** → **System** → **Transparent Cloud Tiering** and then click **Enable Cloud Connection**, as shown in Figure 10-66. The TCT wizard starts and shows the welcome warning.

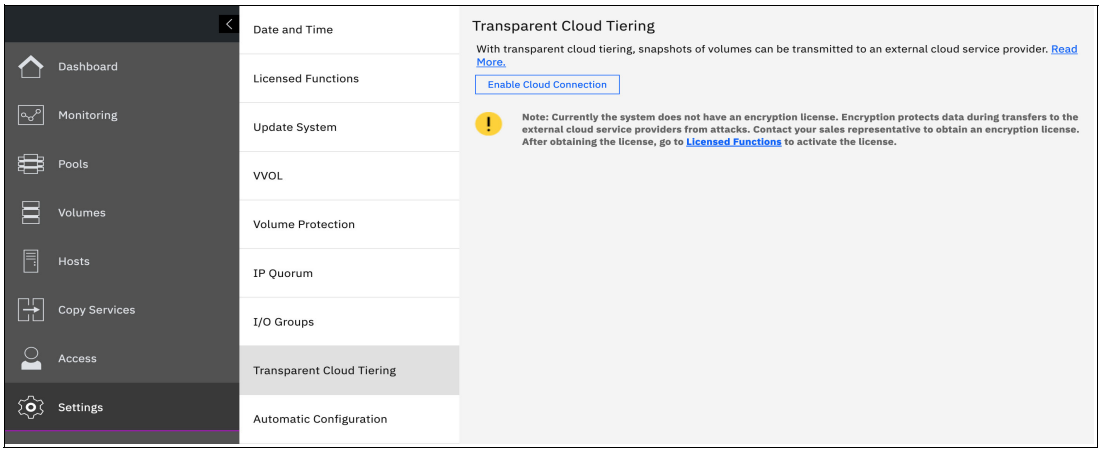


Figure 10-66 Enabling Cloud Tiering

> **Note:** It is important to implement encryption before enabling cloud connecting. Encryption protects your data from attacks during the transfer to the external cloud service. Because the HTTP protocol is used to connect to cloud infrastructure, it is likely to start transactions by using the internet. For the purposes of this writing, encryption is not enabled on our system.

2. The next screen will display a message stating, *"The system is not currently enabled to use encryption."* It is important to note that all data on volumes that are copied to the cloud service provider will be unencrypted, making it potentially vulnerable to exposure.

   Encryption settings cannot be modified after the wizard is completed. Therefore, it is crucial to carefully assess the risks that are associated with unencrypted data before proceeding. To acknowledge the risks and proceed, select the checkbox stating, **I understand the risks and want to continue,** and then click the **Next** button, as shown in Figure 10-67 on page 795.

3. Click **Next** to continue. Select one of four CSPs:

   – IBM Cloud
   – OpenStack Swift
   – Amazon S3
   – Microsoft Azure

   Figure 10-68 on page 795 shows the available options.

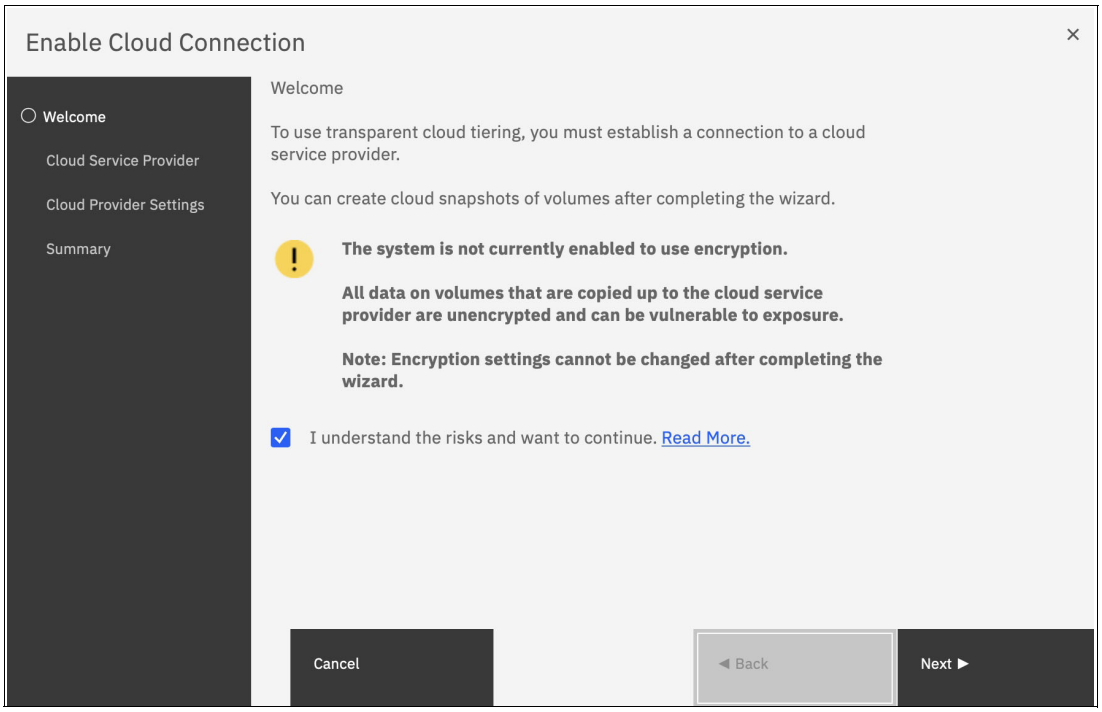


Figure 10-67 Acknowledge that the system is not using encryption

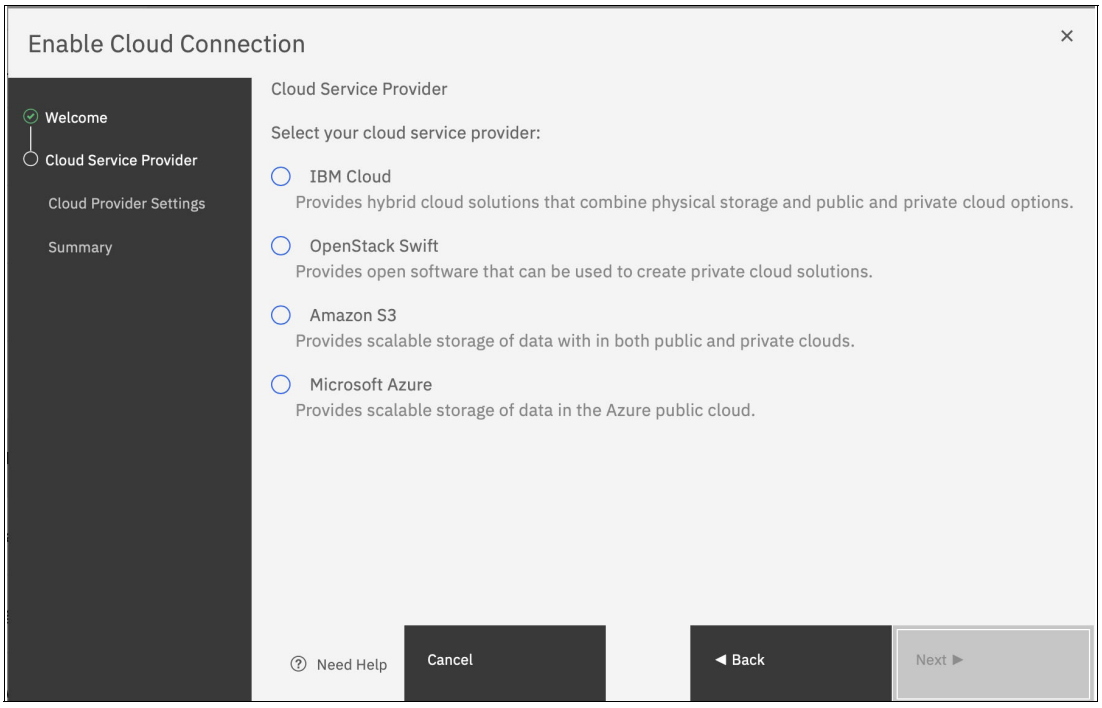


Figure 10-68 Selecting your cloud service provider

4. In the next window, you must complete the settings of the Cloud Provider, credentials, and security access keys. The required settings can change depending on your CSP. An example of an empty form for an IBM Cloud connection is shown in Figure 10-69 on page 796.

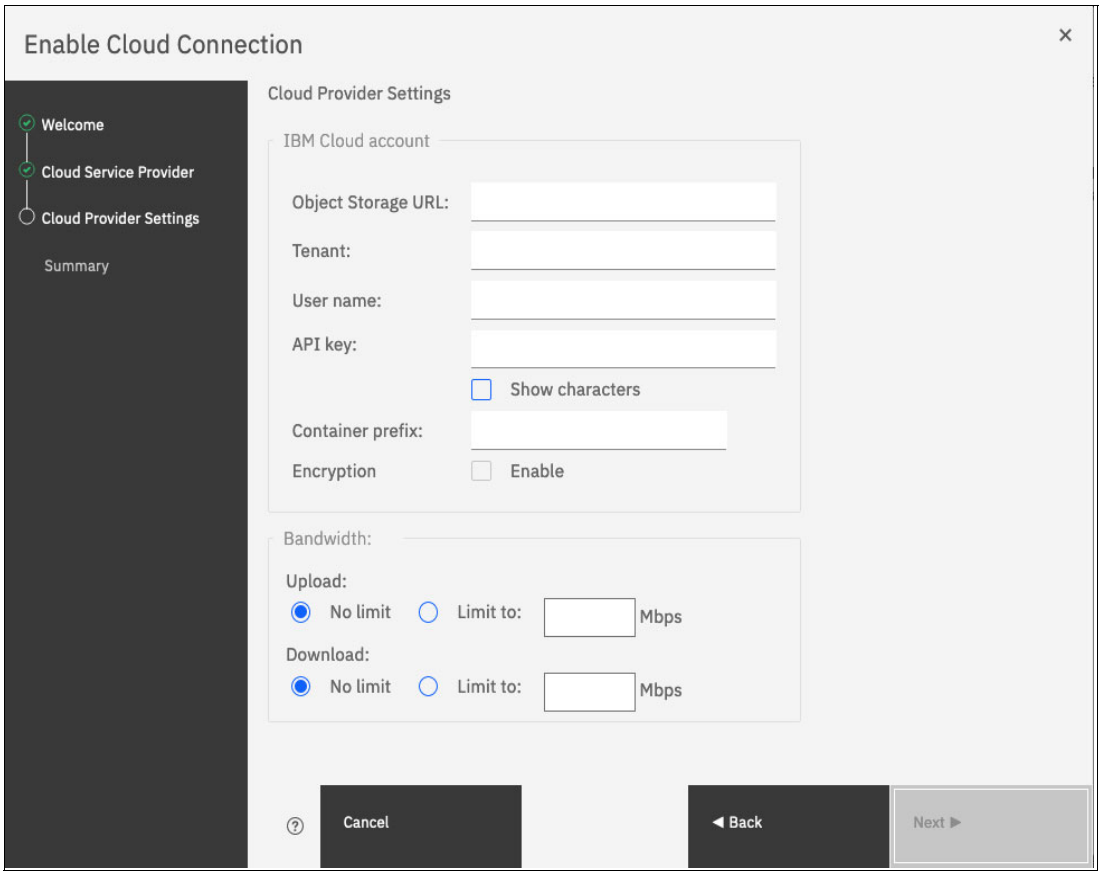


Figure 10-69 Entering cloud service provider information

5. Review your settings and click **Finish**, as shown in Figure 10-70.

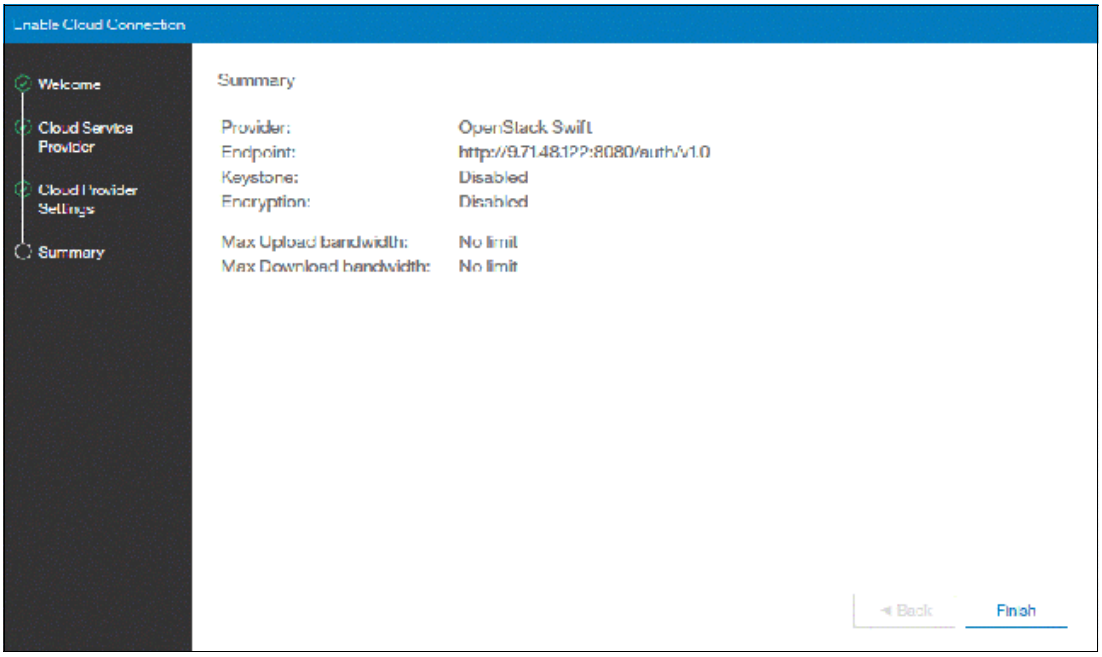


Figure 10-70 Cloud Connection summary

6. The cloud credentials can be viewed and updated at any time by using the function icons in the left side of the GUI and clicking **Settings** → **Systems** → **Transparent Cloud Tiering**. From this window, you can also verify the status, the data usage statistics, and the upload and download bandwidth limits set to support this functionality.

   In the account information window, you can visualize your cloud account information. This window also enables you to remove the account.

   An example of visualizing your cloud account information is shown in Figure 10-71.

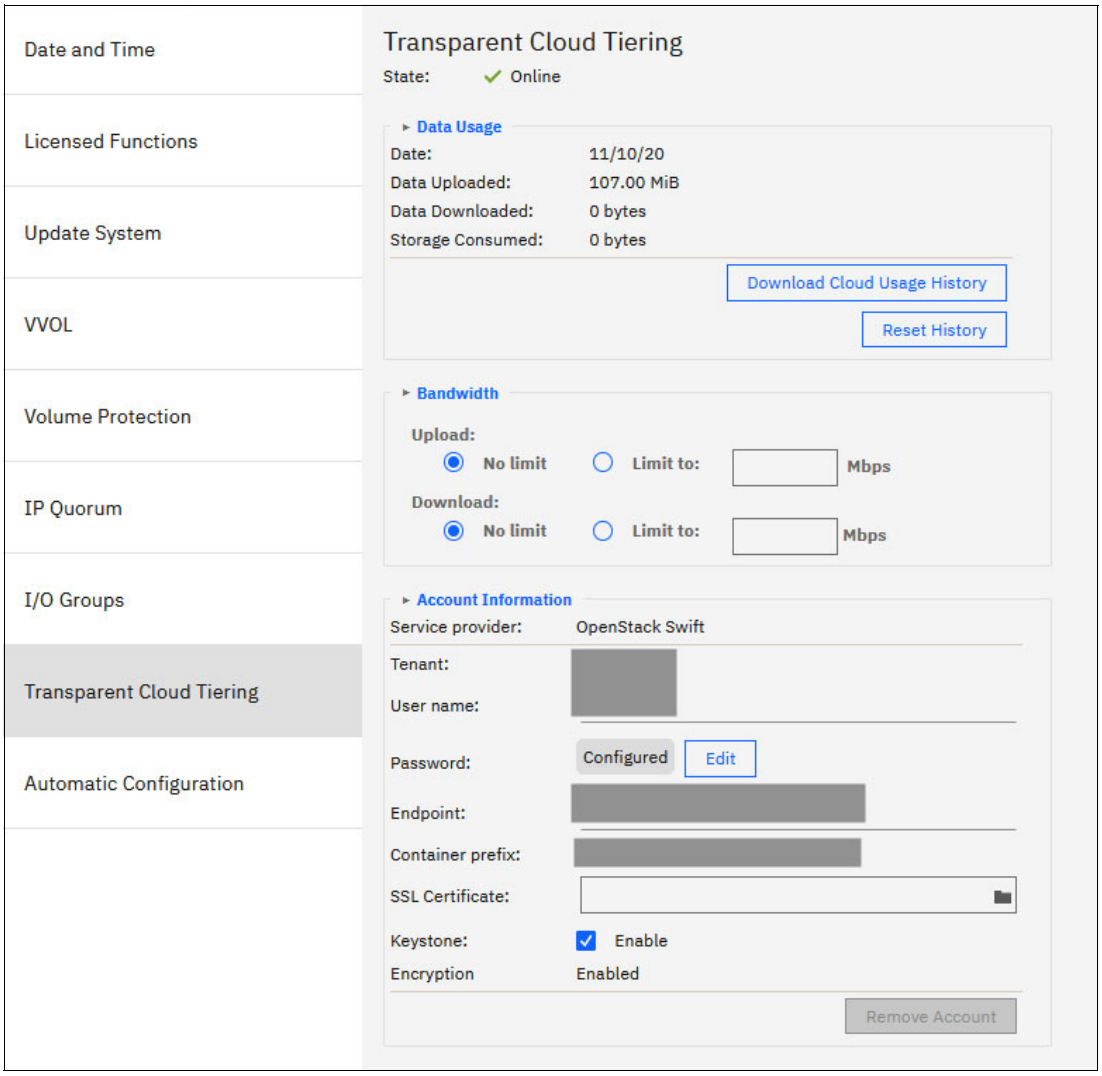


Figure 10-71 Enabled Transparent Cloud Tiering window

10.6.3 Creating cloud snapshots

To manage the cloud snapshots, the IBM Storage Virtualize provides a section in the GUI named Cloud Volumes. This section shows you how to add the volumes that are going to be part of the TCT. As described in 10.5.4, “Transparent Cloud Tiering restrictions” on page 791, cloud snapshot is available only for volumes that do not have a relationship to the list of restrictions previously mentioned.

Any volume can be added to the cloud volumes. However, snapshots work only for volumes that are not related to any other copy service.

To create and cloud snapshots, complete the following steps:

1. Click **Volumes** → **Cloud Volumes**, as shown in Figure 10-72.

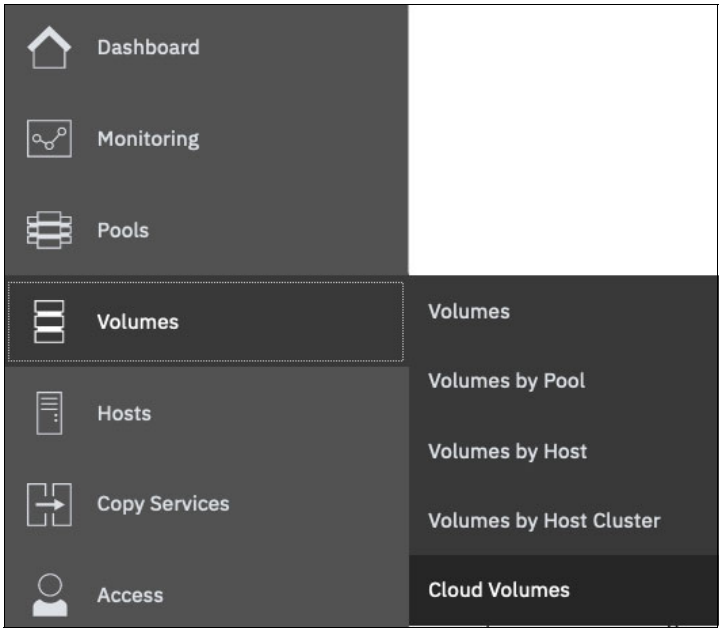


Figure 10-72 Cloud volumes menu

2. A new window opens, and you can use the GUI to select one or more volumes that you need to enable a cloud snapshot or you can add volumes to the list, as shown in Figure 10-73.



Figure 10-73 Cloud volumes window

3. Click **Add Volumes** to enable cloud-snapshot on volumes. A new window opens, as shown in Figure 10-74. Select the volumes that you want to enable Cloud Tiering for and click **Next**.

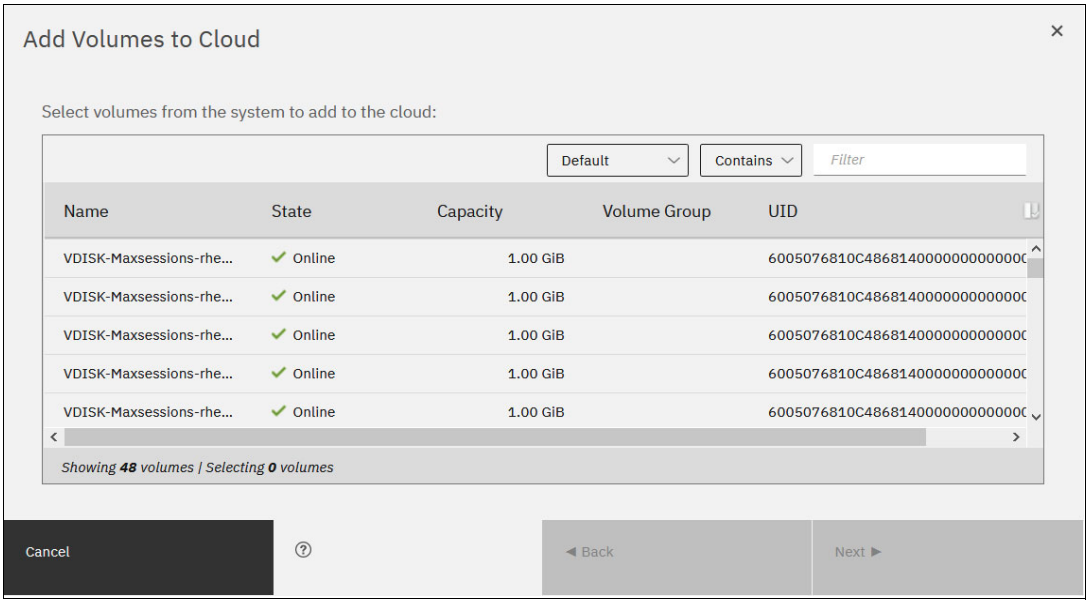


Figure 10-74 Adding volumes to Cloud Tiering

4. IBM Storage Virtualize GUI provides two options for you to select. If the first option is selected, the system decides what type of snapshot is created based on previous objects for each selected volume. If a full copy (full snapshot) of a volume was created, the system makes an incremental copy of the volume.

   The second option creates a full snapshot of one or more selected volumes. You can select the second option for a first occurrence of a snapshot and click **Finish**, as shown in Figure 10-75. You can also select the second option, even if another full copy of the volume exists.

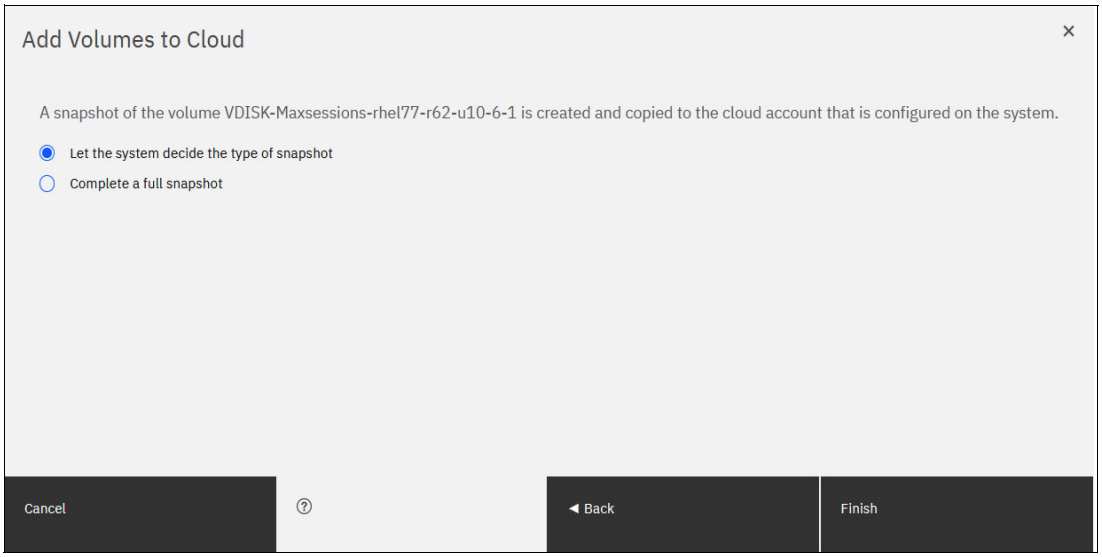


Figure 10-75 Selecting if a full copy is made or if the system decides

The **Cloud Volumes** window shows complete information about the volumes and their snapshots. The GUI shows the following information:

- Name of the volume
- ID of the volume that is assigned by IBM Storage Virtualize
- Snapshot size
- Date and time that the last snapshot was created
- Number of snapshots that are taken for every volume
- Snapshot status
- Restore status
- Volume group for a set of volumes
- Volume unique identifier (UID)

Figure 10-76 shows an example of a Cloud Volumes list.

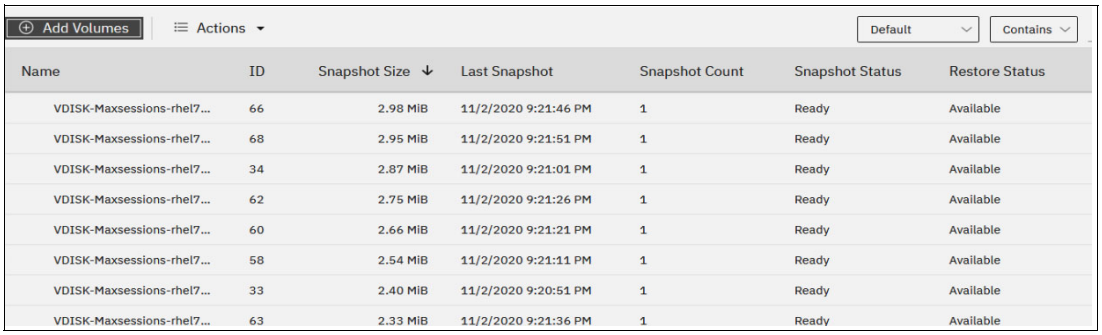


Figure 10-76 Cloud Volumes list example

5. Click the **Actions** menu in the Cloud Volumes window to create and manage snapshots. Also, you can use the menu to cancel, disable, and restore snapshots to volumes, as shown in Figure 10-77.

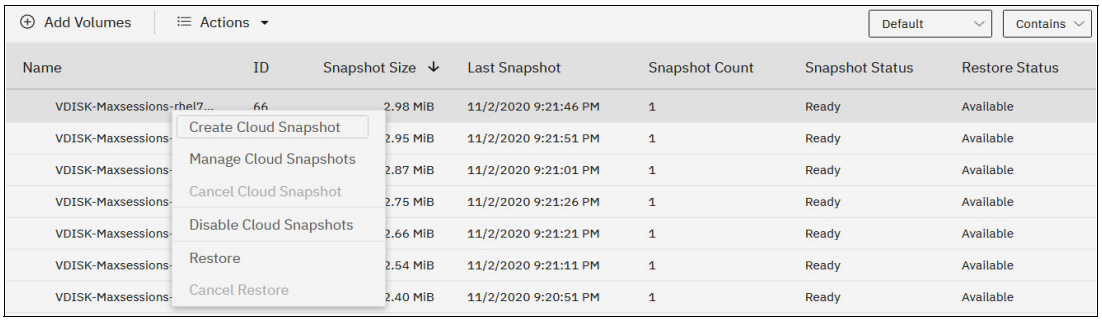


Figure 10-77 Available actions in the Cloud Volumes window

10.6.4 Managing cloud snapshots

To manage volume cloud snapshots, open the Cloud Volumes window, right-click the volume that you want to manage the snapshots from, and select **Manage Cloud Snapshot**.

Managing a snapshot is deleting one or multiple versions. The lists of PiT copies appear and provide details about their status, type, and snapshot date, as shown in Figure 10-78 on page 801.

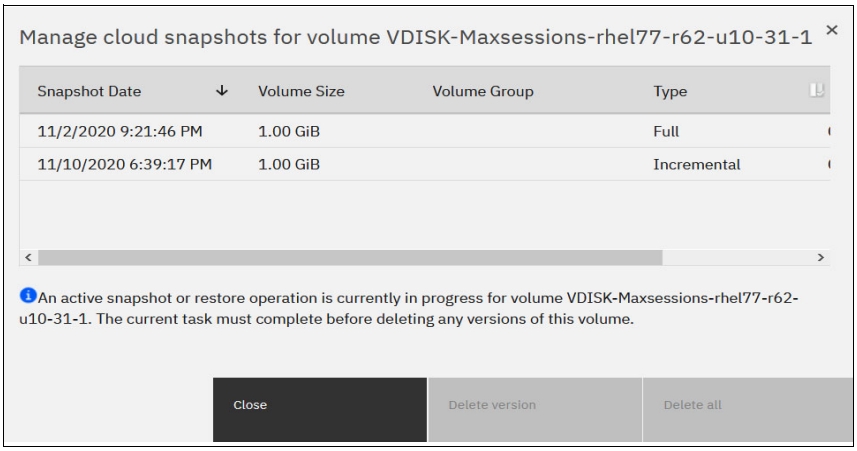


Figure 10-78 Deleting versions of a volume's snapshots

From this window, an administrator can delete old snapshots (old PiT copies) if they are no longer needed. The most recent copy cannot be deleted. If you want to delete the most recent copy, you must first disable Cloud Tiering for the specified volume.

10.6.5 Restoring cloud snapshots

This option allows IBM Storage Virtualize to restore snapshots from the cloud to the selected volumes or to create volumes with the restored data.

If the cloud account is shared among systems, IBM Storage Virtualize queries the snapshots that are stored in the cloud, and enables you to restore to a new volume. To restore a volume's snapshot, complete the following steps:

1. Open the Cloud Volumes window.
2. Right-click a volume and select **Restore**, as shown in Figure 10-79.

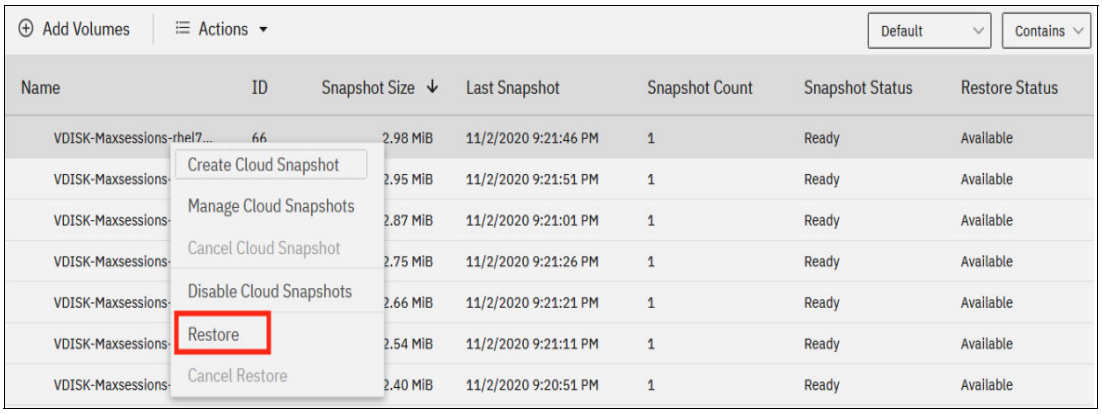


Figure 10-79 Selecting a volume to restore a snapshot from

3. A list of available snapshots is displayed. The snapshots date (PiT), their type (full or incremental), their state, and their size are shown (see Figure 10-80 on page 802). Select the version that you want to restore and click **Next**.

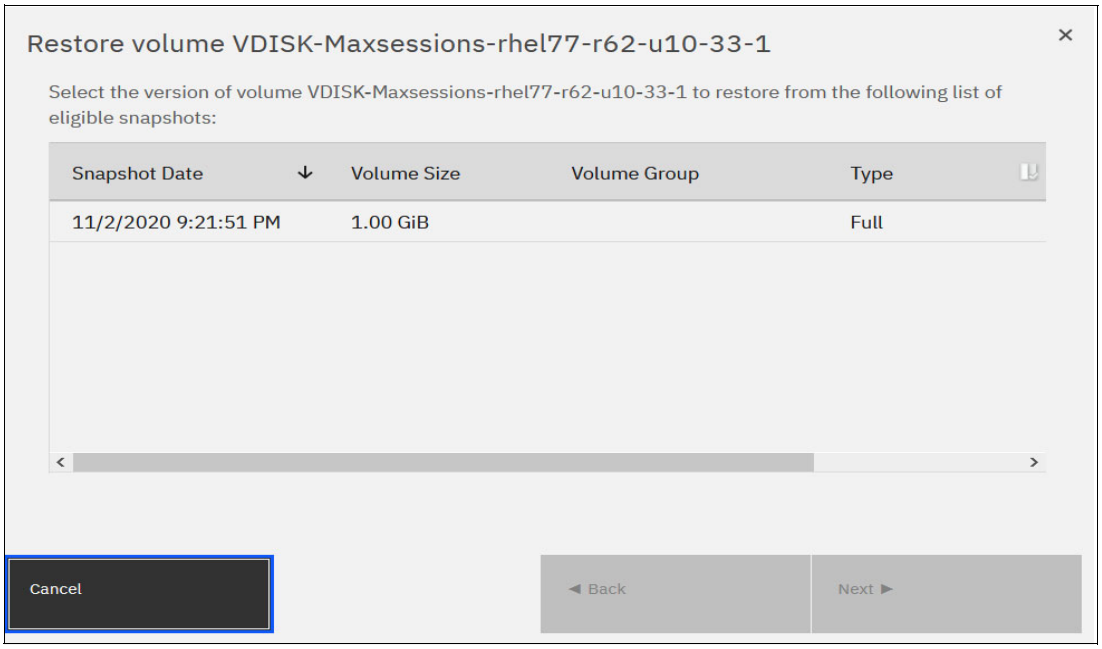


Figure 10-80 Selecting a snapshot version to restore

If the snapshot version that you selected has later generations (more recent Snapshot dates), the newer copies are removed from the cloud.

4. The IBM Storage Virtualize GUI provides two options to restore the snapshot from cloud. You can restore the snapshot from cloud directly to the selected volume, or create a volume to restore the data on, as shown in Figure 10-81. Make a selection and click **Next**.

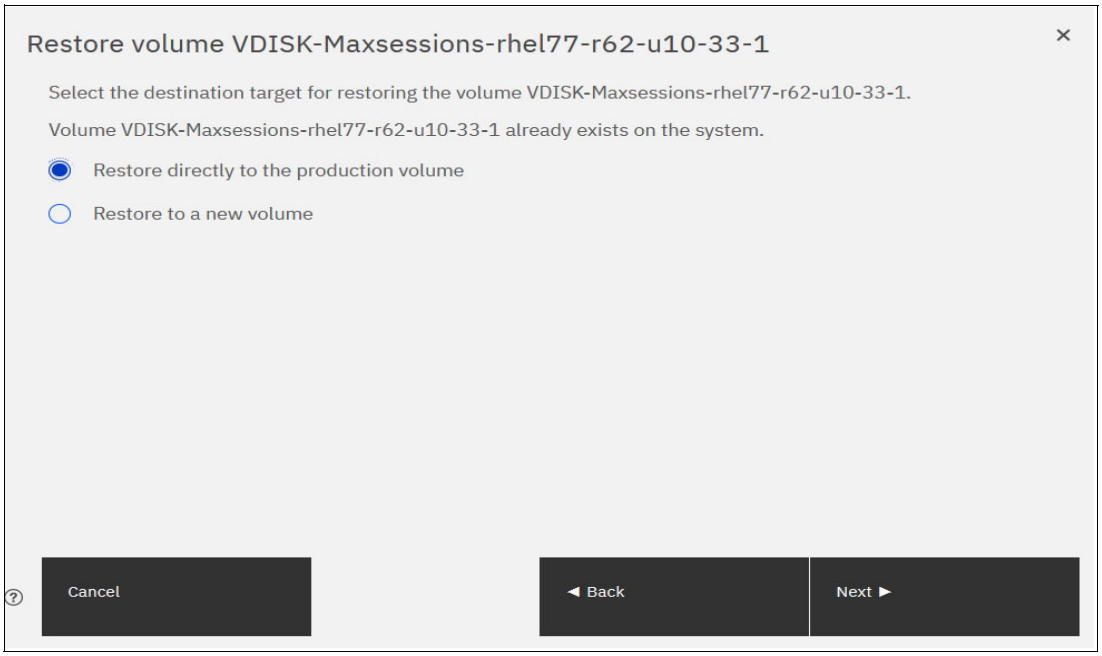


Figure 10-81 Restoring a snapshot on an existing volume or on a new volume

Note: Restoring a snapshot on the volume overwrites the data on the volume. The volume is taken offline (no read or write access), and the data from the PiT copy of the volume is written. The volume returns back online when all data is restored from the cloud.

5. If you selected the **Restore to a new Volume** option, you must enter the following information for the volume to be created with the snapshot data, as shown in Figure 10-82:

   – Name
   – Storage Pool
   – Capacity Savings (None, Compressed or Thin-provisioned)
   – I/O group

   You are not asked to enter the volume size because the new volume's size is identical to the snapshot copy size

   Enter the settings for the new volume and click **Next**.

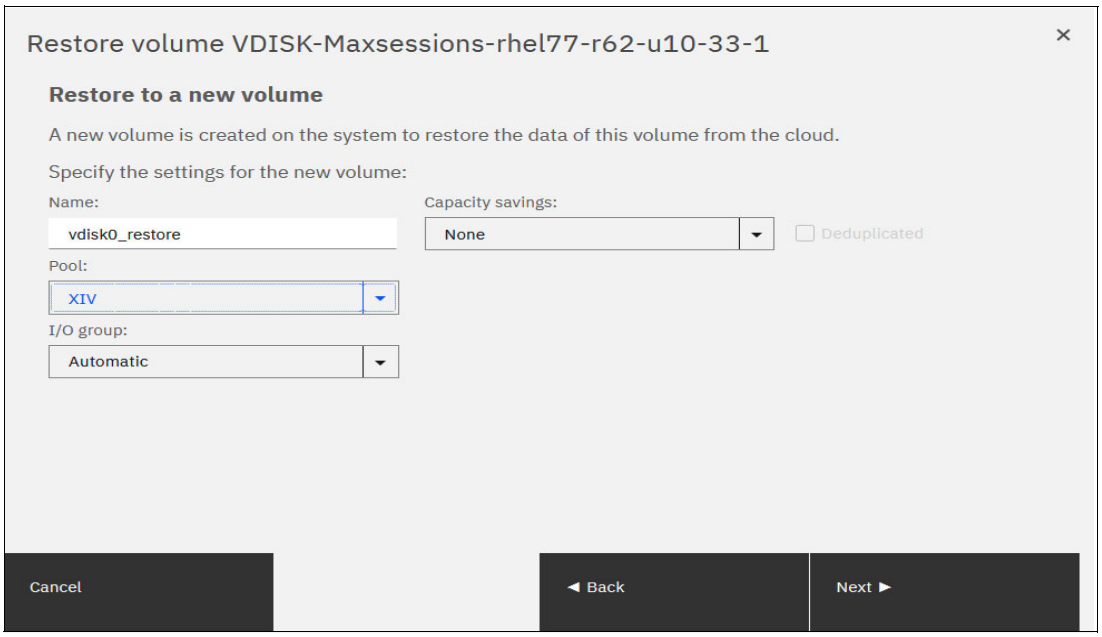


Figure 10-82 Restoring a snapshot to a new volume

6. A Summary window is displayed so you can review the restoration settings, as shown in Figure 10-83 on page 804. Click **Finish**. The system creates a volume or overwrites the selected volume. The more recent snapshots (later versions) of the volume are deleted from the cloud.

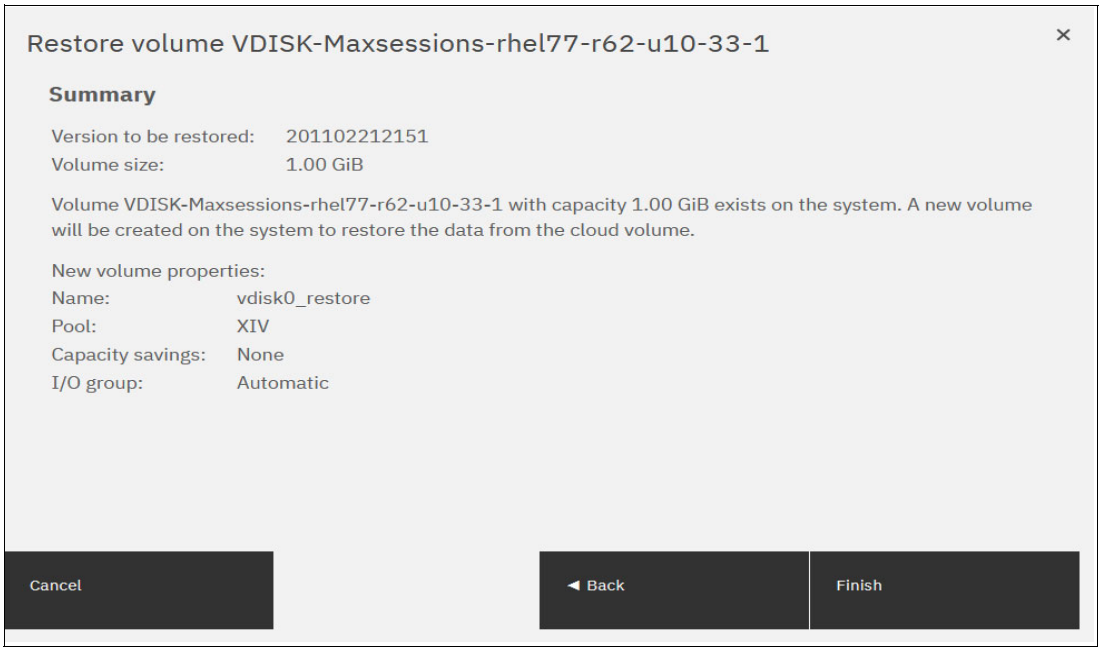


Figure 10-83 Restoring a snapshot summary

If you chose to restore the data from the cloud to a new volume, the new volume appears immediately in the volumes window. However, it is taken offline until all the data from the snapshot is written. The new volume is independent. It is not defined as a target in a FlashCopy mapping with the selected volume, for example. It is also not mapped to a host.

10.7 Volume mirroring and migration options

Volume mirroring is a simple Redundant Array of Independent Disks (RAID) 1-type function that enables a volume to remain online, even when the storage pool that is backing it becomes inaccessible. Volume mirroring is designed to protect the volume from storage infrastructure failures by seamless mirroring between storage pools.

Volume mirroring is provided by a specific volume mirroring function in the I/O stack. It cannot be manipulated like a FlashCopy or other types of copy volumes. However, this feature provides migration functionality, which can be obtained by splitting the mirrored copy from the source or by using the *migrate to* function. Volume mirroring cannot control backend storage mirroring or replication.

With volume mirroring, host I/O completes when both copies are written. This feature is enhanced with a tunable latency tolerance. This tolerance provides an option to give preference to losing the redundancy between the two copies. This tunable timeout value is `Latency` or `Redundancy`.

The `Latency` tuning option, which is set by running the **`chvdisk -mirrorwritepriority latency`** command, is the default. It prioritizes host I/O latency, which yields a preference to host I/O over availability. However, you might need to give preference to redundancy in your environment when availability is more important than I/O response time. Run the **`chvdisk -mirrorwritepriority redundancy`** command to set the redundancy option.

Regardless of which option you choose, volume mirroring can provide extra protection for your environment.

Migration offers the following options:

- Export to Image mode

  By using this option, you can move storage from `managed` mode to `image` mode, which is useful if you use the IBM Storage Virtualize system as a migration device. For example, vendor A's product cannot communicate with vendor B's product, but you must migrate data from vendor A to vendor B. By using Export to Image mode, you can migrate data by using Copy Services functions and then return control to the native array while maintaining access to the hosts.

- Import to Image mode

  By using this option, you can import a storage MDisk or logical unit number (LUN) with its data from an external storage system without putting metadata on it so that the data remains intact. After you import it, all copy services functions can be used to migrate the storage to other locations while the data remains accessible to your hosts.

- Volume migration by using volume mirroring and then by using Split into New Volume

  By using this option, you can use the available RAID 1 functions. You create two copies of data that initially has a set relationship (one volume with two copies, one primary and one secondary) but then break the relationship (two volumes, both primary and no relationship between them) to make them independent copies of data.

  You can use this option to migrate data between storage pools and devices. You might use this option if you want to move volumes to multiple storage pools. Each volume can have two copies at a time, which means that you can add only one copy to the original volume, and then you must split those copies to create another copy of the volume.

- Volume migration by using move to another pool

  By using this option, you can move any volume between storage pools without any interruption to the host access. This option is a quicker version of the Volume Mirroring and Split into New Volume option. You might use this option if you want to move volumes in a single step, or you do not have a volume mirror copy.

Migration: Although these migration methods do not disrupt access, a brief outage does occur to install the host drivers for your IBM Storage Virtualize system if they are not yet installed.

With volume mirroring, you can move data to different MDisks within the same storage pool or move data between different storage pools. The use of volume mirroring over volume migration is beneficial because with volume mirroring, storage pools do not need to have the same extent size as is the case with volume migration.

Note: Volume mirroring does not create a second volume before you split copies. Volume mirroring adds a second copy of the data under the same volume. Therefore, you have one volume that is presented to the host with two copies of data that are connected to this volume. Only splitting copies creates another volume, and then both volumes have only one copy of the data.

Starting with V7.3 and the introduction of the dual-layer cache architecture, mirrored volume performance was significantly improved. Lower cache is beneath the volume mirroring layer, which means that both copies have their own cache.

This approach helps when you have copies of different types; for example, generic and compressed, because both copies use its independent cache and performs its own read prefetch. Destaging of the cache can be done independently for each copy, so one copy does not affect performance of a second copy.

Also, because the IBM Storage Virtualize destage algorithm is MDisk aware, it can tune or adapt the destaging process, depending on MDisk type and usage, for each copy independently.

For more information about volume mirroring, see Chapter 6, “Volumes” on page 419.

10.8 Remote Copy

This section describes the Remote Copy (RC) services, which are a synchronous RC called Metro Mirror (MM), and two asynchronous RC options, called Global Mirror (GM) and Global Mirror with Change Volumes (GMCV). RC in an IBM Storage Virtualize system is similar to RC in the IBM System Storage DS8000 family at a functional level, but the implementation differs.

IBM Storage Virtualize provides a single point of control when RC is enabled in your cluster (regardless of the disk subsystems that are used as underlying storage, if those disk subsystems are supported).

The general application of RC services is to maintain two real-time synchronized copies of a volume. Often, the two copies are geographically dispersed between two IBM Storage Virtualize systems. However, it is possible to use MM or GM within a single system (within an I/O group). If the master copy fails, you can enable an auxiliary copy for I/O operations.

Tips: Intracluster MM/GM uses more resources within the system when compared to an intercluster MM/GM relationship, where resource allocation is shared between the systems. Use intercluster MM/GM when possible. For mirroring volumes in the same system, it is better to use volume mirroring or the FlashCopy feature.

A typical application of this function is to set up a dual-site solution that uses two IBM Storage Virtualize systems. The first site is considered the *primary site* or *production site*, and the second site is considered the *backup site* or *failover site*. The failover site is activated when a failure at the first site is detected.

When MM or GM is used, a specific amount of bandwidth is required for the system intercluster heartbeat traffic. The amount of traffic depends on how many nodes are in each of the two clustered systems.

Table 10-11 lists the amount of heartbeat traffic (in megabits per second) that is generated by various sizes of clustered systems.

Table 10-11 Intersystem heartbeat traffic in Mbps

IBM Storage Virtualize system 1	IBM Storage Virtualize system 2			
	2 nodes	4 nodes	6 nodes	8 nodes
2 nodes	5	6	6	6
4 nodes	6	10	11	12
6 nodes	6	11	16	17
8 nodes	6	12	17	21

10.8.1 IBM SAN Volume Controller and IBM FlashSystem system layers

An IBM Storage Virtualize based system can be in one of the two layers: the *replication* layer or the *storage* layer. The system layer affects how the system interacts with other IBM Storage Virtualize based systems. The IBM SAN Volume Controller is always set to the replication layer. This parameter *cannot* be changed.

In the storage layer, an IBM FlashSystem system has the following characteristics and requirements:

- The system can perform MM and GM replication with other storage-layer systems.
- The system can provide external storage for replication-layer systems or IBM SAN Volume Controller.
- The system cannot use a storage-layer system as external storage.

In the replication layer, an IBM SAN Volume Controller or an IBM FlashSystem system has the following characteristics and requirements:

- Can perform MM and GM replication with other replication-layer systems
- Cannot provide external storage for a replication-layer system
- Can use a storage-layer system as external storage

An IBM FlashSystem family system is in the storage layer by default, but the layer can be changed. For example, you might want to change an IBM FlashSystem 7200 to a replication layer if you want to virtualize other IBM FlashSystem systems or to replicate to an IBM SAN Volume Controller system.

Note: Before you change the system layer, the following conditions must be met on the system at the time of layer change:

- No other IBM Storage Virtualized based system can exist as a backend or host entity.
- No system partnerships can exist.
- No other IBM Storage Virtualize based system can be visible on the SAN fabric.

The layer can be changed during normal host I/O.

In your IBM SAN Volume Controller run the **lssystem** command to check the current system layer, as shown in Example 10-3.

Example 10-3 Output from the lssystem command showing the system layer

```
IBM_FlashSystem:FS9500_Demo:demoUser>lssystem
id 00000204E1A0013A
name FS9500_Demo
...
time_zone 13 Africa/Casablanca
code_level 8.6.0.0 (build 169.9.2306081121000)
...
has_nas_key no
layer storage
rc_buffer_size 256
compression_active no
...
```

Note: Consider the following rules for creating remote partnerships between the IBM SAN Volume Controller and IBM FlashSystem Family systems:

- An IBM SAN Volume Controller is always in the replication layer.
- By default, the IBM FlashSystem systems are in the storage layer, but can be changed to the replication layer.
- A system can form partnerships only with systems in the same layer.
- Starting in V6.4, any IBM Storage Virtualize based system in the replication layer can virtualize an IBM FlashSystem system in the storage layer.

10.8.2 Multiple IBM Storage Virtualize systems replication

Each IBM Storage Virtualize system can maintain up to three partner system relationships, which enables as many as four systems to be directly associated with each other. This system partnership capability enables the implementation of disaster recovery (DR) solutions.

Note: For more information about restrictions and limitations of native IP replication, see 10.9.3, “IP partnership limitations” on page 844.

Figure 10-84 shows an example of a multiple systems mirroring configuration.

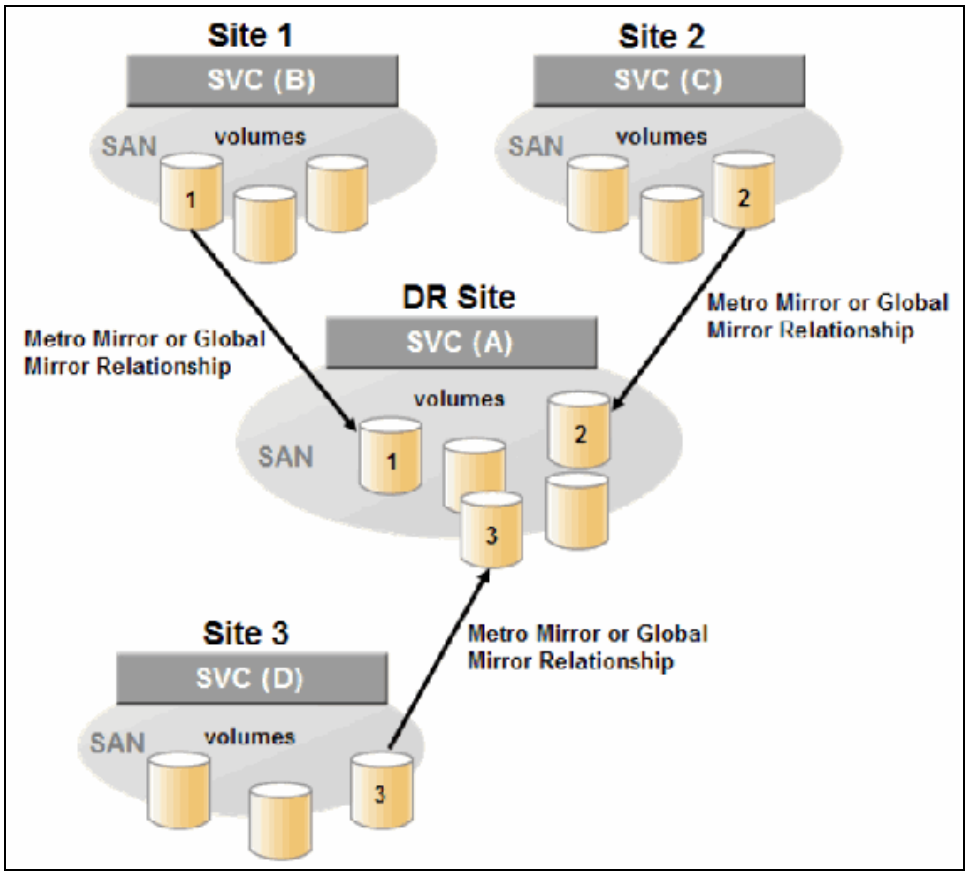


Figure 10-84 Multiple systems mirroring configuration example

Supported multiple systems mirroring topologies

Multiple systems mirroring supports various partnership topologies, as shown in the example in Figure 10-85. This example is a star topology (A → B, A → C, and A → D).

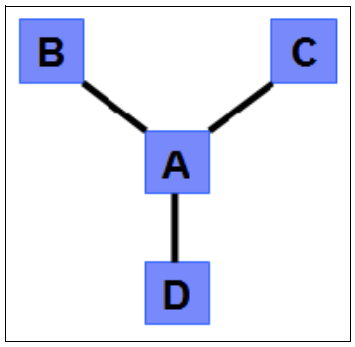


Figure 10-85 Star topology

Figure 10-85 shows four systems in a star topology, with System A at the center. System A can be a central DR site for the three other locations.

By using a star topology, you can migrate applications by using a process, such as the one described in the following example:

1. Suspend application at A.
2. Remove the A → B relationship.
3. Create the A → C relationship (or the B → C relationship).
4. Synchronize to system C, and ensure that A → C is established:
   - A → B, A → C, A → D, B → C, B → D, and C → D
   - A → B, A → C, and B → C

Figure 10-86 shows an example of a triangle topology (A → B, A → C, and B → C).

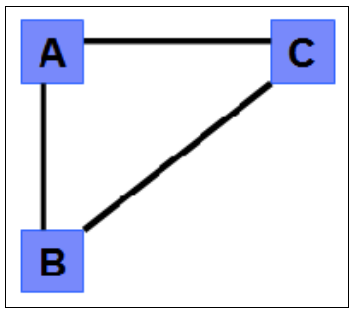


Figure 10-86 Triangle topology

Figure 10-87 shows an example of an IBM Storage Virtualize system fully connected topology (A → B, A → C, A → D, B → D, and C → D).

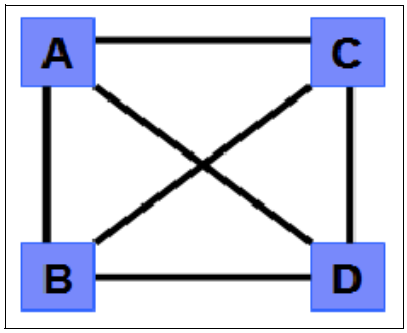


Figure 10-87 Fully connected topology

Figure 10-87 shows a fully connected mesh in which every system has a partnership to each of the three other systems. This topology enables volumes to be replicated between any pair of systems; for example, A → B, A → C, and B → C.

Figure 10-88 shows a daisy-chain topology.

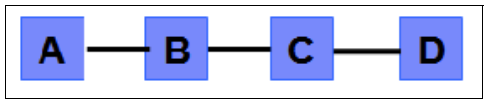


Figure 10-88 Daisy-chain topology

Although systems can have up to three partnerships, volumes can be part of only one RC relationship; for example, A → B.

System partnership intermix: All of these topologies are valid for the intermix of the IBM SAN Volume Controller with IBM FlashSystem if the IBM FlashSystem system is set to the replication layer.

With the IBM Storage Virtualize code version 8.3.1, a 3-site replication solution option was introduced, and was expanded with the 8.4 release. This release allows for active-active implementations while replicating to a third-site.

For more information, see *IBM Spectrum Virtualize 3-Site Replication*, SG24-850.

10.8.3 Importance of write ordering

Many applications that use block storage are required to survive failures, such as loss of power or a software crash, and to not lose data that existed before the failure. Because many applications must perform many update operations in parallel, maintaining write ordering is key to ensure the correct operation of applications after a disruption.

An application that performs many database updates is designed with the concept of dependent writes. With dependent writes, it is important to ensure that an earlier write completed before a later write is started. Reversing or performing the order of writes differently than the application intended can undermine the application's algorithms and can lead to problems, such as detected or undetected data corruption.

The IBM Storage Virtualize MM and GM implementation operate in a manner that it is designed to always keep a consistent image at the secondary site. The GM implementation uses complex algorithms that identify sets of data and number those sets of data in sequence. The data is then applied at the secondary site in this same defined sequence.

Operating in this manner ensures that if the relationship is in a `Consistent_Synchronized` state, the GM target data is at least crash consistent and supports quick recovery through application crash recovery facilities.

For more information about dependent writes, see 10.3.14, "FlashCopy and image mode volumes" on page 745.

Remote Copy consistency groups

An RC consistency group can contain an arbitrary number of relationships up to the maximum number of MM/GM relationships that is supported by the IBM Storage Virtualize system. MM/GM commands can be issued to an RC consistency group.

Therefore, these commands can be issued simultaneously for all MM/GM relationships that are defined within that consistency group, or to a single MM/GM relationship that is not part of an RC consistency group. For example, when a **`startrcconsistgrp`** command is issued to the consistency group, all of the MM/GM relationships in the consistency group are started at the same time.

Figure 10-89 shows the concept of RC consistency groups.

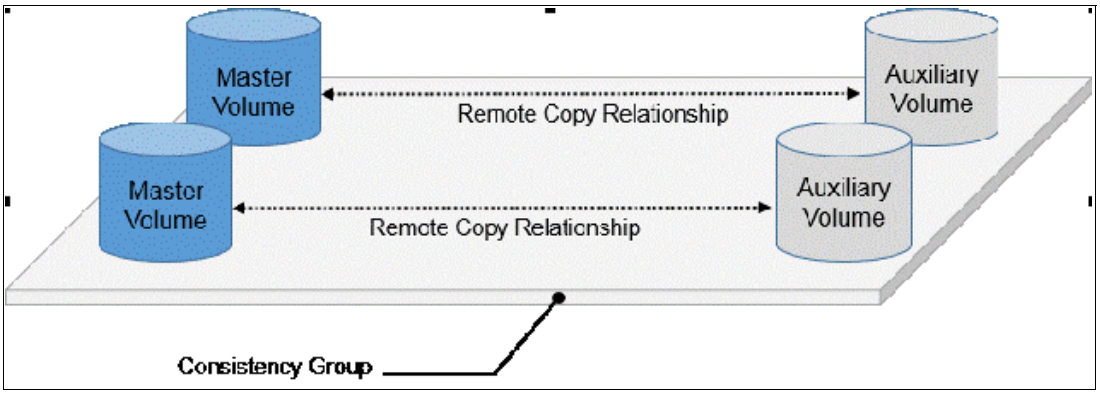


Figure 10-89 Remote Copy consistency group

Certain uses of MM/GM require the manipulation of more than one relationship. RC consistency groups can group relationships so that they are manipulated in unison.

Consider the following points:

- MM/GM relationships can be part of a consistency group, or they can be stand-alone and, therefore, are handled as single instances.
- A consistency group can contain zero or more relationships. An empty consistency group with zero relationships in it has little purpose until it is assigned its first relationship, except that it has a name.
- All relationships in a consistency group must have corresponding master and auxiliary volumes.
- All relationships in one consistency group must be the same type; for example, only MM or only GM.

Although consistency groups can be used to manipulate sets of relationships that do not need to satisfy these strict rules, this manipulation can lead to unwanted side effects. The rules behind a consistency group mean that certain configuration commands are prohibited. These configuration commands are not prohibited if the relationship is not part of a consistency group.

For example, consider the case of two applications that are independent, yet they are placed into a single consistency group. If an error occurs, synchronization is lost and a background copy process is required to recover synchronization. While this process is progressing, MM/GM rejects attempts to enable access to the auxiliary volumes of either application.

If one application finishes its background copy more quickly than the other application, MM/GM still refuses to grant access to its auxiliary volumes, even though it is safe in this case. The MM/GM policy is to refuse access to the entire consistency group if any part of it is inconsistent. Stand-alone relationships and consistency groups share a common configuration and state models. All of the relationships in a non-empty consistency group feature the same state as the consistency group.

10.8.4 Remote Copy intercluster communication

In the traditional Fibre Channel (FC), the intercluster communication between systems in a MM/GM partnership is performed over the SAN. This section describes this communication path.

For more information about intercluster communication between systems in an IP partnership, see 10.9.7, “States of IP partnership” on page 849.

Zoning

At least two FC ports of every node of each system must communicate with each other to create the partnership. Switch zoning is critical to facilitate intercluster communication.

Intercluster communication channels

When an IBM Storage Virtualize system partnership is defined on a pair of systems, the following intercluster communication channels are established:

- A single control channel, which is used to exchange and coordinate configuration information
- I/O channels between each of these nodes in the systems

These channels are maintained and updated as nodes and links appear and disappear from the fabric, and are repaired to maintain operation where possible. If communication between the systems is interrupted or lost, an event is logged (and the MM/GM relationships stop).

Alerts: You can configure the system to raise SNMP traps to the enterprise monitoring system to alert on events that indicate an interruption in internode communication occurred.

Intercluster links

All IBM Storage Virtualize nodes maintain a database of other devices that are visible on the fabric. This database is updated as devices appear and disappear.

Devices that advertise themselves as IBM SAN Volume Controller or IBM FlashSystem nodes are categorized according to the system to which they belong. Nodes that belong to the same system establish communication channels between themselves and exchange messages to implement clustering and the functional protocols of IBM Storage Virtualize.

Nodes that are in separate systems do not exchange messages after initial discovery is complete, unless they are configured together to perform an RC relationship.

The intercluster link carries control traffic to coordinate activity between two systems. The link is formed between one node in each system. The traffic between the designated nodes is distributed among logins that exist between those nodes.

If the designated node fails (or all of its logins to the remote system fail), a new node is chosen to carry control traffic. This node change causes the I/O to pause, but it does not put the relationships in a `ConsistentStopped` state.

Note: Run the **`chsystem`** command with **`-partnerfcportmask`** to dedicate several FC ports to only system-to-system traffic to ensure that RC is not affected by other traffic, such as host-to-node traffic or node-to-node traffic within the same system. Care must also be taken to align the FC ports when **`partnerfcportmask`** is used with portsets.

10.8.5 Metro Mirror overview

MM establishes a synchronous relationship between two volumes of equal size. The volumes in an MM relationship are referred to as the *master* (primary) volume and the *auxiliary* (secondary) volume. Traditional FC MM is primarily used in a metropolitan area or geographical area, up to a maximum distance of 300 km (186.4 miles) to provide synchronous replication of data.

With synchronous copies, host applications write to the master volume, but they do not receive confirmation that the write operation completed until the data is written to the auxiliary volume. This action ensures that both the volumes have identical data when the copy completes. After the initial copy completes, the MM function always maintains a fully synchronized copy of the source data at the target site.

MM has the following characteristics:

- Zero recovery point objective (RPO)
- Synchronous
- Production application performance that is affected by round-trip latency

Increased distance directly affects host I/O performance because the writes are synchronous. Use the requirements for application performance when you are selecting your MM auxiliary location.

Consistency groups can be used to maintain data integrity for dependent writes, which is similar to FlashCopy consistency groups.

IBM Storage Virtualize provides intracluster and intercluster MM, which are described next.

Intracluster Metro Mirror

Intracluster MM performs the intracluster copying of a volume, in which both volumes belong to the same system and I/O group within the system. Because it is within the same I/O group, sufficient bitmap space must exist within the I/O group for both sets of volumes and licensing on the system.

Important: Performing MM across I/O groups within a system is not supported.

Intercluster Metro Mirror

Intercluster MM performs intercluster copying of a volume, in which one volume belongs to a system and the other volume belongs to a separate system.

Two IBM Storage Virtualize systems must be defined in a partnership, which must be performed on both systems to establish a fully functional MM partnership.

By using standard single-mode connections, the supported distance between two systems in an MM partnership is 10 km (6.2 miles), although greater distances can be achieved by using extenders. For extended distance solutions, contact your IBM representative.

Limit: When a local fabric and a remote fabric are connected for MM purposes, the inter-switch link (ISL) hop count between a local node and a remote node cannot exceed seven.

10.8.6 Synchronous Remote Copy

MM is a fully synchronous RC technique that ensures that writes are committed at the master and auxiliary volumes before write completion is acknowledged to the host, but only if writes to the auxiliary volumes are possible.

Events, such as a loss of connectivity between systems, can cause mirrored writes from the master volume and the auxiliary volume to fail. In that case, MM suspends writes to the auxiliary volume and enables I/O to the master volume to continue to avoid affecting the operation of the master volumes.

Figure 10-90 shows how a write to the master volume is mirrored to the cache of the auxiliary volume before an acknowledgment of the write is sent back to the host that issued the write. This process ensures that the auxiliary is synchronized in real time if it is needed in a failover situation.

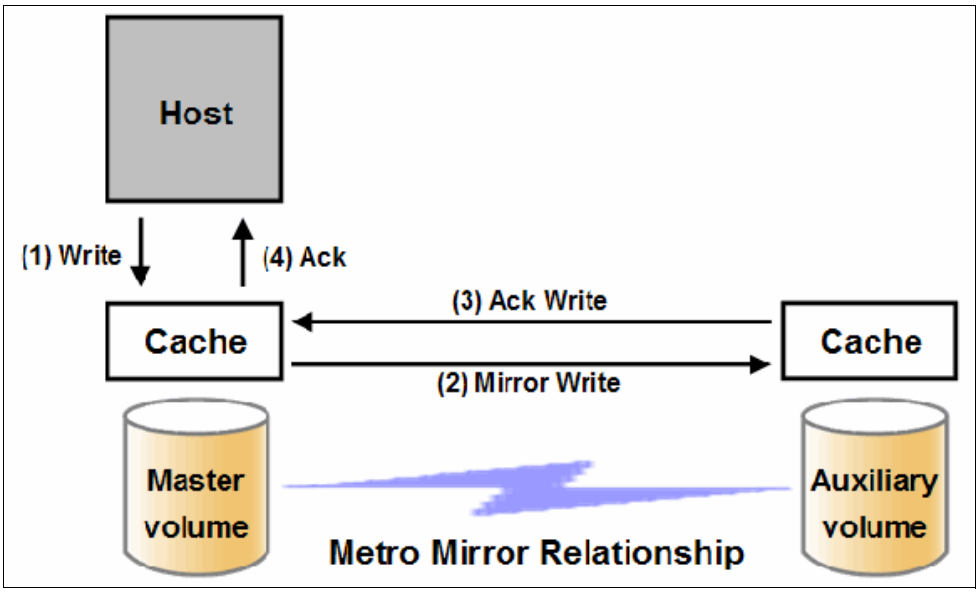


Figure 10-90 Write on volume in Metro Mirror relationship

However, this process also means that the application is exposed to the latency and bandwidth limitations (if any) of the communication link between the master and auxiliary volumes. This process might lead to unacceptable application performance, particularly when placed under peak load. Therefore, the use of traditional FC MM has distance limitations that are based on your performance requirements. IBM Storage Virtualize does not support more than 300 km (186.4 miles).

10.8.7 Metro Mirror features

The IBM Storage Virtualize MM function supports the following features:

- Synchronous RC of volumes that are dispersed over metropolitan distances.
- The MM relationships between volume pairs, with each volume in a pair that is managed by an IBM Storage Virtualize based system (requires V6.3.0 or later).
- Supports intracluster MM where both volumes belong to the same system (and I/O group).
- IBM Storage Virtualize supports intercluster MM where each volume belongs to a separate system. You can configure a specific system for partnership with another system.

All intercluster MM processing occurs between two IBM Storage Virtualize systems that are configured in a partnership.

- Intercluster and intracluster MM can be used concurrently.
- IBM Storage Virtualize does not require that a control network or fabric is installed to manage MM. For intercluster MM, the system maintains a control link between two systems. This control link is used to control the state and coordinate updates at either end. The control link is implemented on top of the same FC fabric connection that the system uses for MM I/O.
- IBM Storage Virtualize implements a configuration model that maintains the MM configuration and state through major events, such as failover, recovery, and resynchronization, to minimize user configuration action through these events.

IBM Storage Virtualize supports the resynchronization of changed data so that write failures that occur on the master or auxiliary volumes do not require a complete resynchronization of the relationship.

10.8.8 Metro Mirror attributes

The MM function in IBM Storage Virtualize features the following attributes:

- A partnership is created between two IBM Storage Virtualize systems that are operating in the replication layer (for intercluster MM).
- An MM relationship is created between two volumes of the same size.
- To manage multiple MM relationships as one entity, relationships can be made part of an MM consistency group, which ensures data consistency across multiple MM relationships and provides ease of management.
- When an MM relationship is started and when the background copy completes, the relationship becomes consistent and synchronized.
- After the relationship is synchronized, the auxiliary volume holds a copy of the production data at the primary, which can be used for DR.
- The auxiliary volume is in read-only mode when relationship is active.
- To access the auxiliary volume, the MM relationship must be stopped with the access option enabled before write I/O is allowed to the auxiliary.
- The remote host server is mapped to the auxiliary volume, and the disk is available for I/O.

10.8.9 Practical use of Metro Mirror

The master volume is the production volume, and updates to this copy are mirrored in real time to the auxiliary volume. The contents of the auxiliary volume that existed when the relationship was created are deleted.

Switching copy direction: The copy direction for an MM relationship can be switched so that the auxiliary volume becomes the master, and the master volume becomes the auxiliary, which is similar to the FlashCopy restore option. However, although the FlashCopy target volume can operate in read/write mode, the target volume of the started RC is always in read-only mode.

While the MM relationship is active, the auxiliary volume is not accessible for host application write I/O at any time. The IBM Storage Virtualize based systems allow read-only access to the auxiliary volume when it contains a consistent image. They also allow boot time operating system discovery to complete without an error, so that any hosts at the secondary site can be ready to start the applications with minimum delay, if required.

For example, many operating systems must read logical block address (LBA) zero to configure a logical unit (LU). Although read access is allowed at the auxiliary in practice, the data on the auxiliary volumes cannot be read by a host because most operating systems write a “dirty bit” to the file system when it is mounted. Because this write operation is not allowed on the auxiliary volume, the volume cannot be mounted.

This access is provided only where consistency can be ensured. However, coherency cannot be maintained between reads that are performed at the auxiliary and later write I/Os that are performed at the master.

To enable access to the auxiliary volume for host operations, you must stop the MM relationship by specifying the **-access** parameter. While access to the auxiliary volume for host operations is enabled, the host must be instructed to mount the volume before the application can be started, or instructed to perform a recovery process.

For example, the MM requirement to enable the auxiliary copy for access differentiates it from third-party mirroring software on the host, which aims to emulate a single, reliable disk regardless of what system is accessing it. MM retains the property that there are two volumes in existence, but it suppresses one volume while the copy is being maintained.

The use of an auxiliary copy demands a conscious policy decision by the administrator that a failover is required, and that the tasks to be performed on the host that is involved in establishing the operation on the auxiliary copy are substantial. The goal is to make this copy rapid (much faster when compared to recovering from a backup copy) but not seamless.

The failover process can be automated through failover management software. The IBM Storage Virtualize software provides SNMP traps and programming (or scripting) commands for the CLI to enable this automation.

10.8.10 Global Mirror overview

This section describes the GM copy service, which is an asynchronous RC service. This service provides and maintains a consistent mirrored copy of a source volume to a target volume.

GM function establishes a GM relationship between two volumes of equal size. The volumes in a GM relationship are referred to as the *master* (source) volume and the *auxiliary* (target) volume, which is the same as MM. Consistency groups can be used to maintain data integrity for dependent writes, which is similar to FlashCopy consistency groups.

GM writes data to the auxiliary volume asynchronously, which means that host writes to the master volume provide the host with confirmation that the write is complete before the I/O completes on the auxiliary volume.

GM has the following characteristics:

- Near-zero RPO
- Asynchronous
- Production application performance that is affected by I/O sequencing preparation time

Intracluster Global Mirror

Intracluster GM is not supported and has no functional value for production use. Intracluster MM provides the same capability with less processor use. However, leaving this functionality in place simplifies testing and supports client experimentation and testing (for example, to validate server failover on a single test system). As with Intracluster MM, you must consider the increase in the license requirement because source and target exist on the same IBM Storage Virtualize system.

Intercluster Global Mirror

Intercluster GM operations require a pair of IBM Storage Virtualize systems that are connected by several intercluster links. The two systems must be defined in a partnership to establish a fully functional GM relationship.

Limit: When a local fabric and a remote fabric are connected for GM purposes, the ISL hop count between a local node and a remote node must not exceed seven hops.

10.8.11 Asynchronous Remote Copy

GM is an asynchronous RC technique. In asynchronous RC, the write operations are completed on the primary site and the write acknowledgment is sent to the host before it is received at the secondary site. An update of this write operation is sent to the secondary site at a later stage, which provides the capability to perform RC over distances that exceed the limitations of synchronous RC.

The GM function provides the same function as MM RC, but over long-distance links with higher latency without requiring the hosts to wait for the full round-trip delay of the long-distance link.

Figure 10-91 shows that a write operation to the master volume is acknowledged back to the host that is issuing the write before the write operation is mirrored to the cache for the auxiliary volume.

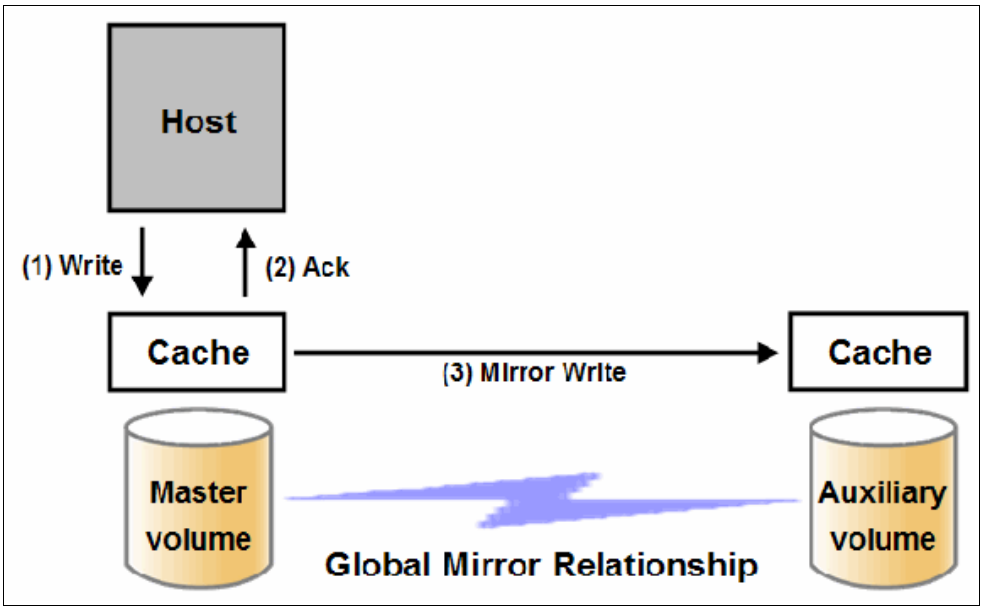


Figure 10-91 Global Mirror write sequence

The GM algorithms maintain a consistent image on the auxiliary. They achieve this consistent image by identifying sets of I/Os that are active concurrently at the master, assigning an order to those sets, and applying those sets of I/Os in the assigned order at the secondary. As a result, GM maintains the features of Write Ordering and Read Stability.

The multiple I/Os within a single set are applied concurrently. The process that marshals the sequential sets of I/Os operates at the secondary system. Therefore, the process is not subject to the latency of the long-distance link. These two elements of the protocol ensure that the throughput of the total system can be grown by increasing system size while maintaining consistency across a growing data set.

GM writes I/O from production system to a secondary system requires serialization and sequence-tagging before being sent across the network to a remote site (to maintain a write-order consistent copy of data).

To avoid affecting the production site, IBM Storage Virtualize supports more parallelism in processing and managing GM writes on the secondary system by using the following methods:

- Secondary system nodes store replication writes in new redundant non-volatile cache
- Cache content details are shared between nodes
- Cache content details are batched together to make node-to-node latency less of an issue
- Nodes intelligently apply these batches in parallel as soon as possible
- Nodes internally manage and optimize GM secondary write I/O processing

In a failover scenario where the secondary site must become the master source of data, specific updates might be missing at the secondary site. Therefore, any applications that use this data must have an external mechanism for recovering the missing updates and reapplying them, such as a transaction log replay.

GM is supported over FC, Fibre Channel over IP (FCIP), Fibre Channel over Ethernet (FCoE), and native IP connections. The maximum distance cannot exceed 80 ms round trip, which is approximately 4000 km (2485.48 miles) between mirrored systems. However, starting with IBM Storage Virtualize V7.4, this distance was increased to 250 ms for specific configurations. Table 10-12 shows the maximum latency for Global Mirror.

Table 10-12 Supported Global Mirror link latencies

System hardware	Partnership		
	FC	**1 Gbps - IP**	**10 Gbps - IP**
FS7300	250 ms	80 ms	10 ms
FS9500	250 ms	80 ms	10 ms
SV3	250 ms	80 ms	10 ms

10.8.12 Global Mirror features

IBM Storage Virtualize GM supports the following features:

- Asynchronous RC of volumes that are dispersed over metropolitan-scale distances.
- IBM Storage Virtualize implements the GM relationship between a volume pair, with each volume in the pair being managed by an IBM Storage Virtualize system.
- IBM Storage Virtualize supports intracluster GM where both volumes belong to the same system (and I/O group).

- An IBM Storage Virtualize system can be configured for partnership with 1 - 3 other systems. For more information about IP partnership restrictions, see 10.9.3, “IP partnership limitations” on page 844.
- IBM Storage Virtualize does not require a control network or fabric to be installed to manage GM. For intercluster GM, the system maintains a control link between the two systems. This control link is used to control the state and to coordinate the updates at either end. The control link is implemented on top of the same FC fabric connection that the system uses for GM I/O.
- IBM Storage Virtualize implements a configuration model that maintains the GM configuration and state through major events, such as failover, recovery, and resynchronization, to minimize user configuration action through these events.
- IBM Storage Virtualize implements flexible resynchronization support, enabling it to resynchronize volume pairs that experienced write I/Os to both disks, and to resynchronize only those regions that changed.
- An optional feature for GM is a delay simulation to be applied on writes that are sent to auxiliary volumes. It is useful in intracluster scenarios for testing purposes.

Colliding writes

The GM algorithm requires that only a single write is active on a volume. I/Os that overlap an active I/O are sequential, which is called *colliding writes.* If another write is received from a host while the auxiliary write is still active, the new host write is delayed until the auxiliary write is complete. This rule is needed if a series of writes to the auxiliary must be tried again and is called *reconstruction*. Conceptually, the data for reconstruction comes from the master volume.

If multiple writes are allowed to be applied to the master for a sector, only the most recent write gets the correct data during reconstruction. If reconstruction is interrupted for any reason, the intermediate state of the auxiliary is inconsistent. Applications that deliver such write activity do not achieve the performance that GM is intended to support. A volume statistic is maintained about the frequency of these collisions.

An attempt is made to allow multiple writes to a single location to be outstanding in the GM algorithm. Master writes must still be sequential, and the intermediate states of the master data must be kept in a non-volatile journal while the writes are outstanding to maintain the correct write ordering during reconstruction. Reconstruction must never overwrite data on the auxiliary with an earlier version. The volume statistic that is monitoring colliding writes is now limited to those writes that are not affected by this change.

Figure 10-92 shows a colliding write sequence example.

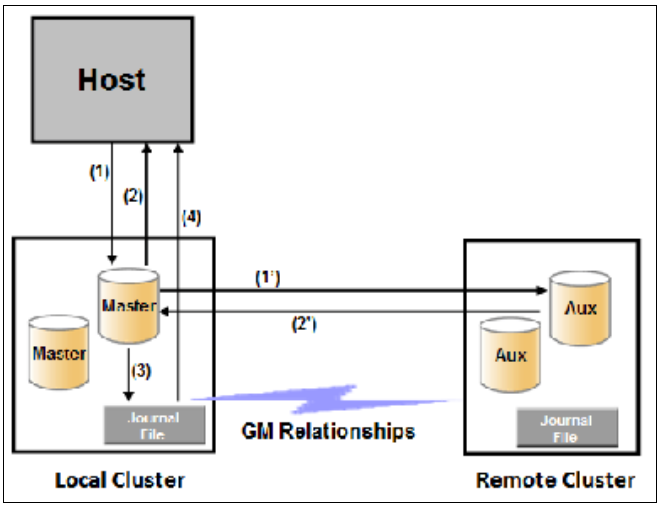


Figure 10-92 Colliding writes example

The following numbers correspond to the numbers that are shown in Figure 10-92:

- (1) The first write is performed from the host to LBA X.
- (2) The host is provided acknowledgment that the write completed, even though the mirrored write to the auxiliary volume is not yet complete.
- (1') and (2') occur asynchronously with the first write.
- (3) The second write is performed from the host also to LBA X. If this write occurs before (2'), the write is written to the journal file.
- (4) The host is provided acknowledgment that the second write is complete.

Delay simulation

GM provides a feature that enables a delay simulation to be applied on writes that are sent to the auxiliary volumes. With this feature, tests can be done to detect colliding writes. It also provides the capability to test an application before the full deployment. The feature can be enabled separately for each of the intracluster or intercluster GMs.

By running the **chsystem** command, the delay setting can be set up and the delay can be checked by running the **lssystem** command. The `gm_intra_cluster_delay_simulation` field expresses the amount of time that intracluster auxiliary I/Os are delayed. The `gm_inter_cluster_delay_simulation` field expresses the amount of time that intercluster auxiliary I/Os are delayed. A value of zero disables the feature.

Tip: If you are experiencing repeated problems with the delay on your link, ensure that the delay simulator was correctly disabled.

10.8.13 Using Global Mirror with change volumes

GM is designed to achieve an RPO as low as possible so that data is as up-to-date as possible. This design places several strict requirements on your infrastructure. In certain situations with low network link quality, congested hosts, or overloaded hosts, you might be affected by multiple 1920 congestion errors.

Congestion errors occur in the following primary situations:

- At the source site through the host or network
- In the network link or network path
- At the target site through the host or network

GM includes functions that address the following conditions, which might negatively affect certain GM implementations:

- The estimation of the bandwidth requirements tends to be complex.
- Ensuring that the latency and bandwidth requirements can be met is often difficult.
- Congested hosts on the source or target site can cause disruption.
- Congested network links can cause disruption with only intermittent peaks.

To address these issues, change volumes were added as an option for GM relationships. Change volumes use the FlashCopy functionality, but they cannot be manipulated as FlashCopy volumes because they are for a special purpose only. Change volumes replicate PiT images on a cycling period. The default is 300 seconds.

Your change rate must include only the condition of the data at the PiT that the image was taken, rather than all the updates during the period. The use of this function can provide significant reductions in replication volume.

Global Mirror with Change Volumes (GMCV) has the following characteristics:

- Larger RPO
- PiT copies
- Asynchronous
- Possible system performance resource requirements because PiT copies are created locally

Figure 10-93 shows a simple Global Mirror relationship without change volumes.

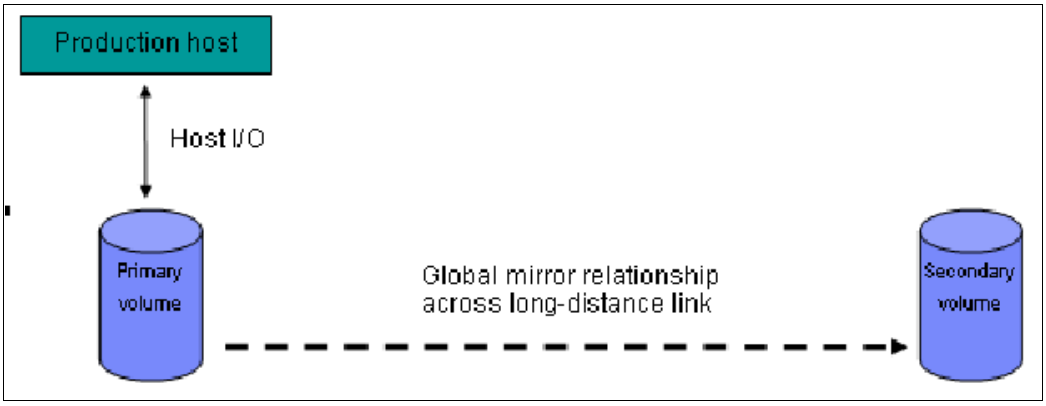


Figure 10-93 Global Mirror without change volumes

With change volumes, this environment looks as it is shown in Figure 10-94.

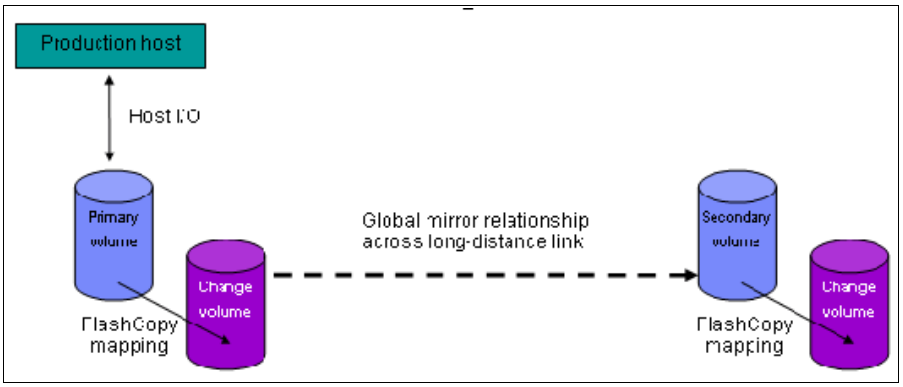


Figure 10-94 Global Mirror with Change Volumes

With change volumes, a FlashCopy mapping exists between the primary volume and the primary change volume. The mapping is updated in the cycling period (60 seconds - 1 day). The primary change volume is then replicated to the secondary GM volume at the target site, which is then captured in another change volume on the target site. This approach provides an always consistent image at the target site and protects your data from being inconsistent during resynchronization.

For more information about IBM FlashCopy, see 10.2, “Safeguarded Copy” on page 718.

You can adjust the cycling period by running the **`chrcrelationship -cycleperiodseconds <60 - 86400>`** command from the CLI. The default value is 300 seconds. If a copy does not complete in the cycle period, the next cycle does not start until the prior cycle completes. For this reason, the use of change volumes gives you the following possibilities for RPO:

- If your replication completes in the cycling period, your RPO is twice the cycling period.
- If your replication does not complete within the cycling period, RPO is twice the completion time. The next cycling period starts immediately after the prior cycling period is finished.

Carefully consider your business requirements versus the performance of GMCV. GMCV increases the intercluster traffic for more frequent cycling periods. Therefore, selecting the shortest cycle periods possible is not always the answer. In most cases, the default must meet requirements and perform well.

Important: When you create your Global Mirror volumes with change volumes, ensure that you remember to select the change volume on the auxiliary (target) site. Failure to do so leaves you exposed during a resynchronization operation.

10.8.14 Distribution of work among nodes

For the best performance, MM/GM volumes must have their preferred nodes that are evenly distributed among the nodes of the systems. Each volume within an I/O group has a preferred node property that can be used to balance the I/O load between nodes in that group. MM/GM also uses this property to route I/O between systems.

If this preferred practice is not maintained, such as if source volumes are assigned to only one node in the I/O group, you can change the preferred node for each volume to distribute volumes evenly between the nodes. You can also change the preferred node for volumes that are in an RC relationship without affecting the host I/O to a particular volume.

The RC relationship type does not matter. The RC relationship type can be MM, GM, or GMCV. You can change the preferred node both to the source and target volumes that are participating in the RC relationship.

10.8.15 Background copy performance

The background copy performance is subject to sufficient RAID controller bandwidth. Performance is also subject to other potential bottlenecks, such as the intercluster fabric, and possible contention from host I/O for the IBM Storage Virtualize system bandwidth resources.

Background copy I/O is scheduled to avoid bursts of activity that might have an adverse effect on system behavior. An entire grain of tracks on one volume is processed at around the same time, but not as a single I/O.

Double buffering is used to try to use sequential performance within a grain. However, the next grain within the volume might not be scheduled for some time. Multiple grains might be copied simultaneously, and might be enough to satisfy the requested rate, unless the available resources cannot sustain the requested rate.

GM paces the rate at which background copy is performed by the appropriate relationships. Background copy occurs on relationships that are in the `InconsistentCopying` state with a status of `Online`.

The quota of background copy (configured on the intercluster link) is divided evenly between all nodes that are performing background copy for one of the eligible relationships. This allocation is made irrespective of the number of disks for which the node is responsible. Each node in turn divides its allocation evenly between the multiple relationships that are performing a background copy.

The default value of the background copy is 25 MBps, per volume.

Important: The background copy value is a system-wide parameter that can be changed dynamically, but only on a *per-system* basis and not on a *per-relationship* basis. Therefore, the copy rate of all relationships changes when this value is increased or decreased. In systems with many RC relationships, increasing this value might affect overall system or intercluster link performance. The background copy rate can be changed to 1 - 1000 MBps.

10.8.16 Thin-provisioned background copy

MM/GM relationships preserve the space-efficiency of the master. Conceptually, the background copy process detects a deallocated region of the master and sends a special *zero buffer* to the auxiliary.

If the auxiliary volume is thin-provisioned and the region is deallocated, the special buffer prevents a write and therefore, an allocation. If the auxiliary volume is not thin-provisioned or the region in question is an allocated region of a thin-provisioned volume, a buffer of “real” zeros is synthesized on the auxiliary and written as normal.

10.8.17 Methods of synchronization

This section describes two methods that can be used to establish a synchronized relationship.

Full synchronization after creation

The full synchronization after creation method is the default method. It is the simplest method in that it requires no administrative activity apart from running the necessary commands. However, in certain environments, the available bandwidth can make this method unsuitable.

Run the following command sequence for a single relationship:

- Run **`mkrcrelationship`** without specifying the **`-sync`** option.
- Run **`startrcrelationship`** without specifying the **`-clean`** option.

Synchronized before creation

In this method, the administrator must ensure that the master and auxiliary volumes contain identical data before creating the relationship by using the following technique:

- Both disks are created with the security delete feature to make all data zero.
- A complete tape image (or other method of moving data) is copied from one disk to the other disk.

With this technique, do not allow I/O on the master or auxiliary before the relationship is established. Then the administrator must run the following commands:

- Run **`mkrcrelationship`** with the **`-sync`** flag.
- Run **`startrcrelationship`** without the **`-clean`** flag.

Important: Failure to perform these steps correctly can cause MM/GM to report the relationship as consistent when it is not. This use can cause loss of a data or data integrity exposure for hosts that are accessing data on the auxiliary volume.

10.8.18 Practical use of Global Mirror

The practical use of GM is similar to MM, as described in 10.8.9, “Practical use of Metro Mirror” on page 816. The main difference between the two RC modes is that GM and GMCV are mostly used on much larger distances than MM. Weak link quality or insufficient bandwidth between the primary and secondary sites can also be a reason to prefer asynchronous GM over synchronous MM. Otherwise, the use cases for MM/GM are the same.

10.8.19 IBM Storage Virtualize HyperSwap topology

The IBM HyperSwap topology is based on IBM Storage Virtualize RC mechanisms. It is also referred to as an “active-active relationship” in this document.

You can create an HyperSwap topology system configuration where each I/O group in the system is physically on a different site. These configurations can be used to maintain access to data on the system when power failures or site-wide outages occur.

In a HyperSwap configuration, each site is defined as an independent failure domain. If one site experiences a failure, the other site can continue to operate without disruption. You must also configure a third site to host a quorum device or IP quorum application that provides an automatic tie-break in case of a link failure between the two main sites. The main site can be in the same room or across rooms in the data center, buildings on the same campus, or buildings in different cities. Different kinds of sites protect against different types of failures.

For more information about HyperSwap implementation and best practices, see *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597.

10.8.20 Consistency Protection for Global Mirror and Metro Mirror

Metro Mirror, Global Mirror, Global Mirror with Change Volumes, and HyperSwap Copy Services functions create RC or remote replication relationships between volumes or consistency groups. If the secondary volume in a Copy Services relationship becomes unavailable to the primary volume, the system maintains the relationship. However, the data might become out of sync when the secondary volume becomes available.

Since V7.8, it is possible to create a FlashCopy mapping (change volume) for an RC target volume to maintain a consistent image of the secondary volume. The system recognizes it as a *Consistency Protection* and a link failure or an offline secondary volume event is handled differently now.

When Consistency Protection is configured, the relationship between the primary and secondary volumes does not stop if the link goes down, or the secondary volume is offline. The relationship does not go in to the consistent stopped status. Instead, the system uses the secondary change volume to automatically copy the previous consistent state of the secondary volume. The relationship automatically moves to the consistent copying status as the system resynchronizes and protects the consistency of the data. The relationship status changes to `consistent synchronized` when the resynchronization process completes. The relationship automatically resumes replication after the temporary loss of connectivity.

Change volumes that are used for Consistency Protection are not visible and manageable from the GUI because they are used for Consistency Protection internal behavior only.

It is not required to configure a secondary change volume on a MM/GM (without cycling) relationship. However, if the link goes down or the secondary volume is offline, the relationship goes in to the `Consistent stopped` status. If write operations occur on the primary or secondary volume, the data is no longer synchronized (Out of sync).

Consistency protection must be enabled on all relationships in a consistency group. Every relationship in a consistency group must be configured with a secondary change volume. If a secondary change volume is not configured on one relationship, the entire consistency group stops with a 1720 error if host I/O is processed when the link is down or any secondary volume in the consistency group is offline. All relationships in the consistency group are unable to retain a consistent copy during resynchronization.

The option to add consistency protection is selected by default when MM/GM relationships are created. The option must be cleared to create MM/GM relationships without consistency protection.

10.8.21 Valid combinations of FlashCopy, Metro Mirror, and Global Mirror

Table 10-13 lists the combinations of FlashCopy and MM/GM functions that are valid for a single volume.

Table 10-13 Valid combination for a single volume

FlashCopy	MM or GM source	MM or GM target
FlashCopy Source	Supported	Supported
FlashCopy Target	Supported	Not supported

10.8.22 Remote Copy configuration limits

Table 10-14 lists the MM/GM configuration limits.

Table 10-14 Metro Mirror configuration limits

Parameter	Value
Number of Metro Mirror or GM consistency groups per system	256
Number of Metro Mirror or GM relationships per system	10000
Number of Metro Mirror or GM relationships per consistency group	10000
Total Metro Mirror and Global Mirror capacity per I/O group	A per I/O group limit of 2 PiB exists on the quantity of master and auxiliary volume address spaces that can participate in Metro Mirror and GM relationships. This maximum configuration uses all 1024 MiB (512 MiB for SV1) of bitmap space for the I/O group and allows 256 MiB of space for all remaining copy services features.

10.8.23 Remote Copy states and events

This section describes the various states of an MM/GM relationship and the conditions that cause them to change. An overview of the status that can apply to a MM/GM relationship in a connected state is shown in Figure 10-95 on page 828.

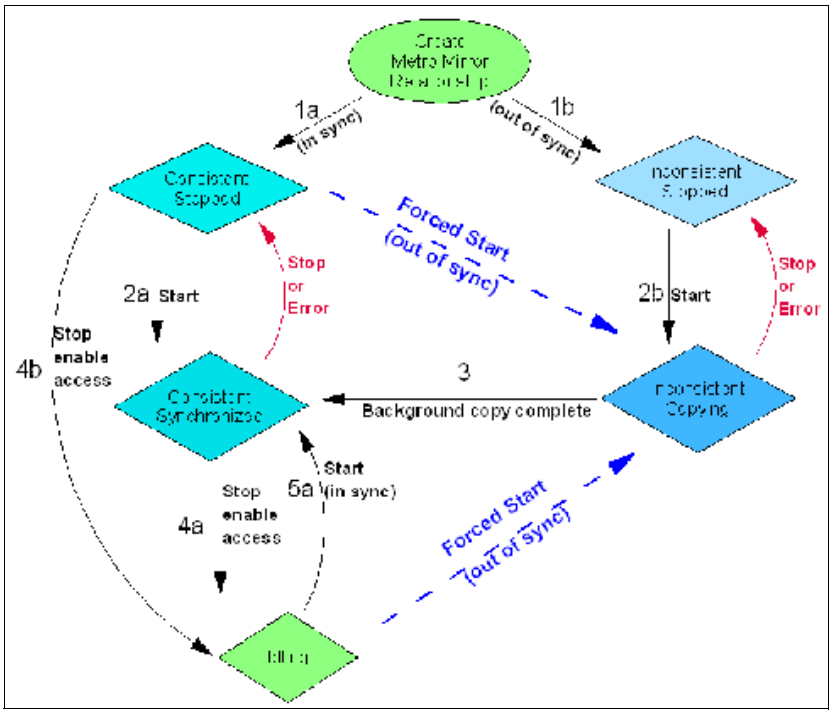


Figure 10-95 Metro Mirror or Global Mirror mapping state diagram

When the MM/GM relationship is created, you can specify whether the auxiliary volume is in sync with the master volume, and the background copy process is then skipped. This capability is useful when MM/GM relationships are established for volumes that were created with the format option.

The following step identifiers are shown in Figure 10-95:

- Step 1:
  a. The MM/GM relationship is created with the **-sync** option, and the MM/GM relationship enters the `ConsistentStopped` state.
  b. The MM/GM relationship is created without specifying that the master and auxiliary volumes are in sync, and the MM/GM relationship enters the `InconsistentStopped` state.
- Step 2:
  a. When an MM/GM relationship is started in the `ConsistentStopped` state, the MM/GM relationship enters the `ConsistentSynchronized` state. Therefore, no updates (write I/O) were performed on the master volume while in the `ConsistentStopped` state. Otherwise, the **-force** option must be specified, and the MM/GM relationship then enters the `InconsistentCopying` state while the background copy is started.
  b. When an MM/GM relationship is started in the `InconsistentStopped` state, the MM/GM relationship enters the `InconsistentCopying` state while the background copy is started.
- Step 3

  When the background copy completes, the MM/GM relationship changes from the `InconsistentCopying` state to the `ConsistentSynchronized` state.

- Step 4:
  a. When a MM/GM relationship is stopped in the `ConsistentSynchronized` state, the MM/GM relationship enters the `Idling` state when you specify the **`-access`** option, which enables write I/O on the auxiliary volume.
  b. When an MM/GM relationship is stopped in the `ConsistentSynchronized` state without an **`-access`** parameter, the auxiliary volumes remain read-only and the state of the relationship changes to `ConsistentStopped`.
  c. To enable write I/O on the auxiliary volume, when the MM/GM relationship is in the `ConsistentStopped` state, run the **`svctask stoprcrelationship`** command, which specifies the **`-access`** option, and the MM/GM relationship enters the `Idling` state.
- Step 5:
  a. When an MM/GM relationship is started from the `Idling` state, you must specify the **`-primary`** argument to set the copy direction. If no write I/O was performed (to the master or auxiliary volume) while in the `Idling` state, the MM/GM relationship enters the `ConsistentSynchronized` state.
  b. If write I/O was performed to the master or auxiliary volume, the **`-force`** option must be specified and the MM/GM relationship then enters the `InconsistentCopying` state while the background copy is started. The background process copies only the data that changed on the primary volume while the relationship was stopped.

Stop on Error

When a MM/GM relationship is stopped (intentionally, or because of an error), the state changes. For example, the MM/GM relationships in the `ConsistentSynchronized` state enter the `ConsistentStopped` state, and the MM/GM relationships in the `InconsistentCopying` state enter the `InconsistentStopped` state.

If the connection is broken between the two systems that are in a partnership, all (intercluster) MM/GM relationships enter a `Disconnected` state. For more information, see “Connected versus disconnected” on page 829.

Common states: Stand-alone relationships and consistency groups share a common configuration and state model. All MM/GM relationships in a consistency group have the same state as the consistency group.

State overview

The following sections provide an overview of the various MM/GM states.

Connected versus disconnected

Under certain error scenarios (for example, a power failure at one site that causes one complete system to disappear), communications between two systems in an MM/GM relationship can be lost. Alternatively, the fabric connection between the two systems might fail, which leaves the two systems that are running but cannot communicate with each other.

When the two systems can communicate, the systems and the relationships that span them are described as *connected*. When they cannot communicate, the systems and the relationships spanning them are described as *disconnected*.

In this state, both systems are left with fragmented relationships and are limited regarding the configuration commands that can be performed. The disconnected relationships are portrayed as having a changed state. The new states describe what is known about the relationship and the configuration commands that are permitted.

When the systems can communicate again, the relationships are reconnected. MM/GM automatically reconciles the two state fragments and considers any configuration or other event that occurred while the relationship was disconnected. As a result, the relationship can return to the state that it was in when it became disconnected, or it can enter a new state.

Relationships that are configured between volumes in the same IBM Storage Virtualize based system (intracluster) are never described as being in a disconnected state.

Consistent versus inconsistent

Relationships that contain volumes that are operating as secondaries can be described as being consistent or inconsistent. Consistency groups that contain relationships can also be described as being consistent or inconsistent. The consistent or inconsistent property describes the relationship of the data on the auxiliary to the data on the master volume. It can be considered a property of the auxiliary volume.

An auxiliary volume is described as *consistent* if it contains data that can be read by a host system from the master if power failed at an imaginary point while I/O was in progress, and power was later restored. This imaginary point is defined as the *recovery point*.

The requirements for consistency are expressed regarding activity at the master up to the recovery point. The auxiliary volume contains the data from all of the writes to the master for which the host received successful completion and that data was not overwritten by a subsequent write (before the recovery point).

Consider writes for which the host did not receive a successful completion (that is, it received bad completion or no completion at all). If the host then performed a read from the master of that data that returned successful completion and no later write was sent (before the recovery point), the auxiliary contains the same data as the data that was returned by the read from the master.

From the point of view of an application, consistency means that an auxiliary volume contains the same data as the master volume at the recovery point (the time at which the imaginary power failure occurred). If an application is designed to cope with an unexpected power failure, this assurance of consistency means that the application can use the auxiliary and begin operation as though it was restarted after the hypothetical power failure. Again, maintaining the application write ordering is the key property of consistency.

For more information about dependent writes, see 10.3.14, “FlashCopy and image mode volumes” on page 745.

If a relationship (or set of relationships) is inconsistent and an attempt is made to start an application by using the data in the secondaries, the following outcomes are possible:

- The application might decide that the data is corrupted and crash or exit with an event code.
- The application might fail to detect that the data is corrupted and return erroneous data.
- The application might work without a problem.

Because of the risk of data corruption, and in particular undetected data corruption, MM/GM strongly enforces the concept of consistency and prohibits access to inconsistent data.

Consistency as a concept can be applied to a single relationship or a set of relationships in a consistency group. Write ordering is a concept that an application can maintain across several disks that are accessed through multiple systems. Therefore, consistency must operate across all of those disks.

When you are deciding how to use consistency groups, the administrator must consider the scope of an application's data and consider all of the interdependent systems that communicate and exchange information.

If two programs or systems communicate and store details as a result of the information that is exchanged, either of the following actions might occur:

- All of the data that is accessed by the group of systems must be placed into a single consistency group.
- The systems must be recovered independently (each within its own consistency group). Then each system must perform recovery with the other applications to become consistent with them.

Consistent versus synchronized

A copy that is consistent and up-to-date is described as *synchronized*. In a synchronized relationship, the master and auxiliary volumes differ only in regions where writes are outstanding from the host.

Consistency does not mean that the data is up-to-date. A copy can be consistent and yet contain data that was frozen at a point in the past. Write I/O might continue to a master but not be copied to the auxiliary. This state arises when it becomes impossible to keep data up-to-date and maintain consistency. An example is a loss of communication between systems when you are writing to the auxiliary.

When communication is lost for an extended period and Consistency Protection was not enabled, MM/GM tracks the changes that occurred on the master, but not the order or the details of such changes (write data). When communication is restored, it is impossible to synchronize the auxiliary without sending write data to the auxiliary out of order. Therefore, consistency is lost.

Note: MM/GM relationships with Consistency Protection enabled use a PiT copy mechanism (FlashCopy) to keep a consistent copy of the auxiliary. The relationships stay in a consistent state (although not synchronized) even if communication is lost. For more information about Consistency Protection, see 10.8.20, "Consistency Protection for Global Mirror and Metro Mirror" on page 826.

Detailed states

The following sections describe the states that are portrayed to the user for consistency groups or relationships. Also described is the information that is available in each state. The major states are designed to provide guidance about the available configuration commands.

InconsistentStopped

`InconsistentStopped` is a connected state. In this state, the master is accessible for read and write I/O, but the auxiliary is not accessible for read or write I/O. A copy process must be started to make the auxiliary consistent. This state is entered when the relationship or consistency group was `InconsistentCopying` and suffered a persistent error or received a **`stop`** command that caused the copy process to stop.

A **`start`** command causes the relationship or consistency group to move to the `InconsistentCopying` state. A **`stop`** command is accepted, but has no effect.

If the relationship or consistency group becomes disconnected, the auxiliary side makes the transition to `InconsistentDisconnected`. The master side changes to `IdlingDisconnected`.

InconsistentCopying

`InconsistentCopying` is a connected state. In this state, the master is accessible for read and write I/O, but the auxiliary is not accessible for read or write I/O. This state is entered after a **`start`** command is issued to an `InconsistentStopped` relationship or a consistency group.

It is also entered when a forced start is issued to an `Idling` or `ConsistentStopped` relationship or consistency group. In this state, a background copy process runs that copies data from the master to the auxiliary volume.

In the absence of errors, an `InconsistentCopying` relationship is active, and the copy progress increases until the copy process completes. In certain error situations, the copy progress might freeze or even regress.

A persistent error or **`stop`** command places the relationship or consistency group into an `InconsistentStopped` state. A **`start`** command is accepted but has no effect.

If the background copy process completes on a stand-alone relationship or on all relationships for a consistency group, the relationship or consistency group changes to the `ConsistentSynchronized` state.

If the relationship or consistency group becomes disconnected, the auxiliary side changes to `InconsistentDisconnected`. The master side changes to `IdlingDisconnected`.

ConsistentStopped

`ConsistentStopped` is a connected state. In this state, the auxiliary contains a consistent image, but it might be out-of-date in relation to the master. This state can arise when a relationship was in a `ConsistentSynchronized` state and experienced an error that forces a Consistency Freeze. It can also arise when a relationship is created with a `CreateConsistentFlag` set to `TRUE`.

Normally, write activity that follows an I/O error causes updates to the master, and the auxiliary is no longer synchronized. In this case, consistency must be given up for a period to reestablish synchronization. You must run a **`start`** command with the **`-force`** option to acknowledge this condition, and the relationship or consistency group changes to `InconsistentCopying`. Enter this command only after all outstanding events are repaired.

In the unusual case where the master and the auxiliary are still synchronized (perhaps following a user stop, and no further write I/O was received), a **`start`** command takes the relationship to `ConsistentSynchronized`. No **`-force`** option is required. Also, in this case, you can run a **`switch`** command that moves the relationship or consistency group to `ConsistentSynchronized` and reverses the roles of the master and the auxiliary.

If the relationship or consistency group becomes disconnected, the auxiliary changes to `ConsistentDisconnected`. The master changes to `IdlingDisconnected`.

An informational status log is generated whenever a relationship or consistency group enters the `ConsistentStopped` state with a status of `Online`. You can configure this event to generate an SNMP trap that can be used to trigger automation or manual intervention to run a **`start`** command after a loss of synchronization.

ConsistentSynchronized

`ConsistentSynchronized` is a connected state. In this state, the master volume is accessible for read and write I/O, and the auxiliary volume is accessible for read-only I/O. Writes that are sent to the master volume are also sent to the auxiliary volume. Successful completion must be received for both writes, the write must be failed to the host, or a state must change out of the `ConsistentSynchronized` state before a write is completed to the host.

A **`stop`** command takes the relationship to the `ConsistentStopped` state. A **`stop`** command with the **`-access`** parameter takes the relationship to the `Idling` state.

A **`switch`** command leaves the relationship in the `ConsistentSynchronized` state, but it reverses the master and auxiliary roles (it switches the direction of replicating data). A **`start`** command is accepted, but has no effect.

If the relationship or consistency group becomes disconnected, the same changes are made as for `ConsistentStopped`.

Idling

`Idling` is a connected state. Both master and auxiliary volumes operate in the master role. Therefore, both master and auxiliary volumes are accessible for write I/O.

In this state, the relationship or consistency group accepts a **`start`** command. MM/GM maintains a record of regions on each disk that received write I/O while they were idling. This record is used to determine what areas must be copied following a **`start`** command.

The **`start`** command must specify the new copy direction. A **`start`** command can cause a loss of consistency if either volume in any relationship received write I/O, which is indicated by the `Synchronized` status. If the **`start`** command leads to loss of consistency, you must specify the **`-force`** parameter.

Following a **`start`** command, the relationship or consistency group changes to `ConsistentSynchronized` if there is no loss of consistency, or to `InconsistentCopying` if a loss of consistency occurs.

Also, the relationship or consistency group accepts a **`-clean`** option on the **`start`** command while in this state. If the relationship or consistency group becomes disconnected, both sides change their state to `IdlingDisconnected`.

IdlingDisconnected

`IdlingDisconnected` is a disconnected state. The target volumes in this half of the relationship or consistency group are all in the master role and accept read or write I/O.

The priority in this state is to recover the link to restore the relationship or consistency.

No configuration activity is possible (except for deletes or stops) until the relationship becomes connected again. At that point, the relationship changes to a connected state. The exact connected state that is entered depends on the state of the other half of the relationship or consistency group, which depends on the following factors:

- The state when it became disconnected
- The write activity since it was disconnected
- The configuration activity since it was disconnected

If both halves are `IdlingDisconnected`, the relationship becomes `Idling` when it is reconnected.

While IdlingDisconnected, if a write I/O is received that causes the loss of synchronization (synchronized attribute transitions from true to false) and the relationship was not already stopped (through a user stop or a persistent error), an event is raised to notify you of the condition. This same event also is raised when this condition occurs for the ConsistentSynchronized state.

InconsistentDisconnected

InconsistentDisconnected is a disconnected state. The target volumes in this half of the relationship or consistency group are all in the auxiliary role, and do not accept read *or* write I/O. Except for deletes, no configuration activity is permitted until the relationship becomes connected again.

When the relationship or consistency group becomes connected again, the relationship becomes InconsistentCopying automatically unless either of the following conditions are true:

- The relationship was InconsistentStopped when it became disconnected.
- The user issued a **stop** command while disconnected.

In either case, the relationship or consistency group becomes InconsistentStopped.

ConsistentDisconnected

ConsistentDisconnected is a disconnected state. The target volumes in this half of the relationship or consistency group are all in the auxiliary role, and accept read I/O but *not* write I/O.

This state is entered from ConsistentSynchronized or ConsistentStopped when the auxiliary side of a relationship becomes disconnected.

In this state, the relationship or consistency group displays an attribute of FreezeTime, which is the point when consistency was frozen. When it is entered from ConsistentStopped, it retains the time that it had in that state. When it is entered from ConsistentSynchronized, the FreezeTime shows the last time at which the relationship or consistency group was known to be consistent. This time corresponds to the time of the last successful heartbeat to the other system.

A **stop** command with the **-access** flag set to true transitions the relationship or consistency group to the IdlingDisconnected state. This state allows write I/O to be performed to the auxiliary volume and is used as part of a DR scenario.

When the relationship or consistency group becomes connected again, the relationship or consistency group becomes ConsistentSynchronized only if this action does not lead to a loss of consistency. The following conditions must be true:

- The relationship was ConsistentSynchronized when it became disconnected.
- No writes received successful completion at the master while disconnected.

Otherwise, the relationship becomes ConsistentStopped. The FreezeTime setting is retained.

Empty

This state applies only to consistency groups. It is the state of a consistency group that has no relationships and no other state information to show.

It is entered when a consistency group is first created. It is exited when the first relationship is added to the consistency group, at which point the state of the relationship becomes the state of the consistency group.

10.8.24 Remote Copy commands

This section describes commands that can be issued to create and operate RC services.

Remote Copy process

The MM/GM process includes the following steps:

1. A system partnership is created between two IBM Storage Virtualize systems (for intercluster MM/GM).
2. A MM/GM relationship is created between two volumes of the same size.
3. To manage multiple MM/GM relationships as one entity, the relationships can be made part of a MM/GM consistency group to ensure data consistency across multiple MM/GM relationships, or for ease of management.
4. The MM/GM relationship is started. When the background copy completes, the relationship is consistent and synchronized. When synchronized, the auxiliary volume holds a copy of the production data at the master that can be used for DR.
5. To access the auxiliary volume, the MM/GM relationship must be stopped with the access option enabled before write I/O is submitted to the auxiliary.

Following these steps, the remote host server is mapped to the auxiliary volume and the disk is available for I/O.

The command set for MM/GM contains the following broad groups:

- Create, delete, and manipulate relationships and consistency group
- Cause state changes

If a configuration command affects more than one system, MM/GM coordinates configuration activity between the systems. Specific configuration commands can be run only when the systems are connected, and fail with no effect when they are disconnected.

Other configuration commands are permitted, even if the systems are disconnected. The state is reconciled automatically by MM/GM when the systems become connected again.

For any command (with one exception), a single system receives the command from the administrator. This design is significant for defining the context for a CreateRelationship **mkrcrelationship** or CreateConsistencyGroup **mkrcconsistgrp** command. In this case, the system that is receiving the command is called the *local system*.

The exception is a command that sets systems into a MM/GM partnership. The **mkfcpartnership** and **mkippartnership** commands must be issued on both the local and remote systems.

The commands in this section are described as an abstract command set, and are implemented by using one of the following methods:

- CLI can be used for scripting and automation.
- GUI can be used for one-off tasks.

Listing available system partners

Run the **lspartnershipcandidate** command to list the systems that are available for setting up a two-system partnership. This command is a prerequisite for creating MM/GM relationships.

Changing the system parameters

When you want to change system parameters specific to any RC or GM only, use the **chsystem** command. The **chsystem** command features the following parameters for MM/GM:

- **-relationshipbandwidthlimit** *cluster_relationship_bandwidth_limit*

  This parameter controls the maximum rate at which any one RC relationship can synchronize. The default value for the relationship bandwidth limit is 25 MBps, but this value can now be specified as 1 - 100,000 MBps.

  The partnership overall limit is controlled at a partnership level by the **chpartnership -linkbandwidthmbits** command, and must be set on each involved system.

  Important: Do not set this value higher than the default without first establishing that the higher bandwidth can be sustained without affecting the host's performance. The limit must never be higher than the maximum that is supported by the infrastructure connecting the remote sites, regardless of the compression rates that you might achieve.

- **-gmlinktolerance** *link_tolerance*

  This parameter specifies the maximum period that the system tolerates delay before stopping GM relationships. Specify values of 60 - 86,400 seconds in increments of 10 seconds. The default value is 300. Do not change this value except under the direction of IBM Support.

- **-gmmaxhostdelay** *max_host_delay*

  This parameter specifies the maximum time delay, in milliseconds, at which the GM link tolerance timer starts counting down. This threshold value determines the extra effect that GM operations can add to the response times of the GM source volumes. You can use this parameter to increase the threshold from the default value of 5 milliseconds.

- **-maxreplicationdelay** *max_replication_delay*

  This parameter sets a maximum replication delay in seconds. The value must be a number 0 - 360 (0 being the default value, no delay). This feature sets the maximum number of seconds to be tolerated to complete a single I/O. If I/O cannot complete within the *max_replication_delay*, the `1920` event is reported. This setting is system-wide and applies to MM/GM relationships.

Run the **chsystem** command to adjust these values, as shown in the following example:

```
chsystem -gmlinktolerance 300
```

You can view all of these parameter values by running the **lssystem** *<system_name>* command.

Focus on the **gmlinktolerance** parameter in particular. If poor response extends past the specified tolerance, a `1920` event is logged, and one or more GM relationships automatically stop to protect the application hosts at the primary site. During normal operations, application hosts experience a minimal effect from the response times because the GM feature uses asynchronous replication.

However, if GM operations experience degraded response times from the secondary system for an extended period, I/O operations queue at the primary system. This queue results in an extended response time to application hosts. In this situation, the **gmlinktolerance** feature stops GM relationships, and the application host's response time returns to normal.

After a 1920 event occurs, the GM auxiliary volumes are no longer in the consistent_synchronized state. Fix the cause of the event and restart your GM relationships. For this reason, ensure that you monitor the system to track when these 1920 events occur.

You can disable the **gmlinktolerance** feature by setting the **gmlinktolerance** value to 0 (zero). However, the **gmlinktolerance** feature cannot protect applications from extended response times if it is disabled. It might be appropriate to disable the **gmlinktolerance** feature under the following circumstances:

- During SAN maintenance windows in which degraded performance is expected from SAN components, and application hosts can stand extended response times from GM volumes.
- During periods when application hosts can tolerate extended response times and it is expected that the **gmlinktolerance** feature might stop the GM relationships. For example, if you test by using an I/O generator that is configured to stress the back-end storage, the **gmlinktolerance** feature might detect the high latency and stop the GM relationships.

  Disabling the **gmlinktolerance** feature prevents this result at the risk of exposing the test host to extended response times.

A 1920 event indicates that one or more of the SAN components cannot provide the performance that is required by the application hosts. This situation can be temporary (for example, a result of a maintenance activity) or permanent (for example, a result of a hardware failure or an unexpected host I/O workload).

If 1920 events are occurring, you might need to use a performance monitoring and analysis tool, such as IBM Spectrum Control, to help identify and resolve the problem.

System partnership

To create a partnership, run one of the following commands, depending on the connection type:

- The **mkfcpartnership** command to establish a one-way MM/GM partnership between the local system and a remote system that are linked over an FC (or FCoE) connection.
- The **mkippartnership** command to establish a one-way MM/GM partnership between the local system and a remote system that are linked over a native IP connection.

To establish a fully functional MM/GM partnership, you must run either of these commands on both of the systems that are included of the partnership. This step is a prerequisite for creating MM or GM relationships between volumes on the IBM Storage Virtualize systems.

When creating the partnership, you must specify the Link Bandwidth and can specify the Background Copy Rate:

- The Link Bandwidth, which is expressed in Mbps, is the amount of bandwidth that can be used for the FC or IP connection between the systems within the partnership.
- The Background Copy Rate is the maximum percentage of the Link Bandwidth that can be used for background copy operations. The default value is 50%.

Background copy bandwidth effect on foreground I/O latency

The combination of the Link Bandwidth value and the Background Copy Rate percentage is referred to as the *Background Copy bandwidth*. It must be at least 8 Mbps. For example, if the Link Bandwidth is set to 10000 and the Background Copy Rate is set to 20, the resulting Background Copy bandwidth that is used for background operations is 2000 Mbps.

The background copy bandwidth determines the rate at which the background copy is attempted for MM/GM. The background copy bandwidth can affect foreground I/O latency in one of the following ways:

- The following results can occur if the background copy bandwidth is set too high compared to the MM/GM intercluster link capacity:
  - The background copy I/Os can back up on the MM/GM intercluster link.
  - There is a delay in the synchronous auxiliary writes of foreground I/Os.
  - The foreground I/O latency increases as perceived by applications.
- If the background copy bandwidth is set too high for the storage at the primary site, background copy read I/Os overload the primary storage and delay foreground I/Os.
- If the background copy bandwidth is set too high for the storage at the secondary site, background copy writes at the secondary site overload the auxiliary storage, and again delay the synchronous secondary writes of foreground I/Os.

To set the background copy bandwidth optimally, ensure that you consider all three resources: Primary storage, intercluster link bandwidth, and auxiliary storage. Provision the most restrictive of these three resources between the background copy bandwidth and the peak foreground I/O workload.

Perform this provisioning by calculation or by determining experimentally how much background copy can be allowed before the foreground I/O latency becomes unacceptable. Then reduce the background copy to accommodate peaks in workload.

The chpartnership command

To change the bandwidth that is available for background copy in the system partnership, run the **`chpartnership -backgroundcopyrate <percentage_of_link_bandwidth>`** command to specify the percentage of whole link capacity to be used by the background copy process.

Creating a Metro Mirror or Global Mirror consistency group

Run the **`mkrcconsistgrp`** command to create an empty MM/GM consistency group.

The MM or GM consistency group name must be unique across all consistency groups that are known to the systems owning this consistency group. If the consistency group involves two systems, the systems must be in communication throughout the creation process.

The new consistency group does not contain any relationships and is in the `Empty` state. You can add MM/GM relationships to the group (upon creation or afterward) by running the **`chrelationship`** command.

Creating a Metro Mirror or Global Mirror relationship

Run the **`mkrcrelationship`** command to create a MM/GM relationship. This relationship persists until it is deleted.

Optional parameter: If you do not use the **`-global`** optional parameter, an MM relationship is created rather than a GM relationship.

The auxiliary volume must be equal in size to the master volume or the command fails. If both volumes are in the same system, they must be in the same I/O group. The master and auxiliary volume cannot be in a relationship, and they cannot be the target of a FlashCopy mapping. This command returns the new relationship (`relationship_id`) when successful.

When the MM/GM relationship is created, you can add it to a consistency group, or it can be a stand-alone MM/GM relationship.

The lsrcrelationshipcandidate command

Run the **lsrcrelationshipcandidate** command to list the volumes that are eligible to form an MM/GM relationship.

When the command is issued, you can specify the master volume name and auxiliary system to list the candidates that comply with the prerequisites to create a MM/GM relationship. If the command is issued with no parameters, all of the volumes that are not disallowed by another configuration state, such as being a FlashCopy target, are listed.

Changing Metro Mirror/Global Mirror relationship

Run the **chrcrelationship** command to modify the following properties of an MM/GM relationship:

- Change the name of an MM/GM relationship.
- Add a relationship to a group.
- Remove a relationship from a group by using the **-force** flag.

Adding an MM/GM relationship: When an MM/GM relationship is added to a consistency group that is not empty, the relationship must have the same state and copy direction as the group to be added to it.

Changing Metro Mirror/Global Mirror consistency group

Run the **chrcconsistgrp** command to change the name of an MM/GM consistency group.

Starting Metro Mirror/Global Mirror relationship

Run the **startrcrelationship** command to start the copy process of an MM/GM relationship.

When the command is run, you can set the copy direction if it is undefined. Optionally, you can mark the auxiliary volume of the relationship as clean. The command fails if it is used as an attempt to start a relationship that is a part of a consistency group.

You can run this command only to a relationship that is connected. For a relationship that is idling, this command assigns a copy direction (master and auxiliary roles) and begins the copy process. Otherwise, this command restarts a previous copy process that was stopped by a **stop** command or by an I/O error.

If the resumption of the copy process leads to a period when the relationship is inconsistent, you must specify the **-force** parameter when the relationship is restarted. This situation can arise if, for example, the relationship was stopped and then further writes were performed on the original master of the relationship.

The use of the **-force** parameter here is a reminder that the data on the auxiliary becomes inconsistent while resynchronization (background copying) occurs. Therefore, this data is unusable for DR purposes before the background copy completes.

In the `Idling` state, you must specify the master volume to indicate the copy direction. In other connected states, you can provide the **-primary** argument, but it must match the existing setting.

Stopping Metro Mirror/Global Mirror relationship

Run the **stoprcrelationship** command to stop the copy process for a relationship. You can also use this command to enable write access to a consistent auxiliary volume by specifying the **-access** parameter.

This command applies to a stand-alone relationship. It is rejected if it is addressed to a relationship that is part of a consistency group. You can issue this command to stop a relationship that is copying from master to auxiliary.

If the relationship is in an inconsistent state, any copy operation stops and does not resume until you run a **startrcrelationship** command. Write activity is no longer copied from the master to the auxiliary volume. For a relationship in the `ConsistentSynchronized` state, this command causes a Consistency Freeze.

When a relationship is in a consistent state (that is, in the `ConsistentStopped`, `ConsistentSynchronized`, or `ConsistentDisconnected` state), you can use the **-access** parameter with the **stoprcrelationship** command to enable write access to the auxiliary volume.

Starting Metro Mirror/Global Mirror consistency group

Run the **startrcconsistgrp** command to start an MM/GM consistency group. You can issue this command only to a consistency group that is connected.

For a consistency group that is idling, this command assigns a copy direction (master and auxiliary roles) and begins the copy process. Otherwise, this command restarts a previous copy process that was stopped by a **stop** command or by an I/O error.

Stopping Metro Mirror/Global Mirror consistency group

Run the **stoprcconsistgrp** command to stop the copy process for an MM/GM consistency group. You can also use this command to enable write access to the auxiliary volumes in the group if the group is in a consistent state.

If the consistency group is in an inconsistent state, any copy operation stops and does not resume until you run the **startrcconsistgrp** command. Write activity is no longer copied from the master to the auxiliary volumes that belong to the relationships in the group. For a consistency group in the `ConsistentSynchronized` state, this command causes a Consistency Freeze.

When a consistency group is in a consistent state (for example, in the `ConsistentStopped`, `ConsistentSynchronized`, or `ConsistentDisconnected` state), you can use the **-access** parameter with the **stoprcconsistgrp** command to enable write access to the auxiliary volumes within that group.

Deleting Metro Mirror/Global Mirror relationship

Run the **rmrcrelationship** command to delete the relationship that is specified. Deleting a relationship deletes only the logical relationship between the two volumes. It does not affect the volumes.

If the relationship is disconnected at the time that the command is issued, the relationship is deleted on only the system on which the command is being run. When the systems reconnect, the relationship is automatically deleted on the other system.

Alternatively, if the systems are disconnected and you still want to remove the relationship on both systems, you can run the **rmrcrelationship** command independently on both of the systems.

A relationship cannot be deleted if it is part of a consistency group. You must first remove the relationship from the consistency group by issuing the **chrcrelationship -noconsistgrp** command.

If you delete an inconsistent relationship, the auxiliary volume becomes accessible, even though it is still inconsistent. This situation is the one case in which MM/GM does not inhibit access to inconsistent data.

Deleting Metro Mirror/Global Mirror consistency group

Run the **rmrcconsistgrp** command to delete an MM/GM consistency group. This command deletes the specified consistency group.

If the consistency group is disconnected at the time that the command is issued, the consistency group is deleted on only the system on which the command is being run. When the systems reconnect, the consistency group is automatically deleted on the other system.

Alternatively, if the systems are disconnected and you still want to remove the consistency group on both systems, you can run the **rmrcconsistgrp** command separately on both of the systems.

If the consistency group is not empty, the relationships within it are removed from the consistency group before the group is deleted. These relationships then become stand-alone relationships. The state of these relationships is not changed by the action of removing them from the consistency group.

Reversing Metro Mirror/Global Mirror relationship

Run the **switchrcrelationship** command to reverse the roles of the master volume and the auxiliary volume when a stand-alone relationship is in a consistent state. When the command is issued, the wanted master must be specified.

Reversing Metro Mirror/Global Mirror consistency group

Run the **switchrcconsistgrp** command to reverse the roles of the master volume and the auxiliary volume when a consistency group is in a consistent state. This change is applied to all of the relationships in the consistency group. When the command is issued, the wanted master must be specified.

Important: By reversing the roles, your current source volumes become targets, and target volumes become source. Therefore, you lose write access to your current primary volumes.

10.9 Native IP replication

IBM Storage Virtualize can implement RC services by using FC connections or IP connections. This section describes the IBM Storage Virtualize IP replication technology and implementation.

10.9.1 Enhancements

IBM Storage Virtualize 8.4.2 included the following enhancements that are related to IP replication:

- Support for multiple IPv4 or IPv6 addresses per port
- VLAN separation for individual IP addresses
- A new PortSet configuration model for IP Connectivity
- A new CLI model for Ethernet network configuration

In addition, IBM Storage Virtualize 8.5.0 and later includes Domain Name Services (DNS) for IP Replication, which allows you to enter a DNS hostname and an IP address of a partner system.

Some of the new features are described in this section. For more information about the new features, see 1.3, “Changes and enhancements in version 8.6.0” on page 13.

Portsets

A portsets group is a set of IP addresses that can be used for iSCSI or iSER host attach, IP replication, or iSCSI storage virtualization. A host can access storage through any of the IP addresses in the portset that is mapped to the host. Multiple hosts can be mapped to a portset; however, multiple portsets cannot be mapped to the same host.

Figure 10-96 shows the new admin model for creating port sets and configuring IP replication.

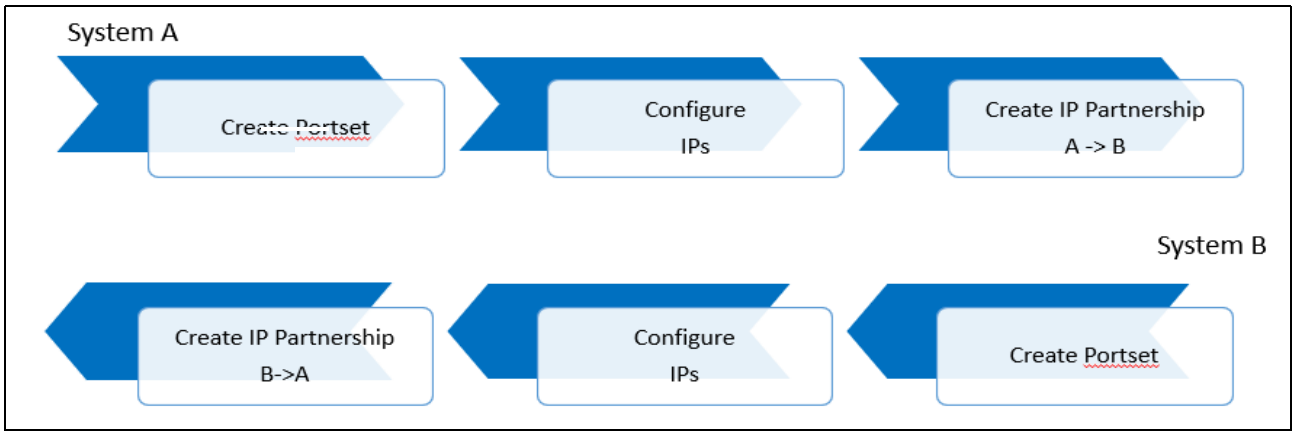


Figure 10-96 New admin model

IBM Storage Virtualize 8.4.2 and later allows a maximum of three IP Replication partnerships per system. The previous limit was one in code versions earlier than 8.4.2. Use cases include 3-site replication and having DR at a remote site or in the cloud. Example 10-4 shows the new commands.

Example 10-4 New commands

```
New Portset CLIs
====================
   # mkportset -name myportset -type replication
    # lsportset
    # rmportset myportset

New IP Configuration CLIs
====================
   # mkip -node node1 -port 1 -portset 4 -ip 10.0.1.1 -prefix 24 (replace
cfgportip)
   # lsip (replace lsportip)
    # rmip 0 (replace rmportip )

New Ethernet Port Configuration CLIs
====================
   # chportethernet -iogrp 0 -mtu 9000 2 (replace cfgportip)
   # lsportethernet (replace lsportip)
```

```
Modified CLIs with portset objects
====================
    # mkhost -iscsiname iqn.localhost.hostid.7f000001 -name hostone -portset
myportset
    # chhost -portset myportset myhost
```

10.9.2 Native IP replication technology

Remote Mirroring over IP communication is supported on the IBM SAN Volume Controller and IBM FlashSystem by using Ethernet communication links. The IBM Storage Virtualize Software IP replication uses innovative Bridgeworks SANSlide technology to optimize network bandwidth and utilization. This function enables the use of a lower-speed and lower-cost networking infrastructure for data replication.

Bridgeworks SANSlide technology, which is integrated into the IBM Storage Virtualize Software, uses artificial intelligence (AI) to help optimize network bandwidth use and adapt to changing workload and network conditions.

This technology can improve remote mirroring network bandwidth usage up to three times. Improved bandwidth usage can enable clients to deploy a less costly network infrastructure, or speed up remote replication cycles to enhance DR effectiveness.

With an Ethernet network data flow, the data transfer can slow down over time. This condition occurs because of the latency that is caused by waiting for the acknowledgment of each set of packets that is sent. The next packet set cannot be sent until the previous packet is acknowledged, as shown in Figure 10-97.

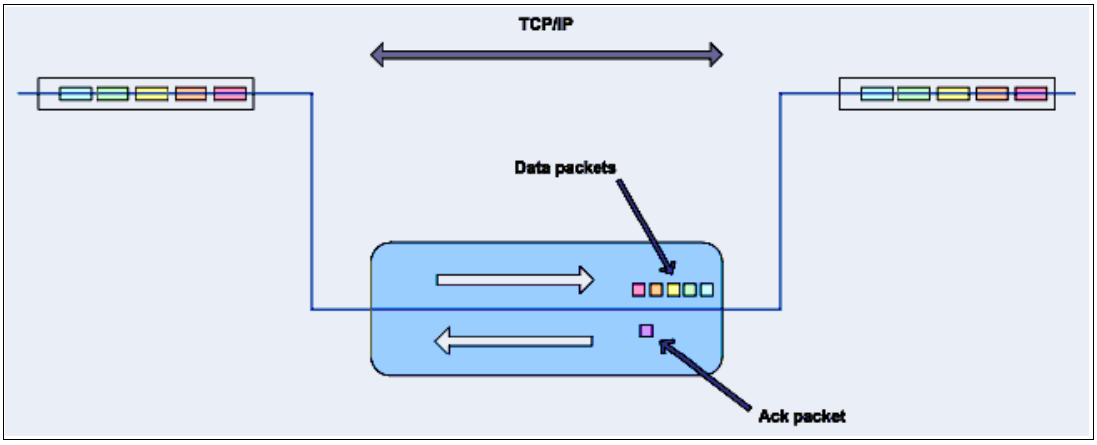


Figure 10-97 Typical Ethernet network data flow

However, by using the embedded IP replication, this behavior can be eliminated with the enhanced parallelism of the data flow by using multiple virtual connections (VC) that share IP links and addresses. The AI engine can dynamically adjust the number of VCs, receive window size, and packet size to maintain optimum performance. While the engine is waiting for one VC's ACK, it sends more packets across other VCs. If packets are lost from any VC, data is automatically retransmitted, as shown in Figure 10-98.

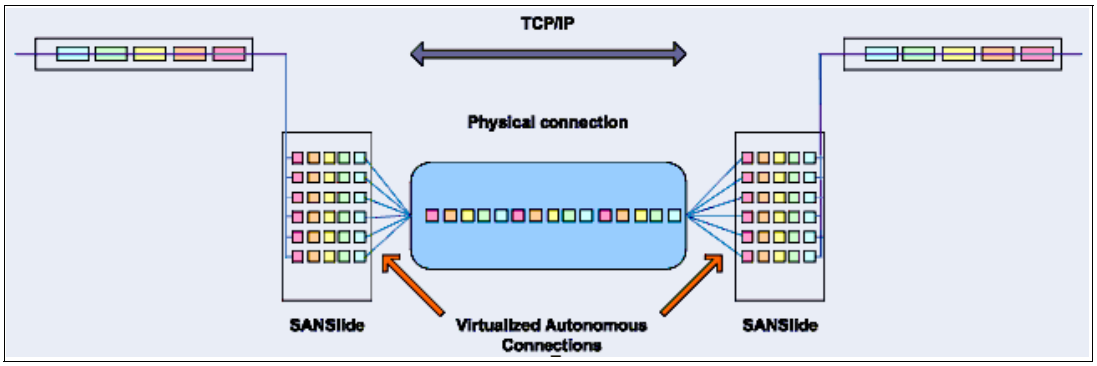


Figure 10-98 Optimized network data flow by using Bridgeworks SANSlide technology

For more information about this technology, see *IBM Storwize V7000 and SANSlide Implementation*, REDP-5023.

With native IP partnership, the following Copy Services features are supported:

- MM

  Referred to as synchronous replication, MM provides a consistent copy of a source volume on a target volume. Data is written to the target volume synchronously after it is written to the source volume so that the copy is continuously updated.

- GM and GMCV

  Referred to as asynchronous replication, GM provides a consistent copy of a source volume on a target volume. Data is written to the target volume asynchronously so that the copy is continuously updated. However, the copy might not contain the last few updates if a DR operation is performed. An added extension to GM is GMCV. GMCV is the preferred method for use with native IP replication.

  Note: For IP partnerships, generally use the GMCV method of copying (asynchronous copy of changed grains only). This method can include performance benefits. Also, GM and MM might be more susceptible to the loss of synchronization.

10.9.3 IP partnership limitations

The following prerequisites and assumptions must be considered before IP partnership between two IBM Storage Virtualize systems can be established:

- The IBM Storage Virtualize systems are successfully installed with V7.2 or later code levels.
- The systems must have the necessary licenses that enable RC partnerships to be configured between two systems. No separate license is required to enable IP partnership.
- The storage SANs are configured correctly and the correct infrastructure to support the IBM Storage Virtualize systems in RC partnerships over IP links is in place.

- The two systems must be able to ping each other and perform the discovery.
- TCP ports 3260 and 3265 are used by systems for IP partnership communications. Therefore, these ports must be open.
- The maximum number of partnerships between the local and remote systems, including both IP and FC partnerships, is limited to the current maximum that is supported, which is three partnerships.
- A total of four partnerships over IP is supported.
- A system can have simultaneous partnerships over FC and IP, but with separate systems. The FC zones between two systems must be removed before an IP partnership is configured.
- IP partnerships are supported on both 10 Gbps links and 1 Gbps links. However, the intermix of both on a single link is not supported.
- The maximum supported round-trip time (RTT) is 80 ms for 1 Gbps links.
- The maximum supported RTT is 10 ms for 10 Gbps links.
- The inter-cluster heartbeat traffic uses 1 Mbps per link.
- Only nodes from two I/O groups can have ports that are configured for an IP partnership.
- Migrations of RC relationships directly from FC-based partnerships to IP partnerships are not supported.
- IP partnerships between the two systems can be over IPv4 or IPv6 only, but not both.
- Virtual local area network (VLAN) tagging of the IP addresses that are configured for RC is supported starting with V7.4.
- Management IP and internet Small Computer Systems Interface (iSCSI) IP on the same port can be in a different network starting with V7.4.
- An added layer of security is provided by using Challenge Handshake Authentication Protocol (CHAP) authentication.
- TCP ports 3260 and 3265 are used for IP partnership communications. Therefore, these ports must be open in firewalls between the systems.
- Only a single RC data session per physical link can be established. It is intended that only one connection (for sending/receiving RC data) is made for each independent physical link between the systems.

  Note: A physical link is the physical IP link between the two sites: A (local) and B (remote). Multiple IP addresses on local system A might be connected (by Ethernet switches) to this physical link. Similarly, multiple IP addresses on remote system B might be connected (by Ethernet switches) to the same physical link. At any time, only a single IP address on cluster A can form an RC data session with an IP address on cluster B.

- The maximum throughput is restricted based on the use of 1 Gbps, 10 Gbps, or 25 Gbps Ethernet ports. It varies based on distance (for example, round-trip latency) and quality of communication link (for example, packet loss):
  - One 1 Gbps port can transfer up to 110 MBps unidirectional, 190 MBps bidirectional
  - Two 1 Gbps ports can transfer up to 220 MBps unidirectional, 325 MBps bidirectional
  - One 10 Gbps port can transfer up to 240 MBps unidirectional, 350 MBps bidirectional
  - Two 10 Gbps port can transfer up to 440 MBps unidirectional, 600 MBps bidirectional

Note: IP Replication is supported by 25 Gbps Melanox and Chelsio adapters, but be aware there is no performance benefit or advantage for IP Replication with these adapters. However, for the purpose of consolidation where these cards are used for other purposes, such as iSCSI Extensions for RDMA (iSER) Host Attach or iSCSI Host Attach/Backend Virtualization, they can be used for IP replication.

The minimum supported link bandwidth is 10 Mbps. However, this requirement scales up with the amount of host I/O that you choose to do. Figure 10-99 shows scaling host I/O.

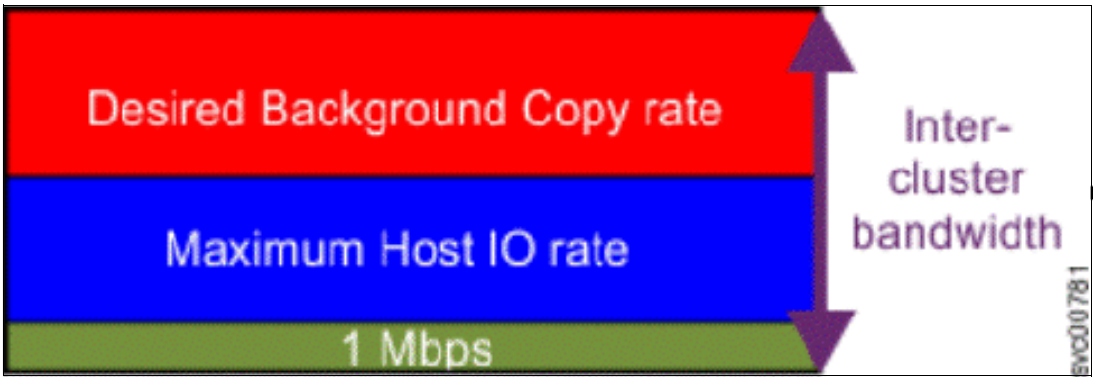


Figure 10-99 Scaling of host I/O

The following equation describes the approximate minimum bandwidth that is required between two systems with < 5 ms RTT and errorless link:

```
Minimum intersite link bandwidth in Mbps > Required Background Copy in Mbps +
Maximum Host I/O in Mbps + 1 Mbps heartbeat traffic
```

Increasing latency and errors results in a higher requirement for minimum bandwidth.

Note: The Bandwidth setting definition when the IP partnerships are created changed in V7.7. Previously, the bandwidth setting defaulted to 50 MiB, and was the maximum transfer rate from the primary site to the secondary site for initial sync/resyncs of volumes.

The Link Bandwidth setting is now configured by using megabits (Mb) not MB. You set the Link Bandwidth setting to a value that the communication link can sustain, or to what is allocated for replication. The Background Copy Rate setting is now a percentage of the Link Bandwidth. The Background Copy Rate setting determines the available bandwidth for the initial sync and resyncs or for GMCV.

10.9.4 IP Partnership and data compression

When creating an IP partnership between two systems, you can specify whether you want to use the data compression feature. When enabled, IP partnership compression compresses the data that is sent from a local system to the remote system and potentially uses less bandwidth than with uncompressed data. It is also used to decompress data that is received by a local system from a remote system.

Data compression is supported for IPv4 or IPv6 partnerships. To enable data compression, both systems in an IP partnership must be running a software level that supports IP partnership compression (V7.7 or later) and both must have the compression feature enabled.

No compression license is needed on any local or remote system.

Volumes that are replicated by using IP partnership compression can be compressed or uncompressed on the system. Volume compression and IP Replication compression are *not* linked features.

For example, the following actions occur to replicate a compressed volume over an IP partnership with the compression feature enabled:

- Read operations in the local system decompress the data when reading from the source volume.
- Uncompressed data is transferred to the Remote Copy code.
- Data is compressed before being sent over the IP partnership link.
- Remote system Remote Copy code decompresses the received data.
- Write operations in the remote system compress the data when writing to the target volume.

10.9.5 VLAN support

Starting with IBM Storage Virtualize V7.4, VLAN tagging is supported for iSCSI host attachment and IP replication. Hosts and remote-copy operations can connect to the system through Ethernet ports. Each traffic type has different bandwidth requirements, which can interfere with each other if they share IP connections. VLAN tagging creates two separate connections on the same IP network for different types of traffic. The system supports VLAN configuration on both IPv4 and IPv6 connections.

When the VLAN ID is configured for IP addresses that is used for iSCSI host attach or IP replication, the VLAN settings on the Ethernet network and servers must be configured correctly to avoid connectivity issues. After the VLANs are configured, changes to the VLAN settings disrupt iSCSI and IP replication traffic to and from the partnerships.

During the VLAN configuration for each IP address, the VLAN settings for the local and failover ports on two nodes of an I/O group can differ. To avoid any service disruption, switches must be configured so that the failover VLANs are configured on the local switch ports and the failover of IP addresses from a failing node to a surviving node succeeds. If failover VLANs are not configured on the local switch ports, no paths are available to the IBM Storage Virtualize system nodes during a node failure and the replication fails.

Consider the following requirements and procedures when implementing VLAN tagging:

- VLAN tagging is supported for IP partnership traffic between two systems.
- VLAN provides network traffic separation at the layer 2 level for Ethernet transport.
- VLAN tagging by default is disabled for any IP address of a node port. You can use the CLI or GUI to optionally set the VLAN ID for port IPs on both systems in the IP partnership.
- When a VLAN ID is configured for the port IP addresses that are used in RC port groups, appropriate VLAN settings on the Ethernet network must also be configured to prevent connectivity issues.

Setting VLAN tags for a port is disruptive. Therefore, VLAN tagging requires that you stop the partnership first before you configure VLAN tags. Restart the partnership after the configuration is complete.

10.9.6 IP partnership and terminology

The IP partnership terminology and abbreviations that are used are listed in Table 10-15.

Table 10-15 Terminology for IP partnership

IP partnership terminology	Description
RC Portset	Only IP addresses that are part of portsets 1 or 2 can participate and form RC connections with a partner system: ► 0: Ports that are not configured for RC ► 1: Ports that belong to RC portset 1 ► 2: Ports that belong to RC portset 2 Each IP address can be shared for iSCSI host attach and RC functions. Therefore, suitable settings must be applied to each portset.
IP partnership	Two systems that are partnered to perform RC over native IP links.
FC partnership	Two systems that are partnered to perform RC over native FC links.
Failover	Failure of a node within an I/O group causes the volume access to go through the surviving node. The IP addresses fail over to the surviving node in the I/O group. When the configuration node of the system fails, management IPs also fail over to an alternative node.
Failback	When the failed node rejoins the system, all failed over IP addresses are failed back from the surviving node to the rejoined node, and volume access is restored through this node.
linkbandwidthmbits	Aggregate bandwidth of all physical links between two sites in Mbps.
IP partnership or partnership over native IP links	These terms are used to describe the IP partnership feature.
Discovery	Process by which two IBM Storage Virtualize systems exchange information about their IP address configuration. For IP-based partnerships, only IP addresses configured for RC are discovered. For example, the first Discovery takes place when the user is running the **mkippartnership** CLI command. Subsequent Discoveries can take place as a result of user activities (configuration changes) or as a result of hardware failures (for example, node failure, ports failure, and so on).

10.9.7 States of IP partnership

The different partnership states in IP partnership are listed in Table 10-16.

Table 10-16 States of IP partnership

State	Systems connected	Support for active RC I/O	Comments
Partially_Configured_Local	No	No	This state indicates that the initial discovery is complete.
Fully_Configured	Yes	Yes	Discovery successfully completed between two systems, and the two systems can establish RC relationships.
Fully_Configured_Stopped	Yes	Yes	The partnership is stopped on the system.
Fully_Configured_Remote_Stopped	Yes	No	The partnership is stopped on the remote system.
Not_Present	Yes	No	The two systems cannot communicate with each other. This state is also seen when data paths between the two systems are not established.
Fully_Configured_Exceeded	Yes	No	There are too many systems in the network, and the partnership from the local system to the remote system is disabled.
Fully_Configured_Excluded	No	No	The connection is excluded because of too many problems, or either system cannot support the I/O work load for the MM and GM relationships.

The process to establish two systems in the IP partnerships includes the following steps:

1. The administrator configures the CHAP secret on both the systems. This step is not mandatory, and users can choose to not configure the CHAP secret.
2. The administrator configures the system IP addresses on both local and remote systems so that they can discover each other over the network.
3. If you want to use VLANs, configure your local area network (LAN) switches and Ethernet ports to use VLAN tagging.
4. The administrator configures the systems ports on each node in both of the systems by using the GUI (or the **mkip** CLI command), and completes the following steps:
   a. Configures the IP addresses for RC data.
   b. Adds the IP addresses in the respective RC port group.
   c. Defines whether the host access on these ports over iSCSI is allowed.
5. The administrator establishes the partnership with the remote system from the local system where the partnership state then changes to `Partially_Configured_Local`.
6. The administrator establishes the partnership from the remote system with the local system. If this process is successful, the partnership state then changes to the `Fully_Configured`, which implies that the partnerships over the IP network were successfully established. The partnership state momentarily remains `Not_Present` before moving to the `Fully_Configured` state.
7. The administrator creates MM, GM, and GMCV relationships.

Partnership consideration: When the partnership is created, no master or auxiliary status is defined or implied. The partnership is equal. The concepts of *master or auxiliary* and *primary or secondary* apply to volume relationships only, not to system partnerships.

10.9.8 Remote Copy portsets

Portsets replace the requirement for creating remote-copy groups for IP partnerships. Dedicated portsets can be created for remote copy traffic. The dedicated portsets provide a group of IP addresses for IP Partnerships.

Each node can have one IP address that is assigned to a portset for remote-copy traffic. If the local system in the IP partnership contains four nodes, a portset can be created that defines four IP addresses, one per node.

During updates of the software, any IP addresses that are assigned to remote copy groups with an IP partnership are automatically moved to a corresponding portset. For example, if remote-copy group 1 is defined on the system before the update then IP addresses from that remote-copy group are mapped to portset 1 after the update. Similarly, the IP address in remote-copy group 2 is mapped to portset 2. Before you can configure a new IP partnership, you must define a portset and assign IP addresses to nodes.

You can configure portsets so that each IP partnership can be mapped to two portsets, one for each WAN link between systems. For network configurations that have a single link between systems in an IP partnership, a single portset can be defined in the Portset Link 1 field on the Create Partnership page from GUI. You can also use the **-link1** attribute in the mkpartnership command for partnerships with a single link. For a partnership with dual links, a second portset must be mapped defined in the Portset Link 2 field. Use the **-link2** attribute to specify the second portset for a dual link configuration.

Remember: IP ports on both partners must be configured with identical RC portset IDs for the partnership to be established correctly.

The IBM Storage Virtualize system IP addresses for replication should be designated with identical RC portset. The system supports two RC portsets: 1 and 2.

Replication portsets 1 and 2 (as well as 0 and 3) are always globally owned and only global administrators can assign and modify IP addresses to these portsets.

You can assign one IPv4 address and one IPv6 address to each Ethernet port on the system platforms. Each of these IP addresses can be shared between iSCSI host attach and the IP partnership. The user must configure the required IP address (IPv4 or IPv6) on an Ethernet port with an RC portset.

The administrator might want to use IPv6 addresses for RC operations and use IPv4 addresses on that same port for iSCSI host attach. This configuration also implies that for two systems to establish an IP partnership, both systems must have IPv6 addresses that are configured.

Administrators can choose to dedicate an Ethernet port for IP partnership only. In that case, host access must be specifically disabled for that IP address and any other IP address that is configured on that Ethernet port.

Note: To establish an IP partnership, each IBM Storage Virtualize controller node must have only a single RC portset configured, portset 1 or portset 2.

10.9.9 Supported configurations

Note: For explanation purposes, this section shows a node with two ports available: 1 and 2. This number generally increments when the latest models of IBM Storage Virtualize systems are used.

The following supported configurations for IP partnership that were in the first release are described in this section:

- Two 2-node systems in IP partnership over a single inter-site link, as shown in Figure 10-100 (configuration 1).

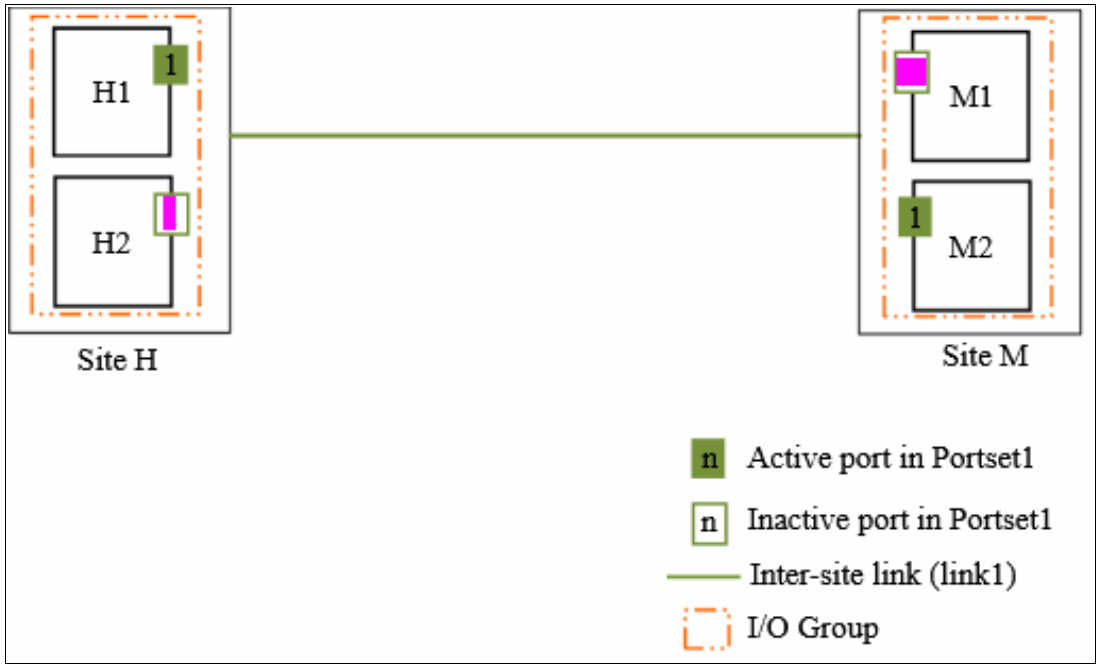


Figure 10-100 Single link with only one Remote Copy portset configured in each system

As shown in Figure 10-100, two systems are available:

- System H
- System M

A single RC portset1 is created on Node H1 on System H and on Node M2 on System M because only a single inter-site link is used to facilitate the IP partnership traffic. An administrator might choose to configure the RC portset on Node M1 on System M rather than Node M2.

At any time, only the IP addresses that are configured in RC portset 1 on the nodes in System H and System M participate in establishing data paths between the two systems after the IP partnerships are created. In this configuration, no failover ports are configured on the partner node in the same I/O group.

This configuration has the following characteristics:

- Only one node in each system has an RC portset that is configured, and no failover ports are configured.
- If the Node H1 in System H or the Node M2 in System M encounter a failure, the IP partnership stops and enters the `Not_Present` state until the failed nodes recover.
- After the nodes recover, the IP ports fail back, the IP partnership recovers, and the partnership state goes to the `Fully_Configured` state.
- If the inter-site system link fails, the IP partnerships change to the `Not_Present` state.
- This configuration is not recommended because it is not resilient to node failures.

- Two 2-node systems in IP partnership over a single inter-site link (with failover ports are configured), as shown in Figure 10-101 (configuration 2).

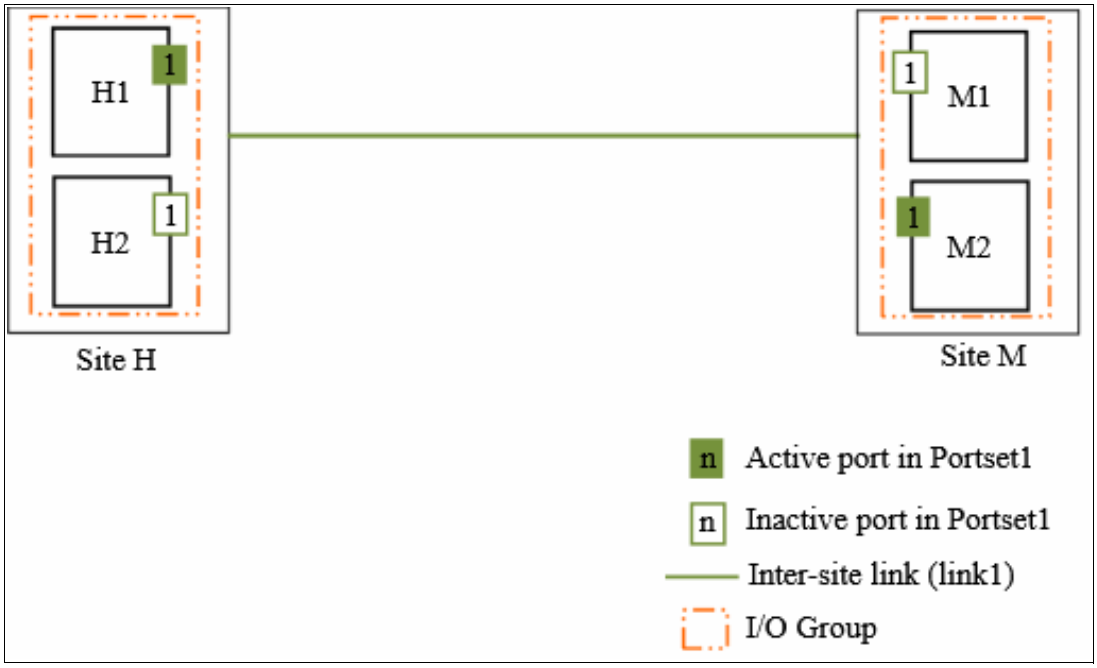


Figure 10-101 One Remote Copy portset on each system and nodes with failover ports configured

As shown in Figure 10-101, two systems are available:

- System H
- System M

A single RC portset 1 is configured on two Ethernet ports, one each on Node H1 and Node H2 on System H. Similarly, a single RC portset is configured on two Ethernet ports on Node M1 and Node M2 on System M.

Although two ports on each system are configured for RC portset 1, only one Ethernet port in each system actively participates in the IP partnership process. This selection is determined by a path configuration algorithm that is designed to choose data paths between the two systems to optimize performance.

The other port on the partner node in the I/O group behaves as a standby port that is used if a node fails. If Node H1 fails in System H, IP partnership continues servicing replication I/O from Ethernet Port 2 because a failover port is configured on Node H2 on Ethernet Port 2.

However, it might take some time for discovery and path configuration logic to reestablish paths post failover. This delay can cause partnerships to change to `Not_Present` for that time. The details of the particular IP port that is actively participating in IP partnership is provided in the **`lsip`** output.

This configuration has the following characteristics:

- Each node in the I/O group has the same RC portset that is configured. However, only one port in that RC portset is active at any time at each system.
- If the Node H1 in System H or the Node M2 in System M fails in the respective systems, IP partnerships rediscovery is triggered and continues servicing the I/O from the failover port.
- The discovery mechanism that is triggered because of failover might introduce a delay where the partnerships momentarily change to the `Not_Present` state and recover.

- Two 4-node systems in IP partnership over a single inter-site link (with failover ports configured), as shown in Figure 10-102 (configuration 3).

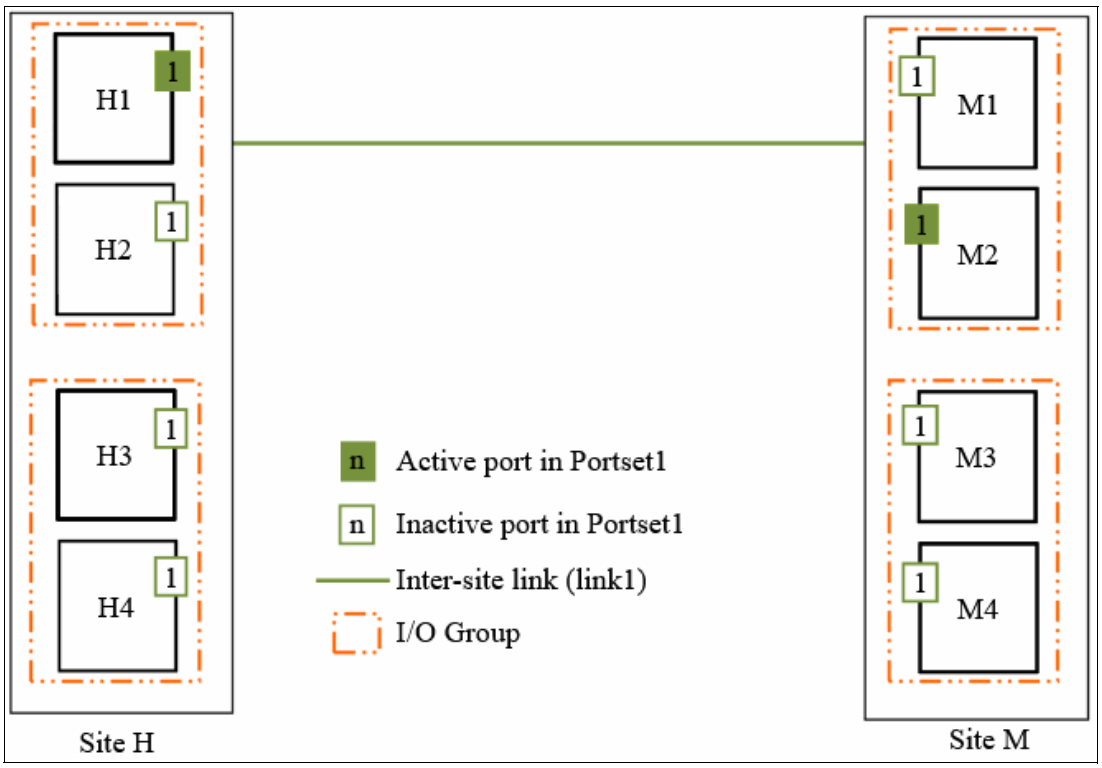


Figure 10-102 Multinode systems single inter-site link with only one RC portset

As shown in Figure 10-102, two 4-node systems are available:

- System H
- System M

A single RC portset 1 is configured on nodes H1, H2, H3, and H4 on System H, Site H; and on nodes M1, M2, M3, and M4 on System M, Site M. Although four ports are configured for RC portset 1, only one Ethernet port in each RC portset on each system actively participates in the IP partnership process.

Port selection is determined by a path configuration algorithm. The other ports play the role of standby ports.

If Node H1 fails in System H, the IP partnership selects one of the remaining ports that is configured with RC portset 1 from any of the nodes from either of the two I/O groups in System H. However, it might take some time (generally seconds) for discovery and path configuration logic to reestablish paths post failover. This process can cause partnerships to change to the `Not_Present` state.

This result causes RC relationships to stop. The administrator might need to manually verify the issues in the event log and start the relationships or RC consistency groups, if they do not autorecover. The details of the particular IP port actively participating in the IP partnership process is provided in the **`lsip`** view.

This configuration has the following characteristics:

- Each node has the RC portset that is configured in both I/O groups. However, only one port in that RC portset remains active and participates in IP partnership on each system.
- If the Node H1 in System H or the Node M2 in System M were to encounter some failure in the system, IP partnerships discovery is triggered and it continues servicing the I/O from the failover port.
- The discovery mechanism that is triggered because of failover might introduce a delay wherein the partnerships momentarily change to the `Not_Present` state and then recover.
- The bandwidth of the single link is used completely.

- Six-node system in IP partnership with six-node system over single inter-site link, as shown in Figure 10-103 on page 855 (configuration 4).

As shown in Figure 10-103 on page 855, a six-node system (System H in Site H1) and a six-node system (System M in Site 2) are used. A single RC portset 1 is configured on nodes HA1, H2, H3, and H4 on System H at Site 1. Similarly, a single RC portset 1 is configured on nodes M1, M2, M3, and M4 on System M.

Although three I/O groups (six nodes) are in System H, any two I/O groups at maximum are supported to be configured for IP partnerships. If Node H1 fails in System H, the IP partnership continues by using one of the ports that is configured in RC portset from any of the nodes from other I/O group in System H.

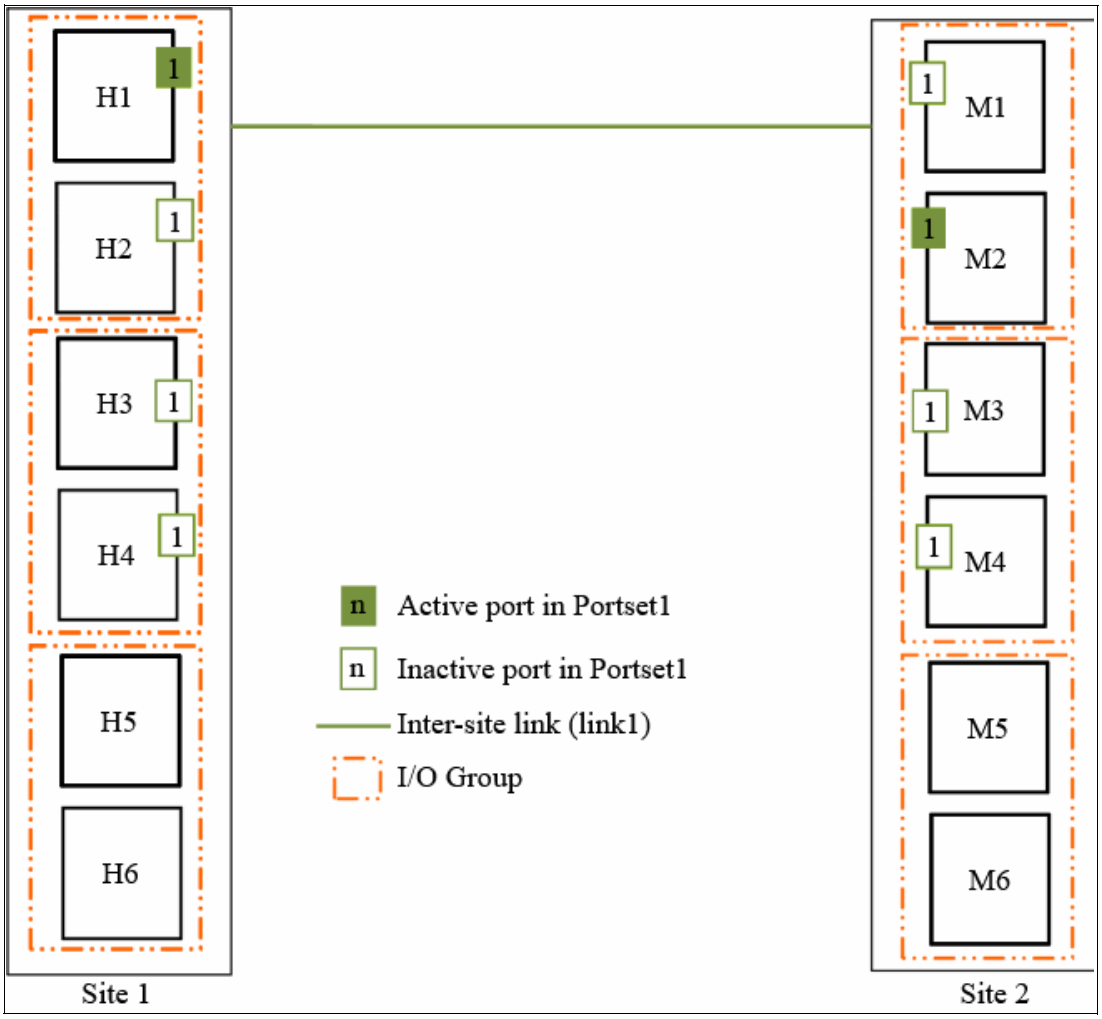


Figure 10-103 Multinode systems single inter-site link with only one Remote Copy portset

However, it might take some time for discovery and path configuration logic to reestablish paths post-failover. This delay might cause partnerships to change to the `Not_Present` state.

This process can lead to RC relationships stopping, and the administrator must manually start them if the relationships do not auto-recover. The details of which specific IP port is actively participating in the IP partnership process is provided in **`lsip`** output.

This configuration features the following characteristics:

- Each node has the RC portset that is configured in both the I/O groups that are identified for participating in IP Replication. However, only one port in that RC port group remains active on each system and participates in IP Replication.
- If the Node H1 in System H or the Node M2 in System M fails in the system, the IP partnerships trigger discovery and continue servicing the I/O from the failover ports.
- The discovery mechanism that is triggered because of failover might introduce a delay wherein the partnerships momentarily change to the `Not_Present` state and then recover.
- The bandwidth of the single link is used completely.
- Two 2-node systems with two inter-site links, as shown in Figure 10-104 on page 856 (configuration 5).

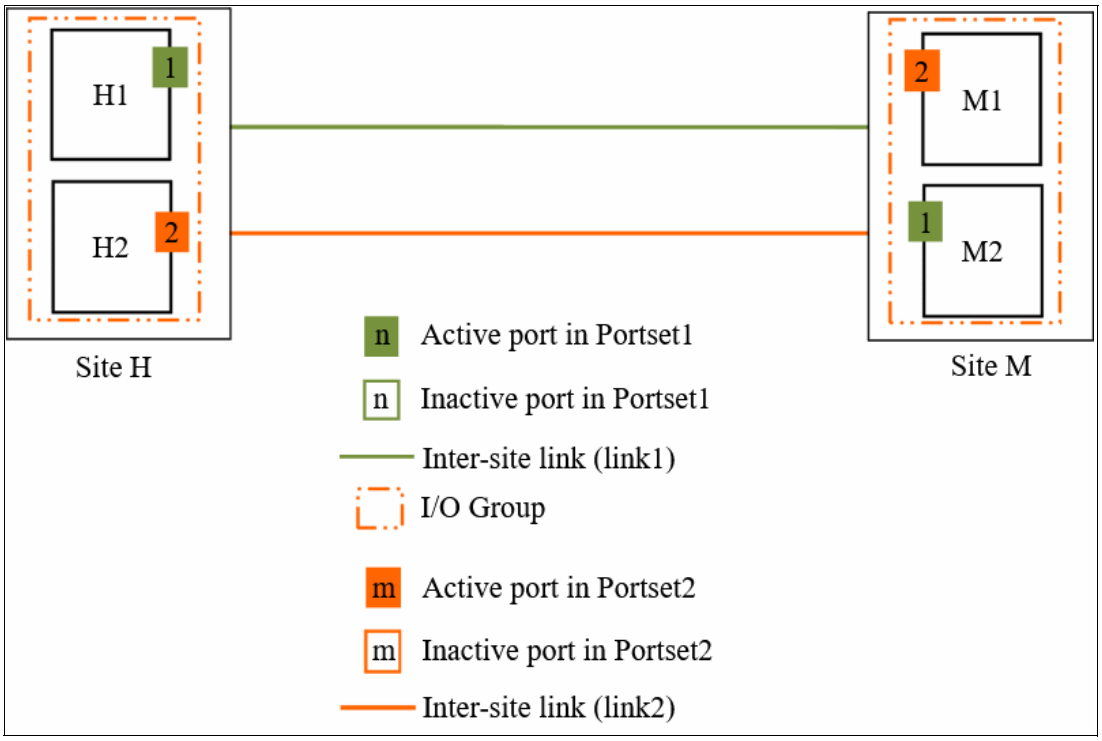


Figure 10-104 Dual links with two Remote Copy portsets on each system configured

As shown in Figure 10-104, RC portsets 1 and 2 are configured on the nodes in System H and System M because two inter-site links are available. In this configuration, the failover ports are not configured on partner nodes in the I/O group. Instead, the ports are maintained in different RC port groups on both of the nodes. They remain active and participate in IP partnership by using both of the links.

However, if either of the nodes in the I/O group fail (that is, if Node H1 on System H fails), the IP partnership continues only from the available IP port that is configured in RC portset 2. Therefore, the effective bandwidth of the two links is reduced to 50% because only the bandwidth of a single link is available until the failure is resolved.

This configuration has the following characteristics:

- Two inter-site links and two RC portset are configured.
- Each node has only one IP port in RC portset 1 or 2.
- Both the IP ports in the two RC portsets participate simultaneously in IP partnerships. Therefore, both of the links are used.
- During node failure or link failure, the IP partnership traffic continues from the other available link and the portset. Therefore, if two links of 10 Mbps each are available and you have 20 Mbps of effective link bandwidth, the bandwidth is reduced to 10 Mbps only during a failure.
- After the node failure or link failure is resolved and failback occurs, the entire bandwidth of both of the links is available as before.

- Two 4-node systems in IP partnership with dual inter-site links, as shown in Figure 10-105 (configuration 6).

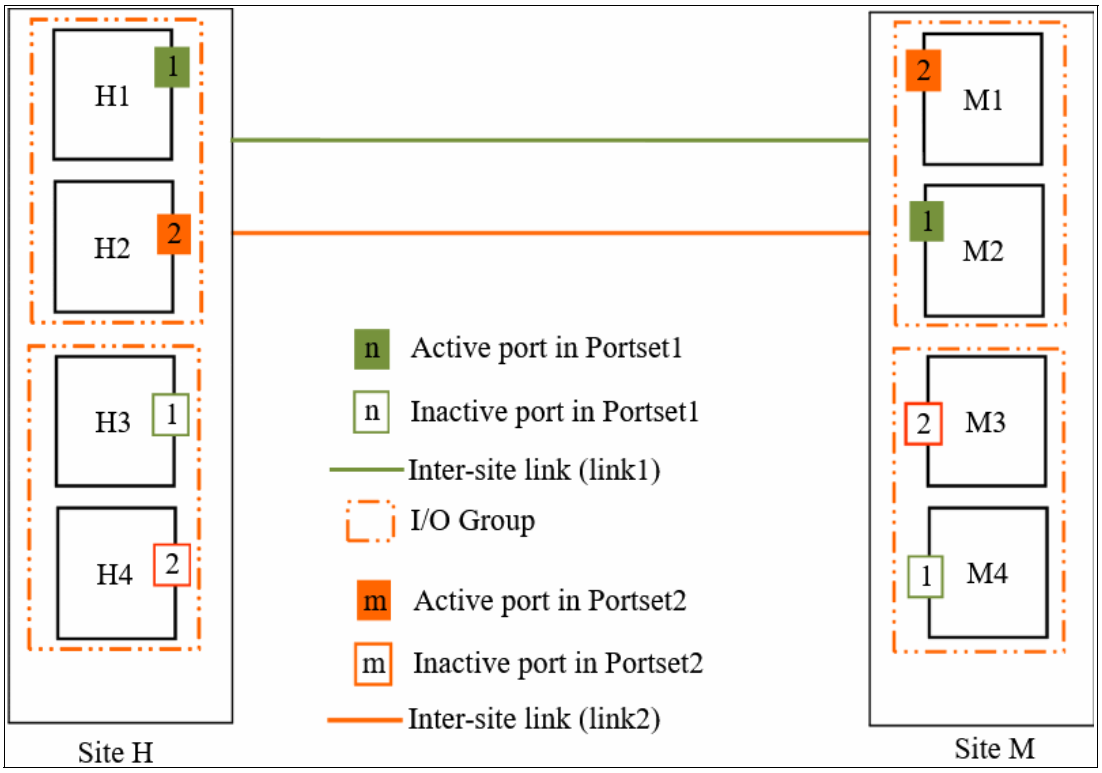


Figure 10-105 Multinode systems with dual inter-site links between the two systems

As shown in Figure 10-105, two 4-node systems are used:

- System H
- System M

This configuration is an extension of Configuration 5 to a multinode multi-I/O group environment. This configuration has two I/O groups, and each node in the I/O group has a single port that is configured in RC portset 1 or 2.

Although two ports are configured in RC portsets 1 and 2 on each system, only one IP port in each RC portset on each system actively participates in IP partnership. The other ports that are configured in the same RC portset act as standby ports in the event of failure. Which port in a configured RC portset participates in IP partnership at any moment is determined by a path configuration algorithm.

In this configuration, if Node H1 fails in System H, IP partnership traffic continues from Node H2 (that is, RC portset 2) and at the same time the failover also causes discovery in RC portset 1.

Therefore, the IP partnership traffic continues from Node H3 on which RC portset 1 is configured. The details of the particular IP port that is actively participating in the IP partnership process is provided in the **`lsip`** output.

This configuration has the following characteristics:

- Each node has the RC portset that is configured in the I/O groups 1 or 2. However, only one port per system in both RC portsets remains active and participates in IP partnership.
- Only a single port per system from each configured RC portset participates simultaneously in the IP partnership. Therefore, both of the links are used.

- During node failure or port failure of a node that is actively participating in IP partnership, IP partnership continues from the alternative port because another port is in the system in the same RC portset but in a different I/O group.
- The pathing algorithm can start discovery of available ports in the affected RC port group in the second I/O group and pathing is reestablished, which restores the total bandwidth, so both of the links are available to support IP partnership.

- Six-node system in IP partnership with a six-node system over dual inter-site links, as shown in Figure 10-106 (configuration 7).

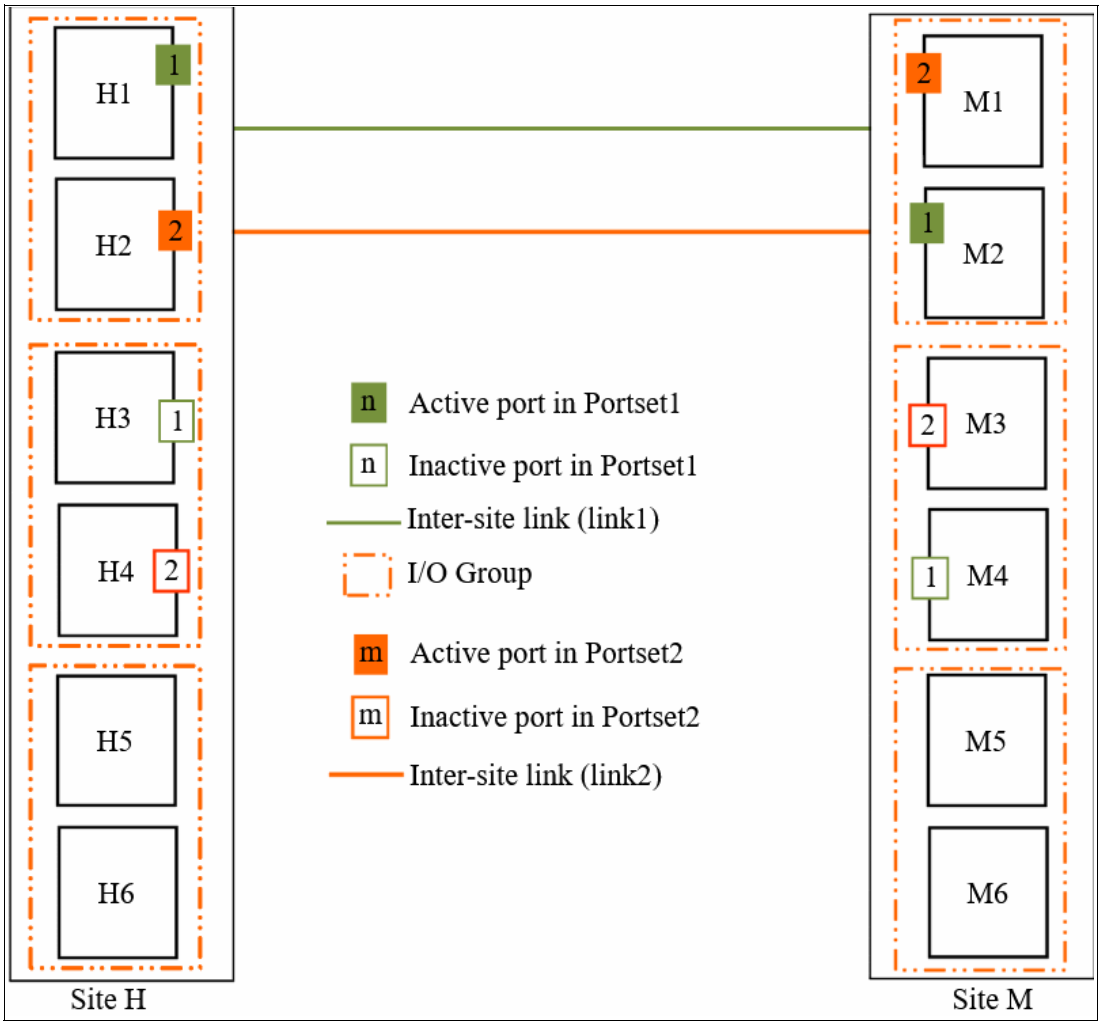


Figure 10-106 Multinode systems with dual inter-site links between the two systems

As shown in Figure 10-106, a six-node System H in Site H and a six-node System M in Site M is used. Because a maximum of two I/O groups in IP partnership is supported in a system (although three I/O groups [six nodes] exist), nodes from only two I/O groups are configured with RC portsets in System H. The remaining or all of the I/O groups can be configured to be RC partnerships over FC.

In this configuration, two links and two I/O groups are configured with RC portsets 1 and 2, but path selection logic is managed by an internal algorithm. Therefore, this configuration depends on the pathing algorithm to decide which of the nodes actively participates in IP partnership. Even if Node H3 and Node H4 are configured with RC portsets correctly, active IP partnership traffic on both of the links might be driven from Node H1 and Node H2 only.

If Node H1 fails in System H, IP partnership traffic continues from Node HA2 (that is, RC portset 2). The failover also causes IP partnership traffic to continue from Node H3 on which RC portset 1 is configured. The details of the specific IP port actively participating in the IP partnership process is provided in the **`lsip`** output.

This configuration has the following characteristics:

- Two I/O groups with nodes in those I/O groups are configured in two RC portsets because two inter-site links are used for participating in IP partnership. However, only one port per system in a particular RC portset remains active and participates in IP partnership.
- One port per system from each RC portset participates in the IP partnership simultaneously. Therefore, both of the links are used.
- If a node or port on the node that is actively participating in IP partnership fails, the RC data path is established from that port because another port is available on an alternative node in the system with the same RC portset.
- The path selection algorithm starts discovery of available ports in the affected RC portset in the alternative I/O groups and paths are reestablished, which restores the total bandwidth across both links.
- The remaining or all of the I/O groups can be in RC partnerships with other systems.

- An example of an *unsupported* configuration for a single inter-site link is shown in Figure 10-107 (configuration 8).

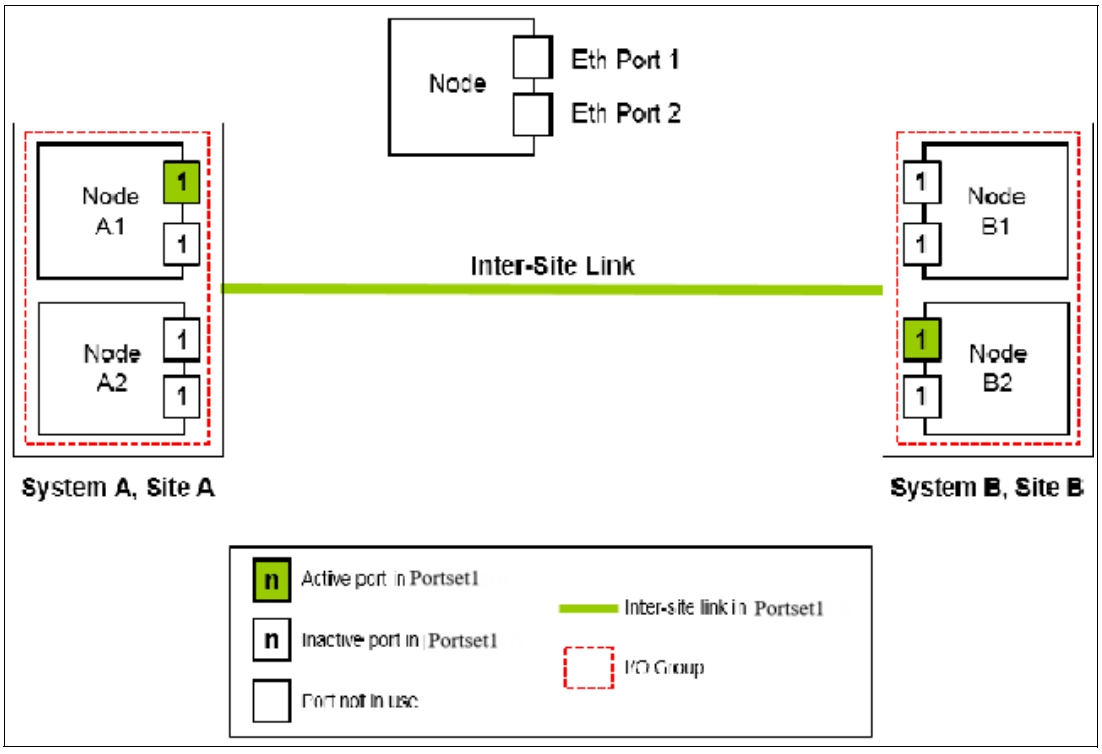


Figure 10-107 Two node systems with single inter-site link and Remote Copy portset configured

As shown in Figure 10-107, this configuration is similar to Configuration 2, but differs because each node now has the same RC portset that is configured on more than one IP port.

On any node, only one port at any time can participate in IP partnership. Configuring multiple ports in the same RC portset on the same node is *not* supported.

- An example of an *unsupported* configuration for a dual inter-site link is shown in Figure 10-108 (configuration 9).

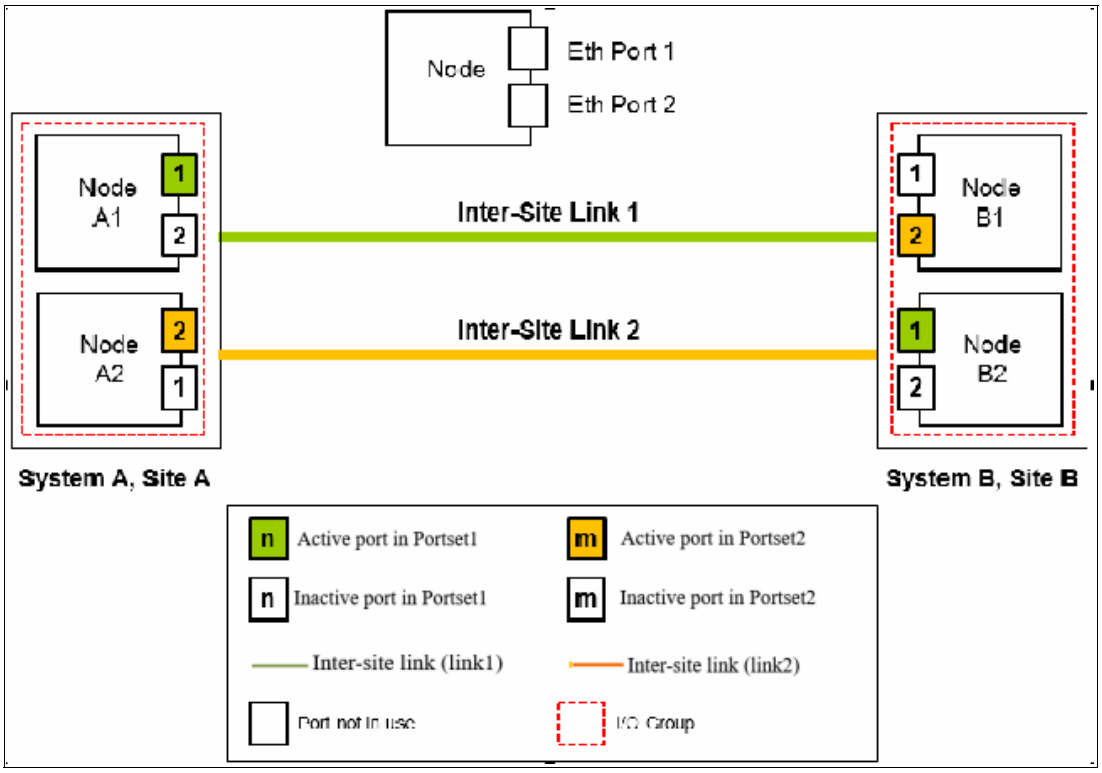


Figure 10-108 Dual Links with two Remote Copy Port Groups with failover Portsets configured

As shown in Figure 10-108, this configuration is similar to Configuration 5, but differs because each node now also has two ports that are configured with RC portsets. In this configuration, the path selection algorithm can select a path that might cause partnerships to change to the `Not_Present` state and then recover, which results in a configuration restriction. The use of this configuration is not recommended until the configuration restriction is lifted in future releases.

- An example deployment for configuration 2 with a dedicated inter-site link is shown in Figure 10-109 (configuration 10).

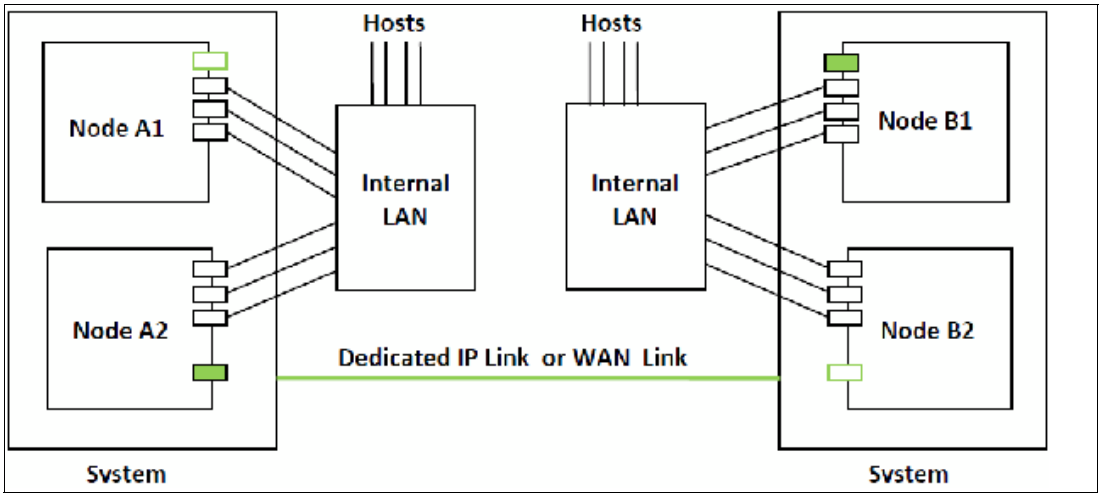


Figure 10-109 Deployment example

In this configuration, one port on each node in System A and System B is configured in RC portset 1 to establish IP partnership and support RC relationships. A dedicated inter-site link is used for IP partnership traffic, and iSCSI host attach is disabled on those ports.

The following configuration steps are used:

a. Configure system IP addresses properly. As such, they can be reached over the inter-site link.

b. Qualify if the partnerships must be created over IPv4 or IPv6, and then assign IP addresses and open firewall ports 3260 and 3265.

c. Configure IP ports for RC on both the systems by using the following settings:

   - Remote copy portset: 1
   - Type: Replication
   - Porttype: Ethernet
   - Assign IP address

d. Check that the maximum transmission unit (MTU) levels across the network meet the requirements as set (default MTU is 1500).

e. Establish IP partnerships from both of the systems.

f. After the partnerships are in the `Fully_Configured` state, you can create the RC relationships.

- Figure 10-109 on page 860 is an example deployment for the configuration that is shown in Figure 10-103 on page 855. Ports that are shared with host access are shown in Figure 10-110 (configuration 11).

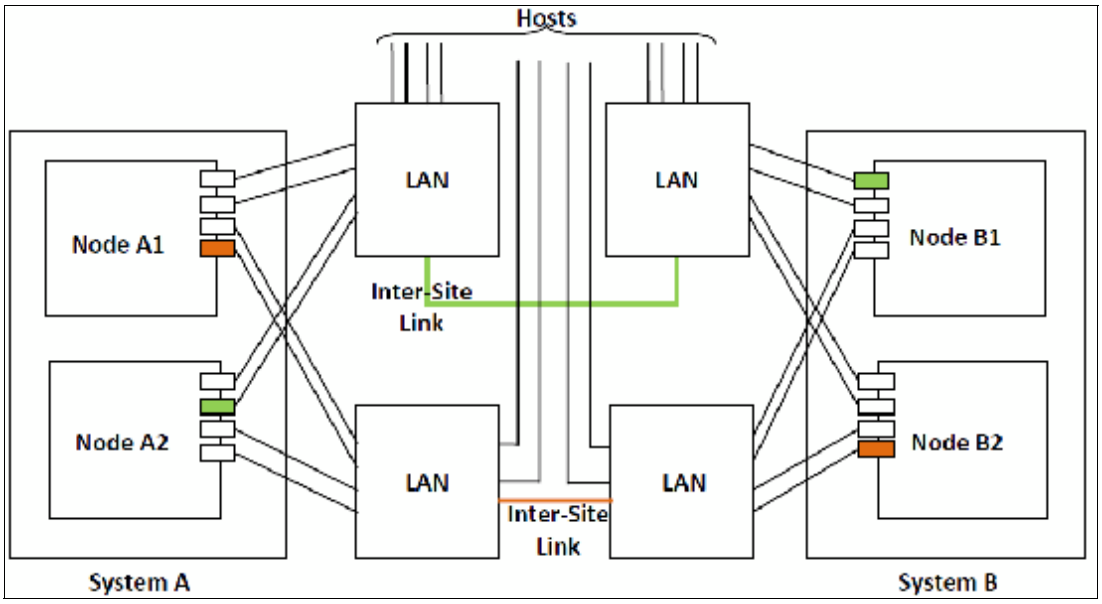


Figure 10-110 Deployment example

In this configuration, IP ports are to be shared by both iSCSI hosts and for IP partnership.

The following configuration steps are used:

a. Configure System IP addresses properly so that they can be reached over the inter-site link.

b. Qualify if the partnerships must be created over IPv4 or IPv6, and then assign IP addresses and open firewall ports 3260 and 3265.

c. Configure IP ports for RC on System A1 by using the following settings:

- Node 1:

  - Port 1, remote copy portset 1
  - Type: Replication
  - Porttype: Ethernet
  - Assign IP address

- Node 2:

  - Port 4, remote copy port group 2
  - Type: Replication
  - Porttype: Ethernet
  - Assign IP address

d. Configure IP ports for RC on System B1 by using the following settings:

- Node 1:

  - Port 1, remote copy portset 1
  - Type: Replication
  - Porttype: Ethernet
  - Assign IP address

- Node 2:

  - Port 4, remote copy port group 2
  - Type: Replication
  - Porttype: Ethernet
  - Assign IP address

e. Check the MTU levels across the network as set (default MTU is 1500 on SAN Volume Controller and Storage Virtualize systems).

f. Establish IP partnerships from both systems.

g. After the partnerships are in the `Fully_Configured` state, you can create the RC relationships.

10.10 Managing Remote Copy by using the GUI

It is often easier to control MM/GM with the GUI if you have few mappings. When many mappings are used, run your commands by using the CLI. This section describes the tasks that you can perform at an RC level.

Note: The **Copy Services** → **Consistency Groups** menu relates to FlashCopy consistency groups only, not RC groups.

The following panels are used to visualize and manage your remote copies:

- Remote Copy panel

  To open the Remote Copy panel, click **Copy Services** → **Remote Copy** in the main menu, as shown in Figure 10-111.

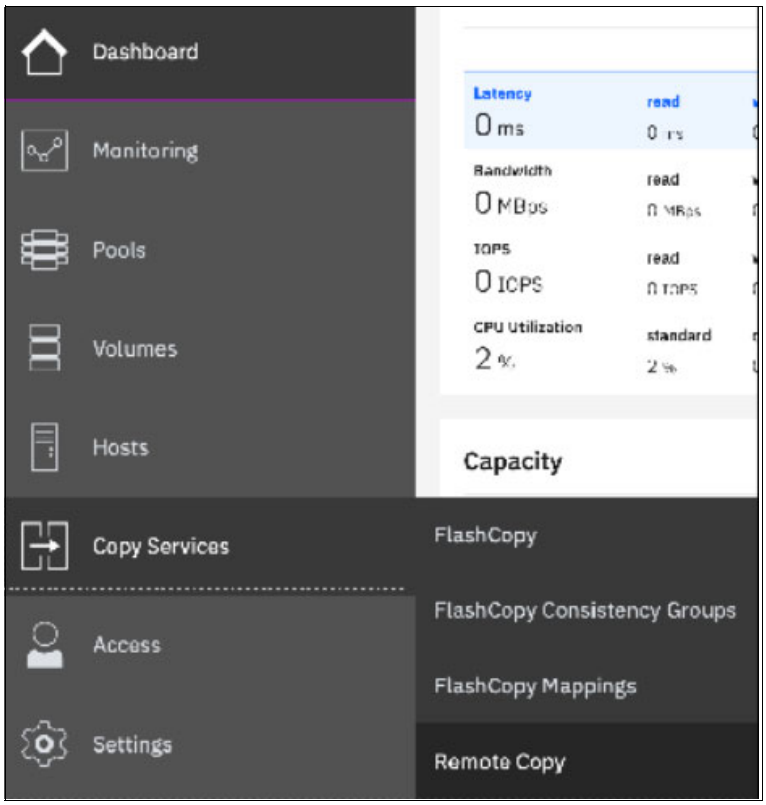


Figure 10-111 Remote Copy menu

The Remote Copy panel is displayed, as shown in Figure 10-112.

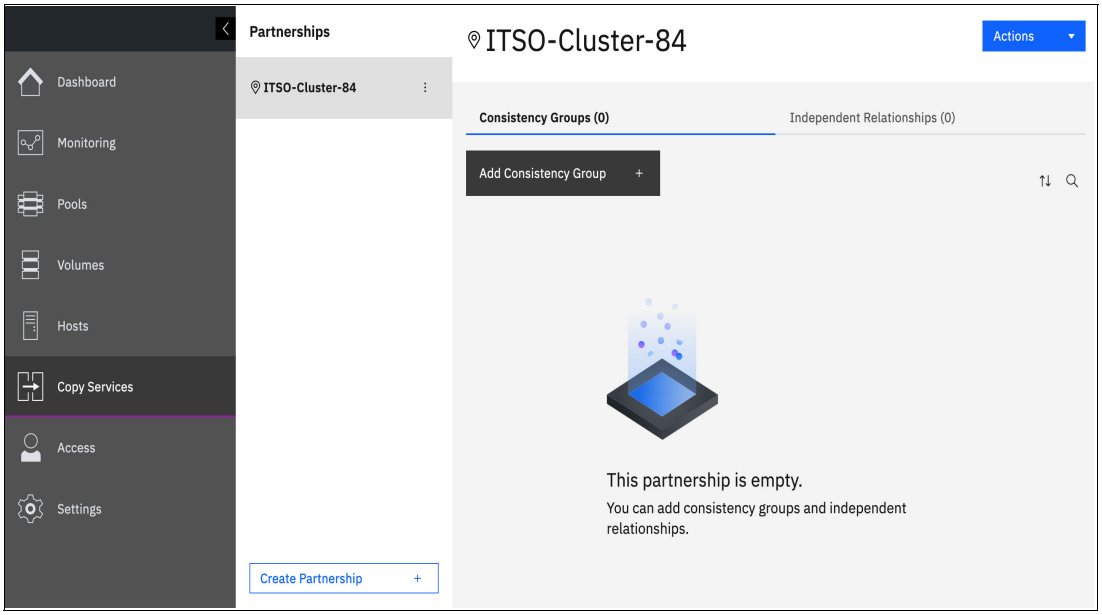


Figure 10-112 Remote Copy panel

10.10.1 Creating Fibre Channel partnership

Intra-cluster MM: If you are creating an intra-cluster MM, do not perform this next process to create the MM partnership. Instead, see 10.10.2, “Creating Remote Copy relationships” on page 866.

To create an FC partnership between IBM Storage Virtualize systems by using the GUI, open the Remote Copy panel shown in the Figure 10-112 on page 863 and click **Create Partnership** to create a partnership.

In the Create Partnership window, complete the following steps:

1. Select the Replication topology between two or three sites, as shown in the Figure 10-113. In this example, we use a two-site partnership.

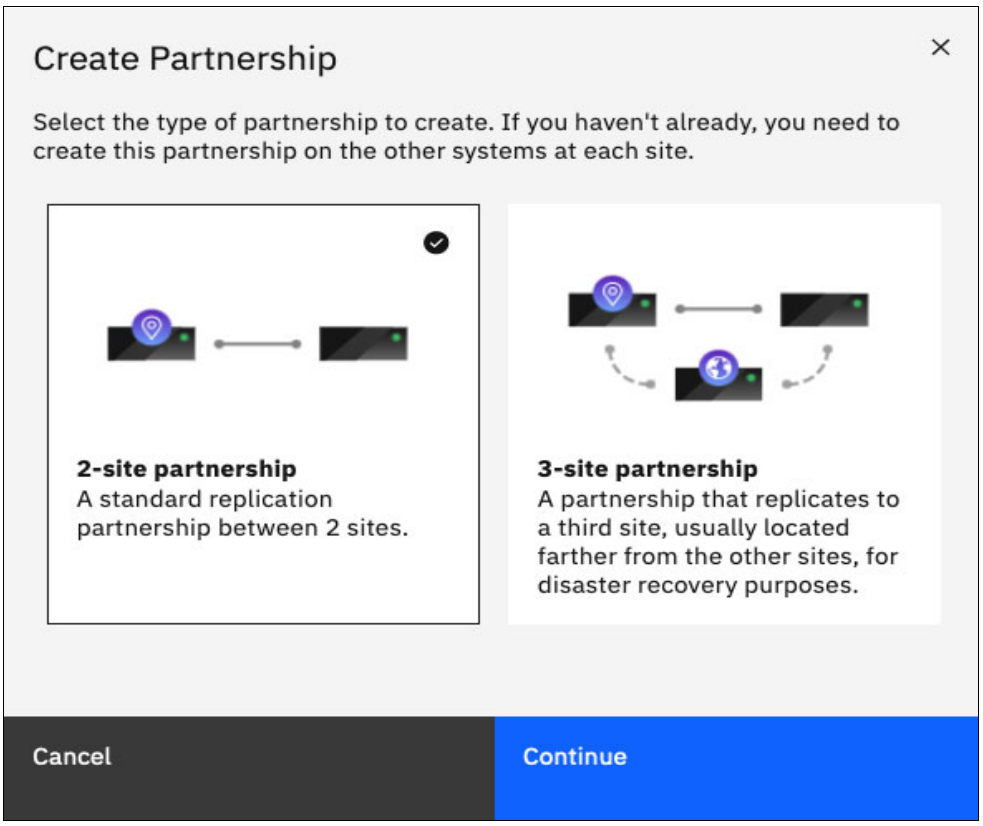


Figure 10-113 Select Remote Copy topology

2. Select the partnership type (**Fibre Channel** or **IP**). If you choose IP partnership, you must provide the IP address of the partner system and the partner system’s CHAP key.

3. If your partnership is based on Fibre Channel Protocol (FCP), select an available partner system from the menu. To select a partner system, the two clusters must be correctly zoned between each other. If no other candidate cluster is available, the following message is displayed (see Figure 10-114):

   `This system does not have any candidates.`

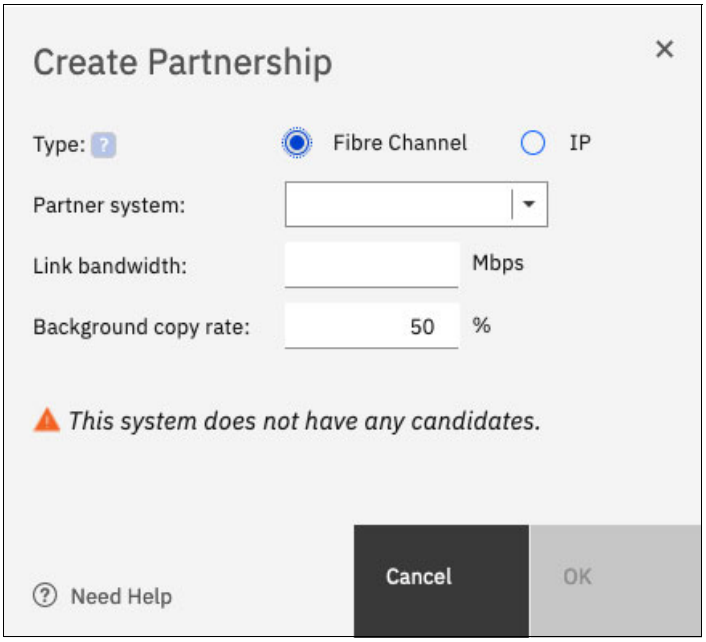


Figure 10-114 Creating a Partnership details

4. Enter a link bandwidth in Mbps that is used by the background copy process between the systems in the partnership.
5. Enter the background copy rate.
6. Click **OK** to confirm the partnership relationship.

To fully configure the partnership between both systems, perform the same steps on the other system in the partnership. If not configured on the partner system, the partnership is displayed as `Partial Local`.

When both sides of the system partnership are defined, the partnership is displayed as `Configured` green status, as shown in Figure 10-115.

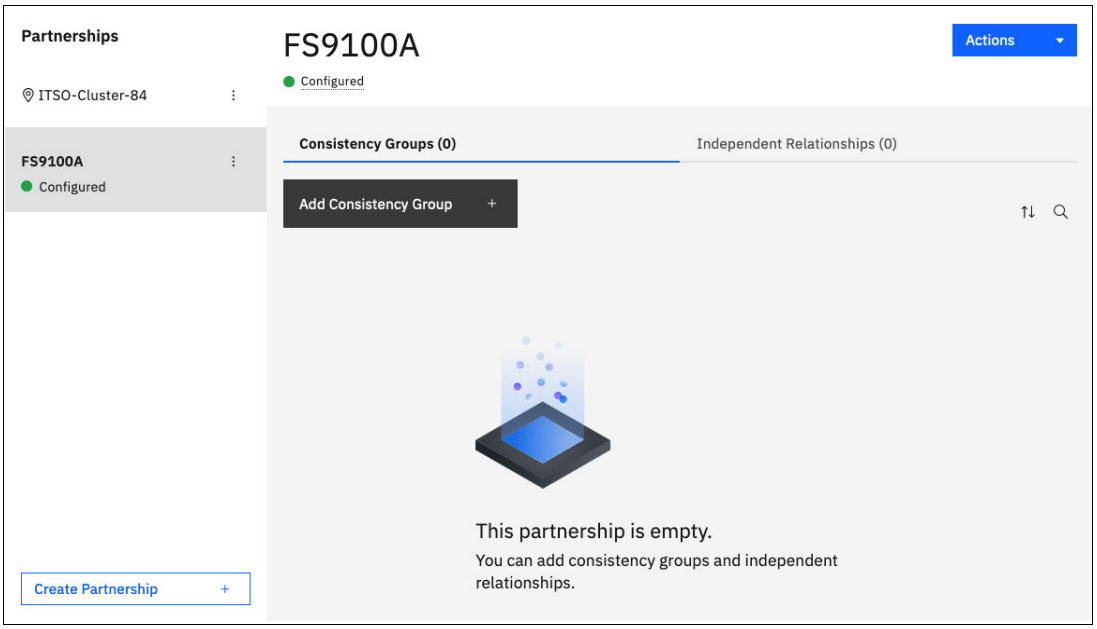


Figure 10-115 Fully configured FC partnership

10.10.2 Creating Remote Copy relationships

This section shows how to create RC relationships for volumes with their respective remote targets. Before creating a relationship between a volume on the local system and a volume on a remote system, both volumes must exist and have the same virtual size.

To create an RC relationship, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Select the target system with which you want to create an RC relationship.
3. If you want to add the relationship to an existing consistency group, select the consistency group for which you want to create the relationship and click **Create Relationship**, as shown in Figure 10-116 on page 867.

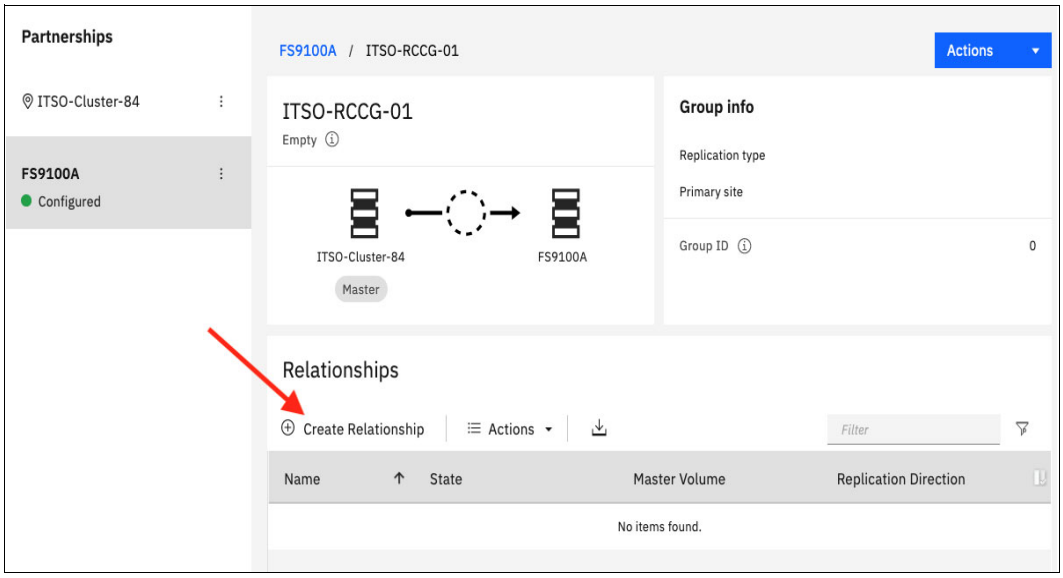


Figure 10-116 Creating a Remote Copy Relationship in an existing consistency group

4. If you want to add a standalone relationship, select the **Independent Relationships** tab and click **Create Relationship**, as shown in Figure 10-117.

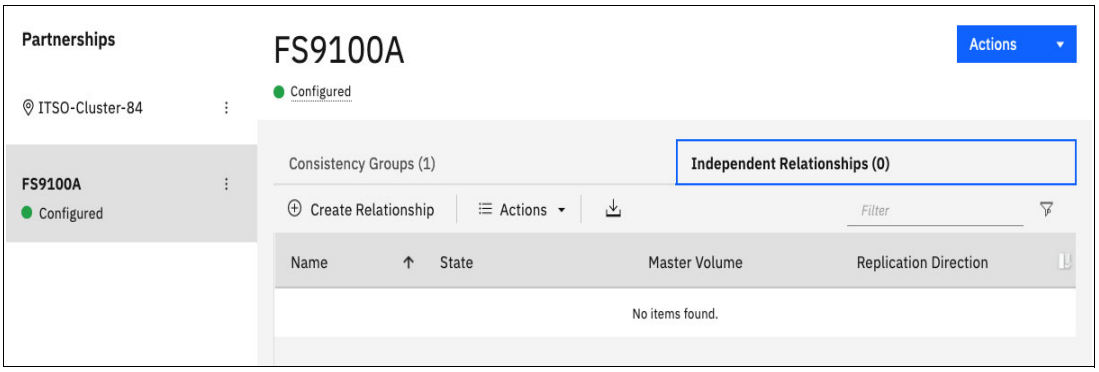


Figure 10-117 Creating Standalone Remote Copy relationships

5. In the Create Relationship window, select one of the following types of relationships that you want to create, as shown in Figure 10-118 on page 868:
   - MM
   - GM (with or without Consistency Protection)
   - GMCV

   Click **Next**.

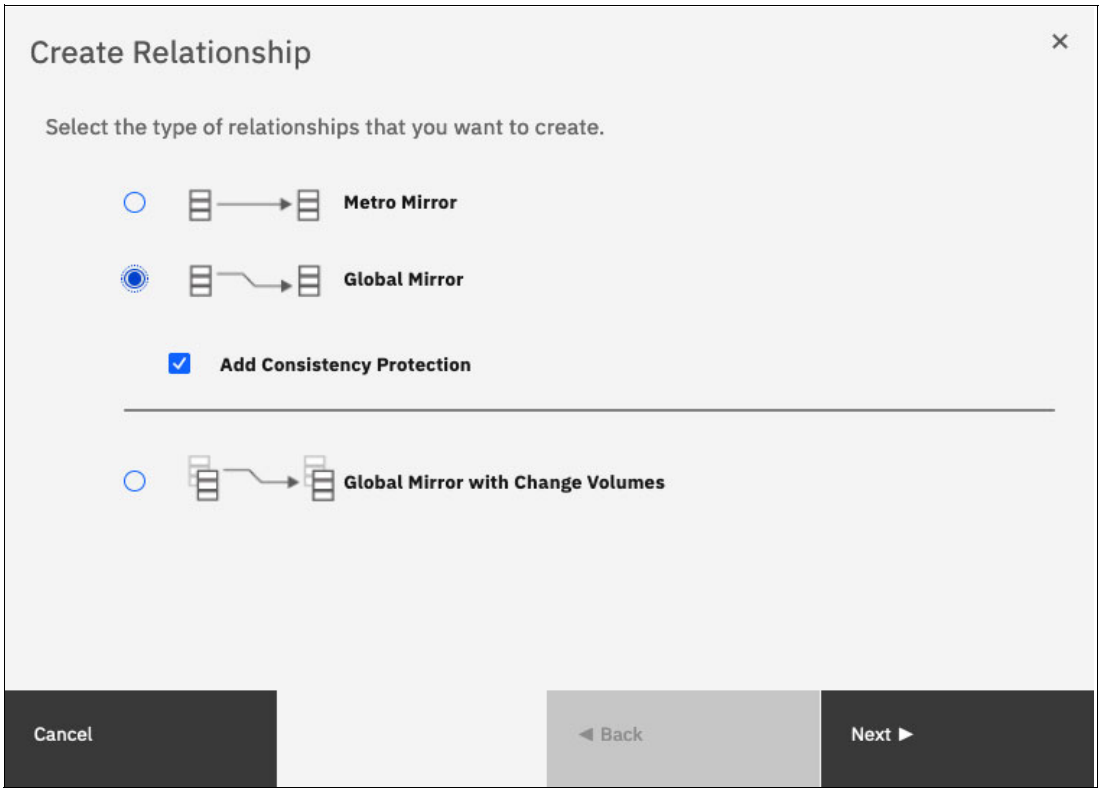


Figure 10-118 Creating a Remote Copy relationship

6. In the next window, select the target system for this RC relationship and click **Next**, as shown in Figure 10-119, “Selecting the target system for the RC relationship” on page 869.
7. Select the master and auxiliary volumes, as shown in Figure 10-120 on page 869. Click **ADD**.

Important: The master and auxiliary volumes must be of equal size. Therefore, only the targets with the suitable size are shown in the list for a specific source volume.

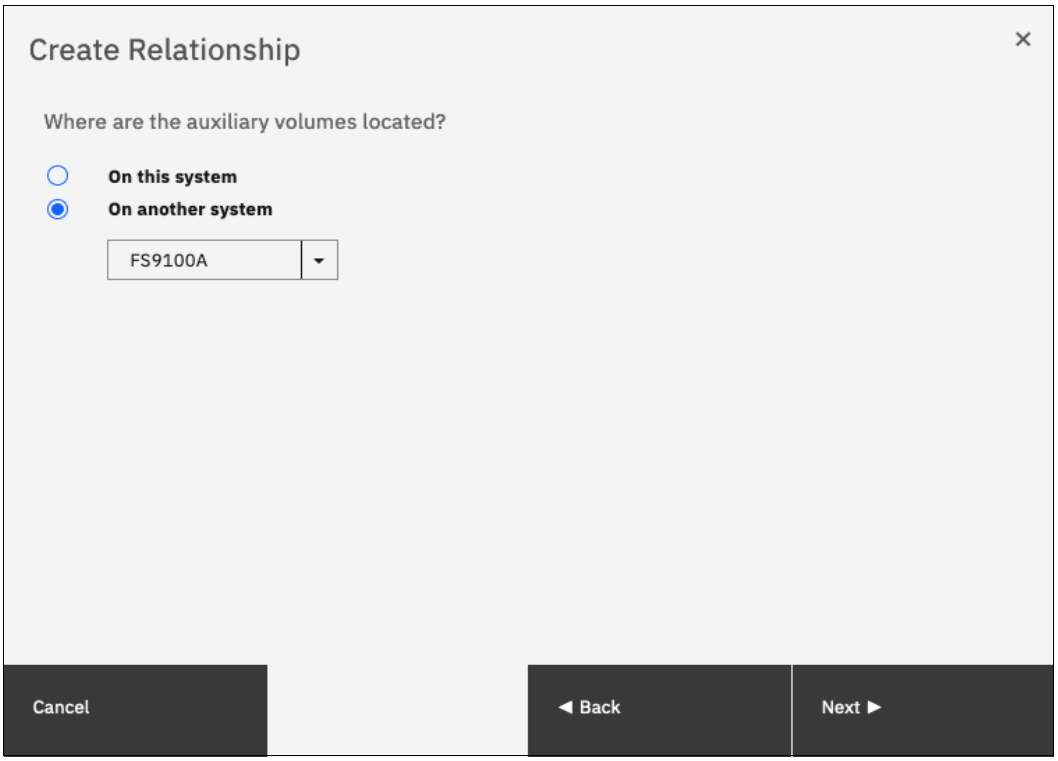


Figure 10-119 Selecting the target system for the RC relationship

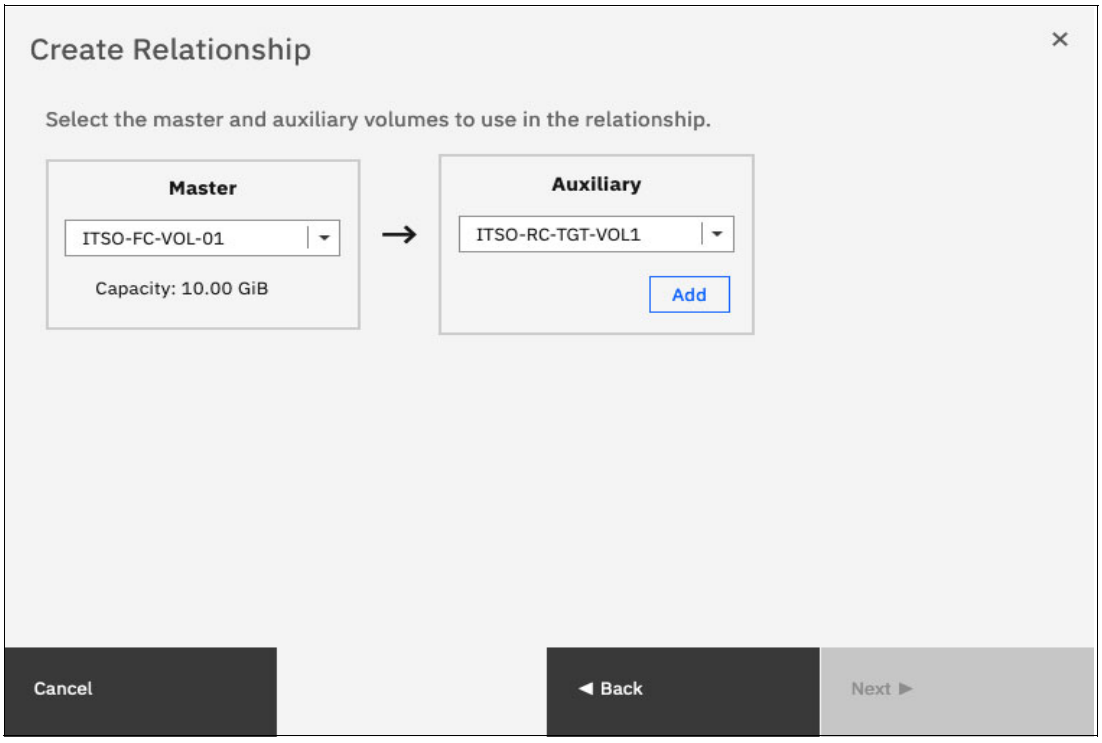


Figure 10-120 Selecting the master and auxiliary volumes

8. In the next window, you can add change volumes if needed, as shown in Figure 10-121. Click **Finish**.

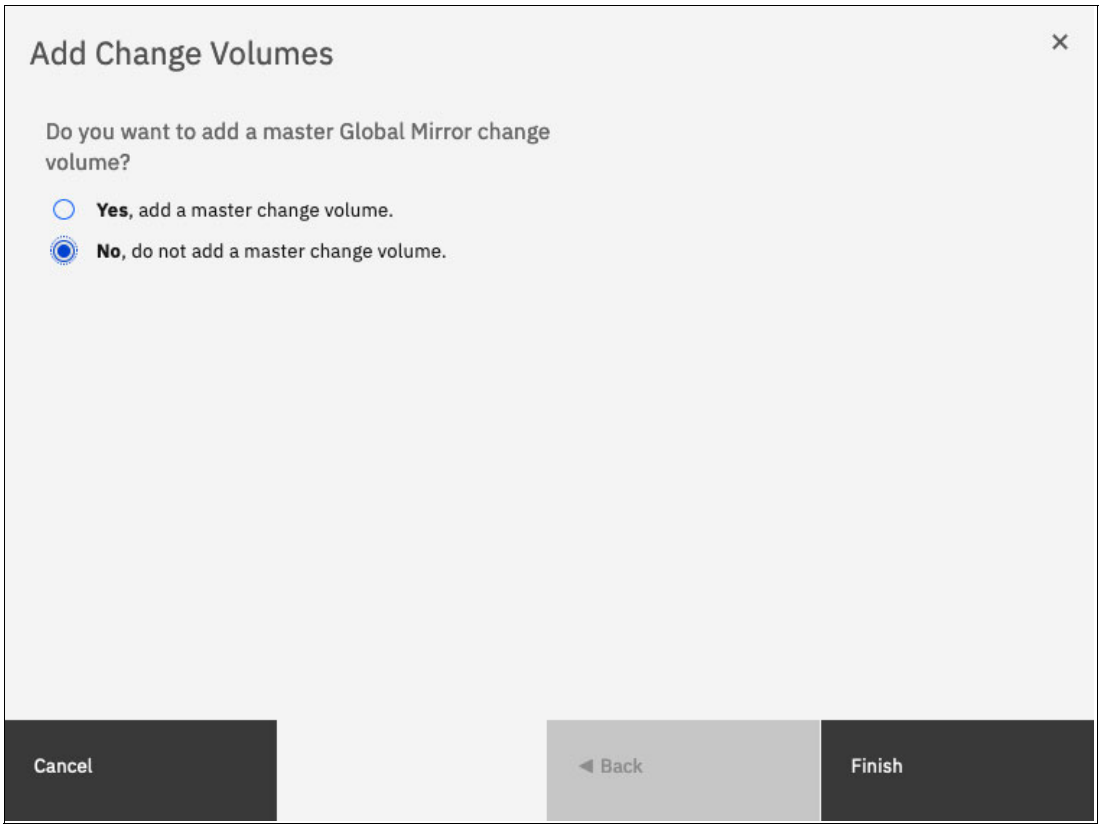


Figure 10-121 Add Change Volumes panel

9. Repeat the steps 7 and 8 to add relationships if needed (see Figure 10-122 on page 871). When you are finished creating relationships, click **Next**.
10. In the next window, select whether the volumes are synchronized so that the relationship is created, as shown in Figure 10-123 on page 871. Click **Next**.

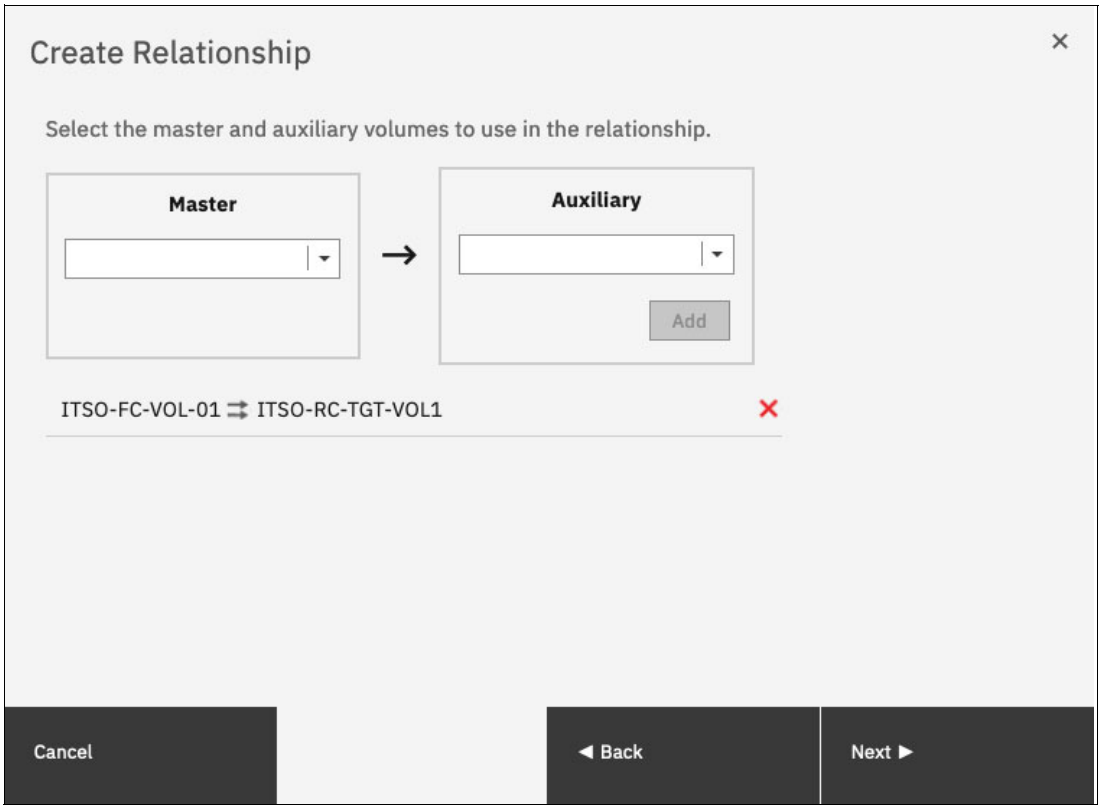


Figure 10-122 Checking and adding the relationship

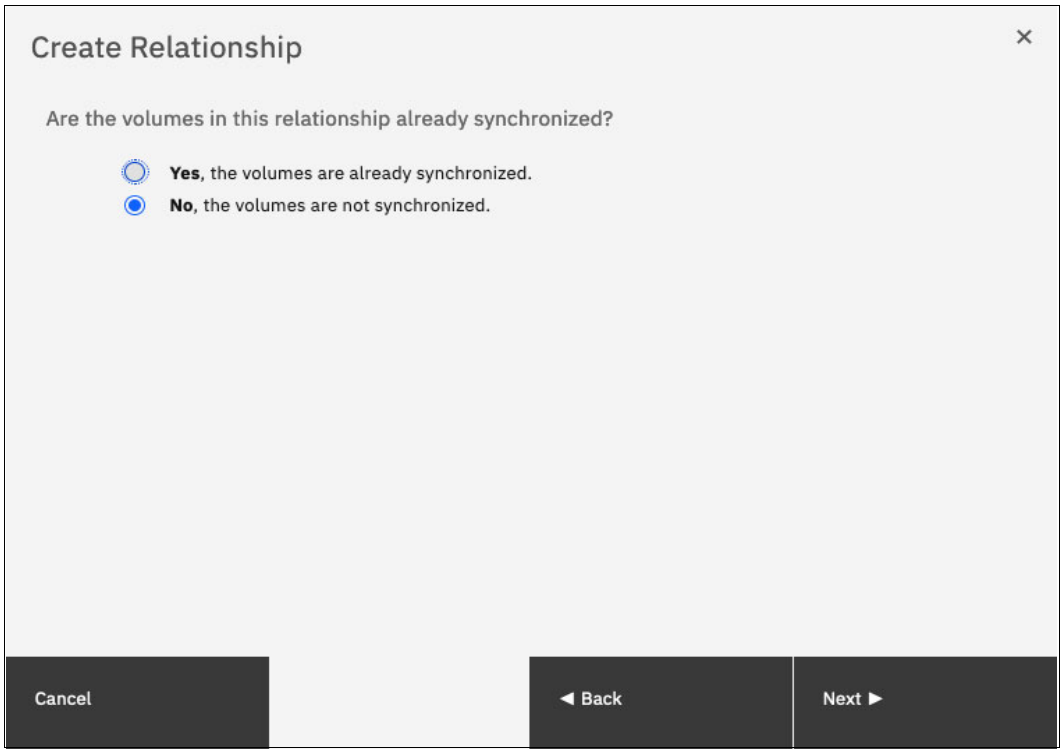


Figure 10-123 Selecting if volumes are synchronized

11. Select whether you want to start synchronizing the Master and Auxiliary volumes when the relationship is created, or if you want to start the copy in a later time, as shown in Figure 10-124. Then click **Finish**.

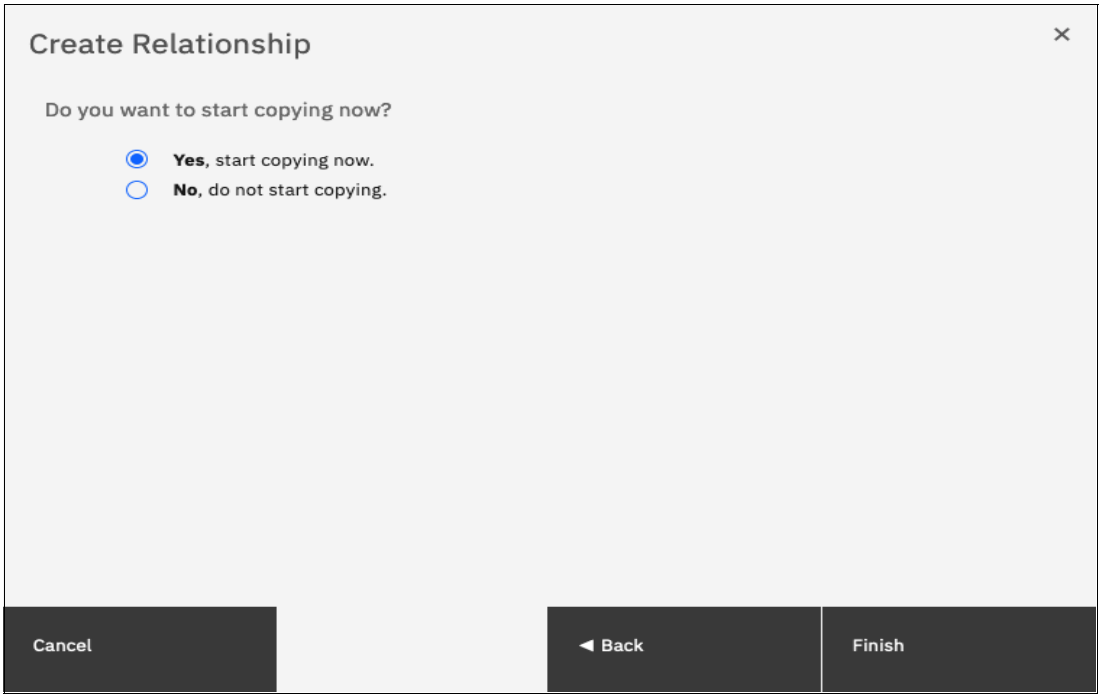


Figure 10-124 Start copying Remote Copy relationship

Note: If the volumes are not synchronized, the initial copy copies the entire source volume to the remote target volume. If you suspect volumes are different or if you have a doubt, synchronize them to ensure consistency on both sides of the relationship.

10.10.3 Creating a consistency group

To create a consistency group, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel and select the target system of the Remote Copy. Then click **Add Consistency Group**, as shown in Figure 10-125.

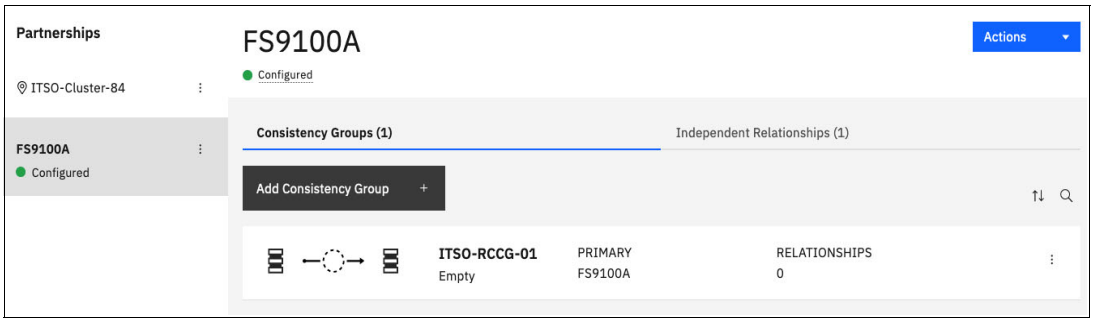


Figure 10-125 Creating a Remote Copy consistency group

2. Enter a name for the consistency group, select the target system and click **Add**, as shown in Figure 10-126. It is then added to the configuration with no relationships.

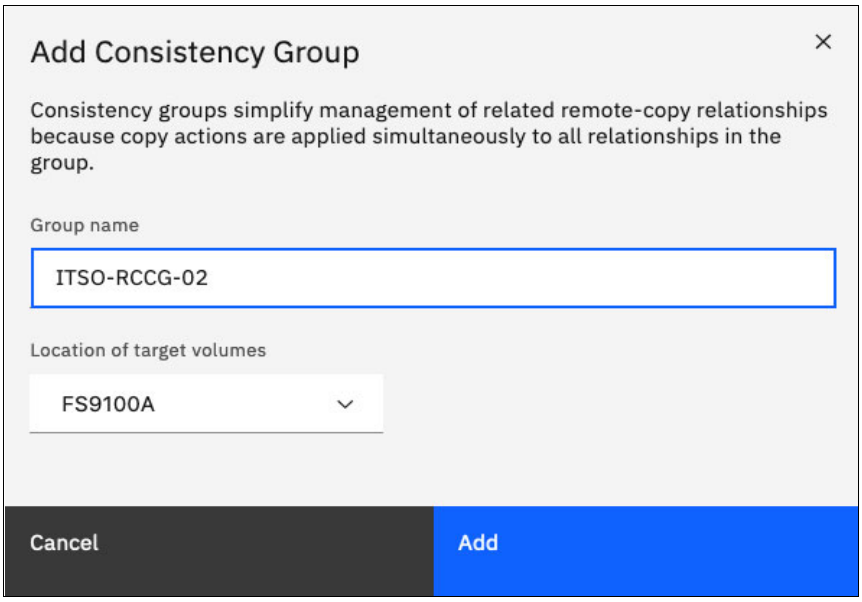


Figure 10-126 Entering a name for the new consistency group

3. You can add standalone relationships to the recently added consistency group by selecting the **Independent Relationships** tab, right-clicking the wanted relationship and clicking **Add to Consistency Group**. You can also create relationships directly for this consistency group by selecting it in the **Consistency Group** tab(Figure 10-127), and then clicking **Create Relationship**.

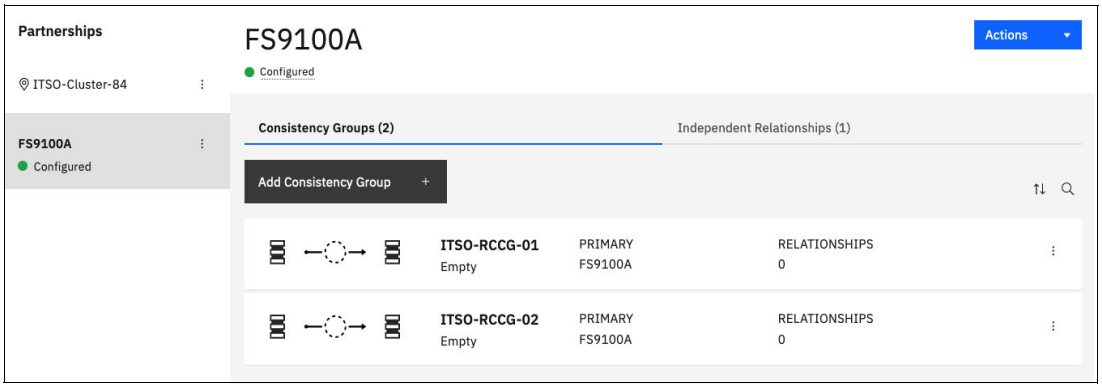


Figure 10-127 Remote Copy Consistency Groups tab

For more information about creating a Remote Copy relationship, see 10.10.2, “Creating Remote Copy relationships” on page 866.

10.10.4 Renaming Remote Copy relationships

To rename one or multiple RC relationships, complete the following steps:

1. Select **Copy Services** → **Remote Copy**.
2. Select the suitable tab for the relationship that you want to rename, whether it is part of a consistency group or an Independent Relationship. If it is part of a consistency group‘s name to view its relationships.
3. Right-click the relationships to be renamed and select **Rename**, as shown in Figure 10-128.

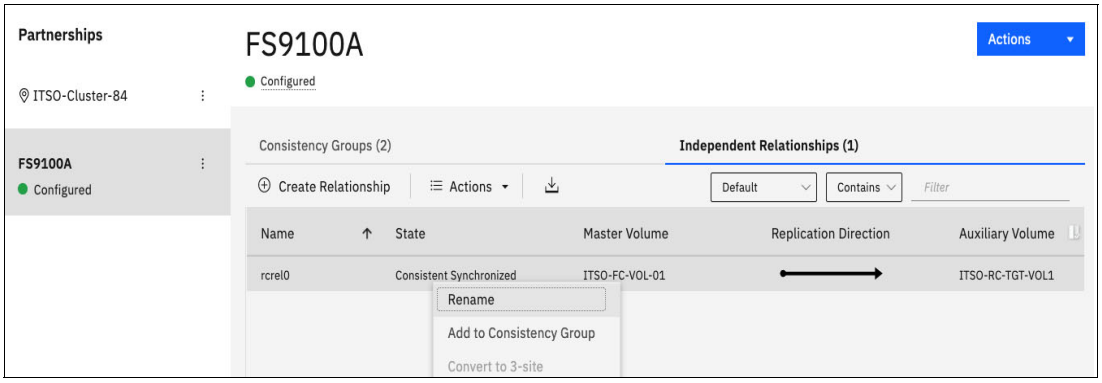


Figure 10-128 Renaming Remote Copy relationships

4. Enter the new name that you want to assign to the relationships and click **Rename**, as shown in Figure 10-129.

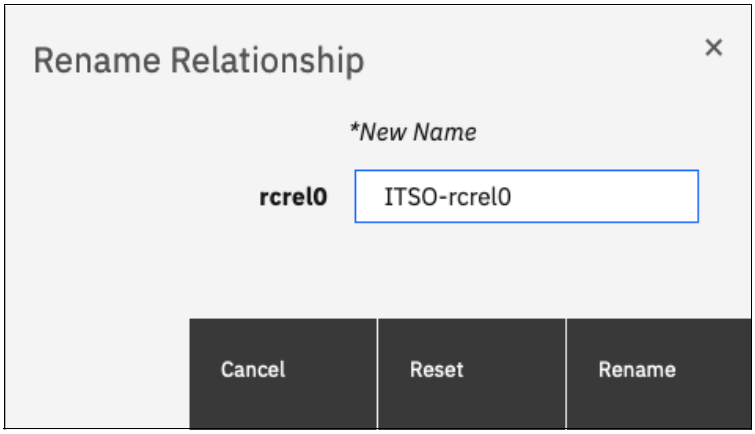


Figure 10-129 Renaming Remote Copy relationships

RC relationship name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The RC name can be 1 - 15 characters. Blanks cannot be used.

10.10.5 Renaming a Remote Copy consistency group

To rename an RC consistency group, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.

2. Select the target system for the RC consistency group you want to rename, click the three dots on the consistency group to be renamed, and select **Rename Group**, as shown in Figure 10-130.

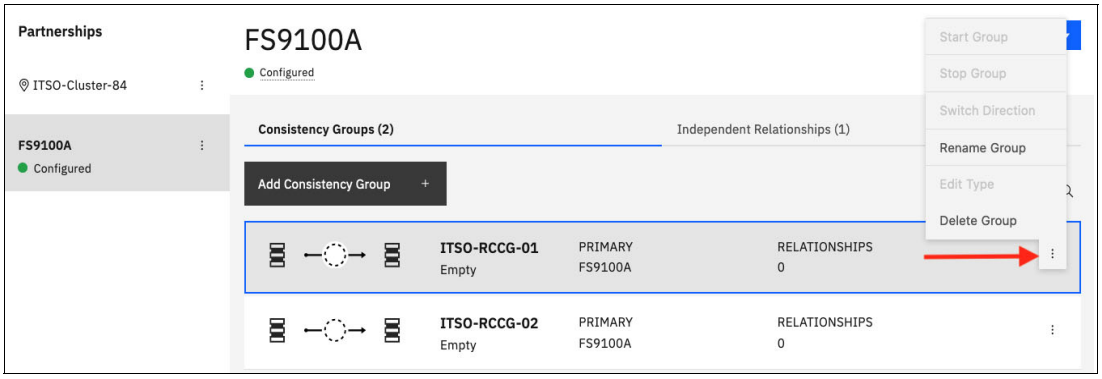


Figure 10-130 Renaming a Remote Copy consistency group

3. Enter the new name that you want to assign to the consistency group and click **Rename**, as shown in Figure 10-131.

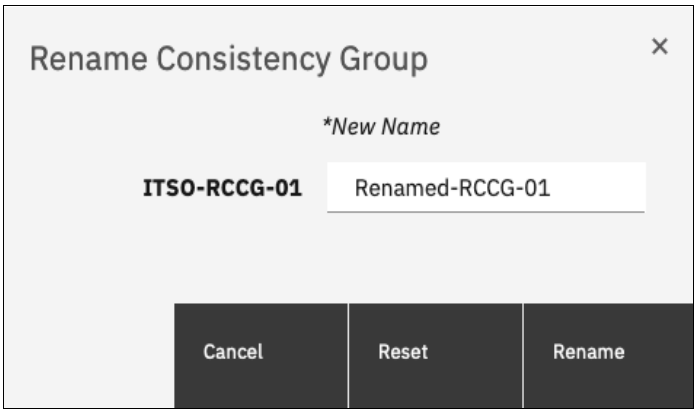


Figure 10-131 Entering new name for consistency group

RC consistency group name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The RC name can be 1 - 15 characters. Blanks cannot be used.

10.10.6 Moving standalone Remote Copy relationships to consistency group

To add one or multiple stand-alone relationships to an RC consistency group, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Select the target system for the relationship to be moved, and go to the **Independent Relationships** tab. Right-click the relationship to be moved and select **Add to Consistency Group**, as shown in Figure 10-132 on page 876.

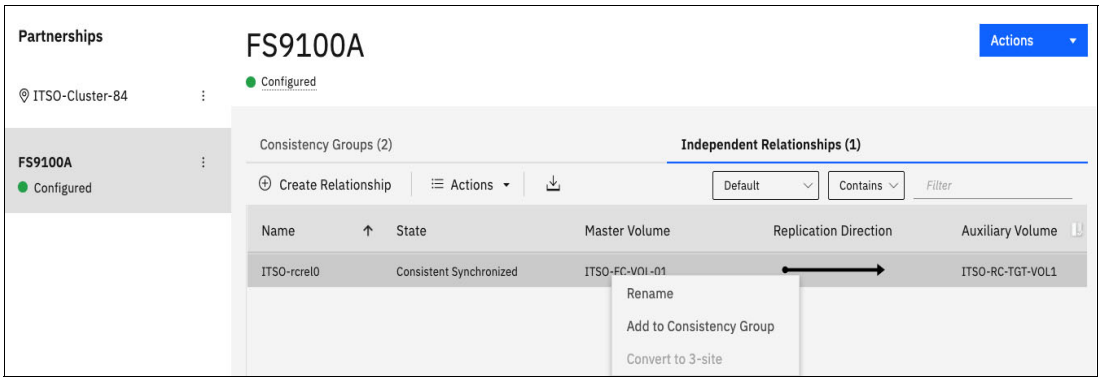


Figure 10-132 Moving relationships to a consistency group

3. Select the consistency group for this RC relationship by using the menu, as shown in Figure 10-133. Click **Add to Consistency Group** to confirm your changes.

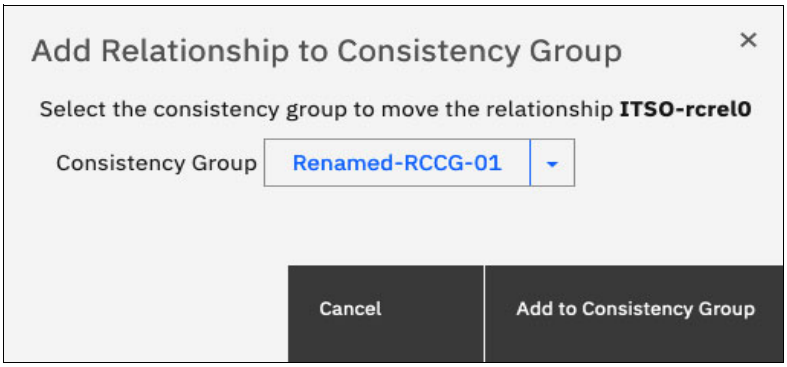


Figure 10-133 Selecting the consistency group to add the relationships to

10.10.7 Removing Remote Copy relationships from consistency group

To remove one or multiple relationships from an RC consistency group, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Select the target RC system and go to the **Consistency Groups** tab. Click the wanted consistency group to view its relationships.
3. Right-click the relationships to be removed and select **Remove from Consistency Group**, as shown in Figure 10-134 on page 877.

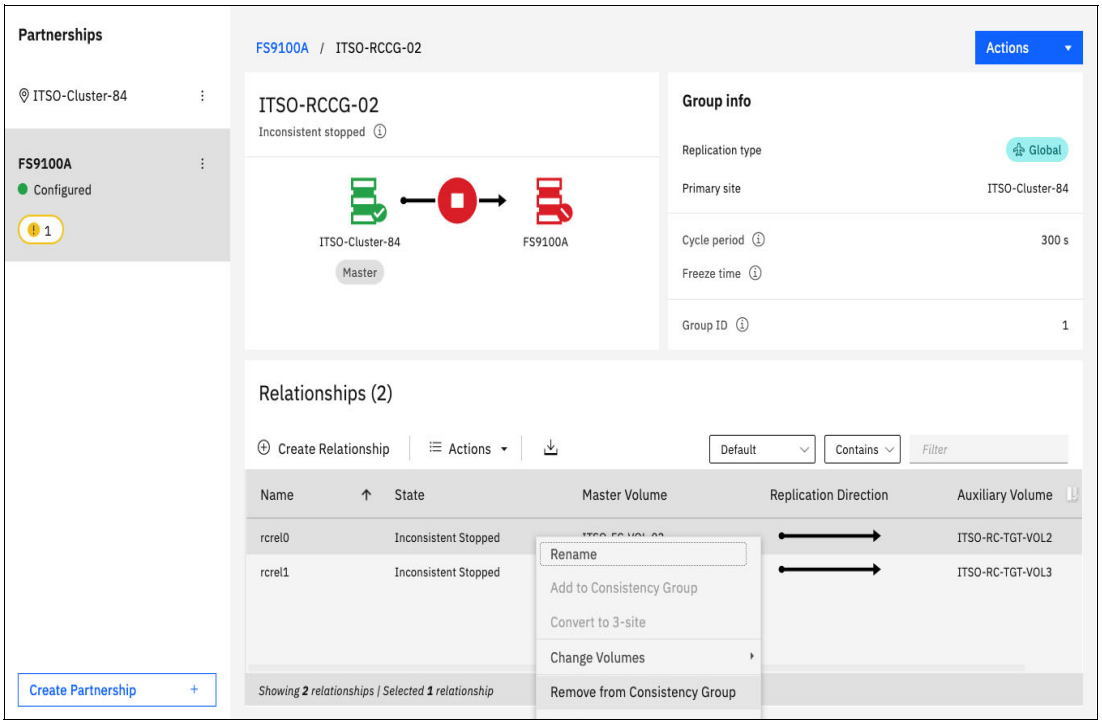


Figure 10-134 Removing relationships from a consistency group

4. Confirm your selection and click **Remove**, as shown in Figure 10-135.

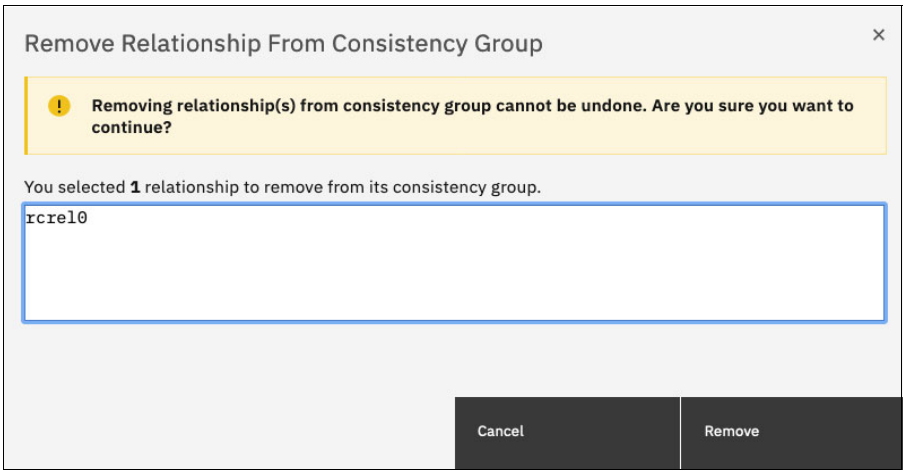


Figure 10-135 Confirm the removal of relationships from a consistency group

10.10.8 Starting Remote Copy relationships

When an RC relationship is created, the RC process can be started. Only relationships that are not members of a consistency group (or the entire consistency group) can be started.

To start one or multiple standalone relationships, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Select the target RC system and go to the **Independent Relationships** tab. Right-click the relationships to be started and select **Start**, as shown in Figure 10-136.

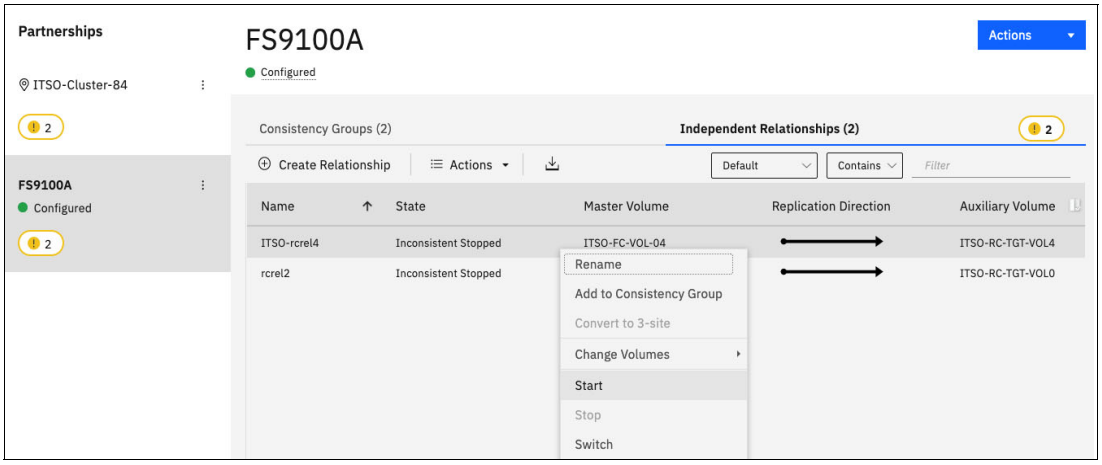


Figure 10-136 Starting Remote Copy relationships

10.10.9 Starting a Remote Copy consistency group

When an RC consistency group is created, the RC process can be started for all the relationships that are part of the consistency groups.

To start a consistency group, open the **Copy Services** → **Remote Copy** panel, select the target RC system and go to the **Consistency Groups** tab. Next, click the three dots on the consistency group to be started, and select **Start Group**, as shown in Figure 10-137.

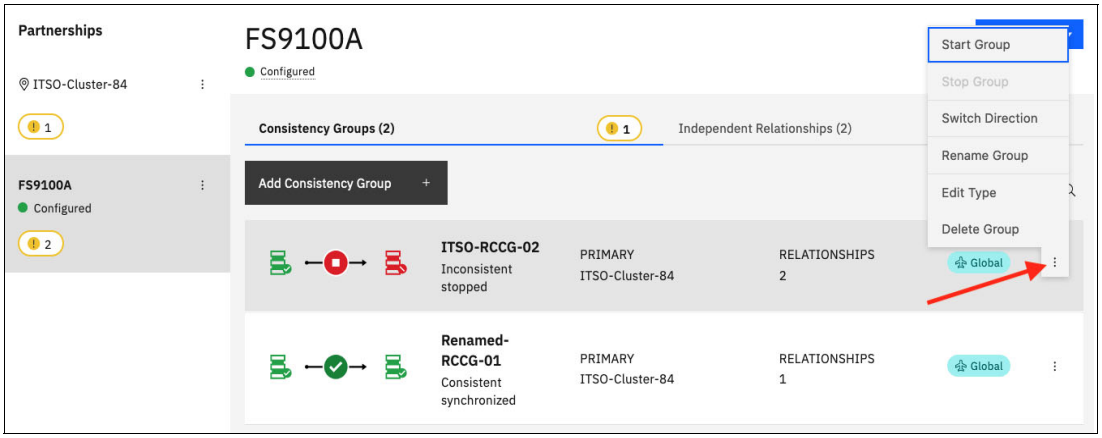


Figure 10-137 Starting a Remote Copy consistency group

10.10.10 Switching a relationship copy direction

When an RC relationship is in the Consistent synchronized state, the copy direction for the relationship can be changed. Only relationships that are not member of a consistency group, or the entire consistency group, can be switched.

Important: When the copy direction is switched, it is crucial that no outstanding I/O exists to the volume that changes from primary to secondary because all of the I/O is inhibited to that volume when it becomes the secondary. Therefore, careful planning is required before you switch the copy direction for a relationship.

To switch the direction of a standalone RC relationship, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Select the target RC system for the relationship to be switched, and go to the **Independent Relationships** tab. Right-click the relationship to be switched and select **Switch**, as shown in Figure 10-138.

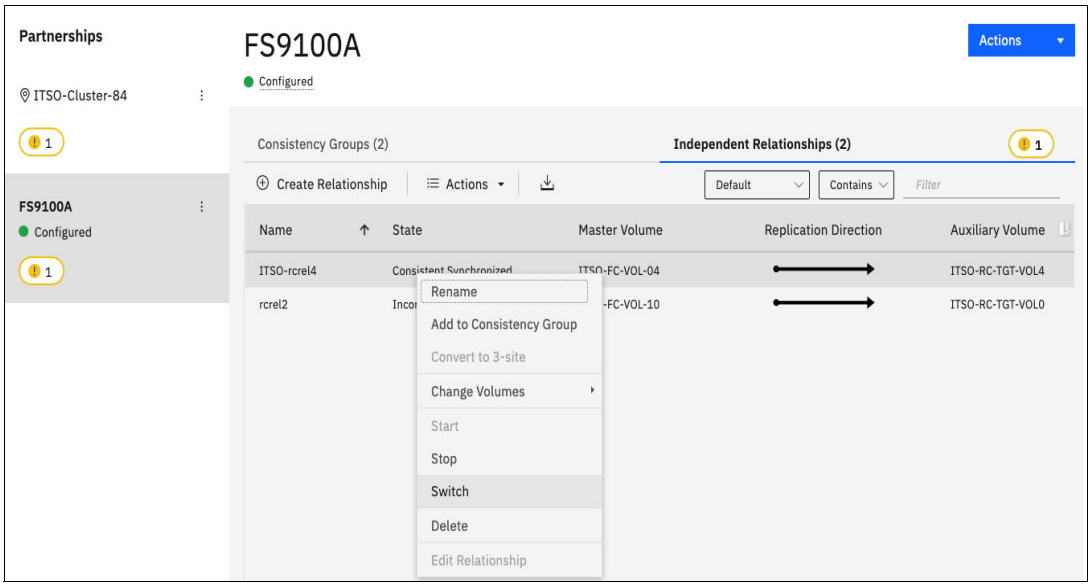


Figure 10-138 Switching Remote Copy relationship direction

3. Because the master-auxiliary relationship direction is reversed, write access is disabled on the new auxiliary volume (former master volume), whereas it is enabled on the new master volume (former auxiliary volume). A warning message is displayed, as shown in Figure 10-139. Click **Yes**.

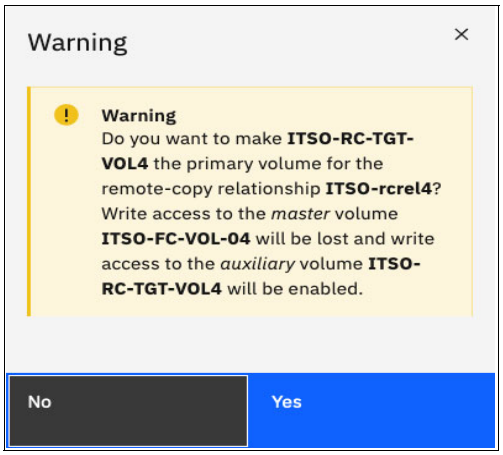


Figure 10-139 Switching master-auxiliary direction of relationships changes the write access

When an RC relationship is switched, an icon is displayed in the Remote Copy panel list, and the Replication Direction is also updated, as shown in Figure 10-140.

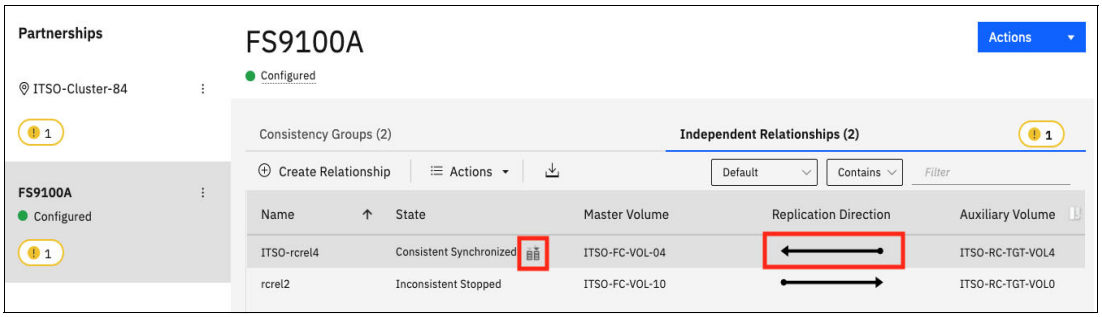


Figure 10-140 Switched Remote Copy Relationship

10.10.11 Switching a consistency group direction

When an RC consistency group is in the consistent synchronized state, the copy direction for the consistency group can be changed.

Important: When the copy direction is switched, it is crucial that no outstanding I/O exists to the volume that changes from primary to secondary because all of the I/O is inhibited to that volume when it becomes the secondary. Therefore, careful planning is required before you switch the copy direction for a relationship.

To switch the direction of an RC consistency group, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.

c
2. Select the target RC system and go to the **Consistency Groups** tab. Click the three dots on the wanted consistency group to be switched and select **Switch Direction**, as shown in Figure 10-141.

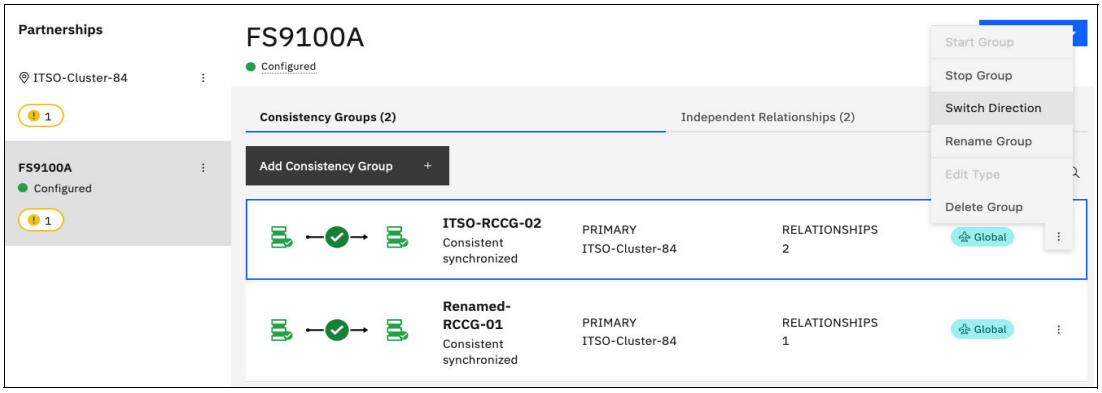


Figure 10-141 Switching a consistency group direction

3. Because the master-auxiliary relationship direction is reversed, write access is disabled on the new auxiliary volume (former master volume), while it is enabled on the new master volume (former auxiliary volume). A warning message is displayed, as shown in Figure 10-142. Click **Yes**.

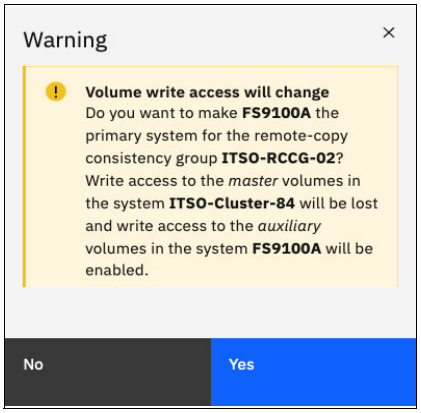


Figure 10-142 Switching direction of Consistency Groups changes the write access

10.10.12 Stopping Remote Copy relationships

When an RC relationship is created and started, the RC process can be stopped. Only relationships that are not members of a consistency group (or the entire consistency group) can be stopped.

To stop one or multiple relationships, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Select the target RC system and go to the **Independent Relationships** tab. Right-click the relationships to be stopped and select **Stop**, as shown in Figure 10-143.

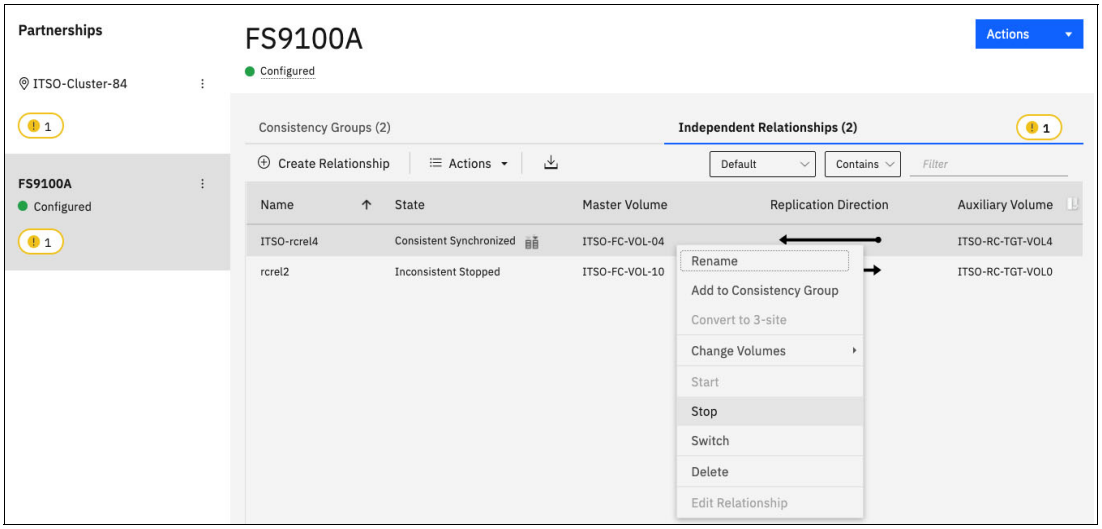


Figure 10-143 Stopping a Remote Copy relationship

3. When an RC relationship is stopped, access to the auxiliary volume can be changed so that it can be read and written by a host. A confirmation message is displayed, as shown in Figure 10-144.

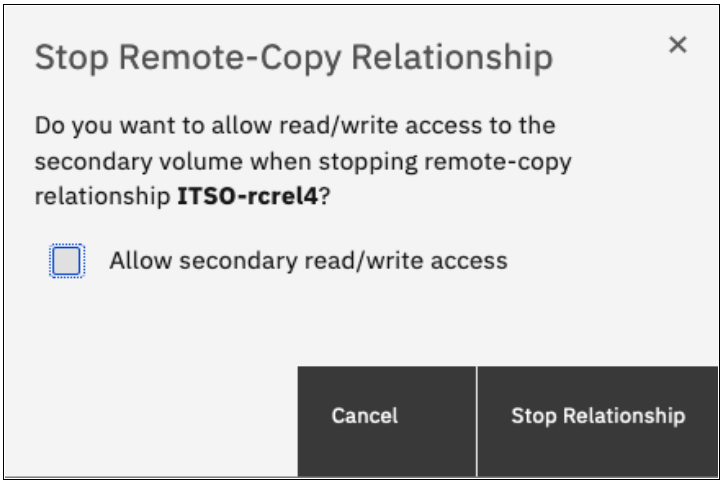


Figure 10-144 Grant access in read and write to the auxiliary volume

10.10.13 Stopping a consistency group

When an RC consistency group is created and started, the RC process can be stopped.

To stop a consistency group, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Select the target RC system and go to the **Consistency Groups** tab. Click the three dots on the consistency group to be stopped and select **Stop Group**, as shown in Figure 10-145.

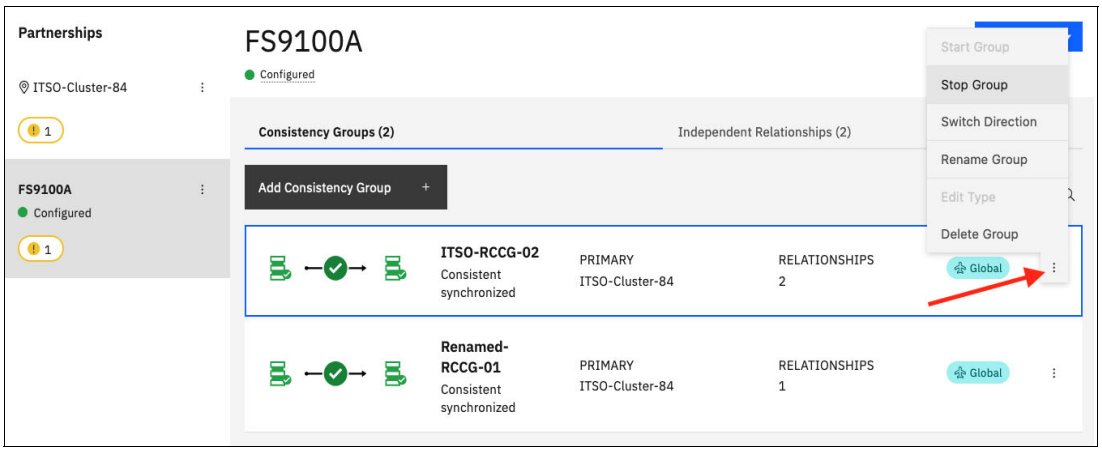


Figure 10-145 Stopping a consistency group

3. When an RC consistency group is stopped, access to the auxiliary volumes can be changed so it can be read and written by a host. A confirmation message is displayed, as shown in Figure 10-146.

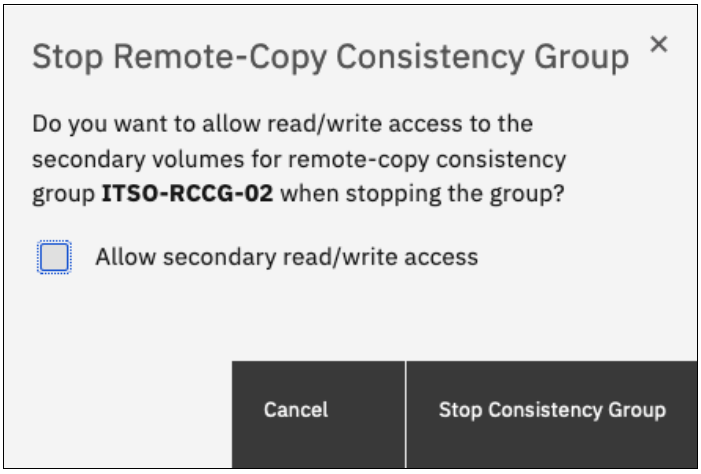


Figure 10-146 Grant access in read and write to the auxiliary volumes

10.10.14 Deleting Remote Copy relationships

To delete RC relationships, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Select the target RC system and go to the **Independent Relationships** tab. Next, right-click the relationships that you want to delete and select **Delete**, as shown in Figure 10-147.

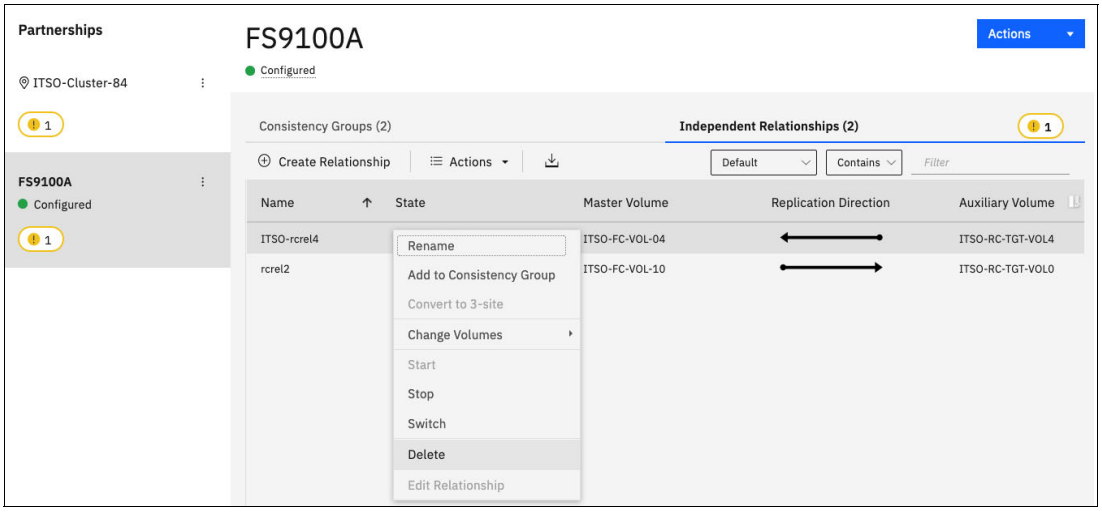


Figure 10-147 Deleting Remote Copy Relationships

3. A confirmation message is displayed that requests that the user enter the number of relationships to be deleted, as shown in Figure 10-148.

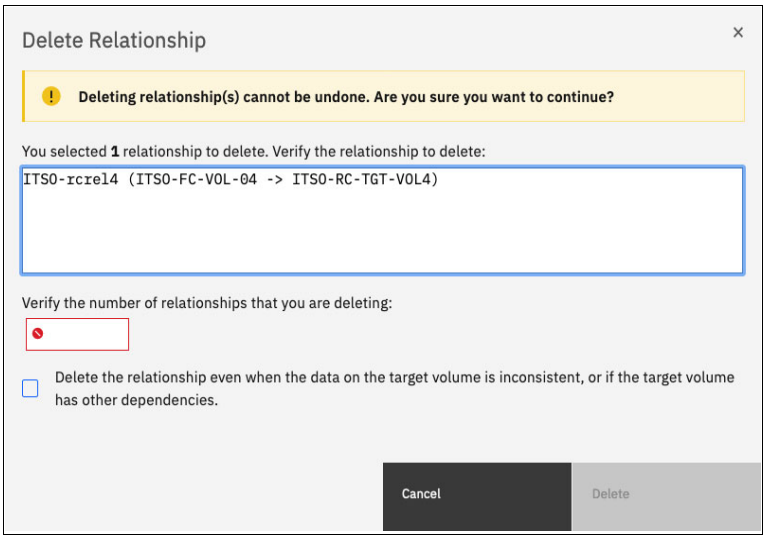


Figure 10-148 Confirmation of relationships deletion

10.10.15 Deleting a consistency group

To delete an RC consistency group, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Select the target RC system and go to the **Consistency Groups** tab. Click the three dots on the consistency group that you want to delete and select **Delete Group**, as shown in Figure 10-149.

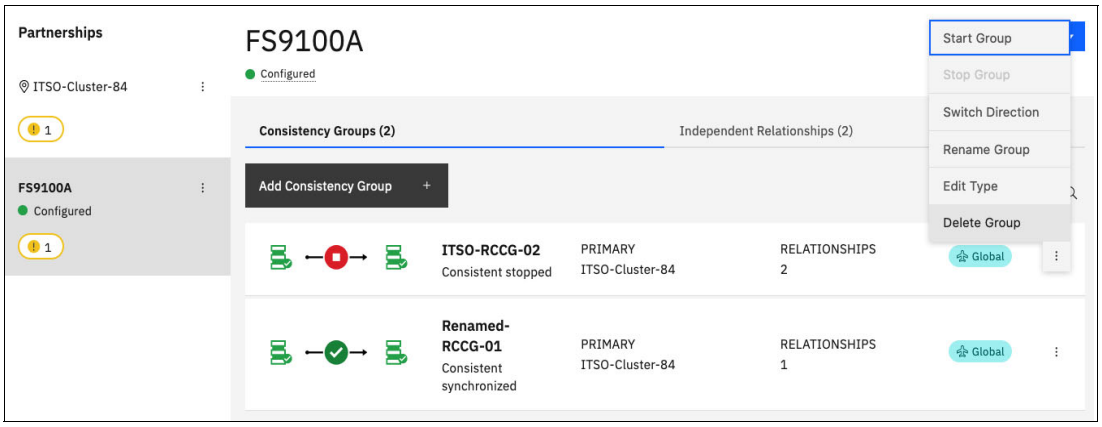


Figure 10-149 Deleting a consistency group

3. A confirmation message is displayed, as shown in Figure 10-150. Click **Yes**.

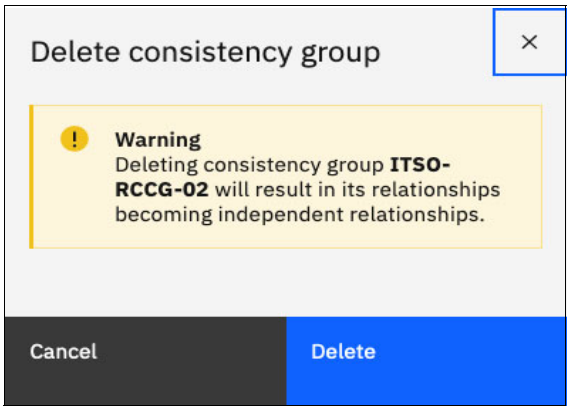


Figure 10-150 Confirmation of a consistency group deletion

Important: Deleting a consistency group does *not* delete its RC mappings.

10.11 Remote Copy memory allocation

Copy Services features require that small amounts of volume cache be converted from cache memory into bitmap memory to allow the functions to operate at an I/O group level. If you do not have enough bitmap space allocated when you try to use one of the functions, the configuration cannot be completed.

The total memory that can be dedicated to these functions is not defined by the physical memory in the system. The memory is constrained by the software functions that use the memory.

For every RC relationship that is created on an IBM Storage Virtualize system, a bitmap table is created to track the copied grains. By default, the system allocates 20 MiB of memory for a minimum of 2 TiB of remote copied source volume capacity. Every 1 MiB of memory provides the following volume capacity for the specified I/O group: for 256 KiB grains size, 2 TiB of total MM, GM, or active-active volume capacity.

Review Table 10-17 to calculate the memory requirements and confirm that your system can accommodate the total installation size.

Table 10-17 Memory allocation for Remote Copy services

Minimum allocated bitmap space	Default allocated bitmap space	Maximum allocated bitmap space	Minimum functionality when using the default values[1]
0	20 MiB	1024 MiB	40 TiB of remote mirroring volume capacity
[1]RC includes MM, GM, and active-active relationships.			

When you configure change volumes for use with GM, two internal FlashCopy mappings are created for each change volume.

Two bitmaps exist for MM, GM, and HyperSwap active-active relationships. For MM/GM relationships, one is used for the master clustered system and one is used for the auxiliary system because the direction of the relationship can be reversed. For active-active relationships, which are configured automatically when HyperSwap volumes are created, one bitmap is used for the volume copy on each site because the direction of these relationships can be reversed.

MM/GM relationships do not automatically increase the available bitmap space. You might need to run the **chiogrp** command to manually increase the space in one or both of the master and auxiliary systems.

You can modify the resource allocation for each I/O group of an IBM SAN Volume Controller system by selecting **Settings** → **System** and clicking the **Resources** menu, as shown in Figure 10-151. At the time of this writing, this GUI option is not available for other IBM Storage Virtualize based systems, so it can be adjusted via `chiogrp` CLI command. For more information about the syntax, see the IBM Documentation website.

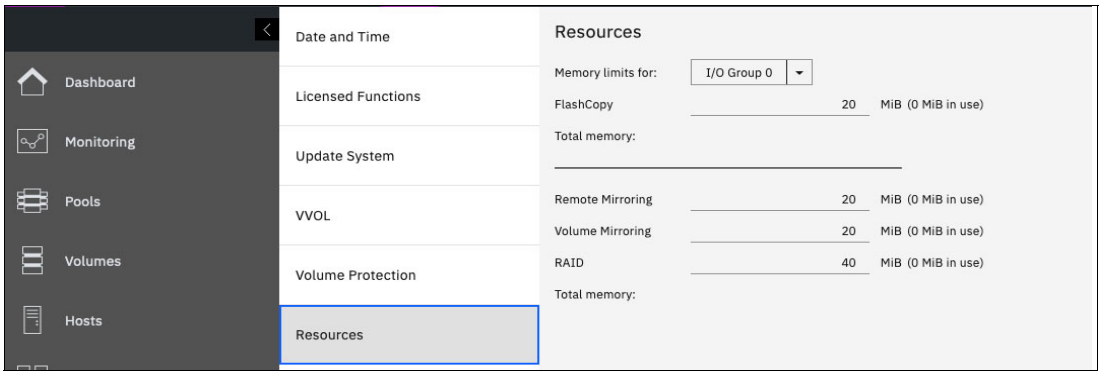


Figure 10-151 Modifying resources allocation

10.12 A real-life implementation of IP replication: Anadolu Sigorta

In this section we present a real-life implementation of IBM Storage Virtualize IP replication.

10.12.1 Customer profile

Anadolu Sigorta is Turkey's first national insurance company operating in all branches, which includes Fire, Transportation, Accident, Engineering, Agriculture, Legal Protection, Personal Accident, Health, and Credit, but excludes life insurance. Providing services since 1925, Anadolu Sigorta has approximately 2,500 professional agents in Turkey.

10.12.2 General information about the infrastructure

Anadolu Sigorta has 2 SVC clusters in the infrastructure operating at an average of 45,000 IOPS with a response time of 1-2 ms. There are more than 20 x86 and IBM Power servers, with over 1,000 virtual servers running on these systems.

Figure 10-152 on page 887 shows the current infrastructure.

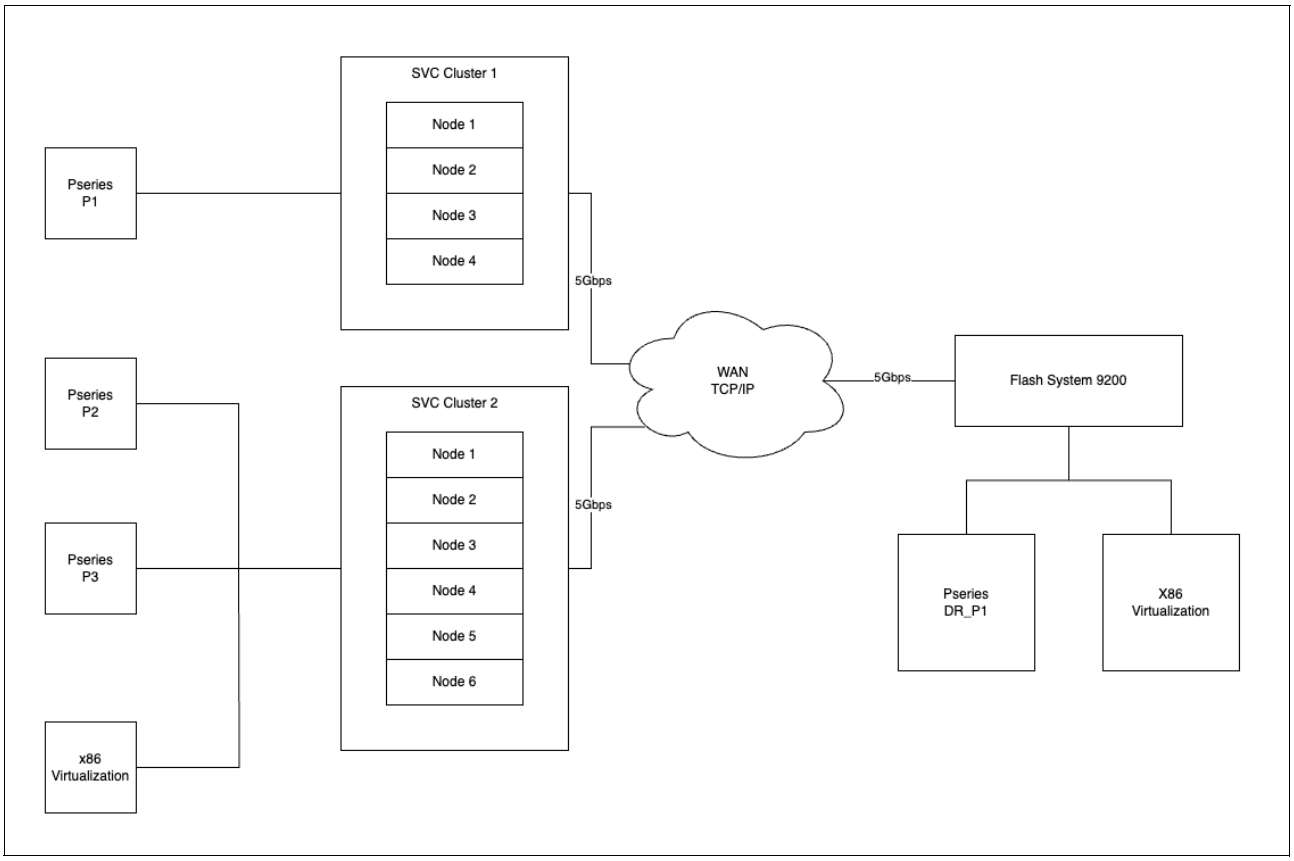


Figure 10-152 Current infrastructure

10.12.3 IP replication implementation

Anadolu Sigorta previously used Fibre Channel over IP (FCIP) routers to replicate data between their Production and Disaster Recovery (DR) centers. However, when the DR center was relocated to a different site, the FCIP devices were more than 5 years old and became too difficult to operate and monitor. As a result, Anadolu Sigorta decided to remove the FCIP devices and use the IP replication feature offered by the SVC devices and IBM Storage Virtualize. This simpler and more reliable solution allowed Anadolu Sigorta to achieve a more efficient disaster recovery process.

Anadolu Sigorta also used FCIP to replicate data between their Production and DR centers. However, when the DR center was upgraded, there were no FCIP devices in the new location. To address this issue, Anadolu Sigorta used IBM Storage Virtualize to move the LUNs from the old DR center to the new one. IBM Storage Virtualize allows for both FC and IP partnerships, so Anadolu Sigorta was able to establish an IP replication partnership between the Production and DR centers. This allowed for a smooth transition to the new DR center with minimal downtime.

10.12.4 Implementation decisions and results

While configuring the IP replication between the Production Data Center and the new Disaster Recovery Center, the implementation team made several changes to the replication infrastructure to improve performance and minimize costs:

1. They connected each node within a cluster to the network switch using a separate cable for replication. This ensured that the replication traffic would not use the same network path as the LUN traffic, which improved performance.
2. They evenly defined the preferred nodes for IO groups that have multiple replicated LUNs configured. This also improved performance by ensuring that the replication traffic was evenly distributed across the nodes.
3. In the old infrastructure, there were a total of 2 FCIP routers with a speed of 2 Gbps for 2 SVC clusters. They upgraded the network infrastructure to support a minimum speed of 4 Gbps per SVC cluster and 1 Gbps per node. This resulted in improved replication performance and minimized IO latency losses.
4. The team used the IBM Storage Virtualize code internally to handle secure data transmission and compression. This eliminated the need for an additional device to perform these functions through the FCIP router, which resulted in cost savings.

Figure 10-153 shows the IP replication solution that was implemented in Anadolu Sigorta.

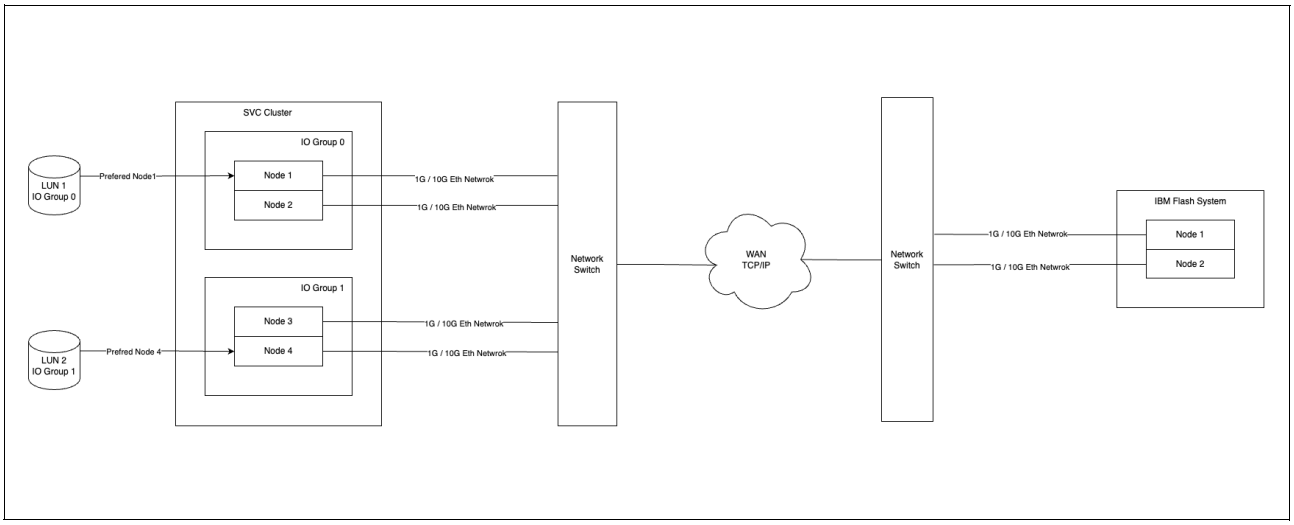


Figure 10-153 IP replication solution in Anadolu Sigorta

10.13 Troubleshooting Remote Copy

Remote Copy (MM and GM) features the following primary error codes:

- A 1920 error can be considered as a voluntary stop of a relationship by the system when it evaluates the replication causes errors on the hosts. A 1920 is a congestion error. This error means that the source, the link between the source and target, or the target cannot keep up with the requested copy rate. The system then triggers a 1920 error to prevent replication from having undesired effects on hosts.
- A 1720 error is a heartbeat or system partnership communication error. This error is often more serious because failing communication between your system partners involves extended diagnostic time.

10.13.1 1920 error

A 1920 error is deliberately generated by the system and is considered as a control mechanism. It occurs after 985003 (“Unable to find path to disk in the remote cluster (system)

within the time-out period") or 985004 ("Maximum replication delay has been exceeded") events. 985003 and 985004 are the IDs of entries in the Storage Virtualize event log. These events normally occur when Remote Copy encounters a problem.

This error can have several triggers, including the following probable causes:

- Primary system or SAN fabric problem (10%)
- Primary system or SAN fabric configuration (10%)
- Secondary system or SAN fabric problem (15%)
- Secondary system or SAN fabric configuration (25%)
- Intercluster link problem (15%)
- Intercluster link configuration (25%)

In practice, the most often overlooked cause is latency. GM has an RTT tolerance limit of 80 or 250 milliseconds, depending on the firmware version and the hardware model. A message that is sent from the source IBM Storage Virtualize system to the target system and the accompanying acknowledgment must have a total time of 80 or 250 milliseconds round trip. That is, it must have up to 40 or 125 milliseconds latency each way.

The primary component of your RTT is the physical distance between sites. For every 1000 kilometers (621.4 miles), you observe a 5-millisecond delay each way. This delay does not include the time that is added by equipment in the path. Every device adds a varying amount of time, depending on the device, but a good rule is 25 microseconds for pure hardware devices.

For software-based functions (such as compression that is implemented in applications), the added delay tends to be much higher (usually in the millisecond plus range.) The following is an example of a physical delay.

Company A has a production site that is 1900 kilometers (1180.6 miles) away from its recovery site. The network service provider uses a total of five devices to connect the two sites. In addition to those devices, Company A uses a SAN FC router at each site to provide FCIP to encapsulate the FC traffic between sites.

Now, there are seven devices and 1900 kilometers (1180.6 miles) of distance delay. All the devices are adding 200 microseconds of delay each way. The distance adds 9.5 milliseconds each way, for a total of 19 milliseconds. Combined with the device latency, the delay is 19.4 milliseconds of physical latency minimum, which is under the 80-millisecond limit of GM until you realize that this number is the best case number.

The link quality and bandwidth play a large role. Your network provider likely ensures a latency maximum on your network link. Therefore, be sure to stay as far beneath the GM RTT limit as possible. You can easily double or triple the expected physical latency with a lower quality or lower bandwidth network link. Then you are within the range of exceeding the limit if high I/O occurs that exceeds the bandwidth capacity.

When you get a 1920 event, always check the latency first. The FCIP routing layer can introduce latency if it is not properly configured. If your network provider reports a much lower latency, you might have a problem at your FCIP routing layer. Most FCIP routing devices have built-in tools to enable you to check the RTT. When you are checking latency, remember that TCP/IP routing devices (including FCIP routers) report RTT by using standard 64-byte ping packets.

Effective transit time must be measured only by using packets that are large enough to hold an FC frame, or 2148 bytes (2112 bytes of payload and 36 bytes of header). Allow estimated resource requirements to be a safe amount because various switch vendors have optional features that might increase this size. After you verify your latency by using the proper packet size, proceed with normal hardware troubleshooting.

Before proceeding, look at the second-largest component of your RTT, which is *serialization delay*. Serialization delay is the amount of time that is required to move a packet of data of a specific size across a network link of a certain bandwidth. The required time to move a specific amount of data decreases as the data transmission rate increases.

The amount of time in microseconds that is required to transmit a packet across network links of varying bandwidth capacity is compared. The following packet sizes are used:

- 64 bytes: The size of the common ping packet
- 1500 bytes: The size of the standard TCP/IP packet
- 2148 bytes: The size of an FC frame

Finally, your path (MTU) affects the delay that is incurred to get a packet from one location to another location. An MTU might cause fragmentation or be too large and cause too many retransmits when a packet is lost.

Note: Unlike 1720 errors, 1920 errors are deliberately generated by the system because it evaluated that a relationship can affect the host's response time. The system has no indication about if or when the relationship can be restarted. Therefore, the relationship cannot be restarted automatically and it must be done manually.

10.13.2 1720 error

The 1720 error (event ID 050020) is the other problem RC might encounter. The amount of bandwidth that is needed for system-to-system communications varies based on the number of nodes. It is important that it is not zero. When a partner on either side stops communication, a 1720 is displayed in your error log. According to the product documentation, there are no likely field-replaceable unit (FRU) breakages or other causes.

The source of this error is most often a fabric problem or a problem in the network path between your partners. When you receive this error, check your fabric configuration for zoning of more than one host bus adapter (HBA) port for each node per I/O group if your fabric has more than 64 HBA ports zoned. The suggested zoning configuration for fabrics is one port for each node per I/O group per fabric that is associated with the host.

For those fabrics with 64 or more host ports, this suggestion becomes a rule. Therefore, you see four paths to each volume discovered on the host because each host must have at least two FC ports from separate HBA cards, each in a separate fabric. On each fabric, each host FC port is zoned to two IBM Storage Virtualize node ports where each node port comes from a different IBM Storage Virtualize node. This configuration provides four paths per volume. Although more than four paths per volume are supported, it is not recommended.

Improper zoning can lead to SAN congestion, which can inhibit remote link communication intermittently. Checking the zero buffer credit timer and port send delay percentage with IBM Spectrum Control and comparing against your sample interval reveals potential SAN congestion. If a zero buffer credit or port send delay percentage is more than 2% of the total time of the sample interval, it might cause problems.

Next, always ask your network provider to check the status of the link. If the link is acceptable, watch for repeats of this error. It is possible in a normal and functional network setup to have occasional 1720 errors, but multiple occurrences might indicate a larger problem.

If you receive multiple 1720 errors, recheck your network connection and then check the system partnership information to verify its status and settings. Then perform diagnostics for every piece of equipment in the path between your two IBM Storage Virtualize systems. It often helps to have a diagram that shows the path of your replication from both logical and physical configuration viewpoints.

Note: With Consistency Protection enabled on the GM relationships, the system tries to resume the replication when possible. Therefore, it is not necessary to manually restart the failed relationship after a 1720 error is triggered.

If your investigations fail to resolve your RC problems, contact your IBM Support representative for a more complete analysis.

10.14 3-Site Replication

The IBM Storage Virtualize software not only provides disaster recovery capabilities between two sites by using Metro Mirror, Global Mirror, and Global Mirror with Change Volumes, but with release v8.3.1.1 came the ability to have another asynchronous copy on a third site with Metro Mirror. IBM Storage Virtualize software version 8.4.0.2 further enhanced this capability by introducing support for HyperSwap volumes. Figure 10-154 shows an overview of the setup.

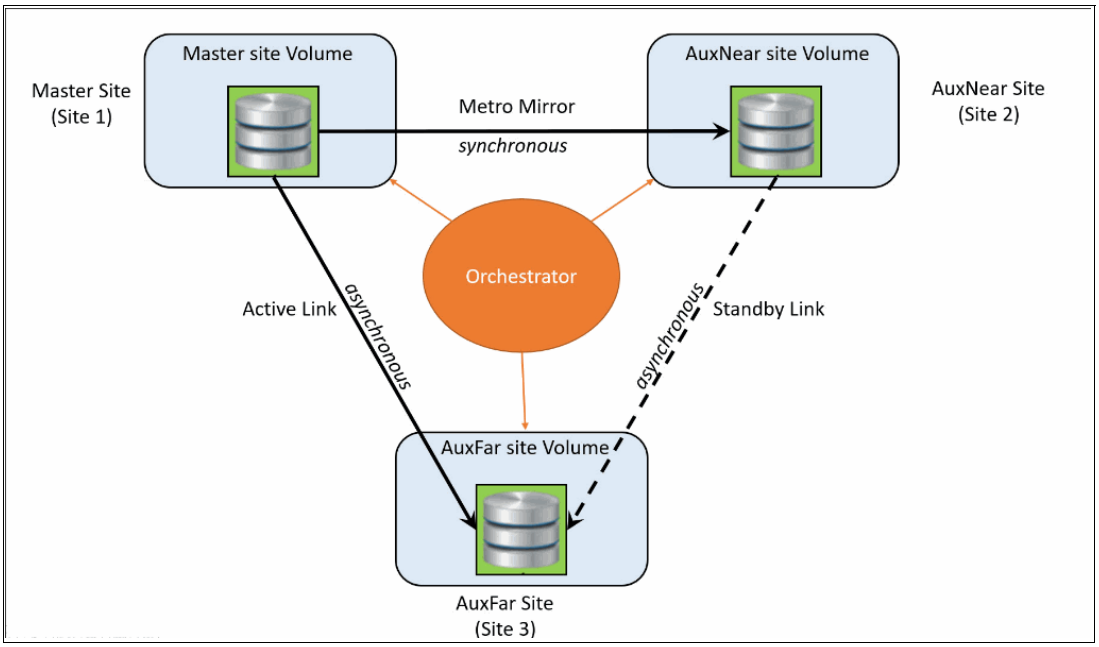


Figure 10-154 3-Site Replication overview

For more information about 3-site replication, see *IBM Spectrum Virtualize 3-Site Replication IBM Redbook*, SG24-8504.

10.14.1 3-Site Replication Orchestrator

The function of the Orchestrator is to coordinate the replication of data between three sites for disaster recovery and high availability scenarios for systems that are geographically dispersed.

The Orchestrator is an independent application that runs on a Linux platform. It configures and manages supported replication configurations on IBM Storage Virtualize products.

Version 4.0 of the Orchestrator introduced the following capabilities:

- Support to add one or more new relationships into an existing 3-site consistency group.
- Support to remove one or more relationships from an existing 3-site consistency group.
- Restricted support for reduced cycle time with the approval of a SCORE request.
- Introduction of a GUI.

To add a relationship to a 3-site consistency group, the following conditions must be met:

- The 3-site consistency group must not be in an error state and all three sites must be reachable.
- The 3-site consistency group that you are adding to must be in a stopped state.
- A state of PARTIAL is supported for roll forward only.
- A stand-alone relationship must be created between Master and Auxnear.
- All attributes of the newly created relationship (MM/HS) must match with that of the target 3-site consistency group.
- The state of the stand-alone relationships must be in `consistent_synchronized` for it to be added to the 3-site consistency group.

To add a stand-alone relationship to a 3-site consistency group, issue the commands in Example 10-5.

Example 10-5 Add stand-alone relationship to a 3-site consistency group

```
IBM_FlashSystem:FS950_YL:demoUser>convertrcrelationship -type 3site -pool 0
-iogrp 0 -consistgrpname cg_1 rcrel4
3-site RC relationship created with volume id [11].
```

The process flow for adding a replication relationship into a 3-way consistency group is shown in Figure 10-155.

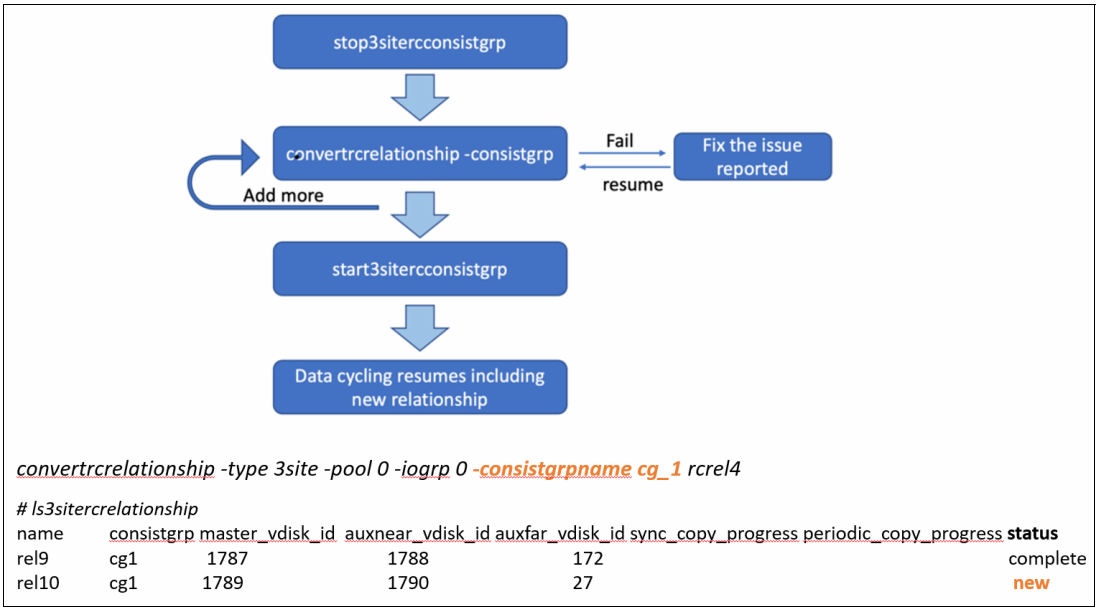


Figure 10-155 Add replication relationship to 3-site consistency group

The Orchestrator also allows for replication relationships to be removed from the 3-site consistency group after they are no longer needed. As with adding relationships, removing relationships also has prerequisites that need to be fulfilled, including the following examples:

- The 3-site consistency group must not be in an error state and all three sites must be reachable.
- The 3-site consistency group that you are adding to must be in a stopped state.
- A state of PARTIAL is supported only for a 3-site consistency group if a failure of **`addnewrelationship cli`** or **`convertrcrelationship`** with data or failure to remove a relationship from 3-site to 2-site occurs.

The process flow for removing a replication relationship from a 3-way consistency group is shown in Figure 10-156.

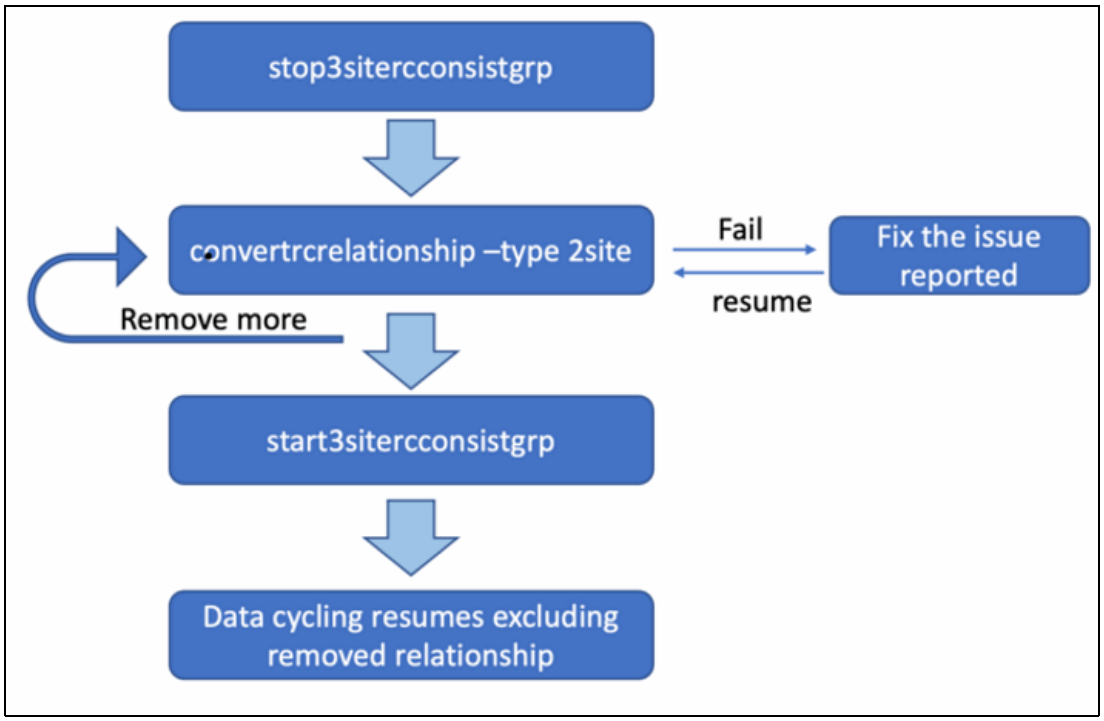


Figure 10-156 Process flow to remove a replication relationship from 3-site consistency group

The final enhancement to the Orchestrator in version 8.5 or later is the ability to reduce the cycle time of a relationship down to a minimum of 60 seconds. However, this reduction can be supported only by an approved SCORE request.

The limitation for this support is that only up to 5 relationships per consistency group are allowed.

The ability to set or change the cycle time is also included in the GUI.

For more information, see *IBM Spectrum Virtualize 3-Site Replication*, SG24-8504.

10.15 IBM Storage volume group and volume group snapshots

Volume groups refer to a collection or grouping of volumes that share common characteristics or are organized together for specific reasons. A volume group is a group of volumes. Volume groups can be formed based on various factors:

- Volumes with similar SLA requirements, such as performance targets, availability, or data protection policies can be grouped together within a volume group. This ensures that the volumes within the group are managed and treated according to the same service level guidelines.
- Some applications or data sets may require mutual consistency among multiple volumes. In such cases, these volumes can be grouped within a volume group to ensure that they are synchronized and maintain consistency in terms of data updates or access.

- Volumes residing on the same server can be grouped together within a volume group. This grouping facilitates efficient management and administration of the volumes within the server environment, which allows for streamlined operations and centralized control.

It is important to note that volume groups are distinct from consistency groups, although in some cases, the underlying system may use a consistency group concept internally when managing volume groups.

Create volume group using CLI

To create a volume group, follow these steps:

- Create volume group using **mkvolumegroup** command as shown in Example 10-6.

Example 10-6 svctask with mkvolumegroup flag

```
IBM_FlashSystem:FS9500_Demo:demoUser>svctask mkvolumegroup -name DB2_VG
Volume Group, id [0], successfully created
```

- Create and add Volumes to the newly created volume group. In this example we created three volumes with 50GiB each as shown in Example 10-7.

Example 10-7 svctask with addsnapshot flag

```
IBM_FlashSystem:FS9500_Demo:demoUser>svctask mkvolume -name DB2Prd_Vol0 -pool 0
-size 53687091200 -unit b -volumegroup 0
Volume, id [12], successfully created

IBM_FlashSystem:FS9500_Demo:demoUser>svctask mkvolume -name DB2Prd_Vol1 -pool 0
-size 53687091200 -unit b -volumegroup 0
Volume, id [13], successfully created

IBM_FlashSystem:FS9500_Demo:demoUser>svctask mkvolume -name DB2Prd_Vol2 -pool 0
-size 53687091200 -unit b -volumegroup 0
Volume, id [14], successfully created
```

A volume group (VG) is a collection of volumes that are subjected to simultaneous Flashcopy operations, providing a simpler approach to achieve this. With volume group Flashcopies, the handling of target volumes, Flashcopy mappings, and the necessary consistency group is automated, which removes the need for manual user intervention.

Each new Flashcopy of a volume group automatically manages the target volumes, mappings, and consistency group. There are different types of copies available within the volume group:

- Snapshots. These are immutable copies of the volumes that cannot be mounted by the host.
- Clones. Clones are created based on snapshots and can be mounted by the host. There are two kinds of clones available:
  - Thick clones. These involve a background copy operation followed by the deletion of the mapping.
  - Thin clones. These remain dependent on the Snapshot and do not involve the creation of a separate copy.

Volumes consist of volume snapshots. Each snapshot corresponds to a volume within the volume group. They are initiated simultaneously through an internal consistency group,

eliminating the need to provide target volumes, mappings, or a separate consistency group. A volume group snapshot represents an immutable and unmountable point-in-time copy of the volume group.

The snapshot volumes within a volume group snapshot cannot be directly mapped to a server. However, it is important to note that users with Administrator authority can delete these Snapshot volumes. Hence, this function does not replace the need for Safeguarded Copy. Also It is not possible to view the volumes or mappings within a volume group Snapshot unless using the command line interface (CLI) as explained in “Volume group snapshot volumes FlashCopy mapping” on page 898.

Internally, a volume snapshot consists of a FlashCopy mapping of the snapshot type, with copy rate and clean rate set to 0. Additionally, a separate volume acts as the target for the internal snapshot FlashCopy mapping. The Volume Snapshot uses the standard FlashCopy dependency chain and requires bitmap space similar to regular FlashCopy mappings.

Note: The volume group snapshot feature is compatible with a wide range of platforms, and offers efficient data management capabilities. However, there are exceptions where the volume group snapshot feature is not available, which includes the following models:

- IBM FlashSystem 5015
- IBM FlashSystem 5035
- IBM FlashSystem 5045

On these platforms, the traditional FlashCopy feature is still available, providing an alternative solution for volume group snapshot-related functionalities.

Starting from version 8.6.0, volume group snapshots in SVC (SAN Volume Controller) Enhanced Stretched Cluster (ESC) clusters include support for mirrored volumes. This enhancement enables SVC ESC clusters to participate in volume group snapshots.

In versions prior to 8.6.0, volume group snapshots were permitted only on non-mirrored volumes. However, the following changes are included in version 8.6.0:

- Snapshots can now be added to volume groups that include mirrored volumes. These snapshots will be automatically mirrored, ensuring data redundancy and protection across the mirrored volumes within the volume group.
- This enhancement extends the support of volume group snapshots to SVC Enhanced Stretched Cluster environments, which allows for the creation and management of snapshots in such configurations.
- Mirrored copies can be added to volumes that already have snapshots. However, it is important to note that in order for the mirrored copies to be associated with the existing snapshots, they must be manually added to the snapshots first.

Mirrored snapshots are compatible with both the Safeguarded Copy feature and the internal scheduler, enabling seamless integration with these functions.

10.15.1 Volume group snapshots

Before the introduction of volume group snapshots, when volumes were dependent on each other, creating point-in-time copies required taking snapshots at the exact same time. This was achieved through the use of consistency groups (CG), which consisted of a group of mappings that had to be started simultaneously. Configuring the mappings and targets for Flashcopies within a consistency group was a complex process that had to be repeated each

time a new point-in-time copy was created. This complexity posed limitations on usage and made it challenging to incorporate new functions.

The purpose of the volume group snapshot management model is to simplify the implementation of standard FlashCopy operations. It achieves this by offering a more straightforward setup process and separating the snapshot and clone features. With volume group snapshots, administrators can create snapshots of volume groups with more ease and efficiency, without the need for complex consistency group configurations.

Demonstration videos: The following demonstration videos are available:

- IBM Storage Virtualize V8.6: Handling Snapshots using the GUI.
- IBM Storage Virtualize V8.6: Handling snapshots using the command line interface.

Table 10-18 lists the snapshot volumes by pool type.

Table 10-18 Snapshot volumes by pool type

Standard pool	DRP pool
► The snapshot volume is thinly provisioned. ► It has the auto-expand feature enabled, ensuring it can dynamically grow as needed. ► The default grain size for the snapshot volume is set to 256KB, providing efficient use of storage.	► The snapshot volume is at least thinly provisioned, offering storage efficiency. ► If the underlying volume is located on FCMs, it will be compressed, reducing storage space requirements while maintaining performance. ► If the source data has been deduplicated, the snapshot volume will also benefit from deduplication, further optimizing storage consumption and maximizing data reduction.

Within a system, storage resources are organized into volumes. These volumes are host-mappable and can be directly accessed by the host system for data operations. This seamless integration simplifies data management because all available volumes are readily usable by the host.

Volume group snapshot volume FlashCopy mapping limitations

There are certain limitations associated with a volume group snapshot FlashCopy mapping:

- The FlashCopy Mapping cannot be directly initiated or prepared. Instead, it is managed implicitly through the volume group snapshot process.
- Similarly, the FlashCopy mapping cannot be directly stopped or deleted. It is controlled by the volume group snapshot process, which manages the associated operations.

Triggering volume group snapshot volumes mappings

The process for triggering volume group snapshot volumes mappings involves a streamlined version of the FlashCopy mapping trigger process. To ensure data consistency and integrity during the Snapshot operation, these mappings are added to an internal consistency group (ICG). The internal consistency group allows for simultaneous management and synchronization of the snapshot mappings. Here's an overview of the triggering process:

- When you initiate the volume group snapshot volumes mappings, the mappings are first included in an internal consistency group. The internal consistency group acts as a container that enables coordinated operations among the snapshot mappings.

- After the mappings are added to the internal consistency group, the group is activated, and the associated snapshot mappings begin their operation simultaneously. This step ensures that all the volumes within the group are in a consistent state during the snapshot process.
- After the snapshot operation is completed, the volumes are taken out of the internal consistency group. This step finalizes the snapshot process, and the mappings return to their independent states.

The triggering of volume group snapshot volumes mappings is done by using the **addsnapshot** command, which initiates the entire process as outlined above as shown in Example 10-8.

Example 10-8 svctask with addsnapshot flag

```
IBM_FlashSystem:FS9500_Demo:demoUser>svctask addsnapshot -volumegroup 0
Snapshot, id [3], successfully created or triggered
```

Volume group snapshot volumes FlashCopy mapping

A volume group snapshot volumes FlashCopy mapping refers to a special configuration that is hidden from regular view, but it can be accessed and examined by using the `--showhidden` flag on the Command Line Interface (CLI), as shown in Example 10-9.

Example 10-9 lsvdisk with showhidden flag

```
IBM_FlashSystem:FS9500_Demo:demoUser>lsvdisk
id name IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity type
FC_id FC_name RC_id RC_name vdisk_UID fc_map_count copy_count fast_write_state
se_copy_count RC_change compressed_copy_count parent_mdisk_grp_id
parent_mdisk_grp_name owner_id owner_name formatting encrypt volume_id volume_name
function volume_group_id volume_group_name protocol is_snapshot snapshot_count
volume_type replication_mode is_safeguarded_snapshot safeguarded_snapshot_count
12 DB2Prd_Vol0 0 io_grp0 online 0 DRP_Pool 50.00GB  striped
6005076813868004E80000000000067D 0 1 not_empty 0 no 0 0 DRP_Pool
yes no 12 DB2Prd_Vol0 0 DB2_VG no 1 no 0
13 DB2Prd_Vol1 0 io_grp0 online 0 DRP_Pool 50.00GB  striped
6005076813868004E80000000000067E 0 1 not_empty 0 no 0 0 DRP_Pool
yes no 13 DB2Prd_Vol1 0 DB2_VG no 1 no 0
14 DB2Prd_Vol2 0 io_grp0 online 0 DRP_Pool 50.00GB  striped
6005076813868004E80000000000067F 0 1 not_empty 0 no 0 0 DRP_Pool
yes no 14 DB2Prd_Vol2 0 DB2_VG no 1 no 0

IBM_FlashSystem:FS9500_Demo:demoUser>lsvdisk -showhidden
id name IO_group_id IO_group_name status  mdisk_grp_id mdisk_grp_name capacity
type FC_id FC_name RC_id RC_name vdisk_UID fc_map_count copy_count
fast_write_state se_copy_count RC_change compressed_copy_count parent_mdisk_grp_id
parent_mdisk_grp_name owner_id owner_name formatting encrypt volume_id volume_name
function volume_group_id volume_group_name protocol is_snapshot snapshot_count
volume_type replication_mode host_mappable is_safeguarded_snapshot
safeguarded_snapshot_count
12 DB2Prd_Vol0 0 io_grp0 online 0 DRP_Pool 50.00GB striped
6005076813868004E80000000000067D 0 1 not_empty 0 no 0 0 DRP_Pool
yes no 12 DB2Prd_Vol0 0 DB2_VG no 1 yes no 0
13 DB2Prd_Vol1 0 io_grp0 online  0 DRP_Pool 50.00GB striped
6005076813868004E80000000000067E 0 1 not_empty 0 no 0 0 DRP_Pool
yes no 13 DB2Prd_Vol1 0 DB2_VG no 1 yes no 0
```

```
14 DB2Prd_Vol2 0 io_grp0 online  0 DRP_Pool 50.00GB striped
6005076813868004E80000000000067F 0 1 not_empty 0 no 0 0 DRP_Pool
yes no 14 DB2Prd_Vol2 0 DB2_VG no 1 yes no 0
64997 vdisk2 0 io_grp0 offline 0 DRP_Pool 50.00GB striped
0 1 not_empty 1 no 0 0 DRP_Pool no no 64997 vdisk2 yes 0 no no 0
64998 vdisk1 0 io_grp0 offline 0 DRP_Pool 50.00GB striped
0 1 not_empty 1 no 0 0 DRP_Pool no no 64998 vdisk1 yes 0 no no 0
64999 vdisk0 0 io_grp0 offline 0 DRP_Pool 50.00GB  striped
0 1 not_empty 1 no 0 0 DRP_Pool no no 64999 vdisk0 yes 0 no no 0
IBM_FlashSystem:FS9500_Demo:demoUser>
```

A FlashCopy mapping of a volume group snapshot includes the following properties:

- Zero CopyRate. This allows for efficient data copying without consuming host resources.
- Zero Clean Rate. Similar to Zero CopyRate, Zero Clean Rate enhances the FlashCopy efficiency by minimizing the need for clean operations.
- AutoDelete off. By default, the AutoDelete feature is disabled, ensuring that the FlashCopy Mapping remains intact even after a FlashCopy operation has been completed.
- Incremental off. Incremental FlashCopy, which only copies changed data between FlashCopy sessions, is disabled by default for the volume group snapshot FlashCopy mapping.
- Grain Size 256k. The Grain Size is set to 256k, which provides an optimal balance between performance and storage efficiency during FlashCopy operations.

Note: Only the Clean Rate can be modified to optimize clean operations if necessary. By default, the Clean Rate is set to 0.

Volume group snapshot availability

Ensuring the availability and protection of volume group snapshot FlashCopy mappings is a critical aspect of data management. To achieve this, a snapshot FlashCopy mapping is marked with the attribute **keeptarget**.

When a snapshot FlashCopy mapping is marked with the **keeptarget** attribute, it indicates that the target volume, which is the snapshot volume, is to be preserved. This means that if the snapshot volume faces offline conditions due to insufficient space, the system will take proactive measures to safeguard the integrity of the snapshot.

If the snapshot volume becomes unavailable due to space constraints, the system will automatically take the parent volume offline to ensure that the snapshot is protected. This offline state of the parent volume is a safeguarding mechanism, preserving the integrity of the snapshot and preventing potential data corruption.

The offline state of the parent volume applies when the parent volume is thinly provisioned or using compression or deduplication features. These data reduction techniques contribute to the efficient use of storage space, but they require careful management to avoid potential data loss.

Removing volume group snapshots

The process of removing volume group snapshots involves asynchronous operations for both the snapshot mapping and the associated snapshot volume. Here are some key points to understand about the removal process:

- When a snapshot is marked for deletion, the removal of both the snapshot mapping and the snapshot volume occurs asynchronously. This means that these operations do not happen simultaneously but are executed independently and in the background.
- By default, snapshots that are in the process of being deleted will be hidden from regular view. This is done to prevent any accidental interference or modifications during the deletion process.
- If a snapshot is being used as the basis for creating a dependent clone or thin clone volume, it will remain visible even while being marked for deletion. This is because the dependent volume relies on the snapshot data for its creation and maintenance.
- When the dependent clone or thin clone no longer requires the Snapshot data, the Snapshot will be automatically removed. This ensures that Snapshots are deleted only when they are no longer needed to maintain data consistency.

To remove a volume group snapshot use the following command, as shown in Example 10-10.

Example 10-10 svctask with rmsnapshot flag

```
IBM_FlashSystem:FS9500_Demo:demoUser>svctask rmsnapshot -snapshot snapshot2
-volumegroup 0
```

Volume group snapshot restrictions

The use of volume group snapshots includes the following restrictions:

- Mirroring VDisks were not supported prior to IBM Storage Virtualize version 8.6.0:
  - In the initial release, users cannot add a Snapshot to a volume that has multiple volume copies (mirroring).
  - However, there is an exception for migration scenarios. If one of the mirrored copies is marked for auto-delete, then it becomes possible to add a snapshot volume copy. This allows for specific use cases where migration and Snapshot functions can coexist.
- Regular FlashCopy and volume group snapshots on the same volume is *not* supported:
  - A volume group snapshot can be added to a volume that serves as the source for a regular FlashCopy mapping.
  - The volume group snapshot can exist within the same dependency chain as the regular mapping, ensuring data consistency.
  - However, adding a volume group snapshot to a volume that acts as the target of a regular FlashCopy mapping is not supported. This restriction prevents potential conflicts and ensures the stability of data operations.
  - When a volume group snapshot is added to a volume that is part of a regular FlashCopy mapping, no new regular mappings can be created or started for that specific volume. This limitation helps facilitate the smooth transition from regular FlashCopy to volume group snapshots, promoting efficient data management and migration.

10.15.2 Volume group thin clones

The volume group thin clone feature enables the creation of a thin clone of a volume group by specifying a source volume group and one of its snapshots. Here are the key points to understand about volume group thin clones:

- By initiating a volume group thin clone, all the necessary thin-clone volumes and related objects are created. These volumes serve as the clones and are prepopulated with data from the selected snapshot.
- Although volume group thin-clones can be assigned a target pool, this assignment only occurs at the group level, which limits the selection to existing pools.
- It is possible to specify a target pool for the volume group of the thin clone.
- The volume thin clone is a host mappable volume that provides access to the snapshot image of the source volume. Storage space is allocated on-demand. The properties of the volume thin clone are similar to those of a snapshot.
- The volume thin clone inherits various characteristics from the source volume group, including the ability to have volume copies, be part of remote copy relationships, and have its own snapshots.
- The volume thin clone cannot be part of a regular FlashCopy mapping. This restriction ensures the independent nature of the thin clone and avoids conflicts with other data operations.
- The volume thin clone can be expanded, allowing for increased storage capacity as needed. Additionally, it can be shrunk, provided that the size does not go below its original size. This flexibility allows for efficient management of storage resources.

To create a volume group thin clone use the following command, as shown in Example 10-11.

Example 10-11 Create thin clone

```
IBM_FlashSystem:FS9500_Demo:demoUser>svctask mkvolumegroup -fromsourcegroup 0
-iogroup 0 -name DB2VG -pool 0 -snapshot snapshot2 -type thinclone
Volume Group, id [2], successfully created
```

10.15.3 Volume group thick clones

The volume group thick clone feature allows for the creation of a thick clone of a volume group by specifying a source volume group and one of its snapshots. Here are the key aspects of volume group thick clones:

- When you initiate a volume group thick clone, all the necessary FlashCopy mappings and targets are created. These mappings and targets facilitate the prepopulation of the clone with data from the specified Snapshot and ensure an accurate copy.
- Although volume group thin-clones can be assigned a target pool, this assignment only occurs at the group level, which limits the selection to existing pools.
- The volume that is created for a volume clone is a host mappable volume. Its properties will match those of the source volumes, ensuring consistency and compatibility within the volume group.
- The volume clone inherits various characteristics from the source volume group, including the ability to have volume copies, be part of remote copy relationships, and have its own snapshots.
- The volume clone cannot be part of a regular FlashCopy mapping until the cloning process is complete and the clone has been detached from the source Snapshot. This ensures data integrity and avoids conflicts during the cloning process.
- The volume clone can be expanded, if needed. Additionally, it can be shrunk, provided that the size does not go below its original size. This flexibility allows for the efficient use of storage resources.

To create a volume group thick clone use the following command, as shown in Example 10-12.

Example 10-12 Create Thick Clone

```
IBM_FlashSystem:FS9500_Demo:demoUser>svctask mkvolumegroup -fromsourcegroup 0
-iogroup 0 -name DB2_Preprod1VG -pool 0 -snapshot snapshot2 -type clone
Volume Group, id [4], successfully created
```

10.15.4 Orphaned volume group snapshots

In certain cases, it might not be feasible or logical to organize volumes into traditional volume groups. Especially because a volume cannot belong to more than one volume group. Orphaned snapshots address this scenario. Orhpaned snapshots are not associated with a specific volume group but still serve as standalone snapshots of individual volumes. Here are some key points to understand about orphaned volume group snapshots:

- Orphaned snapshots are created when the system, or external automation, receives a list of volumes instead of a volume group as input for the *addsnapshot* task. The system treats these specified volumes as if they were part of a volume group and creates an orphaned volume group snapshot.
- Orphaned snapshots possess the same snapshot functionality as regular volume group snapshots. They capture a point-in-time image of the specified volumes, which enables data recovery, backup, or other snapshot-related operations.
- Unlike traditional volume group snapshots associated with specific volume groups, orphaned snapshots exist independently and are not bound by the constraints of a volume group. This flexibility allows for the capture of snapshots from volumes that cannot be organized into volume groups due to limitations or specific requirements.
- Orphaned snapshots can be managed as stand-alone entities, providing flexibility in snapshot administration and maintenance. They can be retained, deleted, restored, or used for various data management purposes.

To create a volume group snapshot for orphaned volumes use the following command, as shown in Example 10-10 on page 900.

Example 10-13 Add snapshot for orphaned volume group

```
IBM_FlashSystem:FS9500_Demo:demoUser>svctask addsnapshot -volumes
SAP_Volume0:SAP_Volume1:SAP_Volume2
Snapshot, id [5], successfully created or triggered
```

10.15.5 Volume group snapshot scheduler

The volume group snapshot scheduler is designed to automate the creation and deletion of snapshots based on predefined schedules and policies, which eliminates the need for external applications. It offers several features and benefits:

- The scheduler automates the process of creating and deleting snapshots and follows the specified schedule defined by the policy. This streamlines the snapshot management process and reduces manual intervention.
- The scheduler is compatible with the latest release of Copy Services Manager (CSM). If desired, CSM can be used to enable advanced functions for snapshot management.

- The snapshots created by the VG Snapshot Scheduler are crash-consistent, ensuring data integrity in the event of unexpected failures or system crashes.
- For enhanced application consistency, the latest release of CSM can call scripts, so you can have application-consistent snapshots, if needed.
- It is important to note that the scheduler is specifically designed for volume group snapshots and *cannot* be used for legacy FlashCopy or SGC volume group snapshot policies.
- The scheduler is associated with a specific volume group and operates independently from other snapshot policies.
- The volume group snapshot scheduler allows the definition of various snapshot policies, and determines the frequency of snapshot creation and the duration for which snapshots are retained.
- Users can choose from predefined snapshot policies or create custom policies tailored to their specific needs.
- The snapshot policies configured with the volume group snapshot scheduler are reusable and can be applied to multiple volume groups as needed.
- The scheduler supports up to 32 snapshot policies to provide flexibility for different snapshot management requirements.
- The volume group snapshot scheduler comes with three default snapshot policy parameters, offering convenient options for most use cases.
- Users can specify the creation frequency of snapshots in minutes, hours, weeks, days, or months. The minimum creation frequency allowed is 60 minutes. Additionally, the retention of snapshots can be specified in terms of days.
- Administrators can define a specific start time for the snapshot creation process using the format: YYMMDDHHMM.

The volume group snapshot scheduler simplifies and enhances snapshot management for volume groups, providing an efficient and automated solution for creating and retaining snapshots as per predefined policies.

Volume group policies and SafeGuarded Copy policies are distinct and separate in their function and implementation. Each type of policy serves a unique purpose within the storage environment and offers different features to address specific data management needs. Table 10-19 provides a comparison table highlighting the differences between volume group and SGC (SafeGuarded Copy) snapshot policies:

Table 10-19 Comparison between volume group snapshot and SafeGuarded Copy snapshot

Aspect	**Volume group snapshot policies**	**SafeGuarded Copy (SGC) snapshot policies**
Policy usage	Can be used in volume groups	Exclusively for SafeGuarded Copies
Predefined snapshot policies	Available	Available
Custom policies	Supported	Supported
Reusable across volume groups	Yes	Yes

Aspect	Volume group snapshot policies	SafeGuarded Copy (SGC) snapshot policies
Snapshot location	Volume group snaps in the source pool or a child pool (default is source volume's pool)	SGC snaps must go to a SafeGuarded child pool snapshot
Deletion	Can be deleted by an Admin user	Can be deleted by only a Secure Admin user
Clones	Available	Not available
Restore	Not available	Available
Thin clones	Similar to SGC recovery volume	Not available

To create a snapshot policy use the command shown in Example 10-14.

Example 10-14 Create snapshot policy

```
IBM_FlashSystem:FS9500_Demo:demoUser>svctask mksnapshotpolicy -backupinterval 4
-backupunit hour -backupstarttime 2307200800 -retentiondays 30
Snapshot Policy, id [6], successfully created

IBM_FlashSystem:FS9500_Demo:demoUser>lssnapshotschedule
policy_id policy_name schedule_id backup_unit backup_interval backup_start_time
retention_days
0         predefinedsspolicy0 1 hour 6 210103000000 7
1         predefinedsspolicy1 1 week 1 210103000000 30
2         predefinedsspolicy2 1 month 1 210103000000 365
3         vols-SGC            1 hour 1 221225102800 1
4         vol5policy          1 hour 1 221225105600 1
5         daily 12            1 hour 1 221225164500 1
6         sspolicy0           1 hour 4 230720080000 30
```

To list snapshot policies use the command shown in Example 10-15.

Example 10-15 List snapshot policy

```
IBM_FlashSystem:FS9500_Demo:demoUser>lssnapshotpolicy
id name volume_group_count
0  predefinedsspolicy0 0
1  predefinedsspolicy1 0
2  predefinedsspolicy2 0
3  vols-SGC            0
4  vol5policy          0
5  daily 12            0
6  sspolicy0           0
```

To assign policy to a volume group snapshot use the command shown in Example 10-10 on page 900.

Example 10-16 Assign policy to a volume group

```
IBM_FlashSystem:FS9500_Demo:demoUser>chvolumegroup -snapshotpolicy 6 0
```

```
IBM_FlashSystem:FS9500_Demo:demoUser>lssnapshotpolicy
id name volume_group_count
0  predefinedsspolicy0 0
1  predefinedsspolicy1 0
2  predefinedsspolicy2 0
3  vols-SGC            0
4  vol5policy          0
5  daily 12            0
6  sspolicy0           1
```

10.16 IBM Storage policy-based replication

Policy-based replication can help the setup, administration, and oversight of replication by employing volume groups and replication policies. This approach offers a simplified means of configuring, managing, and monitoring replication between two systems.

Policy-based replication brings several key advantages to asynchronous replication:

- Policy-based replication uses volume groups to ensure that all volumes are replicated based on the assigned policy.
- By eliminating the need to manage relationships and change volumes manually, policy-based replication can help simplify the overall administration process.
- The feature automatically manages provisioning on the remote system, reducing the administrative burden.
- During a site failover, policy-based replication can enable easier visualization of the replication process, which can facilitate effective management and decision-making.
- IBM Storage Virtualize version 8.5.2 and later provides asynchronous replication over Fibre Channel and IP partnerships.
- Automatic notifications are generated when the recovery point objective (RPO) exceeds the specified threshold, ensuring timely awareness of any deviations.
- Policy-based replication provides easy-to-understand status updates and alerts regarding the overall health of the replication process, enhancing monitoring and troubleshooting capabilities.

Table 10-20 highlights comparison and contrast between Global Mirror and policy-based replication (PBR).

Table 10-20 Comparison and contrast between Global Mirror and policy-based replication (PBR)

Global Mirror (GM)	Policy-based replication (PBR)
It matches or exceeds Global Mirror with Change Volumes in all bench marks.	Policy-based replication significantly outperforms Global Mirror in all bench marks tested.
Global Mirror reactively suspends replication during sustained performance problems.	Policy-based replication proactively avoids causing production performance problems by switching modes.

Global Mirror (GM)	Policy-based replication (PBR)
Remote Copy allows relationships to be created for volumes in all I/O groups in the system, but spans failure domains and has significantly reduced performance.	Policy-based replication requires that all volumes for a volume group be created in the same I/O group and volumes, with a maximum of two I/O groups replicating per system.
Remote Copy requires manual actions on both systems to configure volumes, relationships and consistency groups.	Policy-based replication supports provisioning while disconnected and uses policies to automate configuration.

The use of policy-based replication enables efficient data replication between systems, with minimal manual intervention. This method yields higher throughput and reduced latency compared to the remote-copy function.

A replication policy possesses the following characteristics:

- A replication policy can be assigned to one or more volume groups.
- After they are created, replication policies cannot be modified. If changes are necessary, a new policy can be defined and assigned to the relevant volume group.
- Each system can accommodate a maximum of 32 replication policies.

Note: Policy-based replication is supported in version 8.5.2 and later on the following products: IBM SAN Volume Controller, IBM FlashSystem 9500, IBM FlashSystem 9200, IBM FlashSystem 9100, IBM FlashSystem 7300, IBM FlashSystem 7200, IBM FlashSystem 5200 (requires a minimum of 128 GiB memory in each node canister), and IBM Storage Virtualize for Public Cloud.

For more information, see *Policy-based replication with Policy-based replication with IBM Storage FlashSystem, IBM SAN Volume Controller and IBM Storage Virtualize*, REDP-5704.

Demonstration video: For a demonstration video, see IBM Storage Virtualize V8.6: Policy Based Replication.

Replication approaches

There are two approaches to replication: journaling and snapshot-based. Journaling is used by GM, and GMCV uses snapshots.

- Snapshot-based replication involves periodically capturing and transmitting a snapshot. This method can skip intermediate writes, reducing bandwidth requirements. However, it requires periodic freezing to ensure consistent data collection, especially in clustered systems. It also requires additional storage to maintain snapshots, so the ability to maintain extremely low recovery point objectives (RPO) can be challenging. This approach is used by Storage Virtualize Global Mirror with Change Volumes (GMCV) in cycling mode.
- Journaling-based replication involves sequencing writes and replaying them in order. This method requires additional work for every write, including sequencing, authorization, and hardening. Special considerations are needed for overlapping writes to guarantee proper ordering. Each write is sent individually and is dependent on previous writes. The RPO can be close to half the round-trip time (RTT). This approach is used by Storage Virtualize Global Mirror and is very sensitive to delays. It operates in streaming/journaling mode.

Journaling mode

In journaling mode, writes are stored in a non-mirrored, volatile, in-memory journal. This allows for buffering of several seconds or even minutes of writes, which provides a buffer to absorb any replication performance issues without impacting the host as shown in Figure 10-172 on page 924.

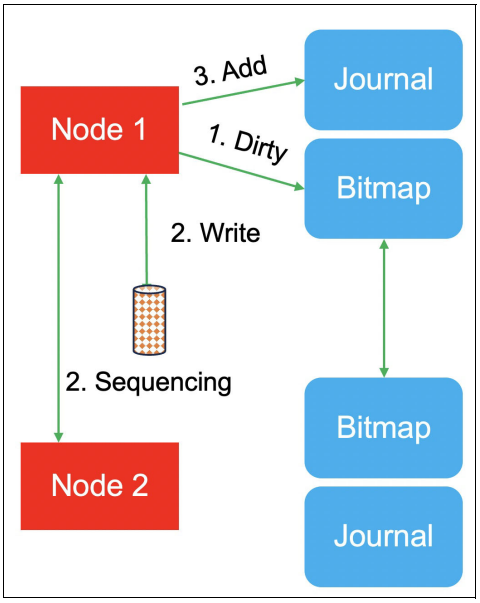


Figure 10-157 Journaling mode

The journal is backed up by a non-volatile bitmap. The size of the journal can vary between 8 GiB and 32 GiB per node.

Cycling mode

Cycling mode is activated when the system is unable to sustain journaling mode. In this mode, change volumes are used to periodically transfer data from the production system to the recovery system as shown in Figure 10-158.

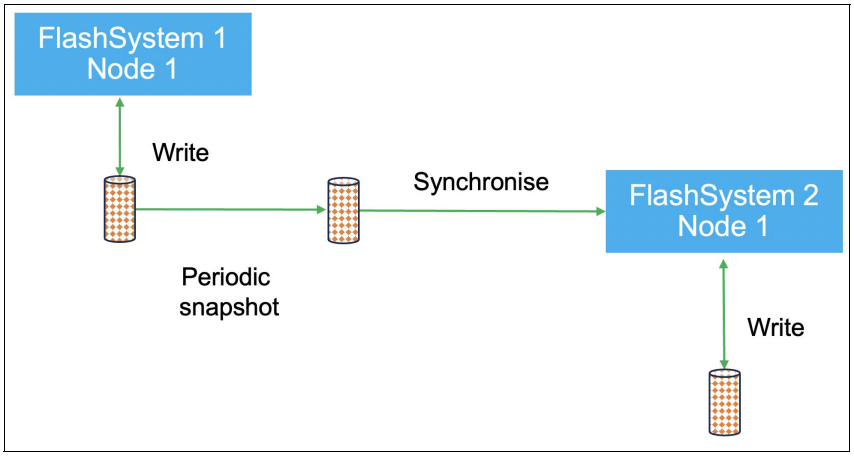


Figure 10-158 Cycling mode

The frequency of cycling is determined by the time it takes to exceed the recovery point objective (RPO). Change volumes ensure the preservation of consistent images at the

recovery location. The duration of the pause during snapshot creation is affected by the number of volumes involved.

Replication mode

The system gives preference to journaling mode due to its shorter recovery point objective (RPO). However, the system automatically strives to keep all volume groups within the RPO warning limit. This means that volume groups with higher RPO warning limits will switch to snapshot mode first.

Volume groups typically prioritize the use of journaling mode because it results in a smaller recovery point. However, if the host throughput surpasses the replication throughput, the volume groups might switch to cycling mode. Cycling mode generally requires less bandwidth compared to journaling mode. Volume groups associated with replication policies that have higher RPOs will switch to cycling mode before others. Regardless of the mode, the system always attempts to meet the RPO for all volume groups.

In rare circumstances where the set replication policy cannot be fulfilled, policy-based replication generates an error message to notify the operator about the discrepancy. This error message specifically highlights situations such as a lack of available line bandwidth between the sites, indicating that the configured value does not align with the current environment.

Replication policies encompass the settings that determine how volume groups are replicated. These policies play a crucial role in ensuring that consistent data is available on both the production and recovery systems.

Assigning a replication policy asynchronously to the sites facilitates automatic synchronization between the production and recovery systems. Any changes made in the production system are reflected in the recovery system through the replication policy. This seamless data copying occurs based on the configured rules and can be assigned, configured, and managed from a single system.

Replication policies offer a significant contribution to maintaining business continuity during outages by keeping the network operational. To configure these policies, certain parameters need to be defined based on your specific requirements:

- Recovery Point Objective (RPO). This denotes the maximum acceptable data loss during a failure. It helps determine the frequency and interval of replication.
- Locations. This refers to the production and recovery systems involved in the replication process.
- Topology. This represents the type of replication chosen. Currently, the supported option is 2-site asynchronous replication.

Table 10-21 lists the key attributes of replication policy.

Table 10-21 Policy-based replication attributes

Attribute	Definition
Set of locations	I/O groups on the partnered systems that contain a replicated copy of the volume group
Topology	Organization of the systems and the type of replication performed between each location
Name	It uniquely identifies the replication policy on both systems

10.16.1 Understanding policy-based replication deployment

This section includes a comprehensive overview of the concepts and architectural framework underlying policy-based replication deployment. We explore the intricacies of implementing asynchronous replication, shedding light on the specific details of its operation. Furthermore, we outline a set of tasks that are typically the responsibility of both storage architects and administrators in effectively managing policy-based replication.

To implement policy-based replication, typically two roles are defined: a storage architect and a storage administrator.

The storage architect is responsible for defining replication policies based on the business requirements. These policies are then applied to volume groups used by specific applications, ensuring that replication is consistently configured as per the defined policies. This allows for standardized replication configurations aligned with the storage architect's guidelines.

On the other hand, the storage administrator is responsible for deploying, managing, and monitoring replication from a centralized interface. Any configuration changes made by the storage administrator are automatically propagated to remote systems, ensuring that data is correctly replicated according to the policies defined by the storage architect. This streamlined approach simplifies the replication management process.

Storage Architect

The following tasks are typically undertaken by the Storage Architect as part of the preparation process:

1. Establish partnerships between systems and facilitates the exchange of certificates. As system certificates are increasingly utilized by various components, it is recommended to update to a valid certificate, preferably during the initial system setup, before configuring any systems or services that rely on these certificates. This is a one-time configuration task that ensures smooth operation and compatibility with certificate-based authentication and security mechanisms as shown in Figure 10-159.

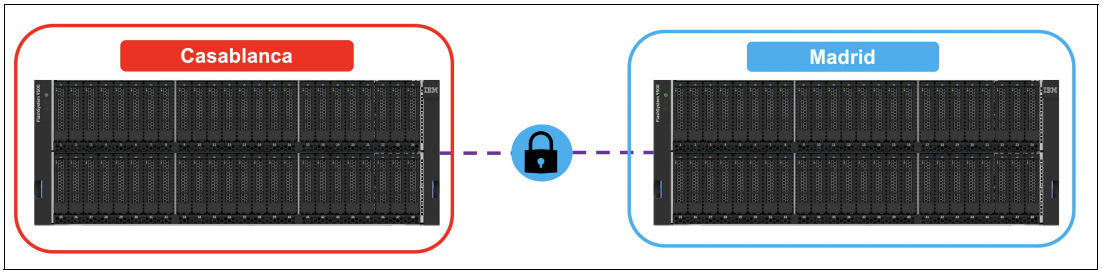


Figure 10-159 Establish partnership

2. Creates storage pools on each system. Creates storage pools and adds capacity on each system. Optionally, creates child pools and allocates capacity from the parent pools as shown in Figure 10-160 on page 910.

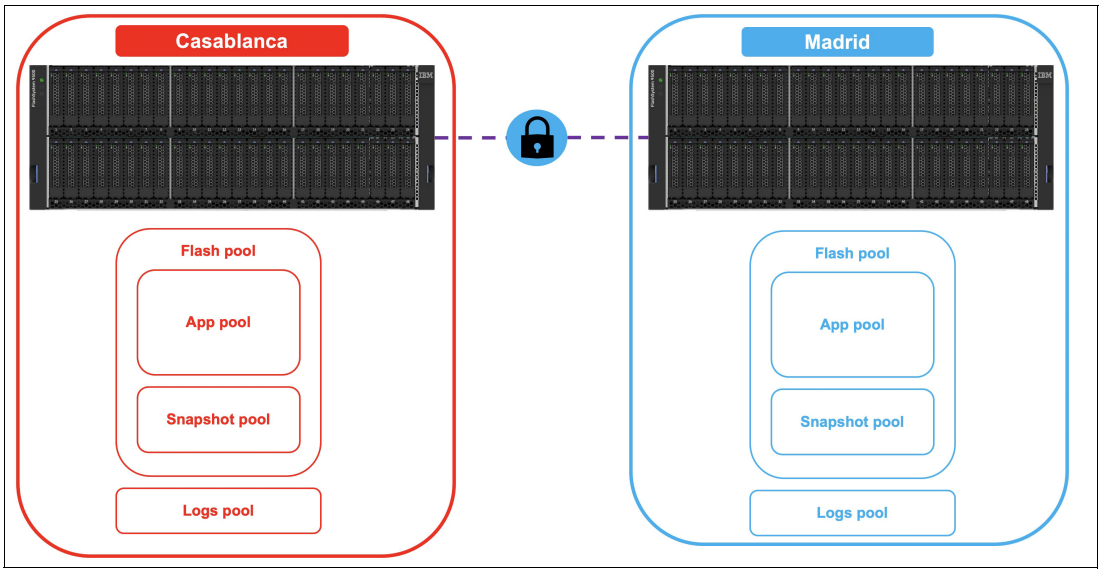


Figure 10-160 Create storage pool

3. Creates provisioning policies:
   - Defines the additional capacity savings used when provisioning new volumes. Savings can be none, thin, compressed, or deduplicated.
   - Different pools can apply different capacity savings.
   - Allows best use of the different capabilities and types of storage on each system, as shown in Figure 10-161.

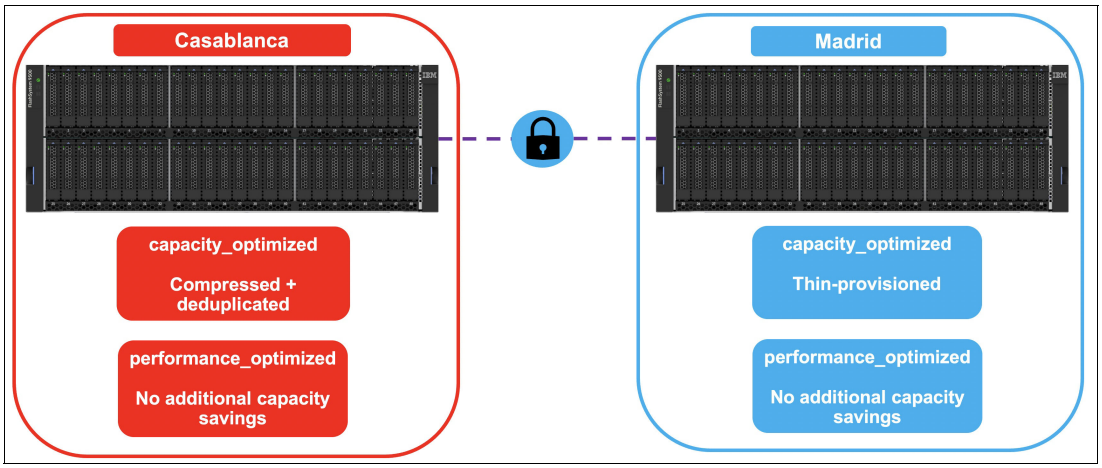


Figure 10-161 Create provisioning policy

4. Assigns a provisioning policy to each pool provisioning:
   - Policy assignment is performed on each system locally, as shown in Figure 10-162 on page 911.
   - The same provisioning policy can be assigned to multiple pools, as shown in Figure 10-163 on page 911.
   - Provisioning policies can be used with any type of pool:
     - parent or child
     - standard or DRP

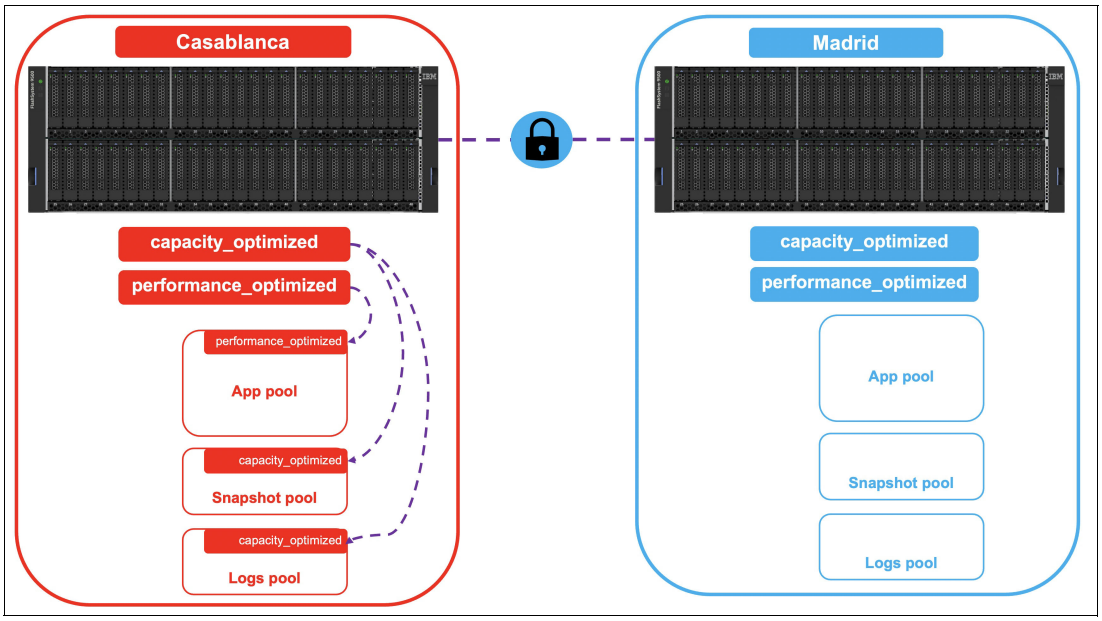


Figure 10-162 Assign provisioning policy on source system

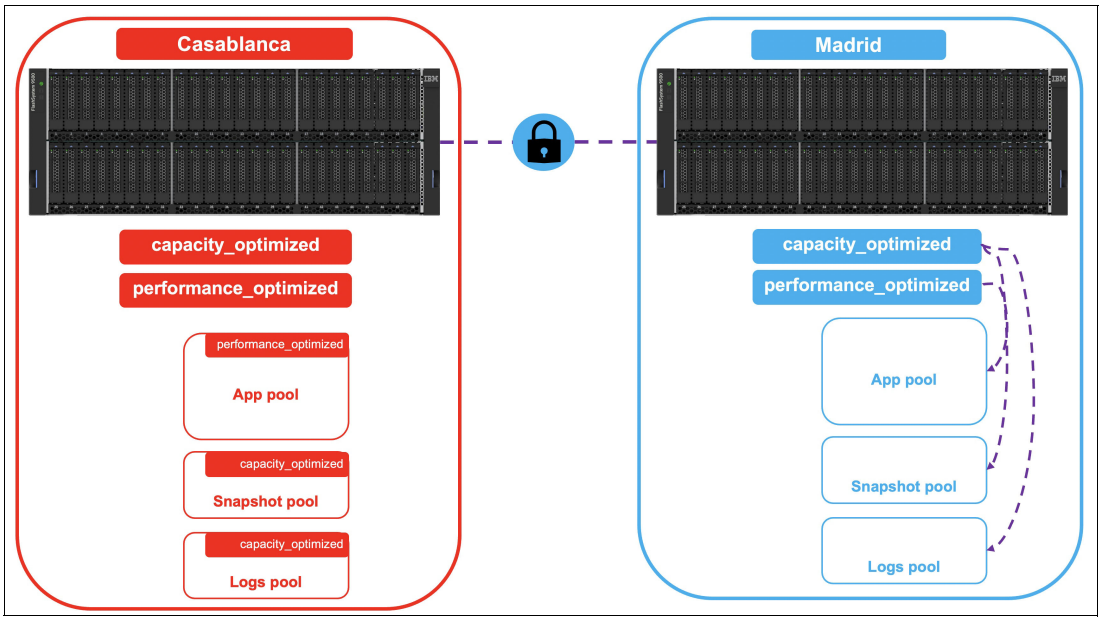


Figure 10-163 Assign provisioning policy on target system

Note: The provisioning policies do not have to be the same between sites.

5. Links pools between systems. Links are required on any pools that contain volumes that use policy-based replication, as shown in Figure 10-164.

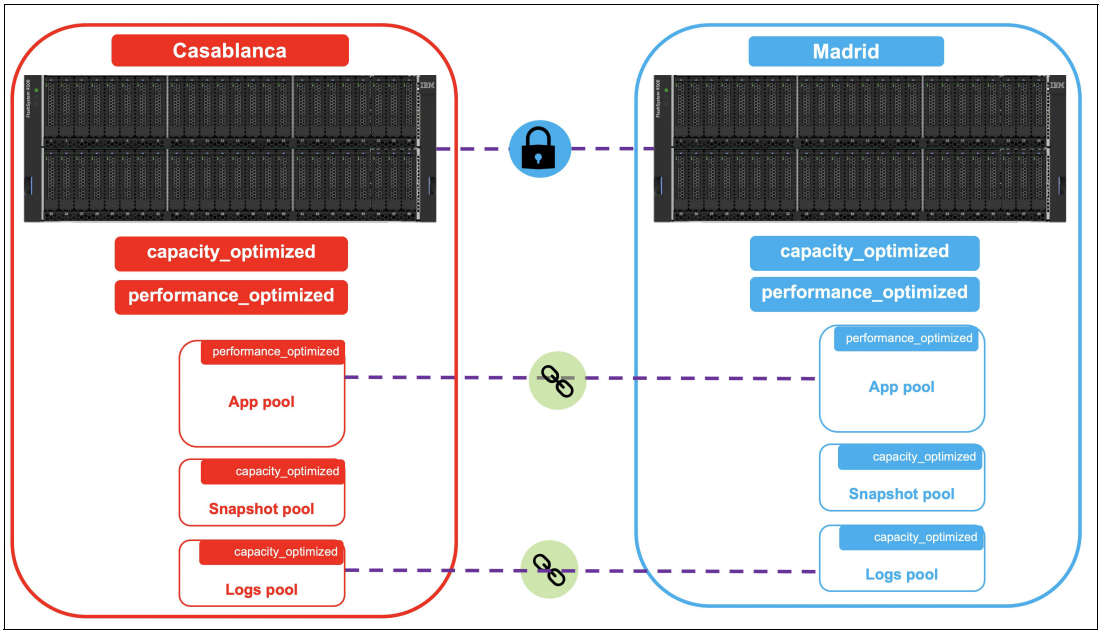


Figure 10-164 Links pools between systems

- Provisioning policy does not need to be symmetric between linked pools.
- Unlinked pools can be linked from either system.
- A pool that has a link and needs another should be linked from the system whose pool is unlinked, the target system.
- The GUI is the easiest way to view and manage pool links and plan policy-based replication.
- If using child pools:
  - The GUI offers a simplified way to create-and-link new child pools with existing pools on a remote system.
  - If child pools exist on the production system, the GUI can simplify the creation of matching linked child pools on the recovery system.

6. Create a replication policy, which can be done on either of the systems, as shown in Figure 10-165.

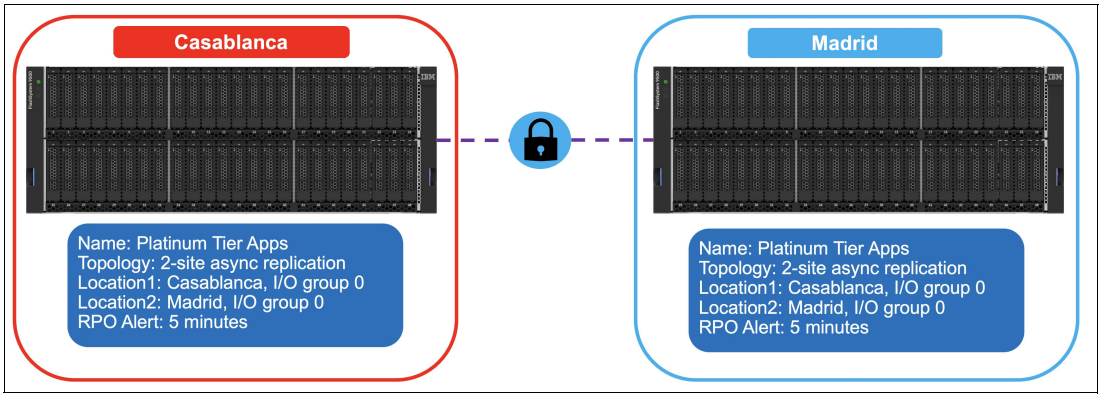


Figure 10-165 Create replication policy

This policy defines how and where volumes should be replicated and sets the acceptable recovery point objective (RPO) for replication. The policy will automatically be replicated to the other systems, ensuring consistency across the environment. You have the flexibility to create multiple policies to meet specific requirements. For example, you can use a second I/O group with policy-based replication or specify different RPOs for different sets of volumes. The primary goal of policy-based replication is to maintain a minimal RPO, but if the system experiences slowdowns, the RPO alert setting will be used to apply Quality of Service (QoS) measures and keep everything within acceptable limits.

With the setup now complete, the storage administrator can proceed to create volume groups and volumes according to the defined policies, as shown in Figure 10-166.

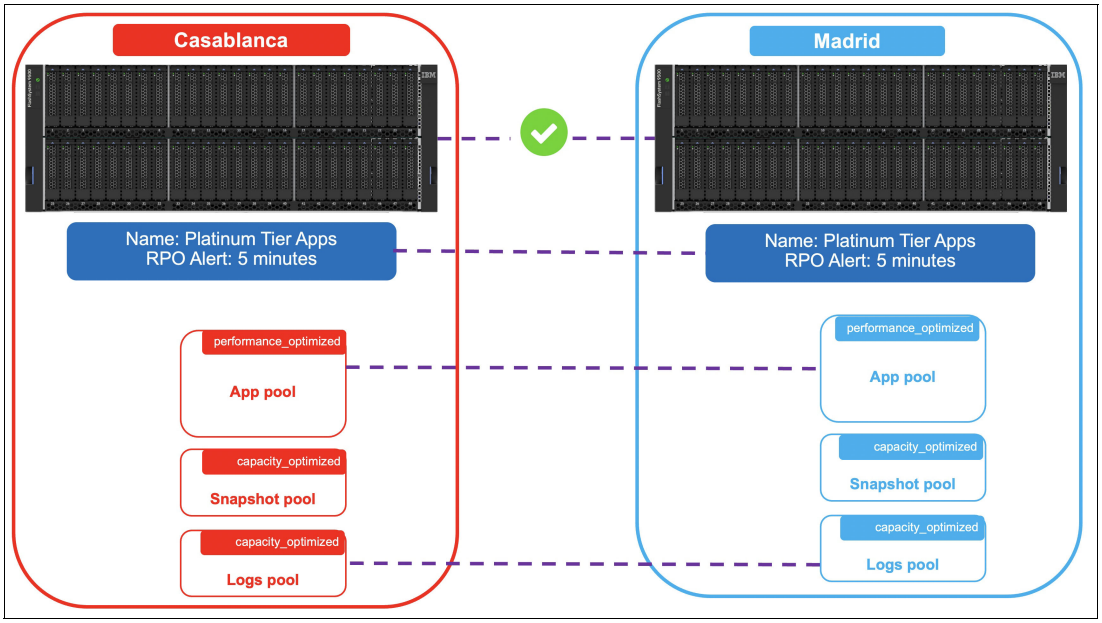


Figure 10-166 Final setup

Storage Administrator

The following tasks are typically undertaken by the Storage Administrator as part of the administration process:

1. Create a new volume group on the production system and assign a replication policy to it. This involves defining the volume group and its properties, such as size and characteristics. Additionally, the administrator assigns a specific replication policy that determines how the data within the volume group will be replicated. After the replication policy is assigned, the system automatically configures the replication settings for the volume group according to the specified policy. This ensures that the data in the volume group is replicated in accordance with the desired replication parameters and requirements, as shown in Figure 10-167 on page 914.

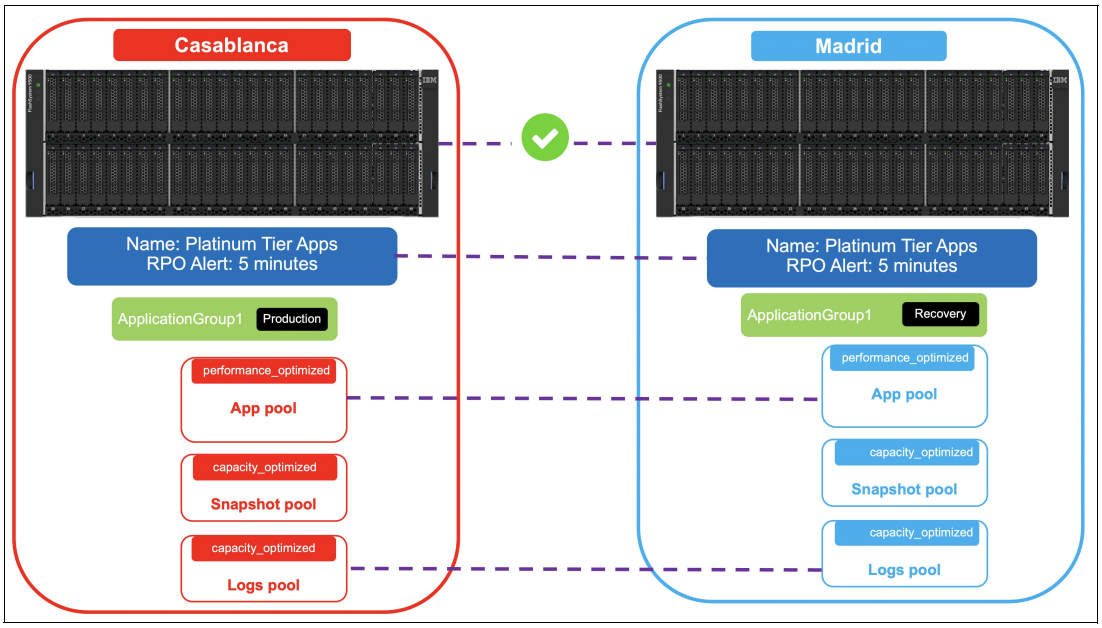


Figure 10-167 Create a new volume group on the production system and assigning a replication policy

2. Initiates the creation of a new volume by specifying its desired size, storage pool, and volume group. The provisioning policy associated with the selected storage pool is then applied to ensure that the volume is created in accordance with the defined provisioning guidelines. This policy defines factors such as allocation method, storage efficiency settings, and performance considerations. Also, the replication policy assigned to the volume group is used to govern the replication of the newly created volume. The replication policy outlines the specific rules and requirements for data replication, including the target location, replication frequency, RPOs, as shown in Figure 10-168.

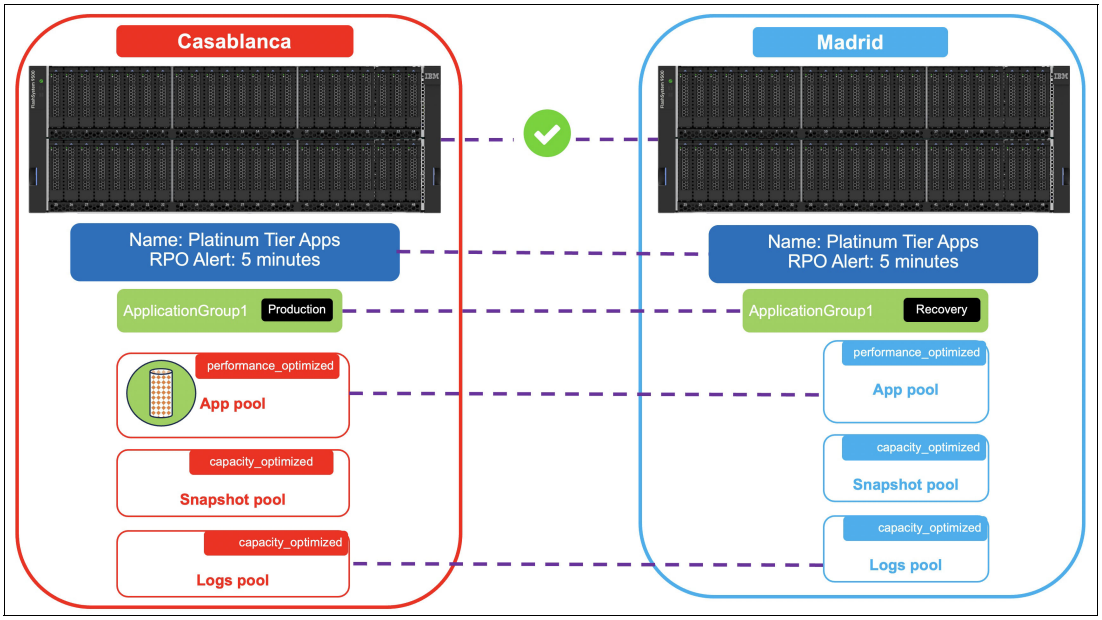


Figure 10-168 Create a new volume group

3. When a new volume is added to the volume group during its creation process, the system recognizes that the source and target volumes are already synchronized. This means that

data replication is not required for these volumes as they are up to date. However, if you have existing volumes that were created prior to the implementation of replication, the system must replicate the data from the primary site to the secondary site. This is necessary because the system cannot guarantee that these volumes are empty or synchronized with the secondary site. Unlike in remote copy scenarios, there is no specific mechanism to inform the system about the state of these volumes. Note that this replication process occurs automatically in the background. The system detects the presence of existing volumes and initiates the replication process to ensure data consistency and protection across both sites. This ensures that newly created and pre-existing volumes are properly replicated to maintain data integrity and availability, as shown in Figure 10-169.

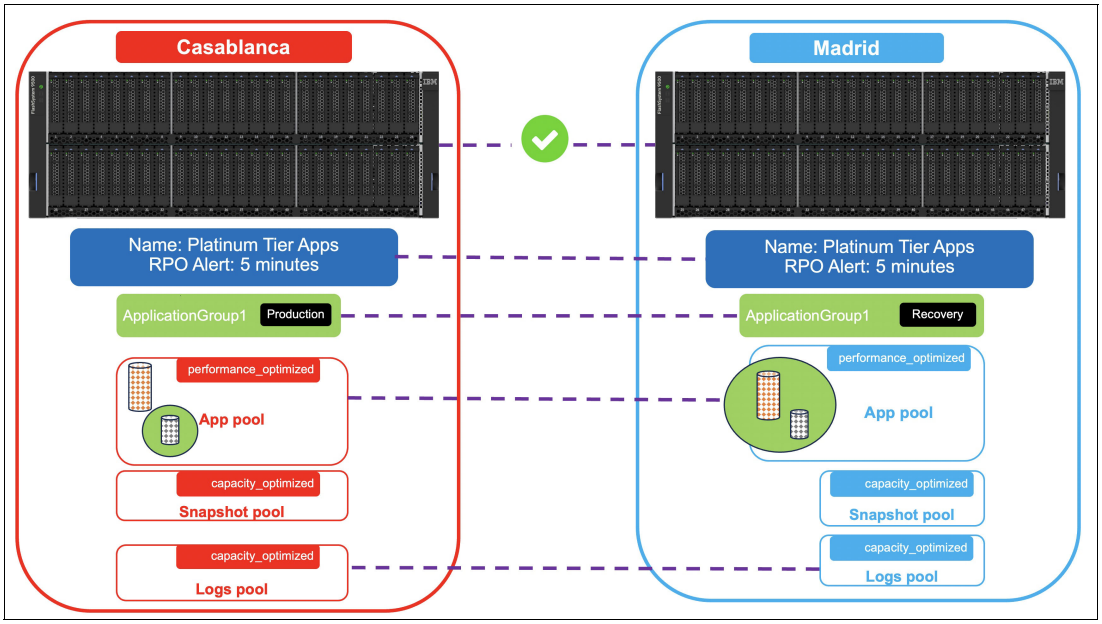


Figure 10-169 Replication is configured automatically

4. If the storage administrator is working in a disconnected environment where the disaster recovery (DR) system is unavailable, they can still create new volumes on the production system. The provisioning process for these volumes will be successful, allowing the administrator to continue their work. During this disconnected state, if the recovery point objective (RPO) for the volume group exceeds the threshold set in the replication policy, the system will raise an alert. This serves as a notification that the replication is not meeting the desired RPO due to the unavailability of the DR system. However, when the DR system becomes accessible again, the system will automatically initiate the replication process. This means that any pending data replication, which could not be performed during the disconnected period, is synchronized between the primary and DR systems to restore the desired RPO, as shown in Figure 10-172 on page 924.

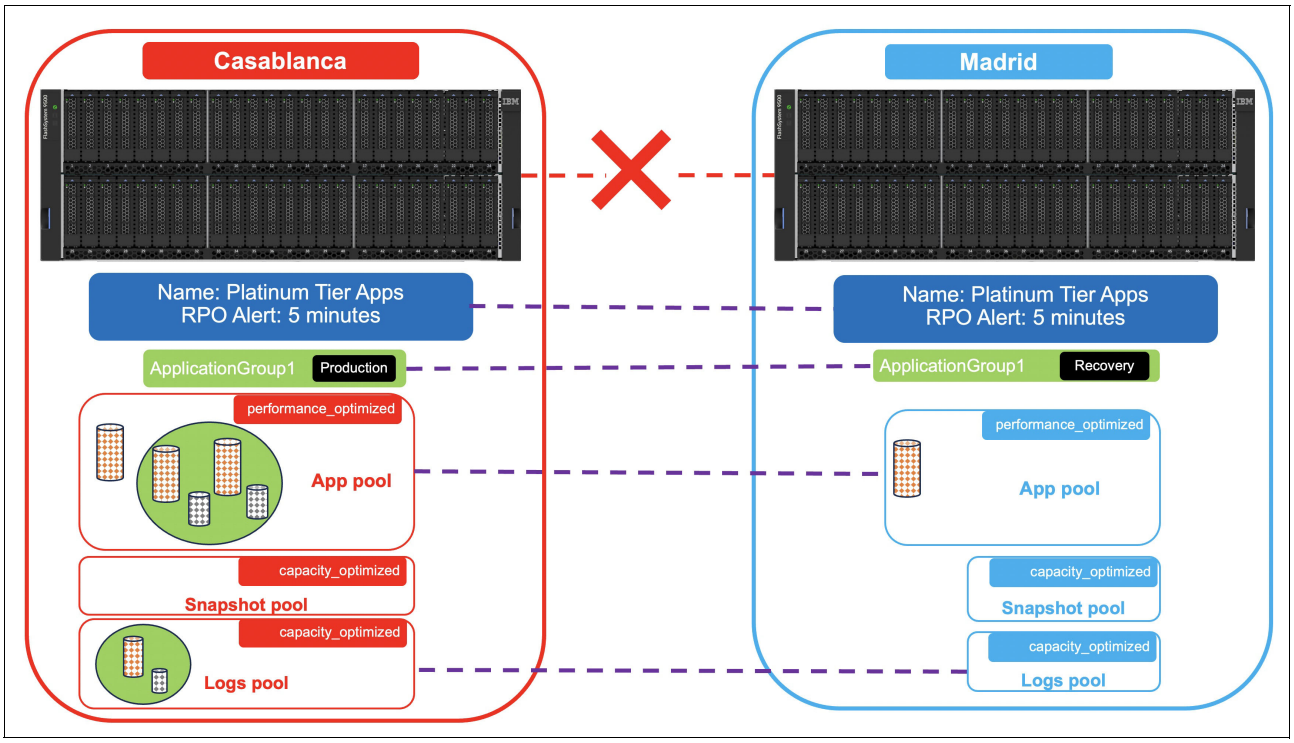


Figure 10-170 Create volume while disconnected

5. The system applies the replication policy to ensure data consistency and maintain a consistent recovery copy even during disconnected periods. When working with new volumes or volumes that are created while disconnected, the system identifies and copies only the data that was written during the disconnected state. During the resynchronization process, the system prioritizes the replication of new data while also ensuring that the existing recovery copy remains consistent for disaster recovery (DR) purposes. This means that while the resynchronization is in progress, the system takes measures to maintain a reliable and up-to-date recovery copy as shown in Figure 10-171 on page 917.

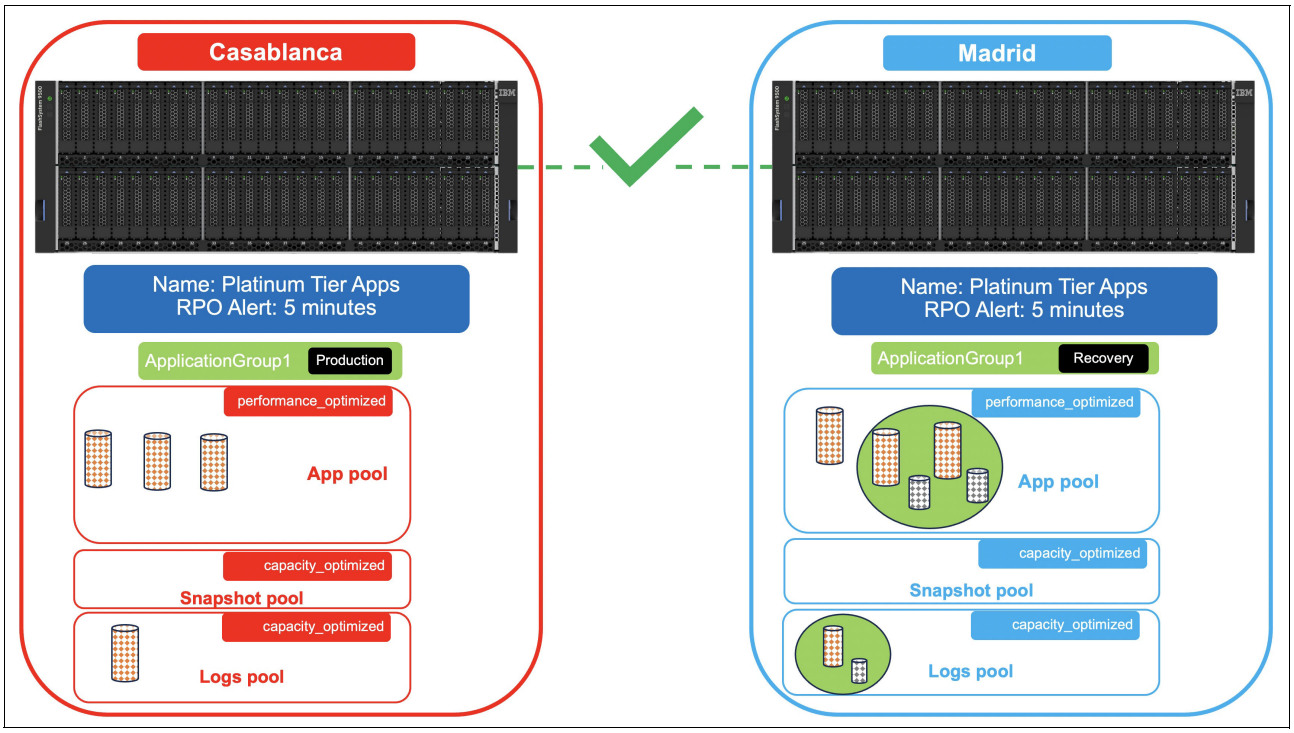


Figure 10-171 Automatically applying replication policy

Planning for policy-based replication

When implementing policy-based replication for applications, it is essential to establish the Recovery Point Objective (RPO) and Recovery Time Objective (RTO). These two metrics play a critical role in defining the organization's data protection and disaster recovery strategy.

The Recovery Point Objective (RPO) determines the acceptable amount of data loss that an application can experience in the event of a failure or disruption. It represents the maximum time interval between the last replicated data point and the occurrence of a failure. For example, if an application has an RPO of one hour, it means that in the event of a failure, the organization can tolerate losing up to one hour's worth of data. In other words, the replicated data will be no older than one hour at the time of recovery.

The Recovery Time Objective (RTO) defines the targeted duration within which an application should be recovered and made operational following a failure. It represents the amount of time allowed for the recovery process before the application needs to be up and running again. For instance, if an application has an RTO of four hours, it means that the organization expects the recovery process to be completed within four hours after a failure, and the application should be fully functional again within that timeframe.

Policy-based replication enforces a specific requirement regarding RPO and RTO. They must be non-zero values. This means that organizations must define measurable RPO and RTO objectives for applications that use policy-based replication. Non-zero values ensure that there are concrete goals in terms of data loss and recovery time, preventing ambiguous or undefined objectives.

When you implement policy-based replication, it is important to consider key factors:

- Number of recovery copies needed
- Geographic distance between the production and recovery systems
- Round-trip time
- Change rate of data
- Allocation of I/O group memory

These factors play a crucial role in designing an effective replication strategy.

Number of recovery copies needed

Determining the number of recovery copies needed involves assessing the level of data redundancy and fault tolerance required for the application. This involves considering factors such as data sensitivity, business criticality, and regulatory compliance. By evaluating these factors, you can decide whether they need a single recovery copy or multiple copies distributed across different locations.

Geographic distance between the production and recovery systems

The geographic distance between the production and recovery systems is a critical consideration. Policy-based replication allows for a larger separation between these systems compared to other replication methods. The production and recovery systems can be located in different geographic regions to provide additional protection against regional disasters or localized disruptions. By increasing the distance between the production and recovery systems, you can minimize the risk of simultaneous failures impacting both environments. The larger separation supported by policy-based replication offers benefits in terms of disaster recovery and business continuity. It allows you to replicate your data to remote sites, providing an additional layer of data protection.

Round-trip time

The round-trip time (RTT) refers to the elapsed time for a packet of data to travel from the source system to the destination system and back. The RTT is affected by various factors, such as network latency, distance between systems, and the quality of the network infrastructure. It plays a crucial role in determining the throughput of replication.

Storage arrays have a finite capacity for transmitting data at any given time. As the RTT increases, the maximum throughput of replication decreases. This is because the increased latency introduces additional time delays for acknowledgments and data transmissions between the systems. Therefore, a higher RTT might impact the replication performance, which can reduce the overall throughput. However, policy-based replication offers an advantage over the remote-copy function when it comes to higher RTT scenarios. Policy-based replication can achieve higher throughput despite the increased RTT. This is because it can optimize the data transmission process by using techniques such as data compression and efficient bandwidth usage. As a result, policy-based replication can deliver better performance in terms of data replication even with a longer RTT.

The Remote Copy function supports a maximum distance between sites, which is determined by the maximum RTT between sites as specified in Table 10-22.

Table 10-22 Longest RTT supported by IBM FlashSystem storage-to-storage connectivity

Maximum *RTT supported by IBM FlashSystem storage-to-storage connectivity*		
Fiber Channel	IP connectivity 1 Gbit Ethernet	IP connectivity 10Gbit Ethernet
250[ms]	80 [ms]	10 [ms]

In scenarios where the distance between two FlashSystem units is significant and the server's write workload is substantial, consider using Fibre Channel over IP (FCIP) connectivity. However, it is essential to first enhance the bandwidth of the Wide Area Network (WAN) connection between the sites to accommodate the increased data transfer requirements. FCIP connectivity typically has higher installation costs compared to other connectivity options. However, given the long distance and high data volume, the improved performance and reliability of FCIP might outweigh the initial installation expenses, especially in scenarios where substantial data transfer is involved.

Change rate of data

The change rate of data for applications storing data on the system refers to the frequency and volume of data modifications or updates occurring within the application. It might be difficult to achieve low RPOs for applications with high change rates but limited available bandwidth.. If the change rate exceeds the bandwidth capacity, the replication process may not be able to keep up with the pace of data modifications, resulting in increased RPOs and potential data loss.

Policy-based replication can be a valuable solution for data protection and disaster recovery, even in scenarios with lower-bandwidth links. However, it is important to understand that lower throughput resulting from limited bandwidth can impact the replication process and potentially lead to a higher recovery point objective (RPO).

When the available bandwidth is lower, the replication process might experience delays in transmitting data between the systems. This delay can result in data not being replicated in real-time, causing a time lag between the production system and the recovery system. As a result, the recovery point might be further behind compared to scenarios with higher throughput.

The higher recovery point means that there is a greater potential for data loss in the event of a failure or disaster. If the RPO is exceeded, then the replicated data might not reflect the latest changes made in the production system. It is important to note that although policy-based replication provides protection against large-scale disasters affecting the site or storage arrays, it does not guarantee complete protection against data corruption or cyberattacks.

There may be instances where data corruption occurs in the production system and is inadvertently replicated to the recovery systems. This can result in the recovery systems also being affected by the corrupted data.

To enhance data protection and safeguard against data corruption or cyberattacks, you can use FlashCopy and Safeguarded Copy functions in conjunction with policy-based replication. You can use FlashCopy to create point-in-time snapshots. Safeguarded Copy ensures that replicated data remains isolated and protected. For more information refer to 10.2, “Safeguarded Copy” on page 718 and 10.3, “IBM FlashCopy” on page 719.

Allocation of I/O group memory

To effectively use policy-based replication on a system, you must allocate a specific amount of I/O group memory. However, if your system is currently configured with remote copy functions like Global Mirror, Metro Mirror, or HyperSwap, you might need to reallocate the I/O group memory to enable policy-based replication. The allocation of I/O group memory is crucial for ensuring optimal performance and resource utilization within the system.

Different replication functions, including Remote Copy functions and policy-based replication, have varying requirements in terms of memory allocation. If your system is already configured with Remote Copy functions and if you want to transition to policy-based replication, you might need to adjust the allocated memory. This is because the memory requirements for policy-based replication may differ from those of remote copy functions. In some cases, you

might need to convert the existing Global Mirror configuration to policy-based replication sites.

If you plan to use policy-based replication with other functions or if you plan to switch from Global Mirror replication to policy-based replication, refer to the following tables to determine the allocated memory that is required for additional features.

Table 10-23 outlines a product comparison indicating the required I/O group memory for policy-based replication.

Table 10-23 Required I/O group memory for policy-based replication

Product	Total I/O group memory (excluding FlashCopy functions)	Required I/O group memory for policy-based replication	Available memory for other features (RAID, volume mirroring, and remote copy)
IBM SAN Volume Controller	3628 MiB	2561 MiB	1067 MiB
IBM FlashSystem 9500	5920 MiB	4609 MiB	1311 MiB
IBM FlashSystem 9200	3628 MiB	2561 MiB	1067 MiB
IBM FlashSystem 9100	3328 MiB	2561 MiB	1067 MiB
IBM FlashSystem 7300	3628 MiB	2561 MiB	1067 MiB
IBM FlashSystem 7200	3628 MiB	2561 MiB	1067 MiB
IBM FlashSystem 5200[a]	2088 MiB	1537 MiB	551 MiB
IBM Storage Virtualize for Public Cloud	2088 MiB	1537 MiB	551 MiB

a. Requires a minimum of 128 GiB memory in each node canister

Table 10-24 illustrates the default setting of the bitmap space within FlashSystem, SVC, and Storage Virtualize for Public Cloud.

Table 10-24 Default and maximum setup of the bitmap space

Copy services	Default allocated bitmap space	Maximum allocated bitmap space
Remote copy	20 MiB	1024 MiB
Volume mirroring	20 MiB	512 MiB
DRAID	40 MiB	800 MiB

Therefore, the total amount of allocated memory for all functions, excluding FlashCopy, should not surpass 1824 MiB. The specific number of functions might vary because of settings such as grain size and strip size, potentially resulting in an increased allocation.

To estimate the necessary bitmap memory (in MiB) for distributed RAID (DRAID), you can use the following formulas:

- For DRAID 1: Drive capacity * (count + 2 * stripe) / (stripe * 256 * 8)
- For DRAID 6: Drive capacity * (count + 3 * stripe) / (stripe * 256 * 8)

Note that the drive capacity is measured in GiB, the count refers to the number of drives. The term *stripe* is the width of the stripe and *strip* is the length of the stripe measured in KiB.

Additionally, when considering FlashCore Modules, use the effective capacity instead of the physical capacity.

Assume you are using the following IBM FlashSystem 7300 configuration:

- Firmware level: 8.6.0
- Drive capacity: 19.2 TB FlashCore Module 3
- Effective Capacity: 57.6TB FCM3
- Count: The number of drives in the DRAID6 configuration is 12 drives.
- Stripe: The width of the stripe is 256 KiB.

A total of 3628 MiB of memory is available for remote copy, volume mirroring, RAID, and policy-based replication. If policy-based replication is activated, it uses 2561 MiB, leaving 1067 MiB for allocation among remote copy, volume mirroring, and RAID. The maximum limits for these allocations are set as 1024 MiB for remote copy, 512 MiB for volume mirroring, and 800 MiB for RAID. However, if policy-based replication is disabled, a maximum of 1824 MiB can be allocated among remote copy, volume mirroring, and RAID, with the same maximum limits of 1024 MiB, 512 MiB, and 800 MiB.

The required bitmap memory in MiB for DRAID is calculated by using the following formula:

Bitmap Memory (MiB) = (57.6 × 1024) GiB × (12 + 3 × 256 KiB) / (256 KiB × 256 × 8)

10.16.2 Planning for hosts, partnerships and connectivity

When preparing for a partnership, the following factors should be considered:

- Data:
  - IP excels in low latency environments and supports a maximum round-trip time (RTT) of 80 ms. FC allows for greater distance with a maximum RTT of 250ms, especially when using I/O group zoning.
  - The data transfer speed between your primary and backup storage (production and recovery I/O groups) depends on the partnership bandwidth and the background copy rate.
  - IP treats Remote Copy and policy-based replication equally, allocating bandwidth based on overall needs, not dividing it between the two.
  - IP partnerships offer the additional benefit of supporting data compression and encryption.
- Management:
  - IP connectivity between management IPs is required, and both IPv4 and IPv6 are supported. This requirement applies whether the replication is FC-based or IP-based.
  - Firewall rules must be configured to allow outbound and inbound traffic between the management IP addresses.
  - All management traffic is authenticated using certificates, which are configured during the partnership setup.
  - Both self-signed and CA-signed certificates are supported.

Before implementing policy-based replication, it is essential to carefully plan the requirements of the host applications that will use this feature. Policy-based replication is compatible with all host operating systems supported by IBM Storage Virtualize systems. For a comprehensive list of supported systems, see IBM System Storage Interoperation Center (SSIC).

To achieve the best performance, it is recommended to configure the multipath drivers on the operating system to use SCSI Asymmetric Logical Unit Access (ALUA) or NVMe Asymmetric Namespace Access (ANA). For detailed instructions for configuring host attachments, see Chapter 7, “Hosts” on page 517.

Note: If you are using the Enhanced Stretched Cluster feature of SAN Volume Controller, define the preferred node at the site that generates the majority of host writes during regular operation.

Also, consider the following factors:

- When you use policy-based replication, you can replicate thin-clones or the target volumes of a FlashCopy mapping to the recovery system, to ensure application consistency. However, it's important to note that Safeguarded copies cannot be replicated using this method.
- Policy-based replication uses the same basic connectivity as remote-copy based replication. However, for managing replication, an additional IP connection is required between the management IP addresses of the partnered systems. This IP connection leverages authentication certificates to ensure secure communication, regardless of whether the primary connection uses Fibre Channel or IP.
- IBM Storage Virtualize replication ensures consistent write-order in all operations, except during synchronization. Policy-based replication uses volume groups to establish a set of volumes that must maintain mutual consistency. Ideally, a volume group should encompass all the volumes associated with the application that requires recovery.
- There are two zoning configurations available for partnerships that use policy-based replication, depending on the RTT between the systems:
  - For an RTT between sites of less than or equal to 80 ms. Zoning and optional port masking can be employed. This configuration enables direct communication between nodes within the specified I/O groups, as defined by the replication policy, and all nodes within the target I/O groups where the replicated volumes reside in the partnered system.
  - For a round-trip latency between sites of greater than 80 ms and less than or equal to 250 ms. Zoning should be implemented with separate intersystem zones created for each local-remote I/O group pair involved in replication. This configuration ensures appropriate isolation and facilitates the replication process.

Configurations limits and restrictions

The usage of policy-based replication is subject to the following configuration limits and restrictions:

- While a replication policy is assigned, it is not possible to change the name of a volume group.
- Similarly, the name of a volume cannot be modified if it belongs to a volume group with an assigned replication policy.
- Policy-based replication does not support the use of ownership groups.
- Systems that use HyperSwap topology do not support policy-based replication.
- Policy-based replication cannot be applied to volumes that have the following characteristics:
  - Configured in Image mode
  - Use HyperSwap functionality

  - Use Transparent Cloud Tiering
  - Use VMware vSphere virtual volumes (vVols)
- When considering the distance for the remote copy function, the Ethernet port speed on both ends of the IBM FlashSystem must be identical when connected through IP. It is not supported to have a 10 Gbit Ethernet port on one IBM FlashSystem and a 1 Gbit Ethernet port on the other. The port speeds must match.
- Volume Actions Restriction. The following actions cannot be performed on a volume if it is part of a volume group with an assigned replication policy:
  - Resizing (expanding or shrinking) the volume.
  - Migrating the volume to image mode or adding an image mode copy.
  - Moving the volume to a different I/O group.

Note: Ports 3260, 3265, and 7443 must be open on the system management IP address to allow the systems to replicate. Fibre Channel and IP replication require port 7443 for REST API access for replication management. IP replication requires ports 3260 and 3265

- I/O group limitations. All volumes within a volume group must belong to the same I/O group, as specified by the replication policy. As a result, replication activities are exclusively conducted with the I/O group partner, leading to notable performance enhancements. This requirement guarantees that all volumes within a volume group share the same failure domain.

Table 10-25 presents the supported targets for policy-based replication along with the corresponding maximum number of volumes and capacity limits.

Table 10-25 Supported targets and volume quantities and capacity limit

Copy services	Maximum number of policy-based replication volumes	Maximum policy-based replication capacity
IBM SAN Volume Controller	7,932	2,048 TiB
IBM FlashSystem 9500	10,000	2,048 TiB
IBM FlashSystem 9200	7,932	2,048 TiB
IBM FlashSystem 9100	7,932	2,048 TiB
IBM FlashSystem 7300	7,932	2,048 TiB
IBM FlashSystem 7200	7,932	2,048 TiB
IBM FlashSystem 5200	7,932	1,024 TiB
IBM Storage Virtualize for Public Cloud[a]	7,932	1,024 TiB

a. Microsoft Azure

10.16.3 Configuring policy-based replication using GUI

To demonstrate the full functionality and capabilities of policy-based replication, this section discusses the roles of the storage architect and the storage administrator.

Define partnership for replication

Before starting the configuration process, it is important to ensure that a partnership has already been established between the two FlashSystem boxes. This partnership is used for data replication and synchronization. To initiate the partnership, perform the initial system setup on both FlashSystem boxes. This setup involves establishing a partnership between the systems, which can be done using either FC or IP connectivity.

Additionally, as part of the partnership setup, a certificate exchange must be performed. This exchange allows each system to have the necessary configuration access to the other system using the REST API.

Use the GUI to verify that you have Storage Virtualize version 8.6.0 or later on both FlashSystems. Go to **Settings**, and select **Update System**, as shown in Figure 10-172.

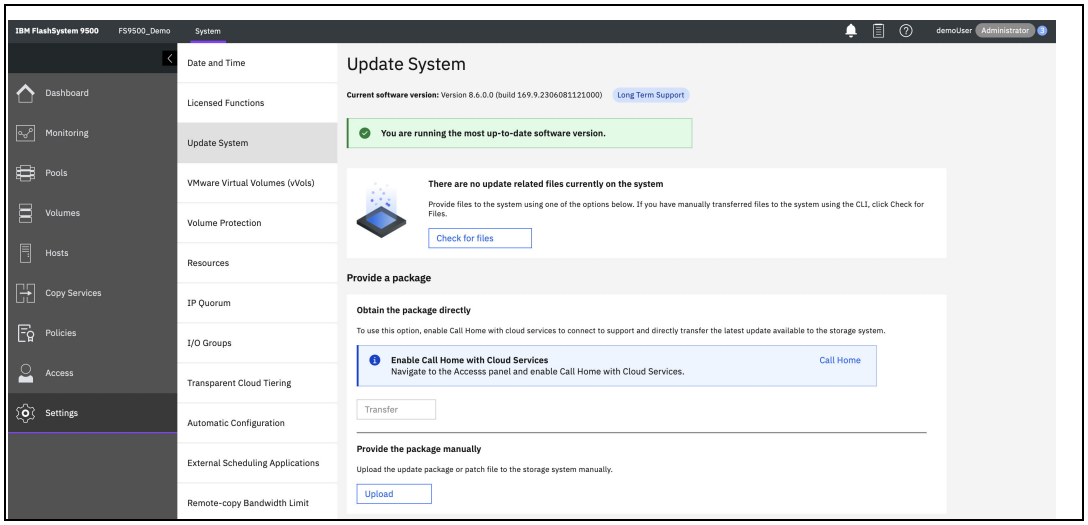


Figure 10-172 Verify software version

Creating linked pool

To establish a linked pool in policy-based replication, the storage pool links determine where the recovery copy of a volume is stored on the recovery system based on the production volume's pool. It is necessary to have links between the pools on the production and recovery systems when using policy-based replication.

At least one linked pool is required on each system for policy-based replication to function properly. There are two approaches to creating a linked pool. The first option involves creating pools on each storage system and then linking them together. Alternatively, you can directly create a linked pool. In the following section, we will focus on the latter method.

1. To begin, initiate the process by creating a child pool on the FS9500.
   a. Access the **Pools** menu.
   b. Right-click the pool.
   c. Select the option to **Create Child Pool** as shown in Figure 10-173 on page 925.

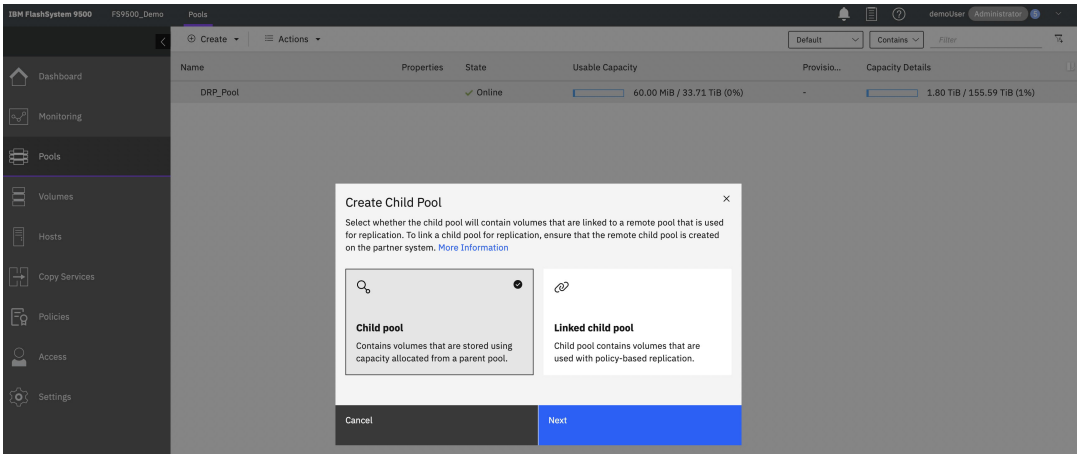


Figure 10-173 Creating a child pool

2. Specify the child pool name and the provisioning policy, as shown in Figure 10-174.

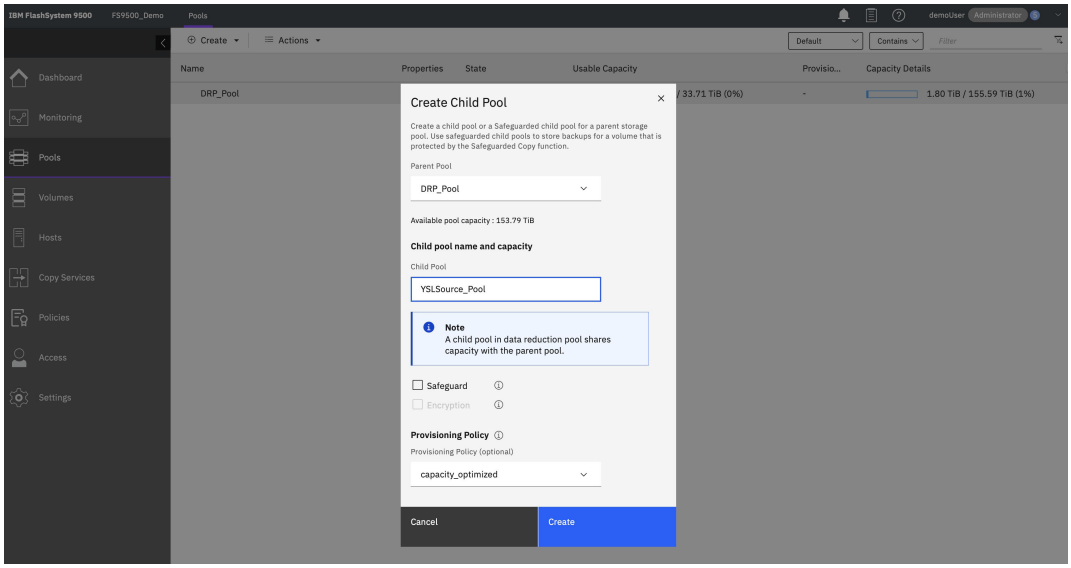


Figure 10-174 Child pool name and provisioning policy

3. To create a linked child pool on your secondary system, IBM FlashSystem 7200, follow these steps:

   a. Access the **Pools** menu on the secondary system.
   b. Right-click the pool.
   c. Select the option to create a child pool, as shown in Figure 10-175 on page 926.

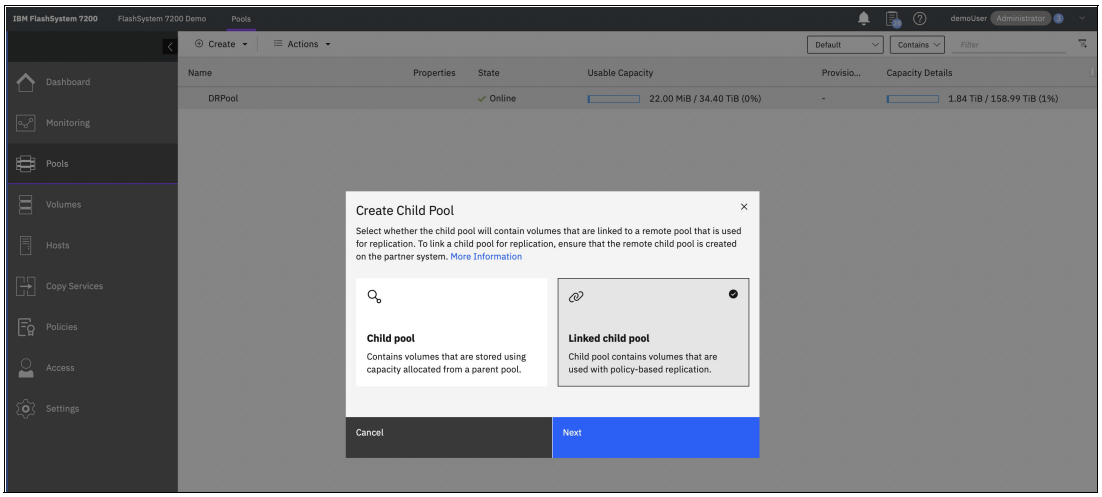


Figure 10-175 Create a linked child pool

d. In the next panel, open the **Select system to link** menu and select the remote primary system and select the specific pool on the primary system that you want to link with the new child pool. Provide a name for the new child pool. Click the **Create** button to complete the creation of the linked child pool, as shown in Figure 10-176.

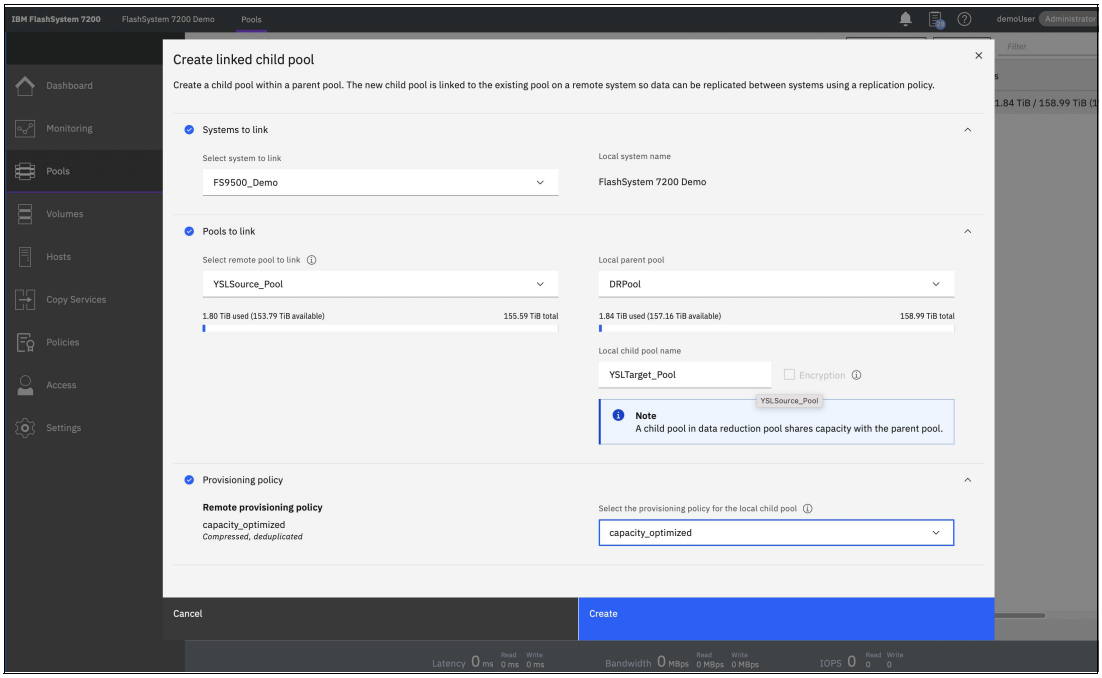


Figure 10-176 Create linked child pool details

Creating replication policies

Replication policies play a crucial role in policy-based replication because they define the configuration of replication between I/O groups in partnered systems and define how it should be applied to volume groups and their associated volumes. These policies are established between two fully configured and partnered systems capable of policy-based replication.

A replication policy can be linked to multiple volume groups. However, each volume group can have a maximum of one replication policy associated with it.

Note: The replication policy must be created on the primary system.

1. To create a replication policy, navigate to the **Policies** menu and select **Replication Policies**, as shown in Figure 10-177.

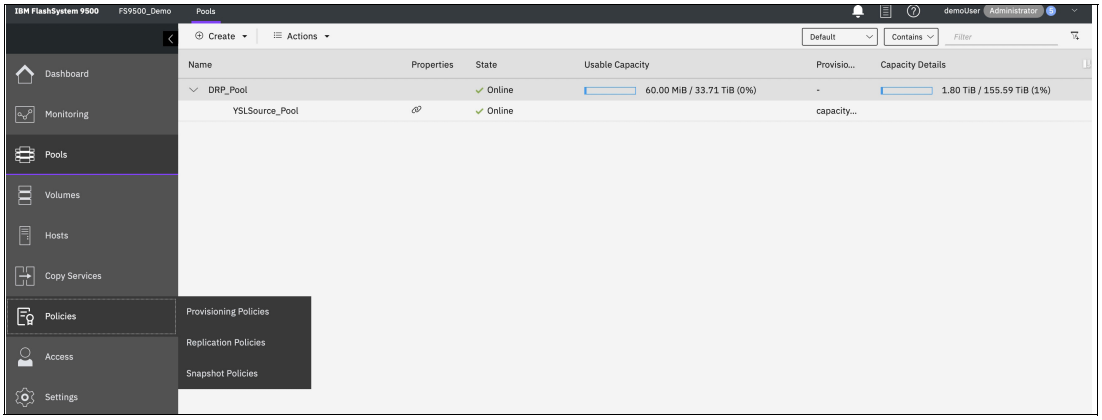


Figure 10-177 Replication policy panel

2. Select **Create Replication Policy**, as shown in Figure 10-178.

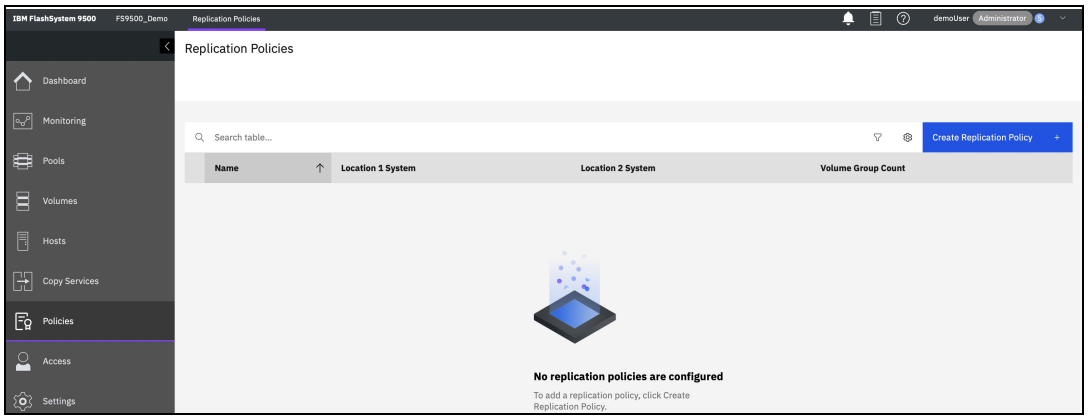


Figure 10-178 Create replication policy

3. Choose a name for the policy and specify the primary and secondary systems involved. Additionally, set the Recovery Point Objective (RPO). Only asynchronous relations are supported in policy-based replication, as shown in Figure 10-179 on page 928.

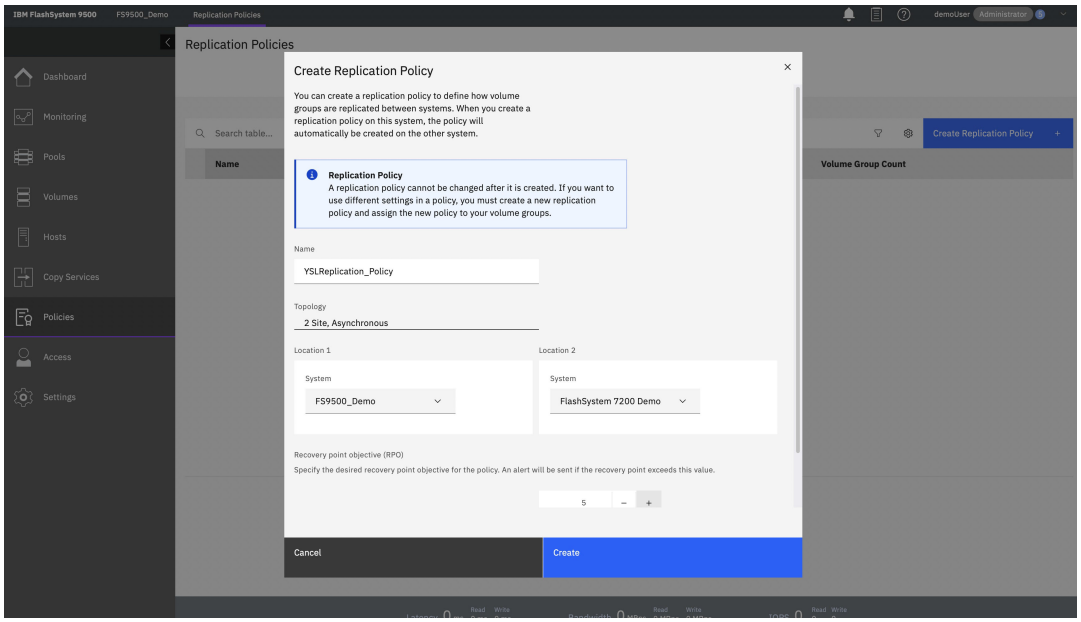


Figure 10-179 Select the primary and secondary systems and RPO

4. Verify that the replication policy is correctly created, as shown in Figure 10-180.

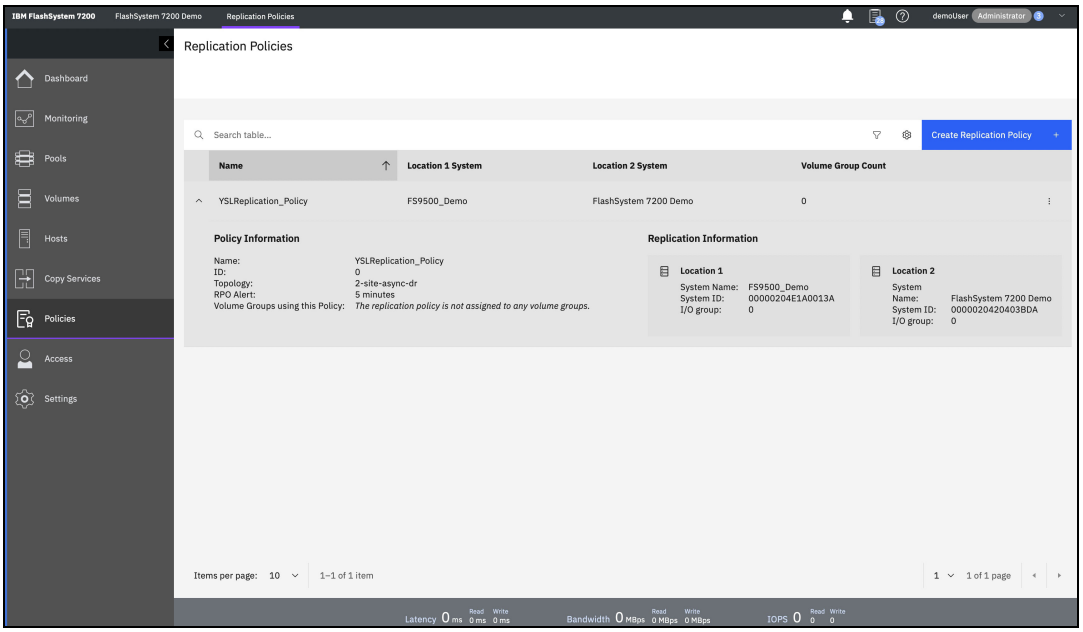


Figure 10-180 Verify that the replication policy is correctly created

Creating a volume group and assigning a replication policy

To deploy and manage replications in policy-based replication, volume groups and replication policies are used. After a replication policy is created, create a volume group and assign it to the replication policy. This process ensures consistent replication by replicating the source volumes together to the recovery system.

The recovery copies of volume groups are immutable, meaning they cannot be modified or altered. Policy-based replication greatly simplifies the configuration, management, and monitoring of replication between two systems.

To create a volume group, follow these steps:

1. In the **Volumes** menu on the left panel, navigate to the **Volume Groups** section, as shown in Figure 10-181.

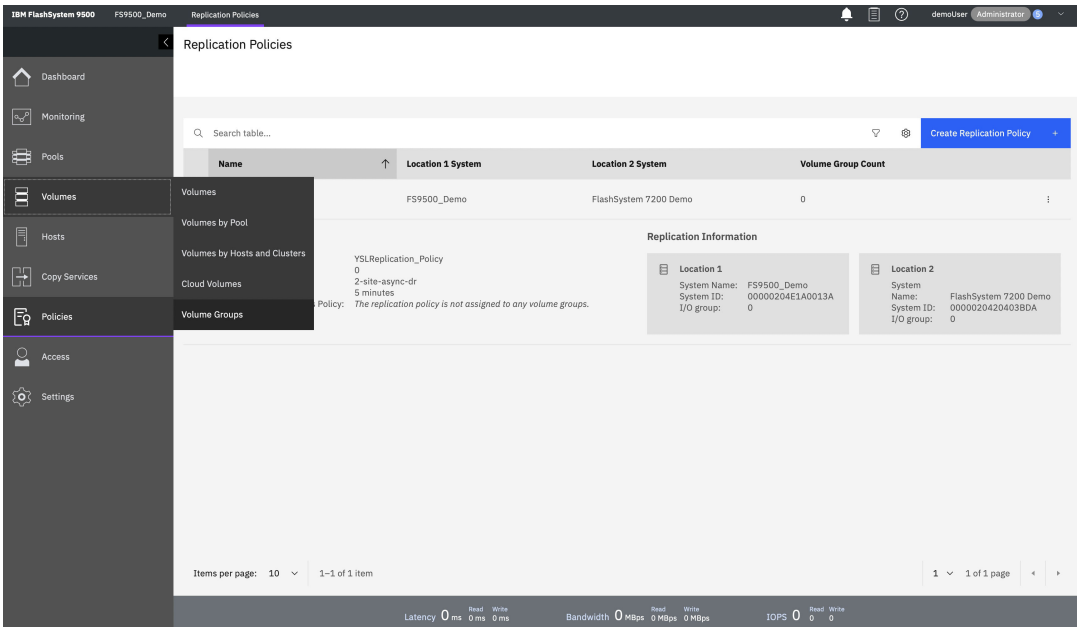


Figure 10-181 Volume group panel

2. If no volume groups exist, click the **Create Volume Group** button at the bottom. Alternatively, click the **"+"** button in the **Create Volume Group** button, as shown in Figure 10-182.

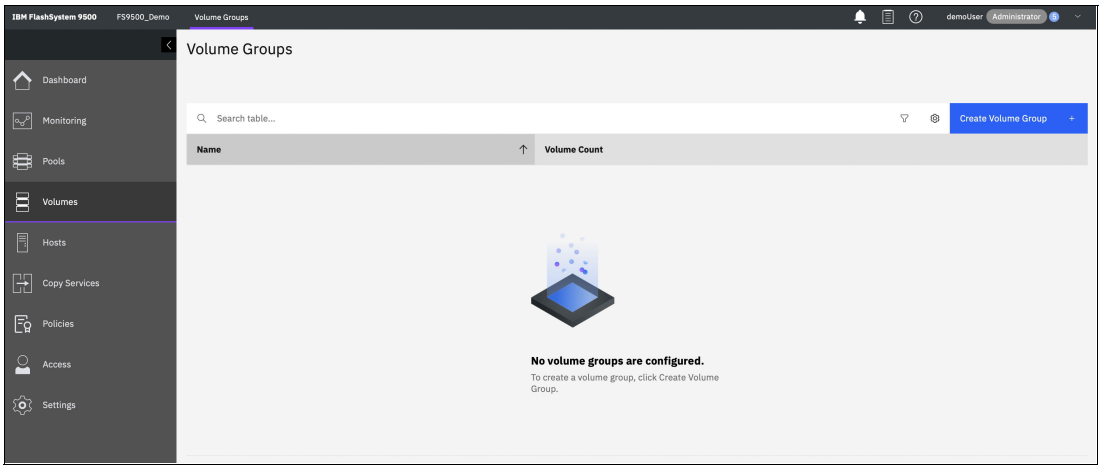


Figure 10-182 Create a volume group

3. Enter a name for your volume group and click the **Create Empty Group** button, as shown in Figure 10-183 on page 930.

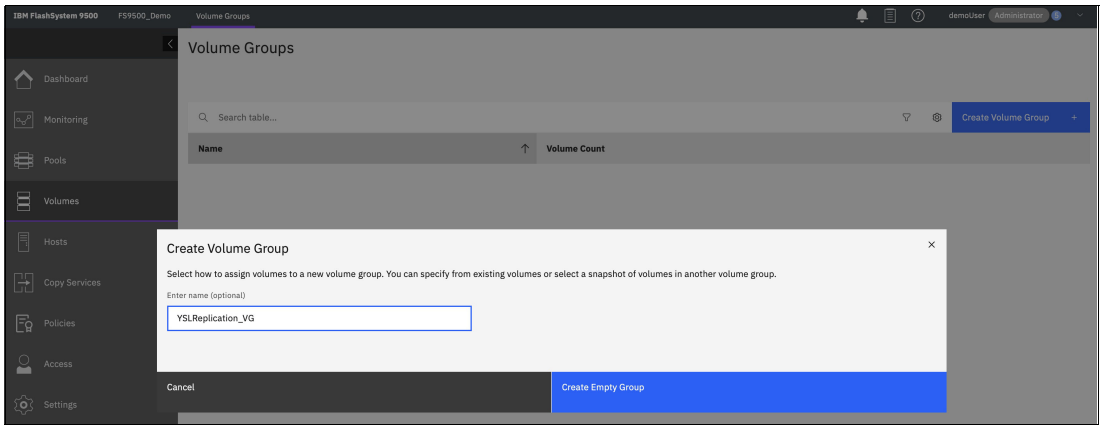


Figure 10-183 Create an Empty Group

4. Select the newly created volume group by clicking its name and click the **Policy** tab, as shown in Figure 10-184.

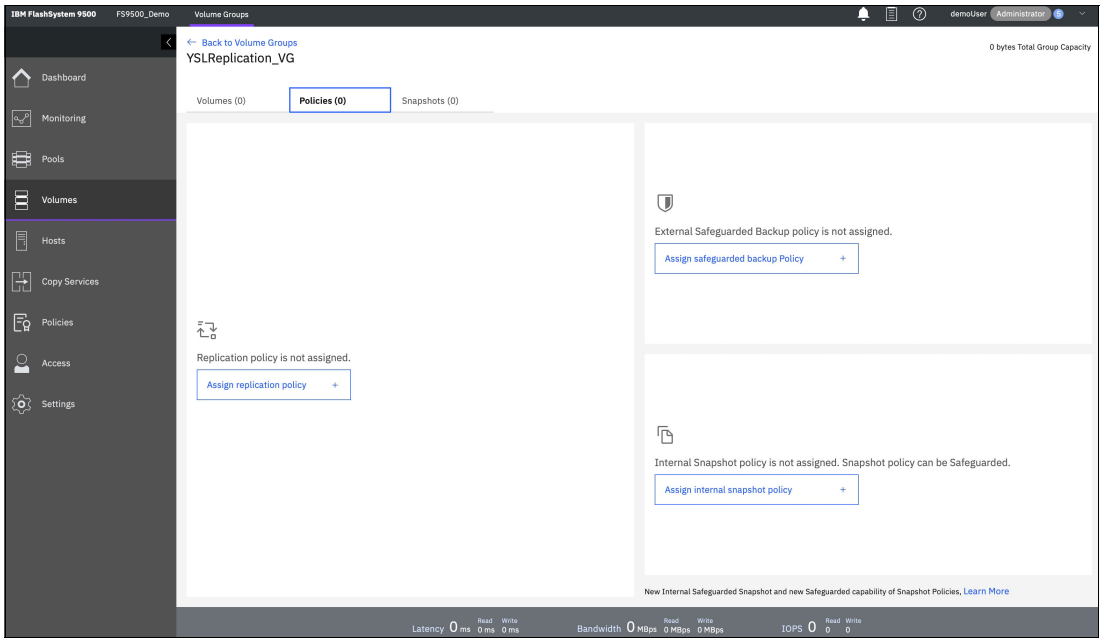


Figure 10-184 Access the volume group and selecting Policy tab

5. Select the replication policy you previously created and click the **Assign** button, as shown in Figure 10-185 on page 931.

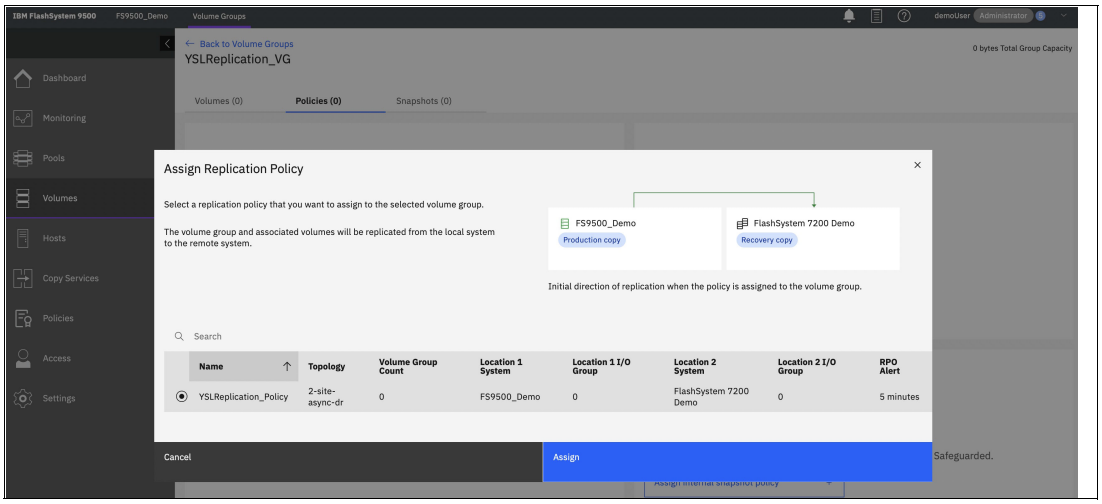


Figure 10-185 Assign replication policy

6. Verify that the replication policy is correctly assigned, as shown in Figure 10-186.

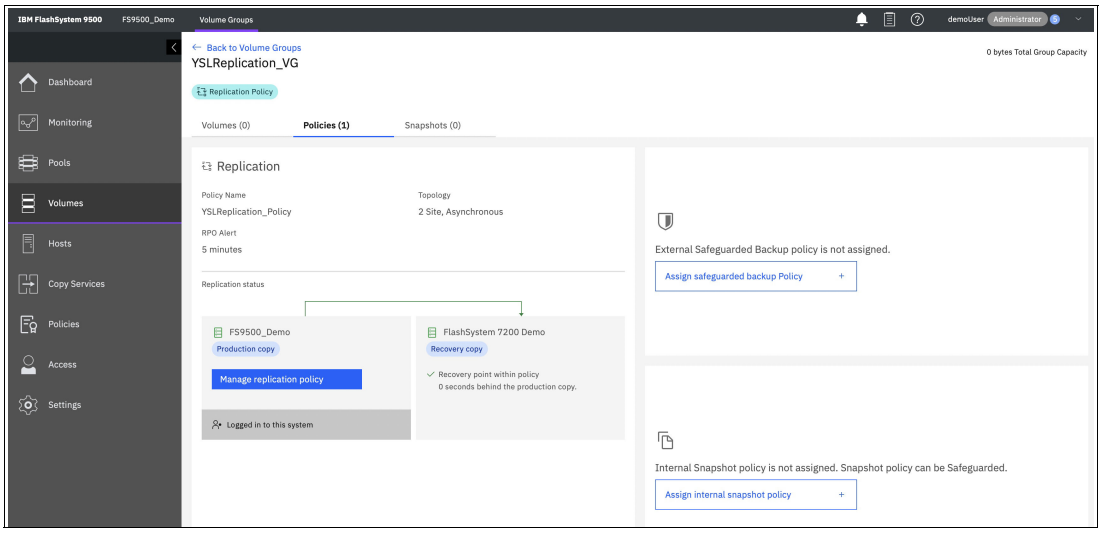


Figure 10-186 Verify that the replication policy is correctly assigned

7. When you access the secondary site, an IBM FlashSystem 7200, note that the volume group was automatically created with the same replication policy, as shown in Figure 10-187 on page 932. Any volumes added to this volume group and its linked child pool are replicated to the secondary site.

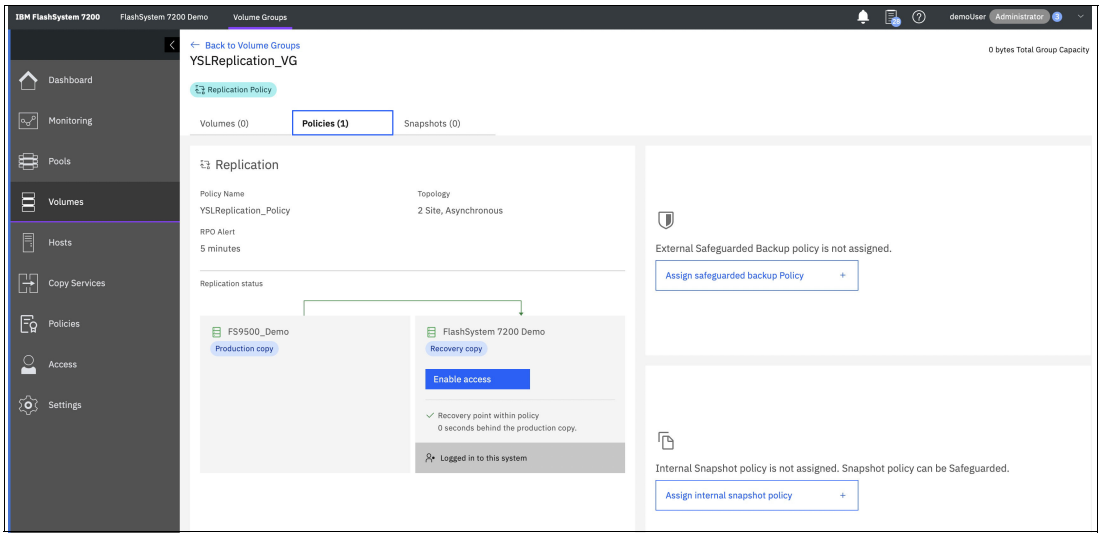


Figure 10-187 Verify that the replication policy is correctly assigned on the target system

Creating replicated volumes

To create replicated volumes perform the following steps:

1. On the primary machine, an IBM FlashSystem 9500, access the **Volumes** menu from the left panels, as shown in Figure 10-188.

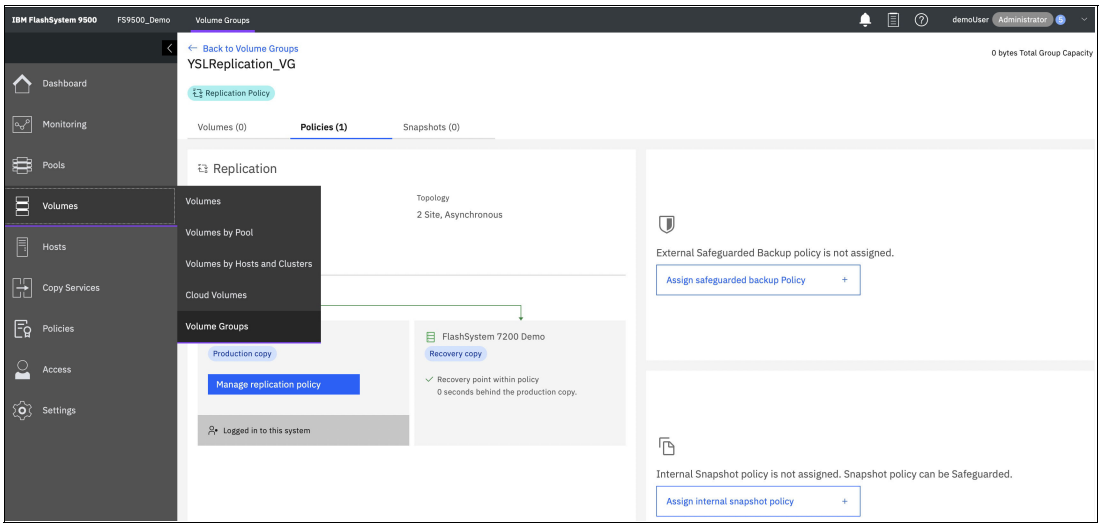


Figure 10-188 Volume panel

2. Any existing volumes are listed. To create new volumes, click **Create Volumes** in the top menu, as shown in Figure 10-189 on page 933.

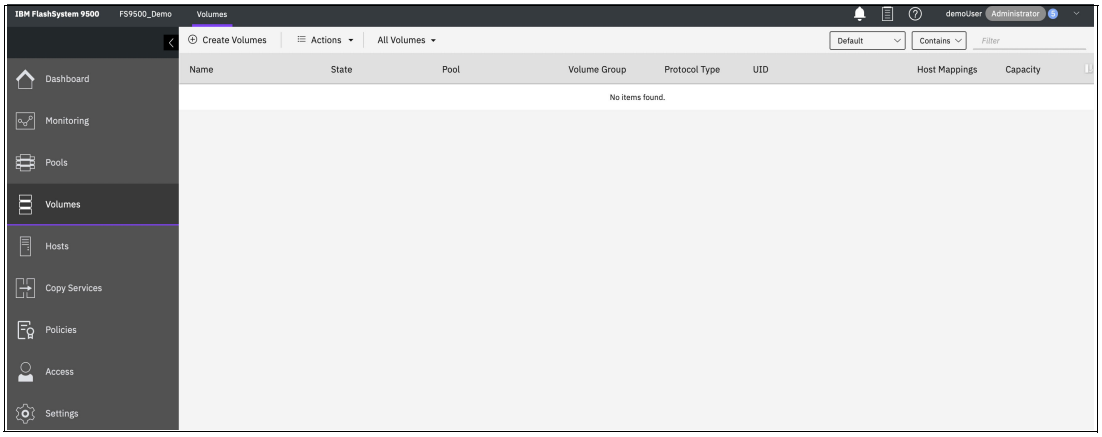


Figure 10-189 Create Volume panel

3. Select the appropriate linked child pool and volume group, and then click the **Define Volume Properties** button. Customize the properties of the volumes according to your preferences, as shown in Figure 10-190.

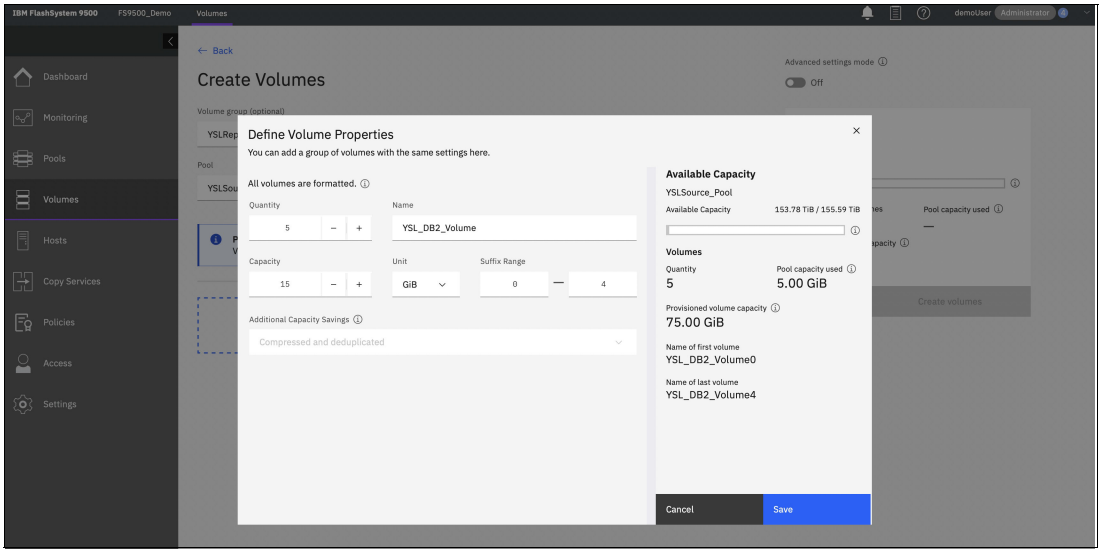


Figure 10-190 Define Volume Properties

4. After the volume properties are defined, click the **Create volume** button, as shown in Figure 10-191 on page 934.

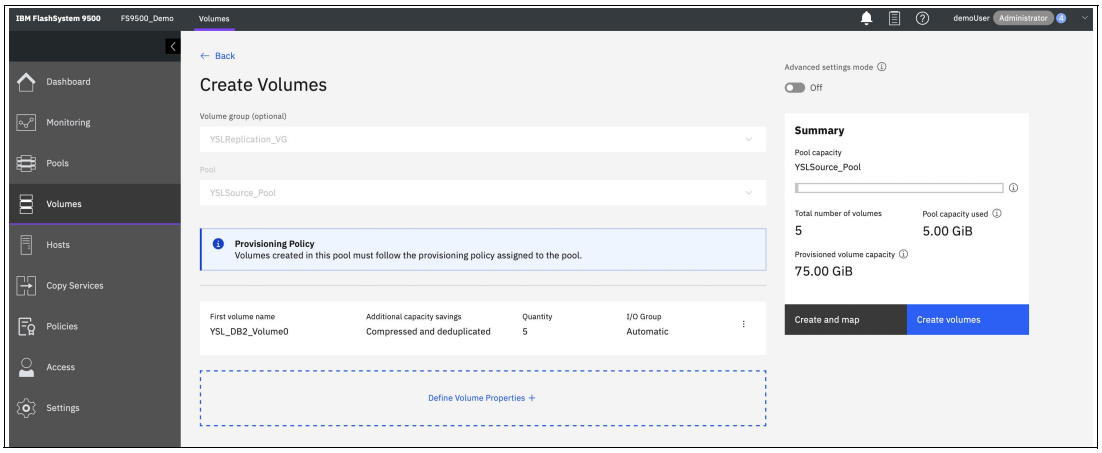


Figure 10-191 Create Volumes

5. A new window might open, which displays the progress of the action. After the volumes are successfully created, you can close this window, as shown in Figure 10-192.

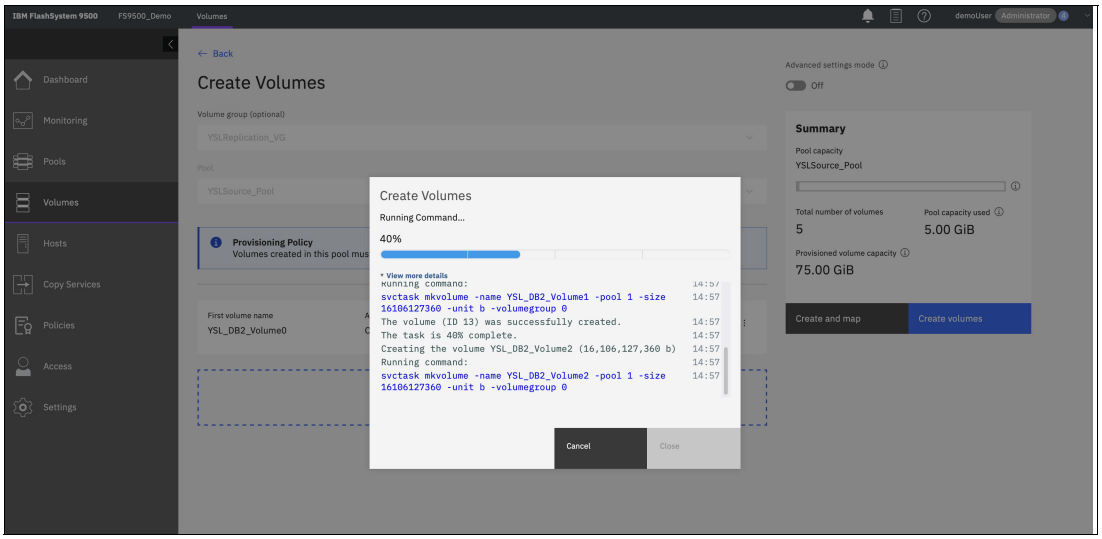


Figure 10-192 Volume creation progress bar

6. On the secondary site, navigate to the **Volume Group** tab. The system has automatically created the volumes, as shown in Figure 10-193 on page 935.

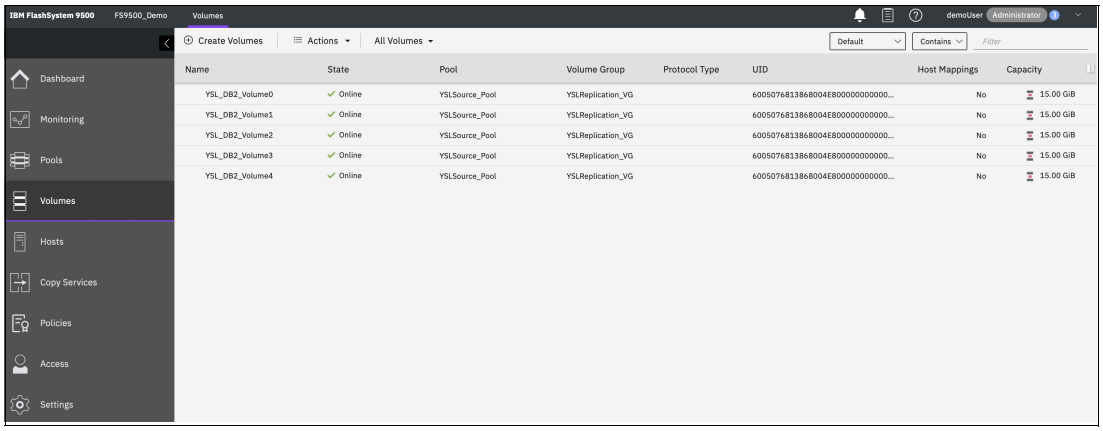


Figure 10-193 List volumes

Using targets of replicated volumes

Although the target volumes can be mapped to hosts, it's important to note that these volumes remain in a read-only state during the replication relation. However, there is an option to **Enable Access** from the secondary site. This can be done in the **Policy** tab of the volume group in the secondary site, as shown in Figure 10-194.

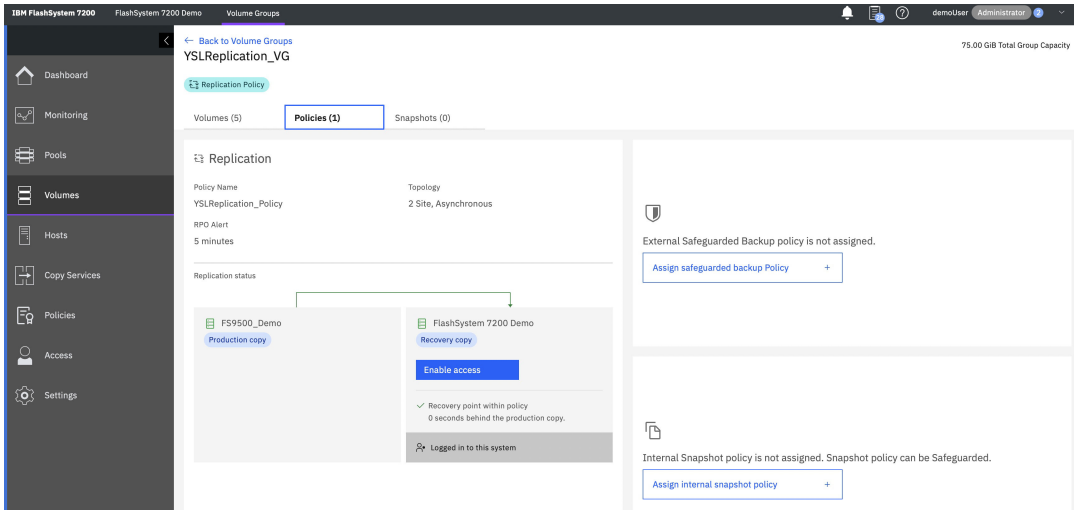


Figure 10-194 Enable access from the secondary site

When you enable access to the recovery copy, it suspends replication, which permits host access and configuration changes for each independent copy. Click the **Enable** button, as shown in Figure 10-195 on page 936.

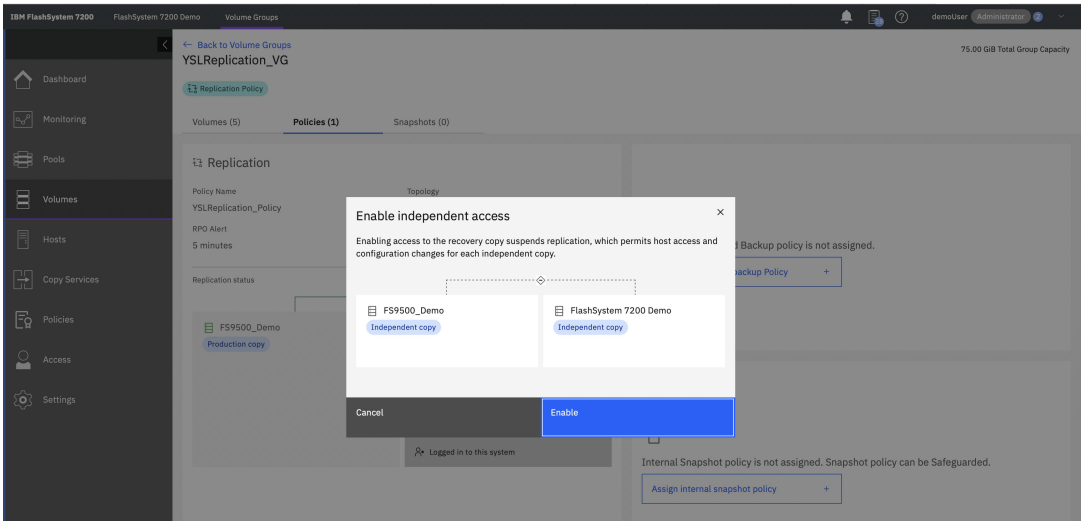


Figure 10-195 Enable access to the recovery copy

You can **Restart Replication.** Changes made to this copy are not replicated until replication is restarted, as shown in Figure 10-196.

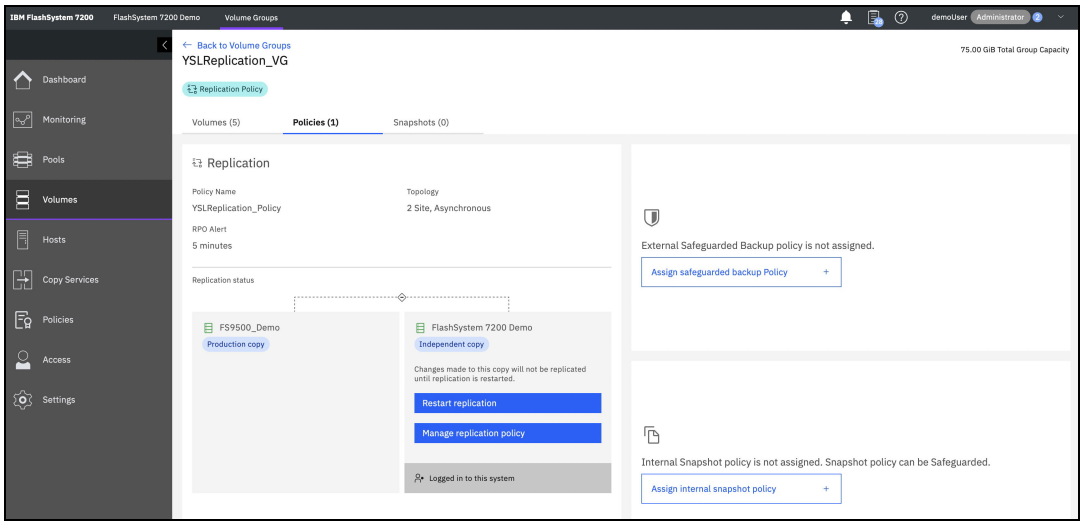


Figure 10-196 Restart replication from the secondary site

Reversing replication

After enabling access on the target and suspending the relation, you can proceed to restart the replication while selecting the desired direction for the relation.

1. To restart the relationship, connect to the storage system that will act as the master for the relationship. In the **Policy tab**, locate and click the **Restart Replication** button, as shown in Figure 10-197 on page 937.

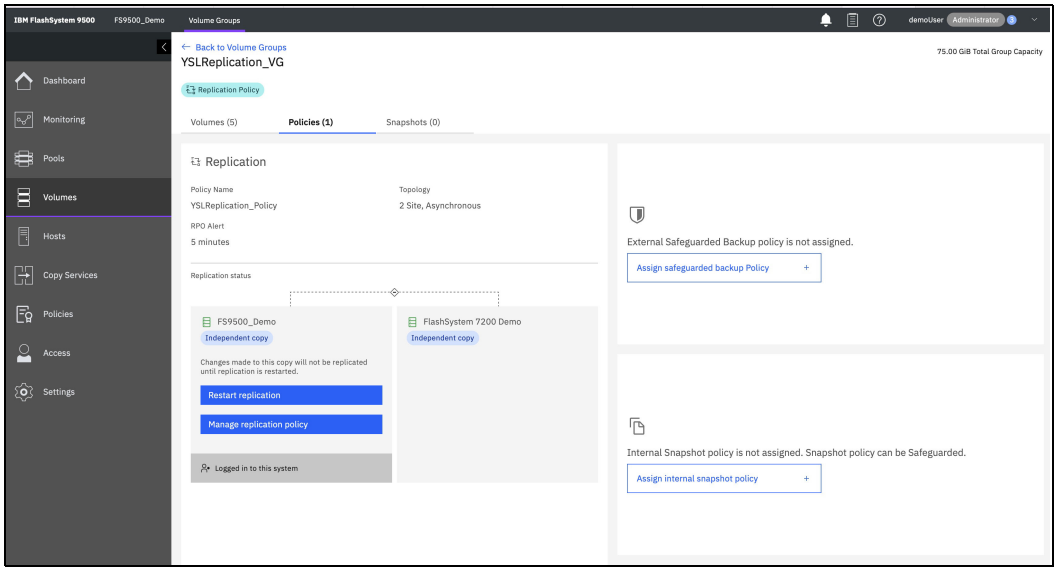


Figure 10-197 Restart replication from the primary site

2. In the following screen, you will be able to review both the previous and preview of the copy direction. After you confirm your choice, click the **Enable** button to initiate the restart of the replication, as shown in Figure 10-198.

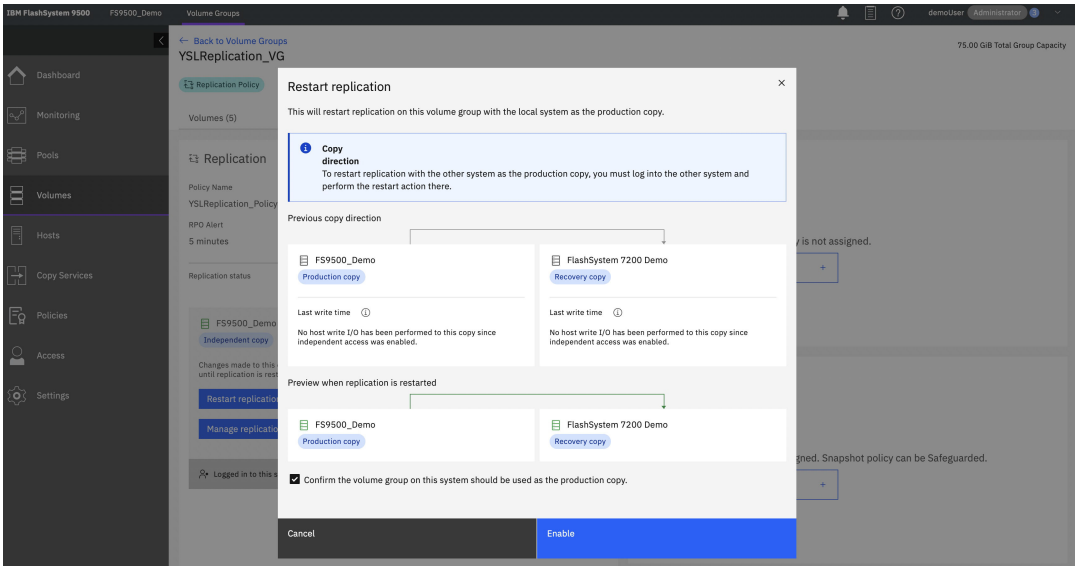


Figure 10-198 Enable Restart replication

3. You will notice that the initial copy is incomplete and the replication status is Red, as shown in Figure 10-199 on page 938.

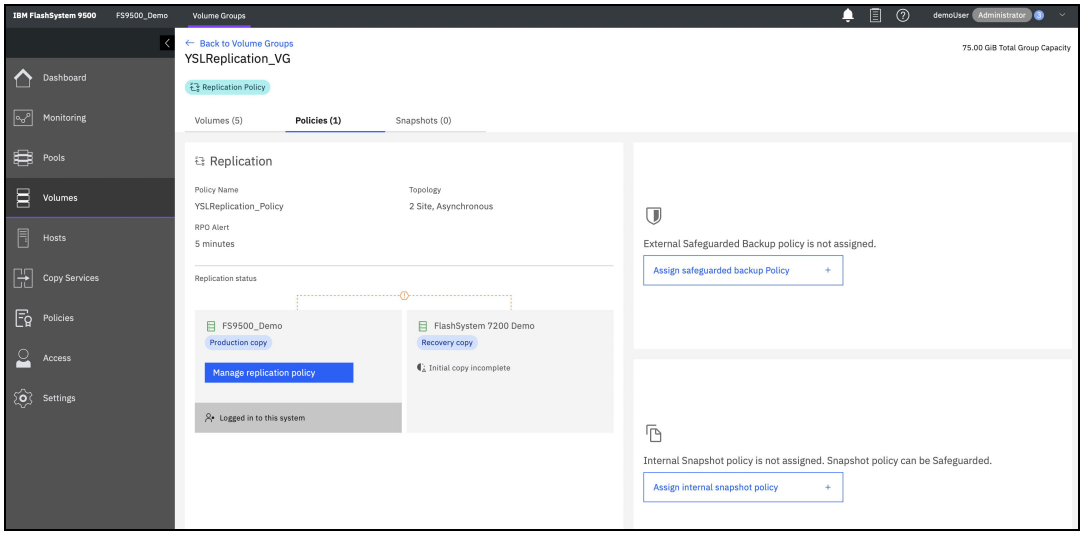


Figure 10-199 Initial copy incomplete status

4. Wait until the initial copy is completed to verify that the recovery point within the policy is 0 seconds behind the production copy, as shown in Figure 10-200.

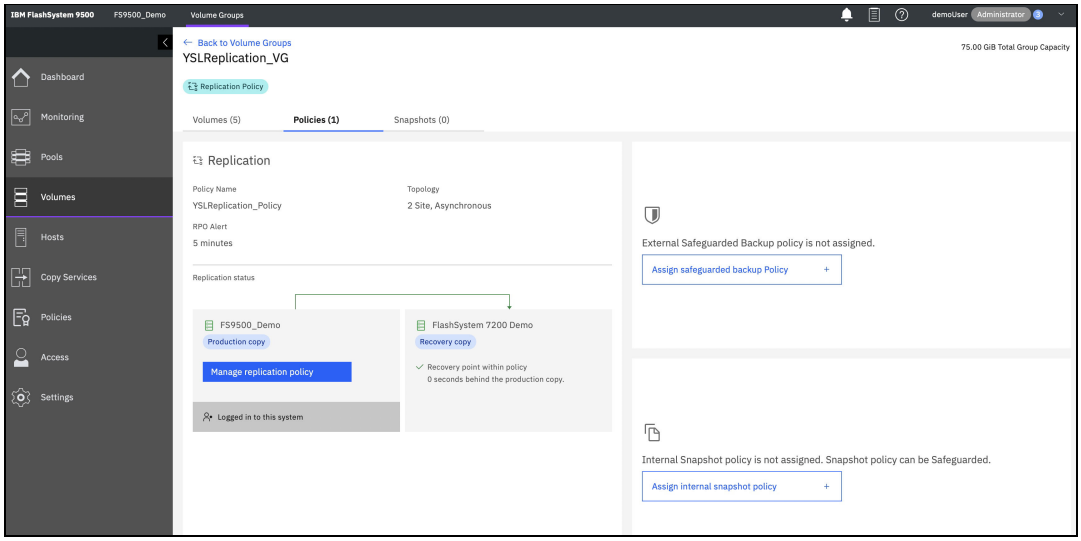


Figure 10-200 Verify that the recovery point within the policy is 0 seconds behind the production copy

10.17 Converting Global Mirror to policy-based replication

For IBM Storage Virtualize version 8.6.0 and later, if you use Global Mirror for data replication between two partnered systems, it is possible to convert the existing setup to policy-based replication. During this conversion process, the remote-copy configuration can be retained for a volume to ensure that a synchronized copy remains available on the disaster recovery (DR) system without any downtime.

Ensure that the following prerequisites are met before configuring a volume, which is currently part of a Global Mirror relationship, to use policy-based replication:

- If the relationship is part of a consistency group, the volume cannot move between I/O groups when the policy-based replication is defined. Use the **movevdisk** command to move the volume as needed.
- The relationship must be either Metro Mirror or Global Mirror.
- The volume being migrated should be the primary volume within the Metro Mirror or Global Mirror relationship.
- Both remote copy and policy-based replication are supported simultaneously on the same partnership, operating on different volumes.
- Volumes using Global Mirror non-cycling and cycling relationships and consistency groups can be manually migrated to use policy-based replication.
- Remote Copy features such as 3-site partnerships, nondisruptive volume migration, and Metro Mirror itself cannot be directly migrated to policy-based replication due to their more complex configuration requirements.
- The remote copy configuration must be removed from a volume before it can be added to a volume group with a replication policy.
- A full resynchronization of the volume is performed.
- No associated change volumes can be linked to the primary volume.
- During the migration process, do not change the direction of the remote-copy relationship or transform it into a secondary relationship.
- The primary volume cannot be designated as a recovery volume for policy-based replication.
- You cannot configure remote copy on a policy-based replication volume. Therefore, if any remote copy replication configuration exists, it must be removed before replicating from the remote copy secondary volumes.
- A volume within a Metro Mirror or Global Mirror relationship can have policy-based replication configured only if it belongs to the same I/O group specified in the replication policy. If it does not match, you can move the volume to the appropriate I/O group using the **movevdisk** command.
- The existing secondary volumes can be retained as a point-in-time copy for disaster recovery until the new recovery copy is established with policy-based replication.
- If you keep a disaster recovery copy, ensure that you have enough resources available on the recovery system to accommodate both sets of copies.

Before running the **movevdisk** command to transfer a volume between I/O groups, several conditions must be satisfied:

- The relationship state must be `consistent_synchronized`. The data in the source and target volumes of the relationship must be fully synchronized and consistent. Resolve any pending changes or discrepancies before proceeding with the volume movement.
- The relationship cannot be in a consistency group. Consistency groups are a collection of relationships that need to maintain data consistency as a group. If the relationship is part of a consistency group, it cannot be moved independently. Restrict the volume movement to relationships that are not associated with any consistency group.
- The relationship type must be Metro Mirror or Global Mirror. The **movevdisk** command is designed specifically for volumes involved in Metro Mirror or Global Mirror relationships. These relationship types enable synchronous or asynchronous replication of data between primary and secondary volumes. Only volumes associated with these types of relationships are eligible for the move operation.

- The relationship must not have a change volume associated with the primary volume. In some replication scenarios a change volume is used to track modifications made to the primary volume. If the relationship has an active change volume associated with the primary volume, the move operation cannot proceed. The change volume should be removed or detached before you start the volume transfer.
- The volume being moved must be the primary volume in the Metro Mirror or Global Mirror relationship. When you move a volume, the volume must be the primary volume in the Metro Mirror or Global Mirror relationship. The primary volume is the source volume where the original data resides, while the secondary volume is the target for replication. Moving the primary volume ensures the appropriate replication of data to the new I/O group.

Convert Global Mirror to policy-based replication

To convert from Global Mirror (GM) replication to policy-based replication, follow these steps:

- Enable policy-based replication on the partnership:
  - Establish a mutual SSL certificate exchange between systems to enable REST API access.
  - Use the GUI for configuration of the existing partnership for policy-based replication.
- Establish link pools between systems and assign necessary provisioning policies:
  - Link pools can be established from either system.
  - The GUI provides a method for configuring pool links and assigning provisioning policies.
- Create replication policies:
  - Depending on the current configuration, it might be necessary to create multiple replication policies. For instance, if Global Mirror with Change Volumes (GMCV) is used with different cycling times, different recovery point objectives can be specified for various sets of volumes.
  - Policy-based replication currently supports replication between two I/O groups per system. If more than two I/O groups are using remote copy, move volumes to the I/O groups defined in the replication policies.
- Create volume groups and assign replication policies:
  - Ensure each consistency group has a corresponding volume group.
  - For policy-based replication, volumes must be placed within volume groups, so one or more volume groups must be created for any previously independent relationships.
- Move volumes into volume groups:
  - Prior to moving volumes into a volume group with a replication policy, remove the remote copy configuration from the volumes.
  - If you want to retain existing secondary volumes, stop relationships and consistency groups with **`-access`** to make the secondary volumes accessible for disaster recovery before removing the remote copy configuration. Do not force delete relationships.
  - Move volumes into volume groups and allow sufficient time for the initial synchronization process to complete.

Convert GM to policy-based replication using the GUI

To convert from Global Mirror replication to policy-based replication for replicated volumes using GUI, follow these steps:

1. Update the existing partnership to support policy-based replication. Access the local system's management interface and navigate to **Copy Services** → **Partnerships and Remote Copy**, as shown in Figure 10-201.

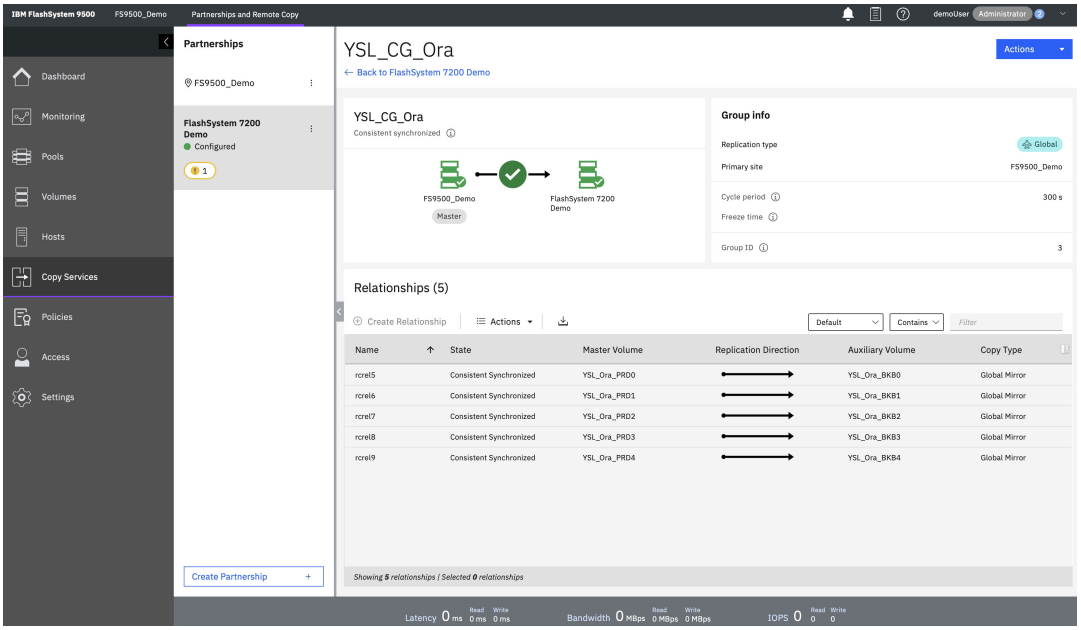


Figure 10-201 Update the existing partnership

2. Choose the current remote copy partnership from the left navigation and select **Actions** → **Partnership Properties**, as shown in Figure 10-202.

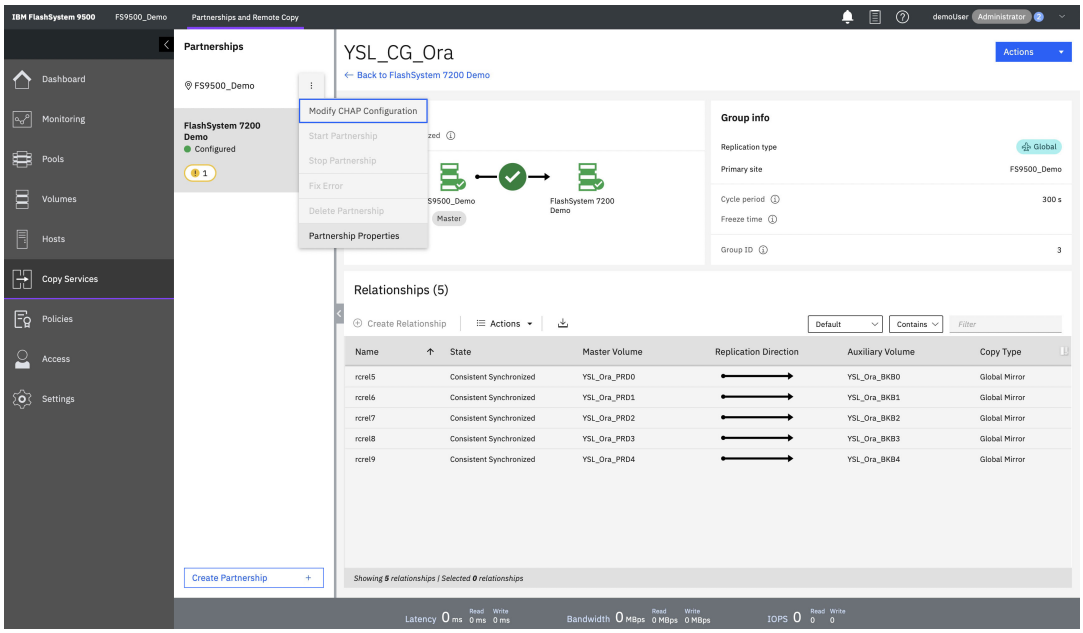


Figure 10-202 Partnership properties

a. On the **Partnership Properties** page, enable policy-based replication to retrieve the certificate from the remote system. The management interface automatically creates a truststore on the local system to store the remote system's certificate.

b. View the remote system certificate to verify it.

c. Save the changes.

3. Repeat steps 1 and 2 on the remote system within the partnership:

   a. Enable policy-based replication on the remote system as well.

   b. Create a truststore for each system in the partnership if using the command line interface.

4. After the remote system is enabled for policy-based replication, confirm that the partnership can be configured to use policy-based replication.

   a. On either of the systems, go to **Copy Services** → **Partnerships and Remote Copy** and select the respective partnership from the left navigation.

   b. Verify the message `This partnership is ready for use with policy-based replication` to ensure configuration is successful.

   c. Verify that the Global Mirror replication is correctly converted to policy-based replication, as shown in Figure 10-203.

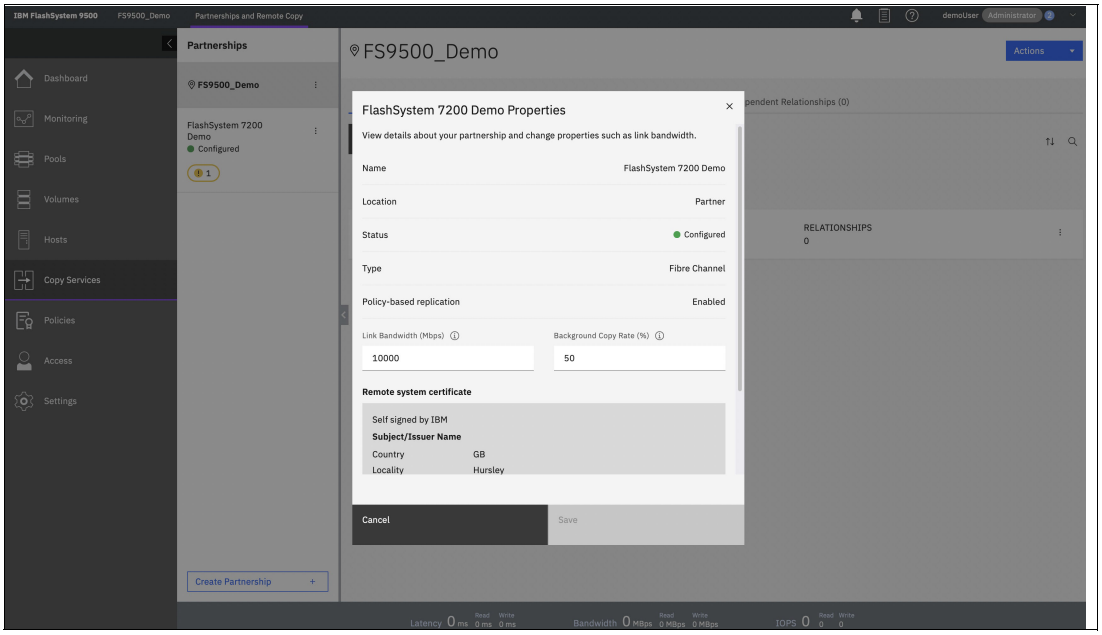


Figure 10-203 Verify that GM replication is converted to policy-based replication

5. Configure a provisioning policy for each linked pool, create one or more replication policies, create an empty volume group for each consistency group and independent relationships and assign a replication policy as described in section “Configuring policy-based replication using GUI” on page 923

Remove remote copy configuration

To remove the remote copy configuration, consider the following points and determine if you want to retain existing secondary volumes as a point-in-time copy for disaster recovery while establishing the new recovery copy with policy-based replication.

If you choose to keep the disaster recovery copy, ensure that sufficient capacity is available on the recovery system to accommodate both sets of copies.

Retaining existing volumes provides data protection in case of an outage and allows you to verify replicated data on the recovery system after configuring policy-based replication.

To remove the remote copy configuration, follow these steps:

1. Access the primary system's management interface and navigate to **Copy Services** → **Partnerships and Remote Copy.**
2. Select the **Consistency Groups** tab and verify that the current state of the consistency group is `Consistent Synchronized` for a Global Mirror consistency group or `Consistency Copying` for a Global Mirror consistency group with change volumes, as shown in Figure 10-204.

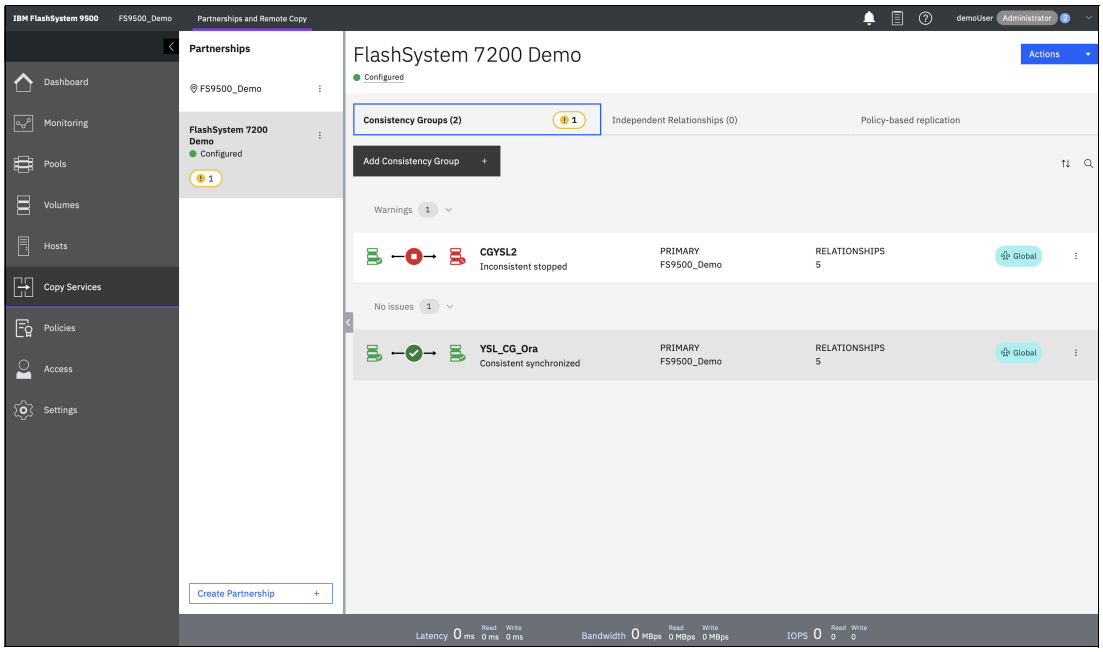


Figure 10-204 Verify the current state of the consistency group

3. Select **Actions** → **Stop Group.** On the **Stop Remote-Copy Consistency Group** page, choose the option **Allow secondary read/write access** to retain the secondary volumes as a disaster recovery copy, as shown in Figure 10-205 on page 944.

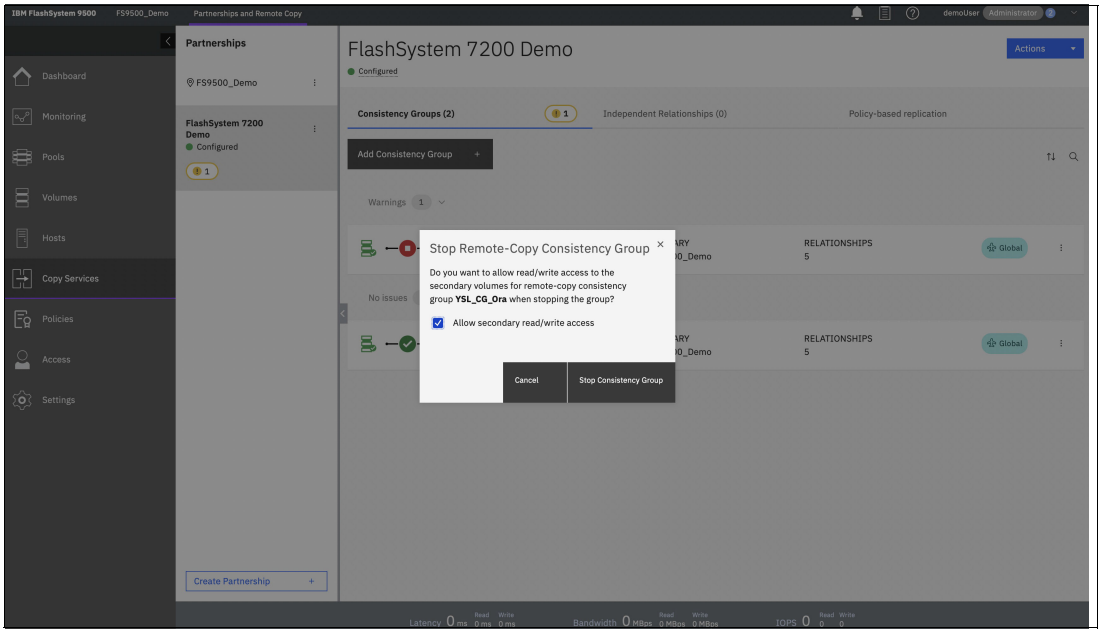


Figure 10-205 Stop Remote-Copy consistency group

4. Click **Stop Consistency Group**. The state of the consistency group changes to `Idling`, as shown in Figure 10-206.
5.

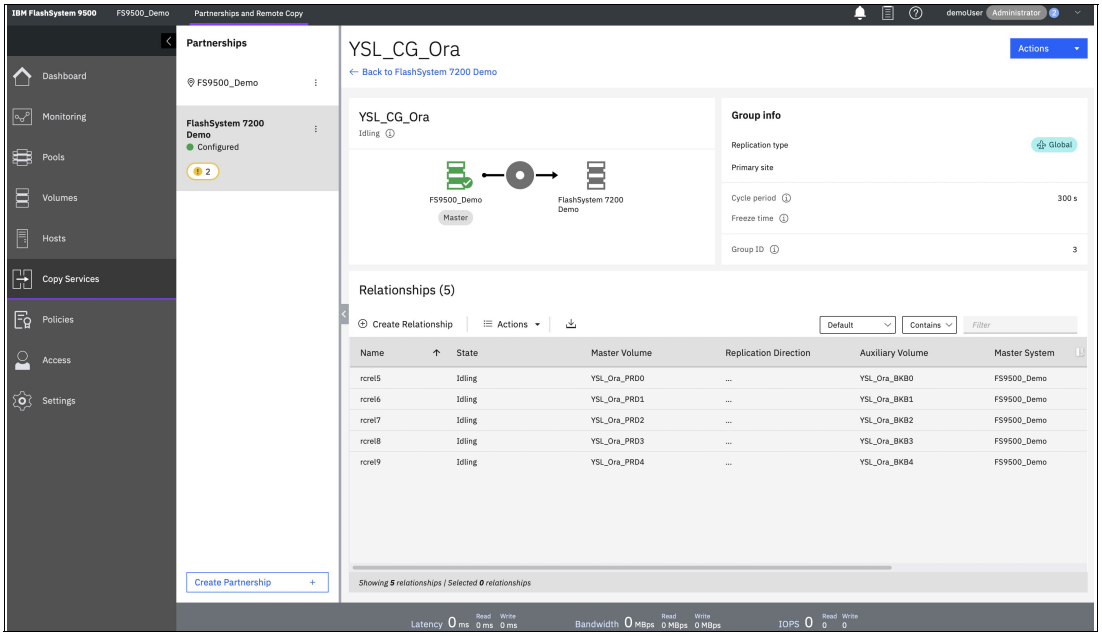


Figure 10-206 State of the consistency group to "Idling"

6. In the **Relationships** section, select all the relationships within the consistency group.
7. Right-click the selected relationships and choose **Delete**, as shown in Figure 10-207 on page 945.

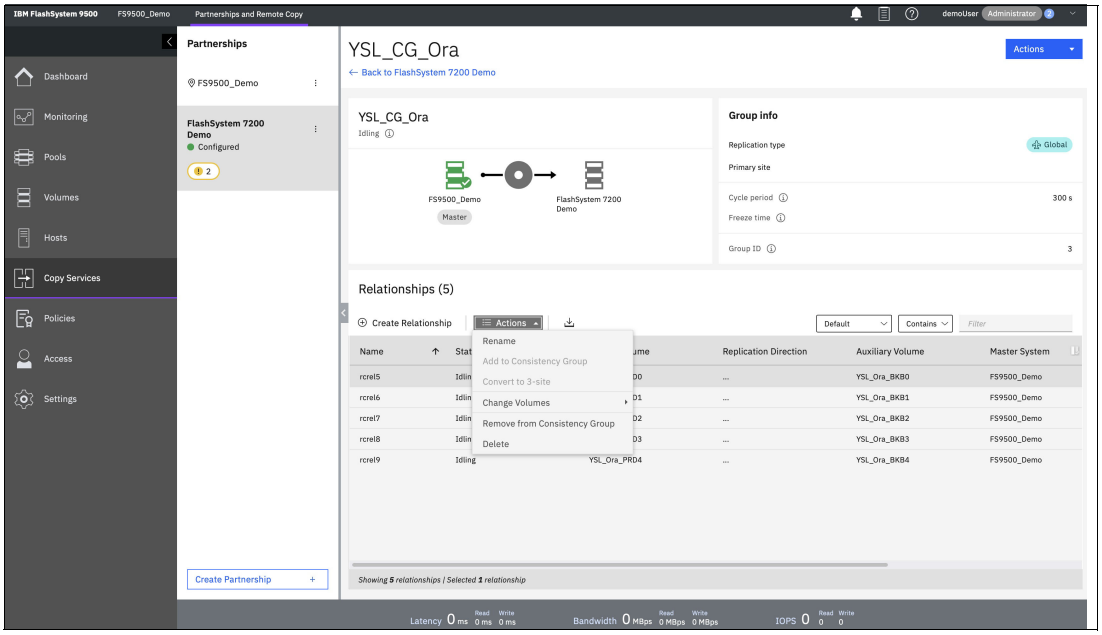


Figure 10-207 Delete relationship

8. On the **Delete Relationship** page, verify the number of relationships being deleted. Verify that the checkbox **Delete the relationship even when the data on the target system is not consistent** is cleared. This cleared checkbox allows the secondary volumes to be retained for disaster recovery until a new recovery point is established using policy-based replication, as shown in Figure 10-208.

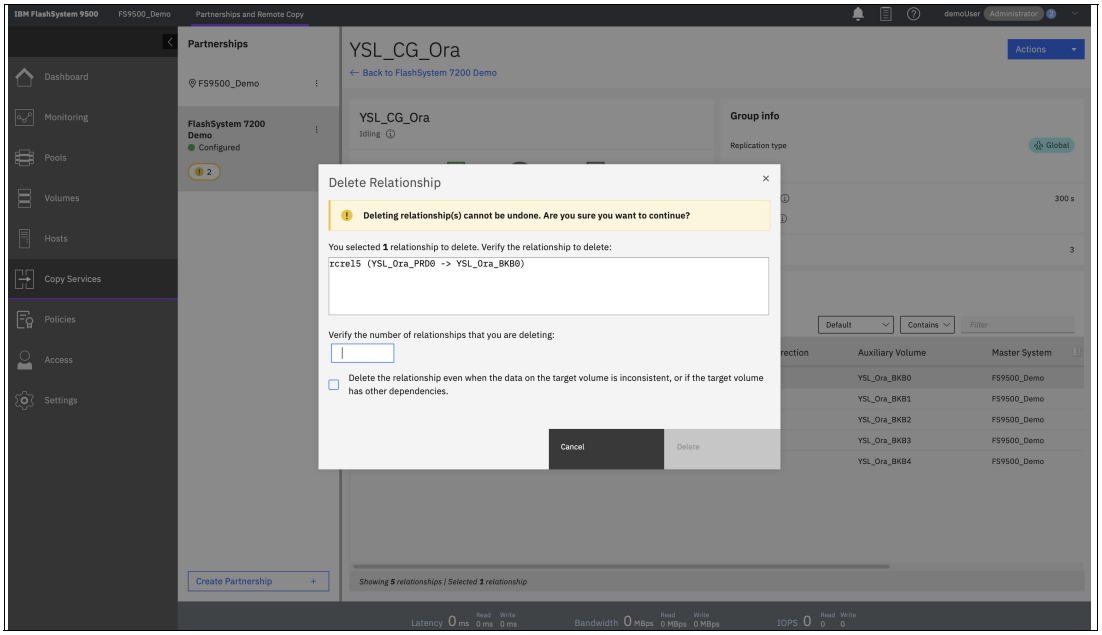


Figure 10-208 Confirm relationship deletion

9. Click **Delete**.

   – After the relationships are deleted from the consistency group, select **Actions** → **Delete Group** to complete the removal process, as shown in Figure 10-209 on page 946.

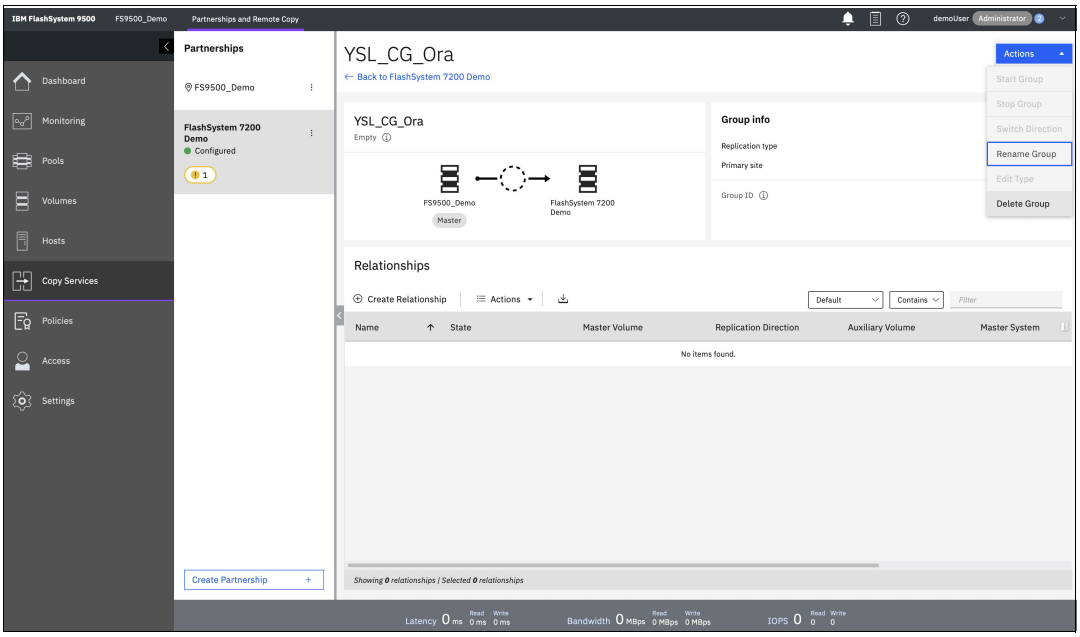


Figure 10-209 Delete group

10.18 Monitoring options for policy-based replication

To ensure effective monitoring of policy-based replication and maintain the integrity of your replication environment, several options are available. These options encompass monitoring volume groups, logging, performance statistics, change volume capacity reporting, and volume limits.

Monitoring volume groups

Monitoring volume groups is a fundamental aspect of policy-based replication monitoring. By closely monitoring volume groups, you can monitor their replication status, health, and overall performance. This allows you to promptly identify any issues or discrepancies and take appropriate action to maintain the desired replication outcomes.

Monitoring volume groups can be done from the main volume groups screen.

To view the replication status, follow these steps:

1. Navigate to **Volumes** → **Volume Groups**.
2. Under this section, you can see a high-level summary of all the volume groups, as shown in Figure 10-210 on page 947.

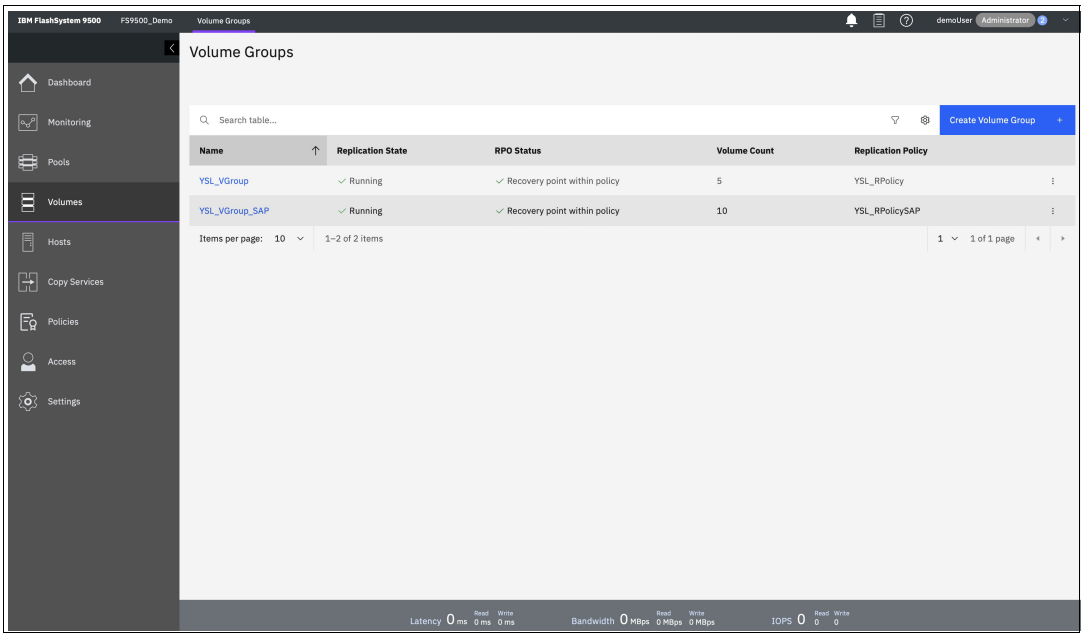


Figure 10-210 Monitoring volume groups

3. For more detailed replication status and actions related to a specific volume group, perform the following:

   a. Go to **Volumes** → **Volume Groups**.

   b. Select the desired volume group.

   c. Within the group, click **Policies**.This will display comprehensive replication status for the selected group and provide links for performing replication actions specific to that group, as shown in Figure 10-211.

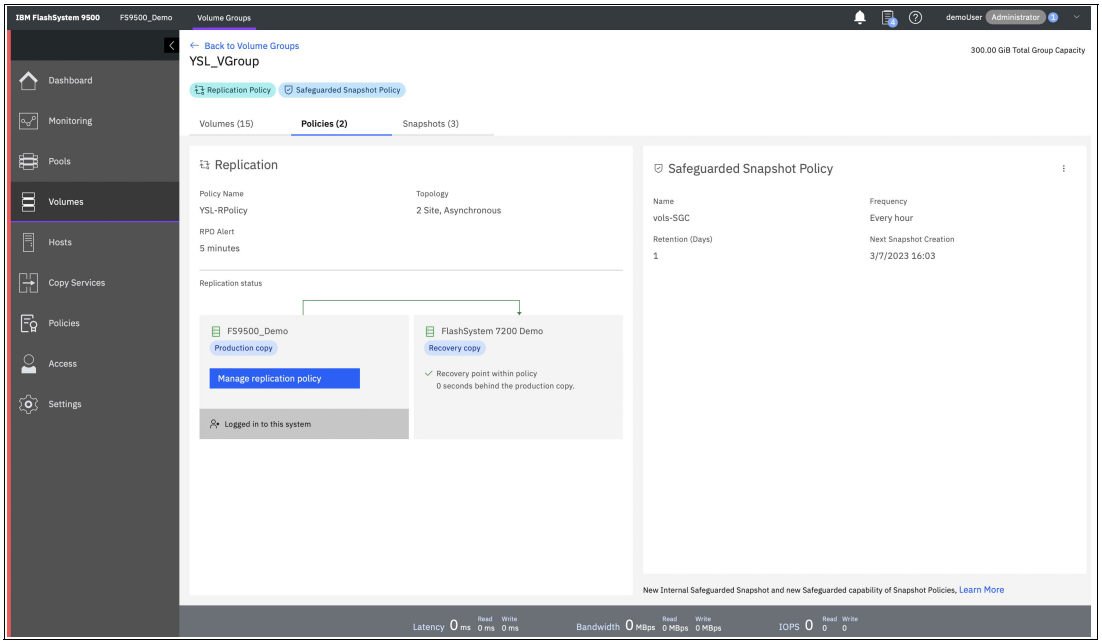


Figure 10-211 Monitor Volume groups policies

In the CLI view, there are two relevant commands:

1. The **lsvolumegroupreplication** command provides information about the status of the volume group in terms of the replication policy, including details such as production and recovery, healthy or unhealthy, and running or suspended status. For recovery copies, it also indicates the current recovery point and whether it is within the RPO specified by the policy.

Example 10-17 Information about the status of volume group using `ls`*replicationpolicy command*

```
IBM_FlashSystem:FS9500_Demo:demoUser>lsvolumegroupreplication
id name           replication_policy_id replication_policy_name
location1_system_name location1_replication_mode location1_within_rpo
location2_system_name location2_replication_mode location2_within_rpo link1_status
0 YSL_VGroup    0 YSL_RPolicy FS9500_Demo production
FlashSystem 7200 Demo recovery yes running
1  YSL_VGroup_SAP 1 YSL_RPolicySAP FS9500_Demo production
FlashSystem 7200 Demo recovery yes running
2 YSL_VGroup_DB2 2          YSL_RPolicy_DB2 FS9500_Demo production
FlashSystem 7200 Demo recovery yes running

IBM_FlashSystem:FS9500_Demo:demoUser>
```

Example 10-18 Details about the status of volume group using lsreplicationpolicy command

```
IBM_FlashSystem:FS9500_Demo:demoUser>lsvolumegroupreplication 2
id 2
name YSL_VGroup_DB2
replication_policy_id 2
replication_policy_name
YSL_RPolicy_DB2
local_location 1
location1_system_id 000002O4E1A0013A
location1_system_name FS9500_Demo
location1_replication_mode production
location1_status healthy
location1_running_recovery_point
location1_fixed_recovery_point
location1_within_rpo
location1_volumegroup_id 2
location1_sync_required
location1_sync_remaining
location1_previous_replication_mode
location1_last_write_time
location2_system_id 0000020420403BDA
location2_system_name FlashSystem 7200 Demo
location2_replication_mode recovery
location2_status healthy
location2_running_recovery_point 0
location2_fixed_recovery_point
location2_within_rpo yes
location2_volumegroup_id 2
location2_sync_required
location2_sync_remaining
location2_previous_replication_mode
location2_last_write_time
link1_status running
IBM_FlashSystem:FS9500_Demo:demoUser>
```

2. The **lsreplicationpolicy** command lists all the replication policies present on the system.

Example 10-19 Information about the status of volume group using lsreplicationpolicy command

```
IBM_FlashSystem:FS9500_Demo:demoUser>lsreplicationpolicy
id name rpo_alert topology volume_group_count location1_system_name
location1_iogrp_id location2_system_name location2_iogrp_id
0  YSL_RPolicy 300 2-site-async-dr 1 FS9500_Demo 0 FlashSystem 7200 Demo 0
1  YSL_RPolicySAP  120 2-site-async-dr 1 FS9500_Demo 0 FlashSystem 7200 Demo 0
2  YSL_RPolicy_DB2 60 2-site-async-dr 1 FS9500_Demo 0 FlashSystem 7200 Demo 0
```

Logging

By enabling detailed logging, you can capture and review important events and activities related to replication. This includes information about replication operations, errors, and other relevant system events. Analyzing the logs can provide valuable insights into the behavior and performance of a policy-based replication environment:

- The audit log captures system commands that were run to enforce the replication policy.
- This includes commands triggered by a partnered system through the REST API.
- Event log entries are generated on a per-volume group basis when the RPO is exceeded.
- Event log entries are also created on a per-partnership or per-volume group basis when issues arise that impact management actions.

  For more detailed log entries, go to **Access** → **Audit Log**, as shown in Figure 10-212.

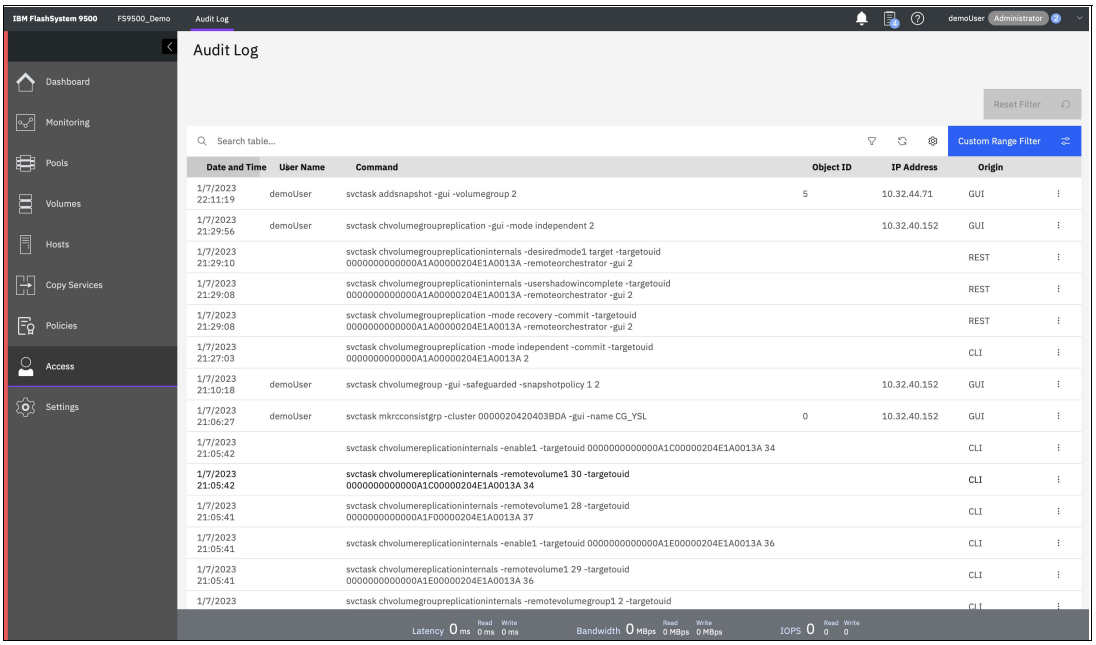


Figure 10-212 Audit logs

Performance statistics

Performance statistics are important for assessing the efficiency and effectiveness of a policy-based replication implementation. By monitoring performance statistics, you can gather data on replication throughput, latency, and other key metrics. By using this

information, you can identify potential bottlenecks or performance issues and make necessary adjustments to optimize the replication process:

- The volume group statistics provide information on the number of replicated operations and blocks, along with the average and worst recovery points during a specified period.
- The node statistics offer insights into the average usage of journal resources, and the peak utilization observed within the designated timeframe.
- For individual volumes, the average and worst recovery points are derived from the volume group statistics, ensuring accurate monitoring of replication progress and performance.

By using these enhanced statistics administrators can assess the replication efficiency and resource uage at different levels, including volume groups, nodes, and individual volumes. For more information refer to “Performance statistics” on page 949.

Volume limits

Volume limits are essential for monitoring the usage and allocation of resources within the policy-based replication setup. You can define thresholds and boundaries for the amount of data and resources that can be replicated. Monitoring these limits ensures that replication operations stay within the defined boundaries, preventing any potential resource overuse or imbalance.

- The **`lssystemlimits`** command helps understand what volume resources are used on the system.
- The CLI can show the number of objects that are configured and available on a system.
- The **`lssystemlimits`** command shows the remaining:
  - Volume copies.
  - Snapshots.
  - FlashCopy mappings.

Example 10-20 `lssystemlimits` *command*

```
IBM_FlashSystem:FS9500_Demo:demoUser>lssystemlimits
type configured max   available warning_threshold
volume_copies      34 15864 15830 0
volume_snapshots   20 64999 64931 0
flashcopy_mappings 7 15864 15857 0
flashcopy_groups   2 500   498 0
volumes            34 15864 15830 0
```

Change volume capacity reporting

Use the change volume capacity reporting to stay informed about the changes in volume capacities within the replication environment by ensuring that the allocated capacity meets the requirements of the replication policies. This helps in maintaining the desired replication performance and avoiding any capacity-related issues.

You can use the CLI to view status to view capcity-related information:

- protection_provisioned_capacity. The amount of capacity provisioned to store point-in-time copies. This includes FlashCopy snapshots and consistency protection for replication.

- protection_written_capacity. The amount of capacity written to store point-in-time copies before any data reduction. This includes volume snapshots and consistency protection for replication.

To see a list of vdisks, use the **lsvdisk** command. See Example 10-21.

Example 10-21 lsvdisk *command*

```
IBM_FlashSystem:FS9500_Demo:demoUser>lsvdisk
id name IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity type
FC_id FC_name RC_id RC_name vdisk_UID fc_map_count copy_count fast_write_state
se_copy_count RC_change compressed_copy_count parent_mdisk_grp_id
parent_mdisk_grp_name owner_id owner_name formatting encrypt volume_id volume_name
function volume_group_id volume_group_name protocol is_snapshot snapshot_count
volume_type replication_mode is_safeguarded_snapshot safeguarded_snapshot_count

12 DB2Vol0 0 io_grp0 online 1 YSL_Child_Pool 15.00GB  striped many  many
6005076813868004E80000000000061C 2 1 not_empty 0 no 1 0 DRP_Pool
no no 12 DB2Vol0 0 YSL_VGroup no 0 production no 0

13 DB2Vol1 0 io_grp0 online 1 YSL_Child_Pool 15.00GB  striped many  many
6005076813868004E80000000000061D 2 1 not_empty 0 no 1 0 DRP_Pool
no no 13 DB2Vol1 0 YSL_VGroup no 0 production no 0

14 DB2Vol2 0 io_grp0 online 1 YSL_Child_Pool 15.00GB  striped many  many
6005076813868004E80000000000061E 2 1 not_empty 0 no 1 0         DRP_Pool
no no 14 DB2Vol2 0 YSL_VGroup no 0 production no 0

15 DB2Vol3 0 io_grp0 online 1 YSL_Child_Pool 15.00GB  striped many  many
6005076813868004E80000000000061F 2 1 not_empty 0 no 1 0 DRP_Pool
no no 15 DB2Vol3 0 YSL_VGroup no 0 production no 0

16 DB2Vol4 0 io_grp0 online 1 YSL_Child_Pool 15.00GB  striped many  many
6005076813868004E800000000000620 2 1 not_empty 0 no 1 0         DRP_Pool
no no 16 DB2Vol4 0 YSL_VGroup no 0 production no 0

17 SAPHANA_Volume4_01 0 io_grp0 online 1 YSL_Child_Pool 10.00GB  striped 0
fcmap30 6005076813868004E80000000000062E 1 1 not_empty 0 no 1
0 DRP_Pool no          no      17 SAPHANA_Volume4_01 no           0 no 0
```

To see a list of volume groups, use the **lsvolumegroup** command. See Example 10-22.

Example 10-22 lsvolumegroup *command*

```
IBM_FlashSystem:FS9500_Demo:demoUser>lsvolumegroup
id name volume_count backup_status last_backup_time owner_id owner_name
safeguarded_policy_id safeguarded_policy_name replication_policy_id
replication_policy_name volume_group_type uid source_volume_group_id
source_volume_group_name parent_uid source_snapshot snapshot_policy_id
snapshot_policy_name
0  YSL_VGroup 5 off
0 YSL_RPolicy 42
0 predefinedsspolicy0
1  YSL_VGroup_SAP 10 off
1 YSL_RPolicySAP 43
0 predefinedsspolicy0
```

```
2  YSL_VGroup_DB2 4 off
2 YSL_RPolicy_DB2 44
1 predefinedsspolicy1
```

To see a list of pools, use the **lsmdiskgrp** command. See Example 10-23.

Example 10-23 lsmdiskgrp *command*

```
IBM_FlashSystem:FS950O_Demo:demoUser>lsmdiskgrp
id name           status mdisk_count vdisk_count capacity extent_size
free_capacity virtual_capacity used_capacity real_capacity overallocation warning
easy_tier easy_tier_status compression_active compression_virtual_capacity
compression_compressed_capacity compression_uncompressed_capacity
parent_mdisk_grp_id parent_mdisk_grp_name child_mdisk_grp_count
child_mdisk_grp_capacity type            encrypt owner_type owner_id owner_name
site_id site_name data_reduction used_capacity_before_reduction
used_capacity_after_reduction overhead_capacity deduplication_capacity_saving
reclaimable_capacity easy_tier_fcm_over_allocation_max provisioning_policy_id
provisioning_policy_name replication_pool_link_uid 0  DRP_Pool     online 1
0 155.59TB 1024 153.7OTB
0.00MB 1.56TB 1.62TB 0 80 auto balanced no
0.00MB 0.00MB 0.00MB 0 DRP_Pool 1 0.00MB parent no none
yes 0.00MB 1.68GB 1.56TB 0.00MB 0.00MB 100%
000000000000009B800000204E1A0013A 1  YSL_Child_Pool online 0 85 155.59TB 1024
153.7OTB 950.00GB 0.00MB 0.00MB 0 80 auto balanced no 0.00MB 0.00MB
0.00MB 0 DRP_Pool 0 0.00MB child_quotaless no none
yes 0.00MB 0.00MB 0.00MB 0.00MB 0.00MB 100% 1
capacity_optimized 000000000000009C500000204E1A0013A

IBM_FlashSystem:FS950O_Demo:demoUser>
```

To view the status of the system, use the **lssystem** command. See Example 10-24.

Example 10-24 lssystem *command*

```
IBM_FlashSystem:FS950O_Demo:demoUser>lssystem
id 00000204E1A0013A
name FS950O_Demo
location local
partnership
total_mdisk_capacity 155.6TB
space_in_mdisk_grps 155.6TB
space_allocated_to_vdisks 1.62TB
total_free_space 154.0TB
total_vdiskcopy_capacity 950.00GB
total_used_capacity 1.56TB
total_overallocation 0
total_vdisk_capacity 950.00GB
total_allocated_extent_capacity 1.64TB
statistics_status on
statistics_frequency 5
cluster_locale en_US
time_zone 14 Africa/Casablanca
code_level 8.6.0.0 (build 169.9.2306081121000)
console_IP 192.168.1.100:443
id_alias 00000204E1A0013A
```

```
gm_link_tolerance 300
gm_inter_cluster_delay_simulation 0
gm_intra_cluster_delay_simulation 0
gm_max_host_delay 5
email_reply y.largou@powerm.ma
email_contact youssef
email_contact_primary
email_contact_alternate
email_contact_location
email_contact2
email_contact2_primary
email_contact2_alternate
email_state stopped
inventory_mail_interval 1
cluster_ntp_IP_address 192.168.1.254
cluster_isns_IP_address
iscsi_auth_method none
iscsi_chap_secret
auth_service_configured yes
auth_service_enabled yes
auth_service_url
auth_service_user_name
auth_service_pwd_set no
auth_service_cert_set no
auth_service_type ldap
relationship_bandwidth_limit 25
tier tier_scm
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier0_flash
tier_capacity 155.59TB
tier_free_capacity 153.95TB
tier tier1_flash
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier_nearline
tier_capacity 0.00MB
tier_free_capacity 0.00MB
easy_tier_acceleration off
has_nas_key no
layer storage
rc_buffer_size 256
compression_active no
compression_virtual_capacity 0.00MB
compression_compressed_capacity 0.00MB
compression_uncompressed_capacity 0.00MB
cache_prefetch on email_organization POWERM
email_machine_address Boulevard Zerktouni
email_machine_city Casablanca
email_machine_state XX
email_machine_zip 27000
email_machine_country MA
```

```
email_from total_drive_raw_capacity 0
compression_destage_mode off
local_fc_port_mask
11111111111111111111111111111111111111111111111111111111111111111111
partner_fc_port_mask
11111111111111111111111111111111111111111111111111111111111111111111
high_temp_mode off
topology standard
topology_status
rc_auth_method none
vdisk_protection_time 15
vdisk_protection_enabled no
product_name IBM FlashSystem 9500
odx off
max_replication_delay 0
partnership_exclusion_threshold 315
gen1_compatibility_mode_enabled
ibm_customer
ibm_component
ibm_country
tier_scm_compressed_data_used 0.00MB
tier0_flash_compressed_data_used 0.00MB
tier1_flash_compressed_data_used 0.00MB
tier_enterprise_compressed_data_used 0.00MB
tier_nearline_compressed_data_used 0.00MB
total_reclaimable_capacity 0.00MB
physical_capacity 33.71TB
physical_free_capacity 33.71TB
used_capacity_before_reduction 0.00MB
used_capacity_after_reduction 1.68GB
overhead_capacity 1.56TB
deduplication_capacity_saving 0.00MB
enhanced_callhome on
censor_callhome on
host_unmap on
backend_unmap on
quorum_mode standard
quorum_site_id
quorum_site_name
quorum_lease short
automatic_vdisk_analysis_enabled on
callhome_accepted_usage no
safeguarded_copy_suspended no
protection_provisioned_capacity 665.00GB
protection_written_capacity 486.75MB
flashcopy_gui_enabled yes
snapshot_policy_suspended no
snapshot_preserve_parent no

IBM_FlashSystem:FS9500_Demo:demoUser>
```

10.19 Troubleshooting policy-based replication

Troubleshooting policy-based replication involves identifying and resolving issues related to the replication process based on predefined policies. When troubleshooting policy-based replication, you must diagnose any discrepancies or errors that might occur during the replication process. Issues can include connectivity problems, configuration mismatches, or performance issues that could impact the replication flow.

Effective troubleshooting often involves verifying the replication policy settings, ensuring that they are correctly configured and aligned with the intended replication requirements. Additionally, monitoring replication logs and status reports can provide valuable information to identify any potential issues or errors that need to be addressed.

Policy-based replication performance analysis

You can leverage the metrics gathered by the system to analyze performance issues in policy-based replication. These statistics are invaluable for troubleshooting and enhancing system performance.

You can use the **lsdumps** command to view a collection of files located within a specific directory for dumps on any of the nodes within the system.

Example 10-25 `lsdumps` *command*

```
IBM_FlashSystem:FlashSystem 7200 Demo:demoUser>lsdumps -prefix /dumps/iostats 0
id  filename
0   Nn_stats_78E3004-2_230702_201743
1   Nv_stats_78E3004-2_230702_201743
2   Nd_stats_78E3004-2_230702_201743
3   Ng_stats_78E3004-2_230702_201743
4   Nm_stats_78E3004-2_230702_201743
5   Ng_stats_78E3004-2_230702_202244
6   Nd_stats_78E3004-1_230702_202244
7   Nd_stats_78E3004-2_230702_202244
8   Nn_stats_78E3004-2_230702_202244
9   Nm_stats_78E3004-1_230702_202244
10  Nm_stats_78E3004-2_230702_202244
11  Nv_stats_78E3004-1_230702_202244
12  Ng_stats_78E3004-1_230702_202244
13  Nv_stats_78E3004-2_230702_202244
(…)
157 Ng_stats_78E3004-1_230702_213755
158 Nn_stats_78E3004-1_230702_213755
159 Nv_stats_78E3004-1_230702_213755
160 Nd_stats_78E3004-1_230702_213755
161 Nv_stats_78E3004-2_230702_213755
162 Ng_stats_78E3004-2_230702_213755
IBM_FlashSystem:FlashSystem 7200 Demo:demoUser>
```

- To access the statistics files, retrieve them from the `/dumps/iostats` directory located on the configuration node, you can employ secure copy (scp) to transfer them from the configuration node to your local workstation.

Example 10-26 `Retrieve /dumps/iostats` *directory*

```
scp demoUser@192.168.1.101/dumps/iostats/Nv_stats_78E3004-2_230702_21375520
/tmp/YSLRedbooks
```

Statistics

The system collects statistics that are specifically for policy-based replication.

Volume group statistics

Policy-based replication operates at the volume group level. Consequently, the majority of replication statistics are recorded at the volume group level. These statistics provide insights into the performance of the volume groups.

Table 10-26 presents the statistics collected for a volume group.

Table 10-26 Volume group statistics

Statistic name	Description
avg_recovery_point (rarp)	Average recovery point in seconds for the volume group since the last statistics collection.
replication_writes (rnrw)	Cumulative number of replication writes to the target sent by this node for the volume group.
worst_recovery_point (rwrp)	Worst recovery point in seconds for the volume group since the last statistics collection.
replicated_blocks (rnrb)	Cumulative number of blocks replicated by this node for the volume group.

Example 10-27 Volume group statistics

```
<?xml version="1.0" encoding="utf-8" ?>
<diskStatsColl
xmlns="http://ibm.com/storage/management/performance/api/2021/04/volumegroupstats"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://ibm.com/storage/management/performance/api/2021/04/volu
megroupstats schema/SVCPerfStatsG.xsd" scope="node" id="node1"
cluster="FlashSystem 7200 Demo" node_id="0x0000000000000001"
cluster_id="0x0000020420403bda" sizeUnits="512B" timeUnits="msec"
contains="volumegroupStats" timestamp="2023-07-04 00:46:49" timezone="GMT+10:00"
timestamp_utc="2023-07-03 14:46:49">
<volumegroup idx="0" rarp="0" rwrp="0" rnrw="0" rnrb="0" name="YSL_VGroup">
</volumegroup>
</diskStatsColl>
```

Node statistics

Replication-based policy adds to the existing node XML statistics file to capture node-level values. These statistics offer information about the performance of individual nodes involved in the replication process.

Table 10-27 displays the statistics collected for node statistics.

Table 10-27 Node statistics

Statistic name	Description
journal_resource_usage (rjru)	Percentage of journal resources utilized for replicating data, averaged across all threads. Expressed as a value out of 10000, representing a percentage rounded to two decimal places.
highest_journal_resource_usage (rjrw)	Highest percentage of journal resources utilized across all threads. Expressed as a value out of 10000, representing a percentage rounded to two decimal places.

Example 10-28 shows an output of the statistics that are collected for a node. Table 10-27 on page 956 describes the names of the statistics.

Example 10-28 Nodes statistics

```
<?xml version="1.0" encoding="utf-8" ?> <diskStatsColl
xmlns="http://ibm.com/storage/management/performance/api/2006/01/nodeStats"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://ibm.com/storage/management/performance/api/2006/01/node
Stats schema/SVCPerfStatsN.xsd" scope="node" id="node2" cluster="FlashSystem 7200
Demo" node_id="0x0000000000000002" cluster_id="0x0000020420403bda" sizeUnits="1B"
timeUnits="msec" contains="nodeStats" timestamp="2023-07-02 21:32:54"
timezone="GMT+10:00" timestamp_utc="2023-07-02 11:32:54">
<cpu busy="36744813" limited="0" system="36744813" comp="0"/>
<cpu_core id="0" system="37156839" comp="0"/>
<cpu_core id="1" system="42883316" comp="0"/>
<cpu_core id="2" system="41987692" comp="0"/>
<cpu_core id="3" system="44506708" comp="0"/>
<cpu_core id="4" system="44226347" comp="0"/>
<cpu_core id="5" system="43209215" comp="0"/>
<cpu_core id="6" system="41889503" comp="0"/>
<cpu_core id="7" system="37378663" comp="0"/>
<cpu_core id="8" system="33919738" comp="0"/>
<cpu_core id="9" system="28726166" comp="0"/>
<cpu_core id="10" system="29616987" comp="0"/>
<cpu_core id="11" system="34229583" comp="0"/>
<cpu_core id="12" system="31353294" comp="0"/>
<cpu_core id="13" system="32749398" comp="0"/>
<cpu_core id="14" system="27339384" comp="0"/>
<cpu_core id="15" system="0" comp="0"/>
<dimm id="1" loc="CPU1_C0" manu="Samsung" sn="256D07FB" ce="0"/>
<dimm id="2" loc="CPU1_C1" manu="Samsung" sn="256D07D4" ce="0"/>
<dimm id="3" loc="CPU1_B0" manu="Samsung" sn="256D07F5" ce="0"/>
<dimm id="4" loc="CPU1_B1" manu="Samsung" sn="256D132C" ce="0"/>
<dimm id="5" loc="CPU1_A0" manu="Hynix" sn="43346DD3" ce="0"/>
<dimm id="6" loc="CPU1_A1" manu="Samsung" sn="256D07DA" ce="0"/>
<dimm id="7" loc="CPU1_D1" manu="Samsung" sn="256D07EC" ce="0"/>
<dimm id="8" loc="CPU1_D0" manu="Hynix" sn="43346DD5" ce="0"/>
<dimm id="9" loc="CPU1_E1" manu="Samsung" sn="256D07D8" ce="0"/>
<dimm id="10" loc="CPU1_E0" manu="Samsung" sn="256D07DB" ce="0"/>
<dimm id="11" loc="CPU1_F1" manu="Samsung" sn="256D07DC" ce="0"/>
<dimm id="12" loc="CPU1_F0" manu="Samsung" sn="256D13CC" ce="0"/>
<dimm id="13" loc="CPU2_F0" manu="Samsung" sn="256D07D7" ce="0"/>
<dimm id="14" loc="CPU2_F1" manu="Samsung" sn="256D07E9" ce="0"/>
<dimm id="15" loc="CPU2_E0" manu="Samsung" sn="256D0786" ce="0"/>
<dimm id="16" loc="CPU2_E1" manu="Samsung" sn="256D078A" ce="0"/>
<dimm id="17" loc="CPU2_D0" manu="Hynix" sn="43346DD4" ce="1"/>
<dimm id="18" loc="CPU2_D1" manu="Samsung" sn="256D0812" ce="0"/>
<dimm id="19" loc="CPU2_A1" manu="Samsung" sn="256D07D5" ce="0"/>
<dimm id="20" loc="CPU2_A0" manu="Hynix" sn="43346DDE" ce="0"/>
<dimm id="21" loc="CPU2_B1" manu="Samsung" sn="256D07D6" ce="0"/>
<dimm id="22" loc="CPU2_B0" manu="Samsung" sn="256D1389" ce="0"/>
<dimm id="23" loc="CPU2_C1" manu="Samsung" sn="256D07F7" ce="0"/>
<dimm id="24" loc="CPU2_C0" manu="Samsung" sn="256D084C" ce="0"/>
<node id="node1" cluster="FlashSystem 7200 Demo" node_id="0x0000000000000001"
cluster_id="0x0000020420403bda" ro="518088179" wo="520294077" rb="527552053411"
```

```
lrb="1672303252387" wb="198236934084" lwb="2613946263486" re="11275" we="66341298"
rq="1136960" wq="67571846"/>
<node id="node2" cluster="FS9500_Demo" node_id="0x0000000000000002"
cluster_id="0x00000204e1a0013a" ro="2181589" wo="1841294" rb="227243368"
lrb="43176263640" wb="211016728" lwb="209302152" re="657" we="3800262" rq="13309"
wq="108469"/>
<node id="node1" cluster="FS9500_Demo" node_id="0x0000000000000001"
cluster_id="0x00000204e1a0013a" ro="3583406" wo="1707872" rb="433133836"
lrb="97069577036" wb="181649512" lwb="179934936" re="1101" we="3792950" rq="24804"
wq="100909"/>
<partnership cluster_id="0x0000020420403bda" node_id="0x0000000000000001"
sc="5980593" sbc="108142893184" sl="222009706" sic="0" sibc="0" sil="0" sicl="0"
rc="5652857" rbc="97361968232" rl="6056216"/>
<partnership cluster_id="0x00000204e1a0013a" node_id="0x0000000000000001"
sc="712287" sbc="106931072" sl="46444914" sic="0" sibc="0" sil="0" sicl="0"
rc="2297396" rbc="75470347584" rl="4565971"/>
<partnership cluster_id="0x00000204e1a0013a" node_id="0x0000000000000002"
sc="711903" sbc="106746752" sl="45250952" sic="0" sibc="0" sil="0" sicl="0"
rc="985663" rbc="32338992832" rl="1881515"/>
<messages_stats node_id="1">
<messages type="sent" period="all">
<msg name="lookup_lock" count="19442" total_time="1504615" total_size="4977152"/>
<msg name="bt_update" count="33846" total_time="32294220" total_size="8664576"/>
<msg name="bt_update_data" count="268" total_time="549686" total_size="1166336"/>
<msg name="unlock" count="514544" total_time="29433148" total_size="131723264"/>
<msg name="backup_write_done" count="2884" total_time="954658"
total_size="738304"/>
<msg name="soften" count="0" total_time="0" total_size="0"/>
<msg name="soften_done" count="2542822" total_time="0" total_size="650962432"/>
<msg name="dedup_ref_lookup_lock" count="0" total_time="0" total_size="0"/>
<msg name="dedup_confirm_lookup_lock" count="2122" total_time="1184821"
total_size="543232"/>
<msg name="dedup_db_lookup" count="0" total_time="0" total_size="0"/>
<msg name="dedup_db_add" count="0" total_time="0" total_size="0"/>
<msg name="update_tracker" count="25418" total_time="2412155"
total_size="6507008"/> <msg name="dedup_dec_lookup" count="739"
total_time="878483" total_size="189184"/>
<msg name="bt_update_multi" count="0" total_time="0" total_size="0"/>
<msg name="invals_update" count="0" total_time="0" total_size="0"/>
<msg name="bt_update_multi2" count="1244" total_time="79529"
total_size="5413888"/>
<msg name="invals_update2" count="787" total_time="45524" total_size="201472"/>
<msg name="dedup_db_multi_add" count="7576" total_time="0" total_size="1939456"/>
<msg name="dedup_db_multi_lookup" count="504" total_time="90061"
total_size="129024"/> </messages>
<messages type="received" period="all"> <msg name="lookup_lock" count="18651"
total_time="0" total_size="4774656"/> <msg name="bt_update" count="110272"
total_time="1164933" total_size="28229632"/> <msg name="bt_update_data"
count="1187300" total_time="10709916" total_size="5167129600"/> <msg name="unlock"
count="0" total_time="0" total_size="0"/> <msg name="backup_write_done" count="0"
total_time="0" total_size="0"/> <msg name="soften" count="0" total_time="0"
total_size="0"/> <msg name="soften_done" count="0" total_time="0" total_size="0"/>
<msg name="dedup_ref_lookup_lock" count="0" total_time="0" total_size="0"/> <msg
name="dedup_confirm_lookup_lock" count="6056" total_time="234547"
total_size="1550336"/>
```

```
<msg name="dedup_db_lookup" count="0" total_time="0" total_size="0"/>
<msg name="dedup_db_add" count="0" total_time="0" total_size="0"/>
<msg name="update_tracker" count="18688" total_time="140837"
total_size="4784128"/>
<msg name="dedup_dec_lookup" count="1965" total_time="2793687"
total_size="503040"/>
<msg name="bt_update_multi" count="0" total_time="0" total_size="0"/>
<msg name="invals_update" count="0" total_time="0" total_size="0"/>
<msg name="bt_update_multi2" count="5800750" total_time="85751526"
total_size="25244864000"/>
<msg name="invals_update2" count="3791525" total_time="33474284"
total_size="970630400"/>
<msg name="dedup_db_multi_add" count="997" total_time="2801" total_size="255232"/>
<msg name="dedup_db_multi_lookup" count="323" total_time="2940"
total_size="82688"/>
</messages>
</messages_stats>
<gc>
<repository id="0" mdg="65535"><cm mbs="0"/><nm mbs="0"/><rm mbs="0"/>
<mm mbs="0"/><ext col="0"/><rec mbs="0"/><rec apr="0"/>
</repository> <repository id="1" mdg="65535"><cm mbs="0"/><nm mbs="0"/><rm
mbs="0"/>
<mm mbs="0"/><ext col="0"/><rec mbs="0"/><rec apr="0"/></repository> <repository
id="2" mdg="65535"><cm mbs="0"/><nm mbs="0"/><rm mbs="0"/>
<mm mbs="0"/><ext col="0"/><rec mbs="0"/><rec apr="0"/></repository> <repository
id="3" mdg="65535"><cm mbs="0"/><nm mbs="0"/><rm mbs="0"/>
<mm mbs="0"/><ext col="0"/><rec mbs="0"/><rec apr="0"/></repository>
</gc>
<fc relr="102283" lelr="145964" rslr="3327" lslr="6152" rblr="8865653"
lblr="864156" elgs="2" slgs="3664" blgs="0" eqag="0" eqcn="0" eqmx="0" sqag="0"
sqcn="0" sqmx="0" bqag="0" bqcn="0" bqmx="0" />
<uca><ca dav="0" dcn="0" dmx="0" dmn="0" sav="0" scn="0" smx="0" smn="0" pav="0"
pcn="0" pmx="0" pmn="0" wfav="0" wfmx="0" wfmn="0" rfav="6" rfmx="6" rfmn="6"
pp="0" hpt="0" ppt="0" opt="0" npt="0" apt="0" cpt="0" bpt="0" hrpt="0" hpt_a="0"
ppt_a="0" opt_a="0" npt_a="0" apt_a="0" cpt_a="4192" bpt_a="32432" hrpt_a="0"
m="0" mn="0"/>
<partition id="0" mdg="0"><ca dav="0" dcn="0" dmx="0" dmn="0" fav="0" fmx="0"
fmn="0" pp="0"/> </partition>
</uca>
<lca><ca dav="0" dcn="0" dmx="0" dmn="0" sav="0" scn="0" smx="0" smn="0" pav="0"
pcn="0" pmx="0" pmn="0" wfav="0" wfmx="0" wfmn="0" rfav="0" rfmx="0" rfmn="0"
pp="0" hpt="0" ppt="0" opt="0" npt="0" apt="0" cpt="0" bpt="0" hrpt="0" hpt_a="0"
ppt_a="0" opt_a="0" npt_a="0" apt_a="0" cpt_a="9568" bpt_a="26176" hrpt_a="0"
m="0" mn="0"/>
<partition id="0" mdg="0"><ca dav="0" dcn="0" dmx="0" dmn="0" fav="0" fmx="0"
fmn="0" dfav="0" dfmx="0" dfmn="0" dtav="0" dtmx="0" dtmn="0" pp="0"/>
</partition>
</lca>
<replication rjru="0.00" rjrw="0.00" />
<odx otrec="0"/>
</diskStatsColl>
```

Volume statistics

Volume statistics play a crucial role in understanding the performance and behavior of individual volumes within a volume group. Replication statistics focus primarily on the overall performance of the volume group. Volume-level statistics can provide more granular data that is related to the performance of each volume.

Table 10-28 illustrates the statistics collected for volume statistics.

Table 10-28 Volume statistics

Statistic name	Description
avg_recovery_point (varp)	Average recovery point in seconds for the specific volume within the volume group since the last statistics collection.
worst_recovery_point (vwrp)	Worst recovery point in seconds for the specific volume within the volume group since the last statistics collection.

Example 10-29 shows an output of the statistics collected for a volume. Table 10-28 explains the names of the statistics.

Example 10-29 Volume statistics

```
<?xml version="1.0" encoding="utf-8" ?>
<diskStatsColl
xmlns="http://ibm.com/storage/management/performance/api/2005/08/vDiskStats"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://ibm.com/storage/management/performance/api/2005/08/vDis
kStats schema/SVCPerfStatsV.xsd" scope="node" id="node1"
cluster="FlashSystem 7200 Demo" node_id="0x0000000000000001"
cluster_id="0x0000020420403bda" sizeUnits="512B" timeUnits="msec"
contains="virtualDiskStats" timestamp="2023-07-02 21:37:55" timezone="GMT+10:00"
timestamp_utc="2023-07-02 11:37:55">
<vdsk idx="0" ctps="0" ctrhs="0" ctrhps="0" ctds="0" ctwfts="0" ctwwts="0"
ctwfws="0" ctwhs="0" cv="0" cm="0" ctws="0" ctrs="0" ctr="0" ctw="0" ctp="0"
ctrh="0" ctrhp="0" ctd="0" ctwft="0" ctwwt="0" ctwfw="0" ctwfwsh="0" ctwfwshs="0"
ctwh="0" varp="0" vwrp="0" gwot="0" gwo="0" gws="0" gwl="0"
id="FCM_Fully_allocated0" ro="0" wo="0" wou="0" rb="0" wb="0" rl="0" wl="0"
rlw="0" wlw="0" xl="0" wxl="0" rxl="0" oro="0" owo="0" orl="0" owl="0" oiowp="0"
uo="0" ub="0" uou="0" ul="0" ulw="0">
<fc wlag="0" wlcn="0" wlmx="0" rlag="0" rlcn="0" rlmx="0" bcag="0" bccn="0"
bcmx="0" clag="0" clcn="0" clmx="0" rwag="0" rwcn="0" rwmx="0" twag="0" twcn="0"
twmx="0" trag="0" trcn="0" trmx="0" bwag="0" bwcn="0" bwmx="0" brag="0" brcn="0"
brmx="0" hwsg="0" hwdg="0" srcp="0" twbl="0" trbl="0" bwbl="0" brbl="0" />
<ca rh="0" d="0" ft="0" wt="0" fw="0" wh="0" v="0" m="0" ri="0" wi="0" r="0"
dav="0" dcn="0" sav="0" scn="0" teav="0" tsav="0"  tav="0"  entav="0" entcn="0"
entmx="0" entmn="0" pp="0"/> <cl bup="0" bdn="0"/><cpy idx="0"> <ca p="0" rh="0"
ph="0" d="0" ft="0" wt="0" fw="0" wh="0" v="0" m="0" pm="0" ri="0" wi="0" r="0"
dav="0" dcn="0" sav="0" scn="0" pav="0" pcn="0" teav="0"  tsav="0"  tav="0"
entav="0" entcn="0" entmx="0" entmn="0" pp="0"/>
</cpy>
</vdsk>
```

Synchronization errors in policy-based replication

The volume groups associated with a replication policy might encounter errors that slow the replication process between partnered systems. These errors can be caused by various factors such as offline volumes, configuration issues, or a disconnected partnership.

During the synchronization process for policy-based replication, the following types of errors can occur:

- If read errors are detected within volumes or volume groups during synchronization, the replication of volume data within the affected volume group is halted. The system logs an error related to the volume group. After the read error is resolved, it is marked as fixed, and replication is restarted. In case of a disconnected partnership, the synchronization process is interrupted but automatically resumes when the connection is restored.
- Configuration errors occur when a configuration task for policy-based replication is blocked either locally or on the remote system. These errors can be caused by the following conditions:
  - Errors logged against the partnership indicate disruptions in the IP link between the partnered systems or errors in Rest API calls to the remote system. Such errors block configuration tasks on the remote system. After connectivity to the remote system is reestablished, the error is automatically marked as fixed. The system then retries any pending configuration tasks.
  - If an error is logged against a specific local volume group, it blocks configuration tasks on the corresponding remote volume group. This can happen when the remote system lacks the necessary resources to complete the requested action. In such cases, an error is raised against the local volume group. To mark the error as fixed, you must follow the associated fix procedure, which provides instructions on resolving the issue. When the error is resolved, the system will retry any pending configuration tasks.
- If the system loses access to an entire I/O group or requires a T3 recovery, replication is automatically suspended. To recover volumes, enter either of the following commands:
  - **`recovervdiskbyiogrp`** *io_group_name* | i*o_group_id* to acknowledge data loss for all volumes in the specified I/O group with a fast_write_state of corrupt and brings the volumes back online.
  - **`recovervdiskbysystem`** to acknowledge data loss for all volumes in the system with a fast_write_state of corrupt and bring the volumes back online.

Manual restart of replication is necessary for each volume group, and a full resynchronization is performed. To mitigate capacity impacts, you can prioritize volume groups for resynchronization or opt to remove the existing replication policy and assign a new policy with the same settings.

11

Reliability, availability, and serviceability

Ensuring that storage resources are always available is critical for the success of businesses. This chapter introduces useful and common procedures to maintain and monitor the system. It includes the following topics:

- 11.1, “Reliability, availability, and serviceability” on page 964
- 11.2, “Shutting down an IBM Storage Virtualize System” on page 975
- 11.3, “Removing a node or adding a node” on page 976
- 11.4, “Configuration backup” on page 979
- 11.5, “Updating software” on page 984
- 11.6, “Health checker feature” on page 1000
- 11.7, “Troubleshooting and fixing procedures” on page 1001
- 11.8, “Monitoring and Event Notification” on page 1006
- 11.9, “Audit log” on page 1022
- 11.10, “Collecting support information by using the GUI, CLI, and USB” on page 1024
- 11.11, “Service Assistant Tool” on page 1031
- 11.12, “IBM Storage Insights monitoring” on page 1035

© Copyright IBM Corp. 2024. All rights reserved.

11.1 Reliability, availability, and serviceability

Reliability, availability, and serviceability (RAS) are important concepts in the design of the IBM Storage Virtualize system. Hardware features, software features, design considerations, and operational guidelines all contribute to make the IBM Storage Virtualize systems reliable.

Fault tolerance and high levels of availability are achieved by using the following methods:

- Distributed redundant array of independent disks (DRAID) capabilities ensures rapid rebuild and high levels of data protection
- Auto-restart of hung nodes or node canisters
- Integrated battery backup units (BBUs) to provide write cache protection if a site power failure occurs
- Host system failover capabilities by using N_Port ID Virtualization (NPIV)
- Deploying advanced multi-site configurations, such as IBM HyperSwap and Stretch cluster for IBM SAN Volume Controller implementations

High levels of serviceability are available by using the following methods:

- Cluster error logging
- Asynchronous error notification
- Automatic dump capabilities to capture software-detected issues
- Concurrent:
  - Diagnostic procedures
  - Log analysis and memory dump data recovery tools
  - Maintenance of IBM Storage Virtualize System components
  - Upgrade of IBM Storage Virtualize software and firmware
  - Addition or deletion of nodes or node canisters in the clustered system
- Directed maintenance procedures (DMPs) with guided online replacement processes
- Automatic software version leveling when replacing a node or a node canister
- Detailed status and error conditions that are displayed by light-emitting diode (LED) indicators
- Error and event notification through Simple Network Management Protocol (SNMP), syslog, and email
- Enhanced support by using Call Home and Remote Support functions
- Optional Remote Support Assistant (RSA)
- IBM Storage Insights

The primary part of the IBM Storage Virtualize system is a pair of nodes. For an IBM SAN Volume Controller, these nodes are separate physical appliances. For an IBM FlashSystem, each system contains two *node canisters* that run inside a single enclosure.

Generally, the nodes or the node canisters provide similar functions. Unless specified, a *node* refers to a physical node or a node canister. When differences exist, they are highlighted. Two nodes share the read and write data workload from the attached hosts and to the disk arrays.

This section examines the RAS features of the systems, monitoring, and troubleshooting.

11.1.1 Hardware information

IBM FlashSystem and IBM SAN Volume Controller use a clustered node hardware design. In a FlashSystem, each node is a node canister and both are in the same enclosure. For IBM SAN Volume Controller, each node is in a separate enclosure.

As shown in Figure 11-1, for IBM FlashSystem 7300 the top node canister is inverted above the bottom canister. The control enclosure also contains two power supply units (PSUs) that operate independently of each other. The PSUs are visible from the back of the control enclosure.

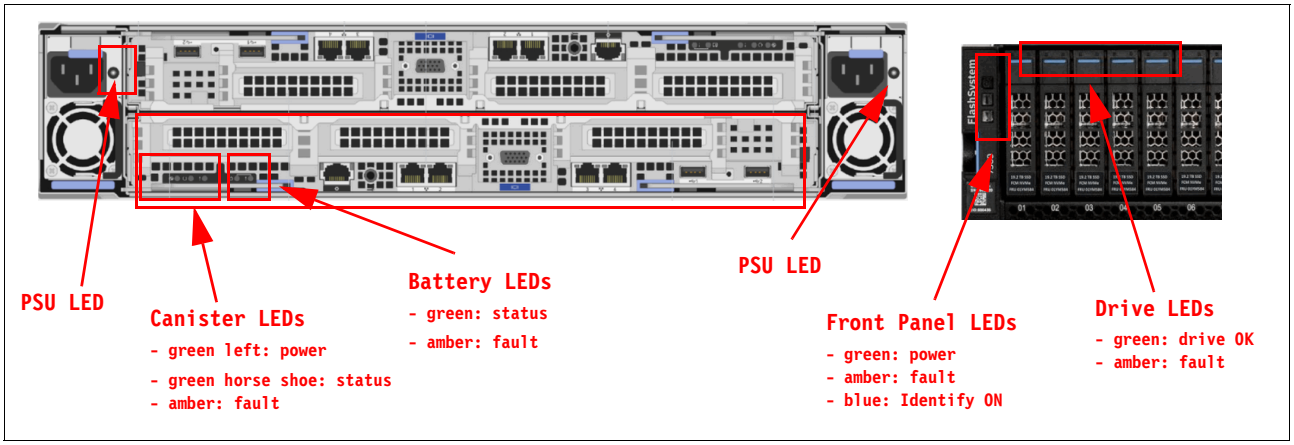


Figure 11-1 LEDs on each node canister

The connections of a FlashSystem 7300 single node canister (upper) are shown in Figure 11-2. These are the same for an IBM SAN Volume Controller SV3 in which both have the same connections and interface card slots.

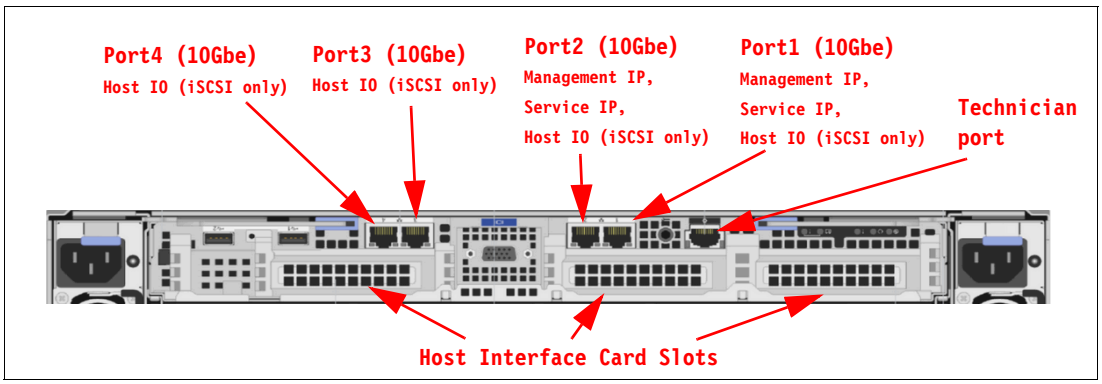


Figure 11-2 FlashSystem 7300 node canister

For an IBM FlashSystem 9500, the node canisters are not inverted, and each one is 2U. Each node has three cold swappable cages to hold host interface cards. Each node also has two hot pluggable battery backup units (BBUs) and two hot pluggable boot drives, as shown in Figure 11-3.

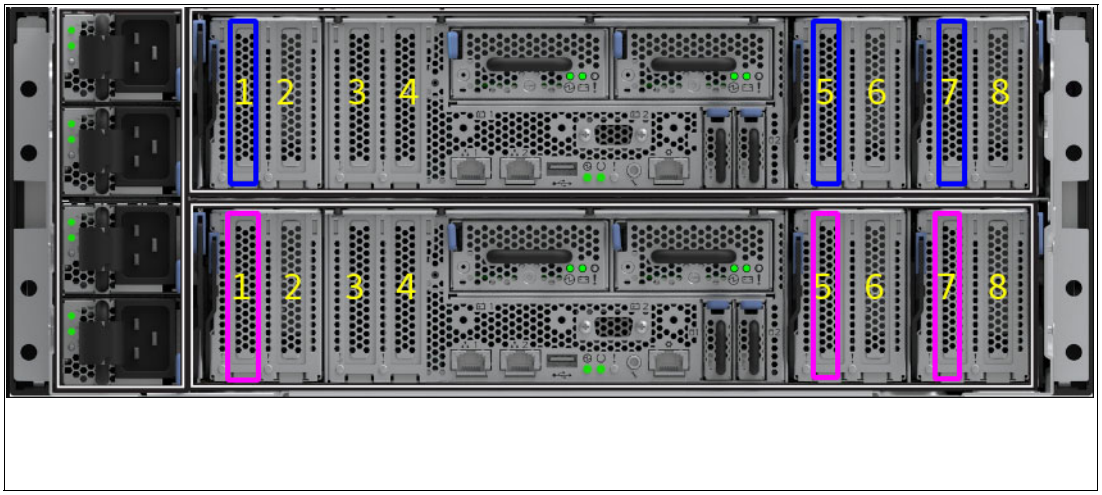


Figure 11-3 IBM FlashSystem 9500 rear view

Host interface cards

For the IBM FlashSystem 7300 each canister features three host interface card (HIC) slots. Depending on the system, a 4-port serial-attached Small Computer System Interface (SCSI) SAS-card might be installed in each node canister, which leaves two HIC slots that can be populated with a range of cards, as shown in Table 11-1. Node canisters in the same I/O group must have the same HIC configuration.

Table 11-1 Supported card configurations for IBM FlashSystem 7300

Supported number of cards	Ports per card	Protocol	Slot positions	Note
0–3	4	16 Gb Fibre Channel (FC)	1, 2, 3	FC, NVMeoF
0–3	4	32 Gb FC	1, 2, 3	FC, NVMeoF
0–3	2	25 Gb Ethernet (iWARP)	1, 2, 3	Internet Wide Area Remote Direct Memory Access (RDMA) Protocol (iWARP)
0–3	2	25 GbE (RoCE)	1, 2, 3	RDMA over Converged Ethernet (RoCE)
0–3	2	Maximum of twelve 100 Gbps Ethernet	1, 2, 3	iSCSI and NVMe RDMA
0–1	2	12 Gb SAS Expansion	3	► Four-port card, but only two are active ► Expansion only; no SAS host attachment

Note: The systems include onboard compression cards. No compression-assist cards are included, as in previous models.

For FlashSystem 5200, only two card slots are available. The supported card configurations are listed in Table 11-2.

Table 11-2 Supported card configurations for IBM FlashSystem 5200

Supported number of cards	Ports per card	Protocol	Slot positions	Note
0–2	4	16 Gb FC	1, 2	Not supported with SAS Host adapter
0 –2	2	32 Gb FC	1, 2	Not supported with SAS Host adapter
0–2	2	25 GbE (iWARP)	1, 2	No FCoE nor Ethernet NVMeoF support
0–2	2	25 GbE (RoCE)	1, 2	No FCoE nor Ethernet NVMeoF support
0–1	2	12 Gb SAS Expansion	2	► Four-port card, but only two are active ► Expansion only, no SAS host attachment

The IBM FlashSystem 9500 has other cards available, including 100 Gb Ethernet. Supported card configurations for the 9500 are described in Table 11-3.

Table 11-3 Supported card configurations for IBM FlashSystem 9500 and SV3

Supported number of cards	Ports per card	Protocol	Cage/Slot positions	Note
1-6	4	32 Gb FC	any slot	FC, NVMeoF
1-6	2	25 Gb Ethernet (iWARP)	any slot	Internet Wide Area Remote Direct Memory Access (RDMA) Protocol (iWARP)
1-6	2	25 Gb Ethernet (RoCE)	any slot	RDMA over Converged Ethernet (RoCE)
1-3	2	100Gb Ethernet	Left slot in each cage	iSCSI and NVMe RDMA
1-2	2	12 Gb SAS expansion	any slot	► Four-port card, but only two are active ► Expansion only; no SAS host attachment ► Not available for SV3

If you want to cluster two or more systems, an FC card is required. Clustering enables management of multiple FlashSystems and multiple I/O groups for IBM SAN Volume Controller through a single interface.

For FC configurations, the meanings of the port LEDs are listed in Table 11-4.

Table 11-4 Fibre Channel link LED statuses

Port LED	Color	Meaning
Link status	Green	Link is up, connection established.
Speed	Amber	Link is not up or speed fault.

USB ports

Two active USB connectors are available in the horizontal position in the rear of the node canister. They have no numbers, and no indicators are associated with them. These ports can be used for initial cluster setup, encryption key backup, and node canister status or log collection.

Ethernet and LED status

On the IBM FlashSystem 7300 four 10 GbE ports and one 1-Gigabit Ethernet port are on each canister. Their functions are listed in Table 11-5.

Table 11-5 Ethernet ports and their functions

Onboard Ethernet port	Speed	Function
1	10 Gbps	Management IP, Service IP, and Host I/O (iSCS only)
2	10 Gbps	Secondary Management IP, and Host I/O (iSCSI only)
3	10 Gbps	Host I/O (iSCSI only)
4	10 Gbps	Host I/O (iSCSI only)
T	1 Gbps	Technician Port: DHCP / domain name server (DNS) for direct attach service management

For the IBM FlashSystem 9500, two 1 Gbps Ethernet management ports and a technician port are available.

Each port has two LEDs, and their status values are listed in Table 11-6. However, the T port is dedicated to technician actions for initial configuration and emergency configuration by local support personnel.

Table 11-6 Ethernet LED statuses

LED	Color	Meaning
Link state	Green	On when an Ethernet link exists
Activity	Amber	Flashing when activity is detected on the link

Serial-attached SCSI ports

When a 4-port SAS interface card is installed, it is possible to connect the 2U and 5U expansion enclosures. However, only ports 1 and 3 are used for SAS connections, with the SAS chain from port 1 installed below the lower node canister, and the SAS chain from port 3 installed above the upper node canister. The SAS card must be installed in PCIe slot 3 of each node canister. Two LEDs are used for each SAS port with statuses, as shown in Table 11-7.

Table 11-7 SAS LED statuses

LED	Meaning
Green	Link is connected and up.
Orange	Fault on the SAS link (disconnected, wrong speed, and errors).

Node canister status LEDs

Three LEDs are in a row at the left of the canister that indicates the status and the functions of the node canister (see Table 11-8).

Table 11-8 Node canister LEDs

Position	Color	Name	State	Meaning
Left	Green	Power	On	The node is started and active. It might not be safe to remove the canister. If the fault LED is off, the node is an active member of a cluster or candidate. If the fault LED is also on, the node is in the service state or in error, which prevents the software from starting.
			Flashing (2 Hz)	Canister is started and in standby mode.
			Flashing (4 Hz)	Node is running a power-on self-test (POST).
			Off	No power to the canister, or it is running on battery.
Middle	Green	Status	On	The node is a member of a cluster.
			Flashing (2 Hz)	The node is a candidate or in a service state.
			Flashing (4 Hz)	The node is performing a dump. Never unplug the canister during this time.
			Off	No power to the canister or the canister is in standby mode.
Right	Amber	Fault	On	The canister is in a service state or in error, for example, a **POST** error that is preventing the software from starting.
			Flashing (2 Hz)	Canister is being identified.
			Off	Node is in the candidate or active state.

Battery LEDs

Immediately to the right of the canister LEDs, with a short gap between them, are the Battery LEDs, which provide the status of the battery (see Table 11-9).

Table 11-9 Battery LEDs

Position	Color	Name	State	Meaning
Left	Green	Status	On	Indicates that the battery is fully charged and has sufficient charge to complete two dumps.
			Flashing (2 Hz)	Indicates that the battery has sufficient charge to complete a single dump.
			Flashing (4 Hz)	Indicates that the battery is charging and has insufficient charge to complete a single dump.
			Off	Indicates that the battery is not available for use. For example, it is missing or contains a fault.
Right	Amber	Fault	On	Indicates that a battery has a fault or a condition occurred. The node enters the service state.
			Off	Indicates that there are no known battery faults or conditions. An exception is when a battery has insufficient charge to complete a single dump. Refer to the Status LED.

Expansion canisters

Although expansion drawers are not supported on available versions of IBM SAN Volume Controller, they are supported on IBM FlashSystems. Figure 11-4 shows two 12 Gbps SAS ports are side by side on the canister of every enclosure. They are numbered 1 on the left and 2 on the right. The expansion canisters have vertical ports located next to each other on the back of the enclosure.

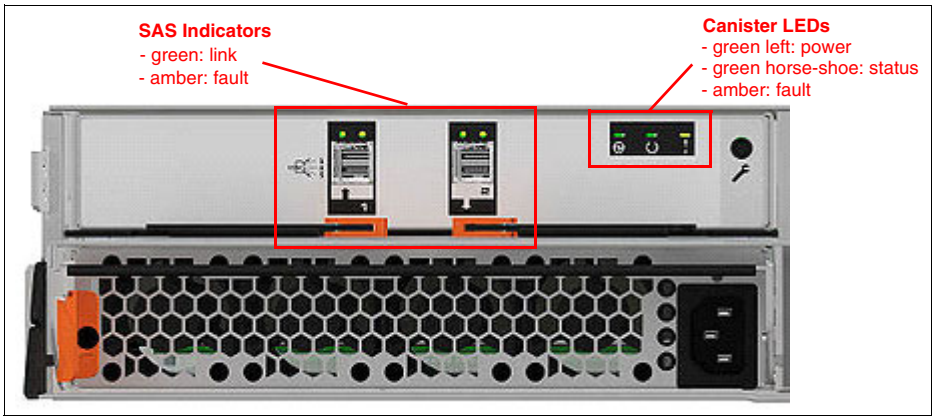


Figure 11-4 Expansion canister status LEDs

The SAS status LEDs have the same meaning as the LED indicators of SAS ports in the control enclosure (see Table 11-7 on page 969).

Table 11-10 lists the LED status values of the expansion canister.

Table 11-10 Expansion canister LEDs statuses

Position	Color	Name	State	Meaning
Left	Green	Power	On	The canister is powered on.
			Off	No power is available to the canister.
Middle	Green	Status	On	The canister is operating normally.
			Flashing	There is an error with the vital product data (VPD).
Right	Amber	Fault	On	There is an error that is logged against the canister or the system is not running.
			Flashing	Canister is being identified.
			Off	No fault, the canister is operating normally.

11.1.2 Dense Drawer Enclosures LED

As shown in Figure 11-5, two 12 Gbps SAS ports are side by side on the canister of every enclosure. They are numbered 1 on the right and 2 on the left. Each Dense Drawer has two canisters side by side, although they are inverted when compared to the 2U enclosures.

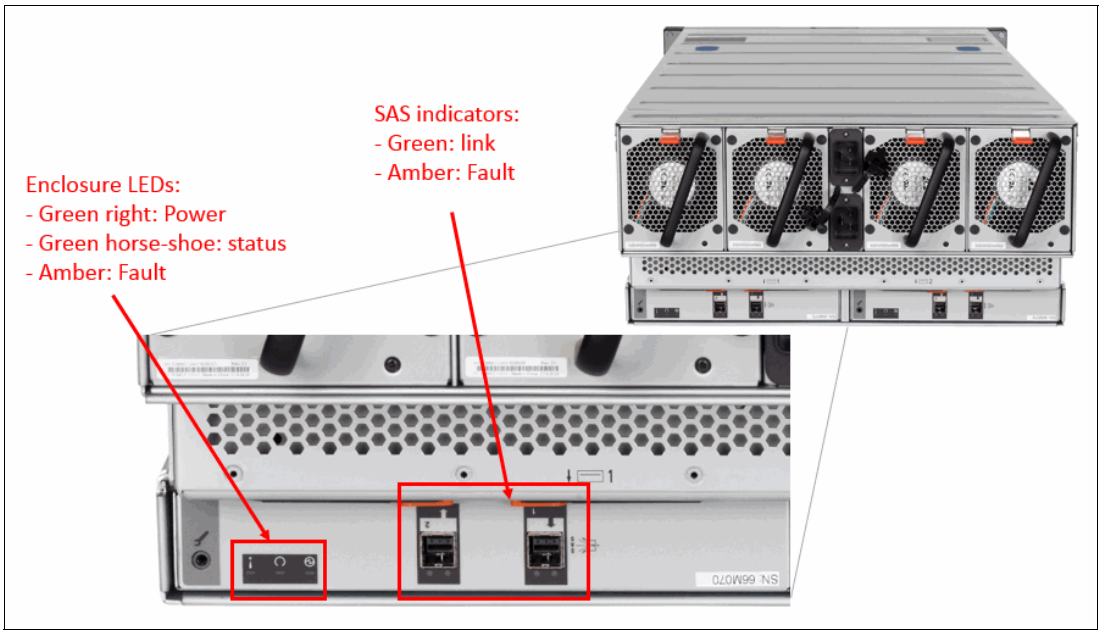


Figure 11-5 Dense Drawer LEDs

The SAS status LEDs have the same meaning as the LED indicators of SAS ports that are described in Table 11-10.

Table 11-11 shows the LED status values of the expansion canister.

Table 11-11 Expansion canister LEDs statuses

Position	Color	Name	State	Meaning
Right	Green	Power	On	The canister is powered on.
			Off	No power is available to the canister.
Middle	Green	Status	On	The canister is operating normally.
			Flashing	There is an error with the VPD.
Left	Amber	Fault	On	There is an error that is logged against the canister or the system is not running (OSES).
			Flashing	Canister is being identified.
			Off	No fault, the canister is operating normally.

11.1.3 Enclosure SAS cabling

Expansion enclosures are attached to control enclosures through 12 Gbps SAS cables. The IBM FlashSystem control enclosure attaches up to 20 expansion enclosures or up to eight Dense Drawer enclosures.

A strand starts with a SAS initiator chip inside an IBM FlashSystem node canister and progresses through SAS expanders, which connect disk drives. Each canister contains an expander. Each drive has two ports, each connected to a different expander and strand. This configuration ensures that both node canisters in the I/O group have direct access to each drive, and that no single point of failure (SPOF) exists.

Figure 11-6 shows how the SAS connectivity works inside the node canister and expansion canisters.

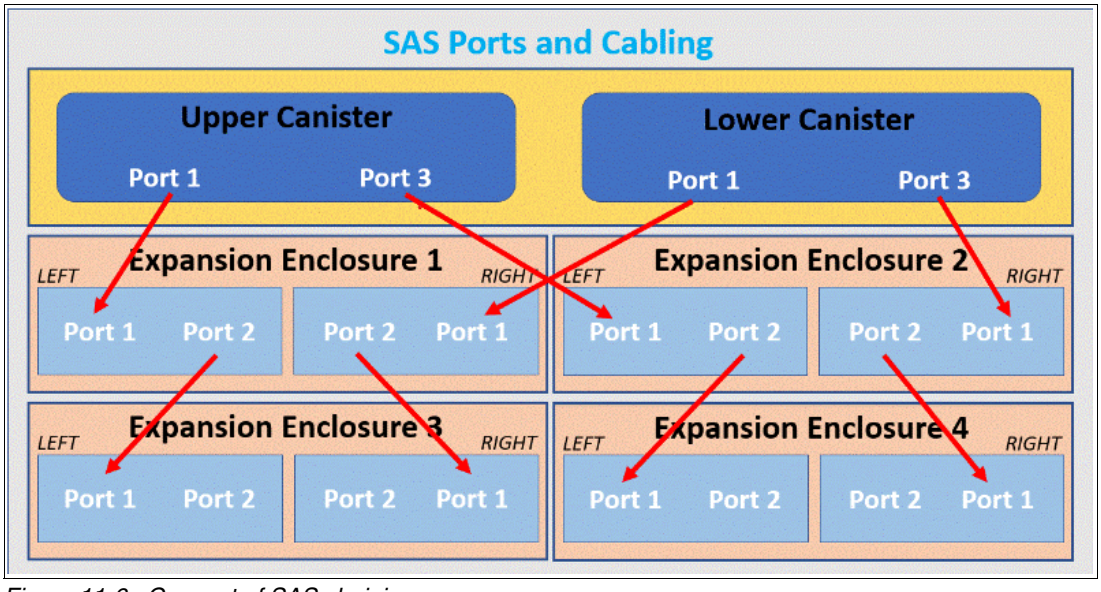


Figure 11-6 Concept of SAS chaining

Note: The last expansion enclosure in a chain must not have cables in port 2 of canister 1 or port 2 of canister 2. Therefore, if you add two enclosures to the setup that is shown in Figure 11-6 on page 972, you connect a cable to port 2 of the existing enclosure canisters and port 1 of the new enclosure canisters.

A *chain* consists of a set of enclosures that are correctly interconnected (see Figure 11-7). Chain 1 of an I/O group is connected to SAS port 1 of both node canisters. Chain 2 is connected to SAS port 3. This configuration means that chain 2 includes the SAS expander and drives of the control enclosure.

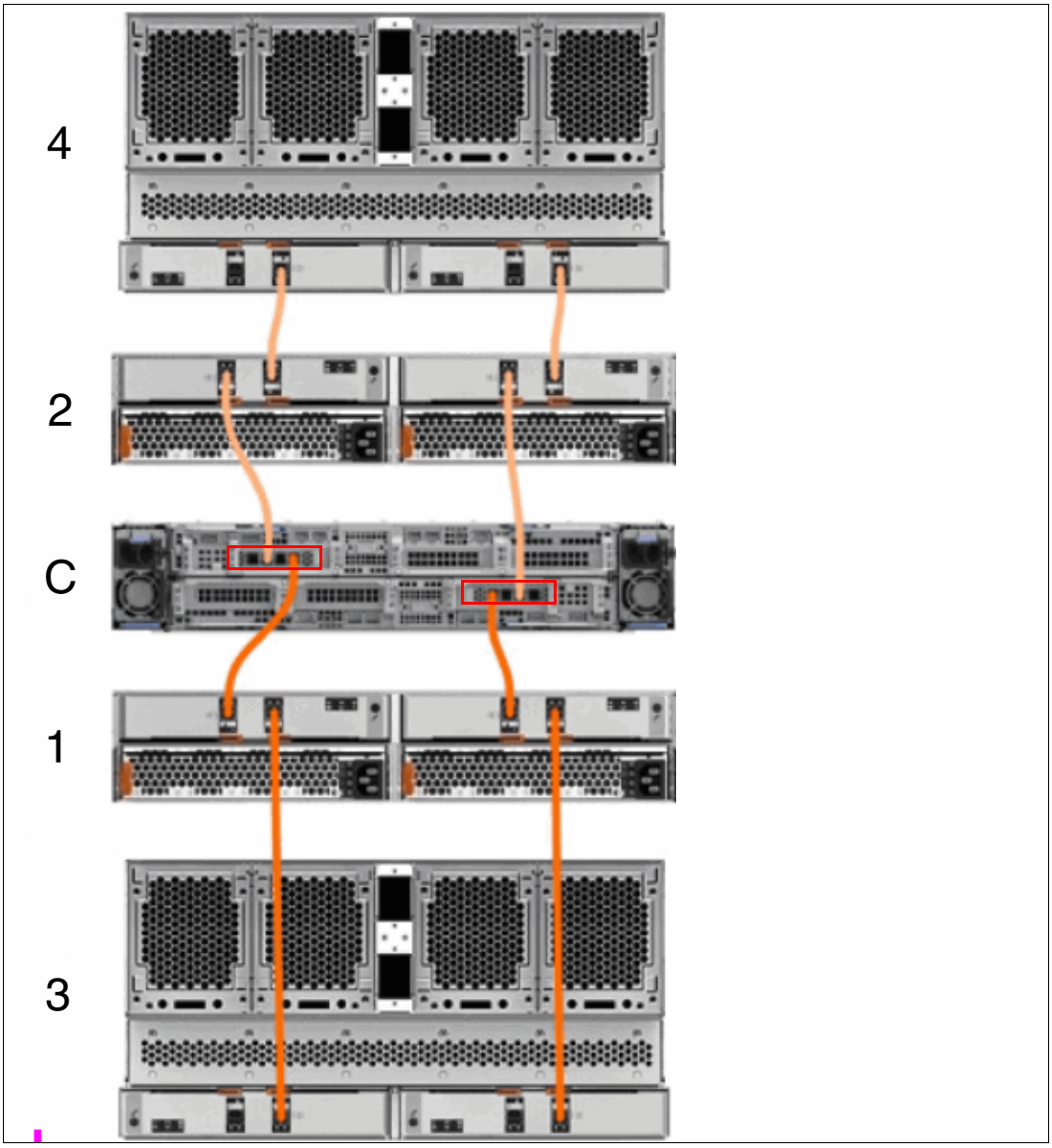


Figure 11-7 SAS cabling with numbered enclosures

At system initialization, the system performs a discovery process to update the state of the drive and enclosure objects when devices are added to or removed from strands.

11.1.4 Drive modules

FlashSystems support standard NVMe drives, IBM FlashCore Modules, and Storage Class Memory drives. Consider the following points about RAS regarding all drive types:

- A drive that is a member of a RAID array must not be reseated unless directed by IBM Support. Re-seating drives that are still in use by an array can cause unwanted consequences.
- When removing an array, drives might show as offline for some time because of formatting. They automatically come back online after the task completes.

11.1.5 Power

All enclosures except the IBM FlashSystem 9500 accommodate two PSUs. The IBM FlashSystem 9500 has four PSUs. Any enclosure can operate with only half of the PSUs, which provides protection from partial power failures. For this reason, it is highly advised to supply AC power to each PSU from different power distribution units (PDUs).

For IBM FlashSystem control enclosures and IBM SAN Volume Controller enclosures, the integrated battery continues to supply power to the node if a power loss occurs. It holds power for 5 seconds before starting a graceful shutdown, which includes de-staging write cache information.

A fully charged battery can perform two critical data de-stages, where a node canister stores cache and system data to an integrated drive if a power failure occurs.

Figure 11-8 shows two FlashSystem 7300 PSUs that are present in the control and expansion enclosure. The controller PSU has one LED that can be green or amber, depending on the status of the PSU. If the LED is off, no AC power is available to the entire enclosure.

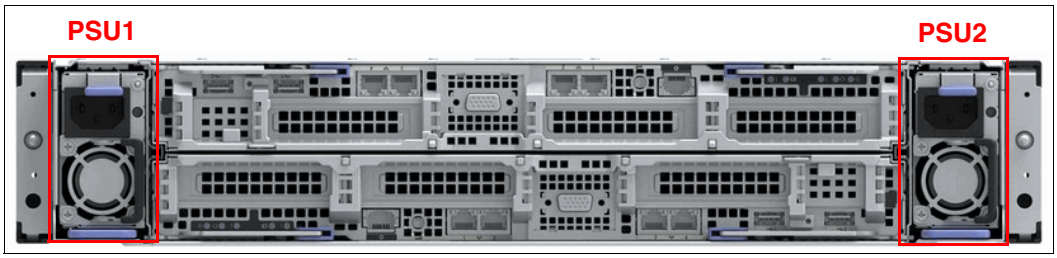


Figure 11-8 Controller enclosure LED status indicator FlashSystem 7300

Figure 11-9 shows an overview of the rear of an expansion enclosure with a PSU. The enclosure is powered on by the direct attachment of a power cable.



Figure 11-9 Expansion enclosure power supply unit

Power supplies in both control and expansion enclosures are hot-swappable and replaceable without shutting down a node canister or cluster. If the power is interrupted in one node canister for less than 5 seconds, the canister does not perform a memory dump and continues operation from the battery. This feature is useful for a case of, for example, maintenance of UPS systems in the data center or replugging the power to a different power source or PDU unit. A fully charged battery can perform two critical data de-stages.

11.2 Shutting down an IBM Storage Virtualize System

You can safely shut down the system by using the GUI or command line interface (CLI).

Important: Never shut down the system by powering off the PSUs, removing all PSUs, or removing power cables from a running system. These actions can lead to inconsistency or loss of the data that is staged in the cache.

Before shutting down the IBM Storage Virtualize system, stop all hosts that include allocated volumes from the device. This step can be skipped for hosts that have volumes that are also provisioned with host-based mirroring from different storage devices. However, doing so can cause errors that are related to lost storage paths and disks on the host.

You can shut down a single node, or you can shut down the entire cluster. When you shut down only one node, all host I/O continues if best practices for fabric connectivity are followed. When you shut down a node or the entire cluster, you must power it on locally to restart.

11.2.1 Shutting down and powering on a complete infrastructure

When you shut down or power on the entire infrastructure, which include storage, servers, and applications, follow a particular sequence for both the shutdown and the power-on actions. This section provides an example sequence of a shutdown sequence, and then a power-on sequence of an infrastructure that includes an IBM FlashSystem system.

Shutting down the infrastructure

To shut down the infrastructure, complete the following steps:

1. Shut down your servers and all applications.
2. Shut down your IBM FlashSystem systems:
   a. Shut down the IBM FlashSystem by using the GUI or CLI.
   b. Power off both switches of the controller enclosure.
   c. Power off both switches of all the expansion enclosures.
3. Shut down your storage area network (SAN) switches.

Powering on

To power on your infrastructure, complete the following steps:

1. Power on your SAN switches and wait until the start completes.
2. Power on your storage systems by completing the following steps:
   a. Power on both power supplies of all the expansion enclosures.
   b. Power on both power supplies of the control enclosure.
   c. When the storage systems are up, power on your servers and start your applications.

11.3 Removing a node or adding a node

In some circumstances, IBM Support might prompt you to remove a node from the system briefly. One typical use case is when a node becomes stuck during a code upgrade.

You might be instructed by IBM Support personnel to remove the node canister from the cluster temporarily to commit the upgrade and complete or cancel the procedure depending on how many node canisters have completed the upgrade. This procedure should be done only under the direction of IBM Support.

The easiest way to complete this task is by running the **svcinfo lsnode** command to display all nodes and their ID and status, as shown in Example 11-1. You can make sure that each I/O group has two nodes online or if you remove a node canister, that one node canister remains in the I/O group to continue serving I/O.

Example 11-1 The lsnode output

```
IBM_FlashSystem:FS7300:superuser>svcinfo lsnode
id name  UPS_serial_number WWNN             status IO_group_id IO_group_name config_node UPS_unique_id hardware
iscsi_name                          iscsi_alias panel_name enclosure_id canister_id enclosure_serial_number site_id
site_name
1  node1                 5005076810000214 online 0           io_grp0       no                          AF7
iqn.1986-03.com.ibm:2145.fs7300.node1             01-1       1            1           78E003K
2  node2                 5005076810000216 online 0           io_grp0       yes                         AF7
iqn.1986-03.com.ibm:2145.fs7300.node2             01-2       1            2           78E003K
```

In this example for an IBM FlashSystem 7300, node 1 is removed from the cluster. The **svctask rmnodecanister 1** command is run, as shown in Example 11-2.

Example 11-2 The rmnodecanister command

```
IBM_FlashSystem:FS7300:superuser>svctask rmnodecanister 1
IBM_FlashSystem:FS7300:superuser>
```

In this example, for an IBM SAN Volume Controller, node 1 is removed from the cluster. The **svctask rmnode 1** command is run, as shown in Example 11-3.

Example 11-3 The rmnode command

```
IBM_2145:superuser>>svctask rmnode 1
IBM_2145:superuser>>
```

A node can also be removed by using the GUI. Complete the following steps:

1. Select **Monitoring** → **System**, and then select the relevant control enclosure of the node that you want to remove, which opens the Enclosure Details window. Select the node and right-click it and click **Remove**, or use the menu in Components Details to remove it (Figure 11-10), which opens a confirmation window.

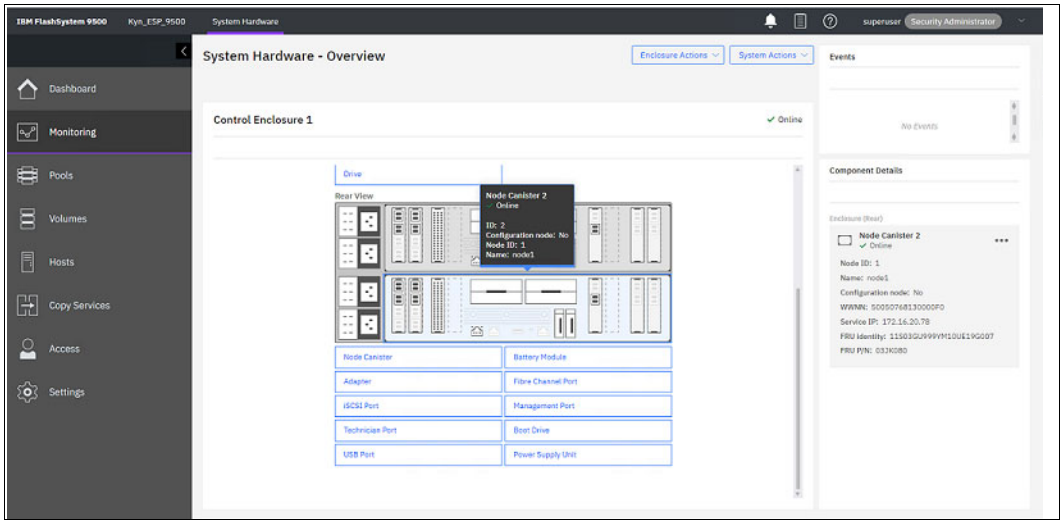


Figure 11-10 Removing a node by using the GUI

After you remove the node canister, if you rerun the **svcinfo lsnode** command, you see that it disappeared from the cluster, as shown in Example 11-4. The Service Assistant Tool (SAT) and GUI also reflect that only one node canister in the cluster now.

Example 11-4 The lsnode output after removing a node canister

```
IBM_FlashSystem:FS7300:superuser>svcinfo lsnode
id name  UPS_serial_number WWNN             status IO_group_id IO_group_name config_node UPS_unique_id hardware
iscsi_name                          iscsi_alias panel_name enclosure_id canister_id enclosure_serial_number
site_id site_name
2  node2                 5005076810000216 online 0           io_grp0       yes                          AF7
iqn.1986-03.com.ibm:2145.fs7300.node2             01-2       1             2           78E003K
IBM_FlashSystem:FS7300:superuser>
```

Note: By default, the cache is flushed before the node canister is deleted to prevent data loss if a failure occurs on the other node in the I/O group. This flush incurs a delay from when you remove a node canister to when it comes back up as candidate status.

2. After a brief period, check the SAT, which shows that the node canister that you removed is in the service or candidate status, as shown in Figure 11-11.

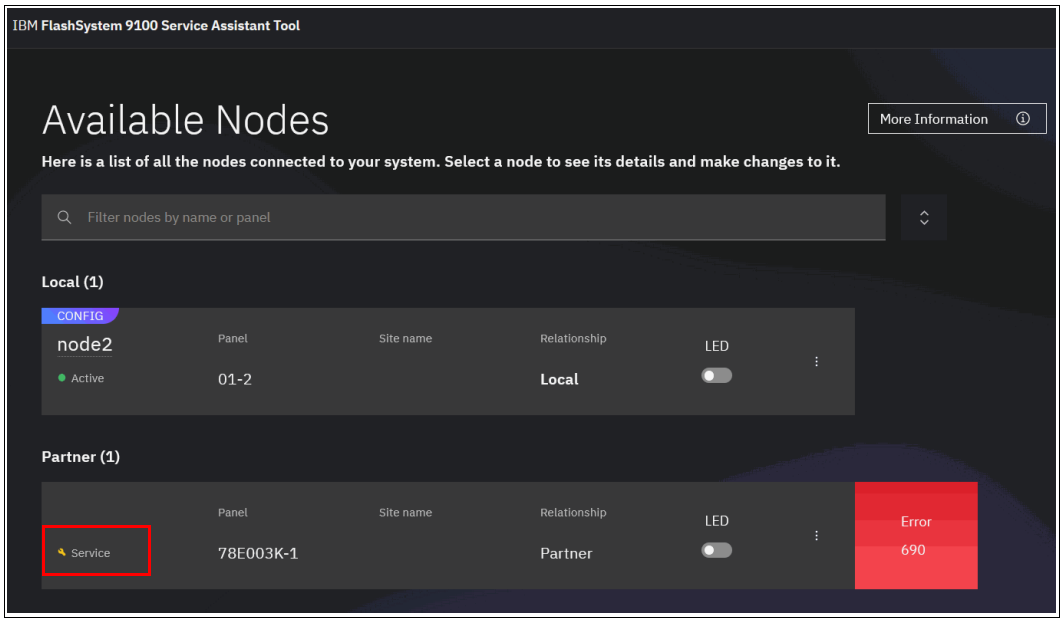


Figure 11-11 Service Assistant Tool post-node removal

3. Select the radio button for the node canister that is in service state and then select **Exit Service State** as shown in Figure 11-12.

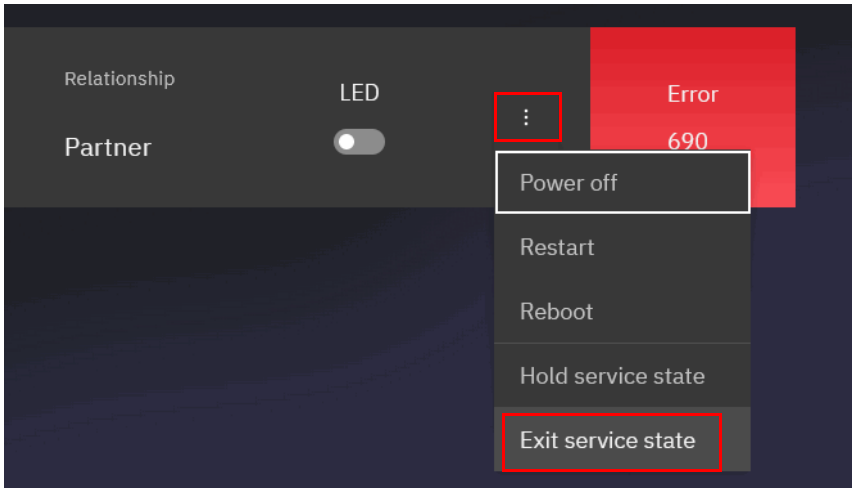


Figure 11-12 Radio button for node actions

4. A confirmation window opens and shows that the node canister exited the service state. Click **Confirm**, or close the window and click **Refresh** under the list of the nodes as shown in Figure 11-13 on page 979.

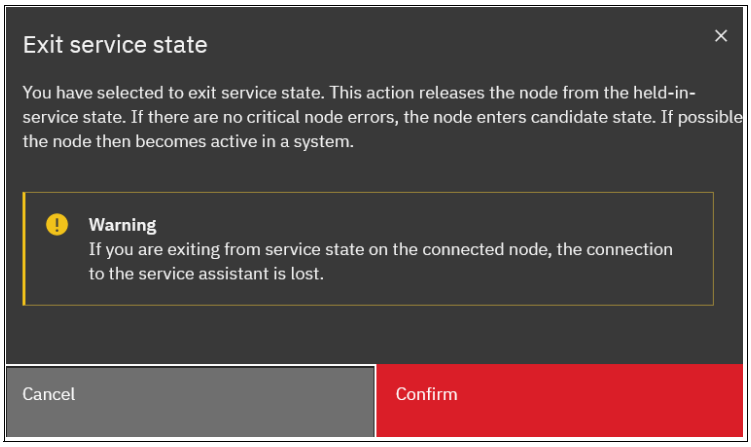


Figure 11-13 Exit service state

5. The node canister should automatically rejoin the cluster. If not, review the numbers in the `Panel` column and return to your CLI session. Run the **`addcontrolenclosure`** command and specify the panel ID to add the node canister back into the cluster, as shown in Example 11-5.

Example 11-5 The addncontrolenclosure command

```
IBM FlashSystem 7300:superuser>svctask addcontrolenclosure -panelname 78E003K-1
-iogrp io_grp0
```

6. Run the **`svcinfo lsnode`** command again or check the SAT to ensure that the node canister was added back, as shown in Example 11-6.

Example 11-6 The svcinfo lsnode command

```
IBM_FlashSystem:FS7300:superuser>svcinfo lsnode
id name  UPS_serial_number WWNN             status IO_group_id IO_group_name config_node UPS_unique_id hardware
iscsi_name                          iscsi_alias panel_name enclosure_id canister_id enclosure_serial_number
site_id site_name
4  node1                 5005076810000214 online 0           io_grp0       yes                          AF7
iqn.1986-03.com.ibm:2145.fs7300.node1             01-1       1             1           78E003K
2  node2                 5005076810000216 online 0           io_grp0       no                           AF7
iqn.1986-03.com.ibm:2145.fs7300.node2             01-2       1             2           78E003K
IBM_FlashSystem:FS7300:superuser>
```

Note: If you want to remove an entire control enclosure from the cluster to reduce the size of the cluster or to decommission it, you can do this task by using the GUI. Go to the Enclosure Overview window, as shown in Figure 11-10 on page 977, but instead of selecting a node, select **Enclosure Actions** and then **Remove.** A confirmation window opens. This action runs the **`rmnodecanister`** command against both nodes in the control enclosure. For more information about removing an enclosure, see IBM Documentation and search for “Removing a control enclosure and its expansion enclosures”.

11.4 Configuration backup

You can download and save the configuration backup file by using the IBM Storage Virtualize GUI or CLI. As needed, manually perform this procedure because it can save the file directly to your workstation. The CLI option requires you to log in to the system and download the dumped file by using Secure Copy Protocol (SCP). The CLI option is a best practice for an automated backup of the configuration. For IBM SAN Volume Controller systems, the

configuration is backed up every day at 1am. This automatically generated backup file is named svc.config.cron.xml_<panel_id> and is in the /dumps directory.

Important: Generally, perform a daily backup of the IBM FlashSystem configuration backup file, for which the best approach is to automate this task. Always perform another backup before any critical maintenance task, such as an update of the IBM Storage Virtualize Software version.

The backup file is updated by the cluster every day. Saving it after any changes to your system configuration is important. It contains configuration data of arrays, pools, volumes, and other items. The backup does not contain any data from the volumes.

To successfully perform the configuration backup, the following prerequisites must be met:

- All nodes are online.
- No independent operations that change the configuration can be running in parallel.
- No object name can begin with an underscore.

Important: A backup of the configuration can be done only from the CLI by using the **svcconfig backup** command. Then the output of the command can be downloaded by using SCP or the GUI.

11.4.1 Backing up by using the CLI

You can use the CLI to trigger configuration backups manually or by a regular automated process. The **svcconfig backup** command generates a new backup file. Triggering a backup by using the GUI is not possible. However, you might choose to save the automated 1 AM cron backup if you have not made any configuration changes.

Example 11-7 shows how to use the **svcconfig backup** command to generate an *ad hoc* backup of the current configuration.

Example 11-7 Saving the configuration by using the CLI

```
IBM_FlashSystem:FS7300:superuser>svcconfig backup
.....................................................................................................
.....................................................
CMMVC6155I SVCCONFIG processing completed successfully
IBM_FlashSystem:FS7300:superuser>
```

The **svcconfig backup** command generates three files that provide information about the backup process and cluster configuration. These files are dumped into the /tmp directory on the configuration node. Run the **lsdumps** command to list them (see Example 11-8).

Example 11-8 Listing the backup files by using the CLI

```
IBM_FlashSystem:FS7300:superuser>lsdumps |grep backup
16  svc.config.backup.bak_78E003K-2
115 svc.config.backup.xml_78E003K-2
116 svc.config.backup.log_78E003K-2
117 svc.config.backup.sh_78E003K-2
IBM_FlashSystem:FS7300:superuser>
```

Note: The svc.config.backup.bak file is a previous copy of the configuration, and not part of the current backup.

Table 11-12 lists the three files that are created by the backup process.

Table 11-12 Files that are created by the backup process

File name	Description
`svc.config.backup.xml`	This file contains your cluster configuration data.
`svc.config.backup.sh`	This file contains the names of the commands that ran to create the backup of the cluster.
`svc.config.backup.log`	This file contains details about the backup, including any error information that might have been reported.

Save the current backup to a secure and safe location. The files can be downloaded by running **scp** (UNIX) or **PSCP** (Microsoft Windows), as shown in Example 11-9. Replace the IP address with the cluster IP address of your system and specify a local folder on your workstation. In this example, the file is saved to `C:\FS7300Backup`.

Example 11-9 Saving the config backup files to your workstation

```
C:\putty>pscp -unsafe
superuser@xx.xx.xx.xx:/dumps/svc.config.backup.* c:\FS7300backup
Using keyboard-interactive authentication.
Password:
svc.config.backup.bak_782 | 133 kB |  33.5 kB/s | ETA: 00:00:00 | 100%
svc.config.backup.log_782 | 16 kB |  16.8 kB/s | ETA: 00:00:00 | 100%
svc.config.backup.sh_7822 | 5 kB |   5.9 kB/s | ETA: 00:00:00 | 100%
svc.config.backup.xml_782 | 105 kB |  52.8 kB/s | ETA: 00:00:00 | 100%

C:\putty>

C:\>dir FS7300backup
 Volume in drive C has no label.
 Volume Serial Number is 0608-239A

 Directory of C:\FS7300backup

24.10.2018  10:57    <DIR>          .
24.10.2018  10:57    <DIR>          ..
24.10.2018  10:57           137.107 svc.config.backup.bak_7822DFF-1
24.10.2018  10:57            17.196 svc.config.backup.log_7822DFF-1
24.10.2018  10:57             6.018 svc.config.backup.sh_7822DFF-1
24.10.2018  10:58           108.208 svc.config.backup.xml_7822DFF-1
               4 File(s)        268.529 bytes
               2 Dir(s) 48.028.662.272 bytes free

C:\>
```

Using the **-unsafe** option enables you to use the wildcard for downloading all the `svc.config.backup` files with a single command.

Tip: If you encounter the `Fatal: Received unexpected end-of-file from server` error, when running the **PSCP** command, consider upgrading your version of PuTTY.

11.4.2 Saving the backup by using the GUI

You can save the backup files by using the GUI. To do so, complete the following steps:

1. Select **Settings** → **Support** → **Support Package**.
2. Click the **Manual Download Instructions** drop-down menu.
3. Click **Download Existing Package**, as shown in Figure 11-14.

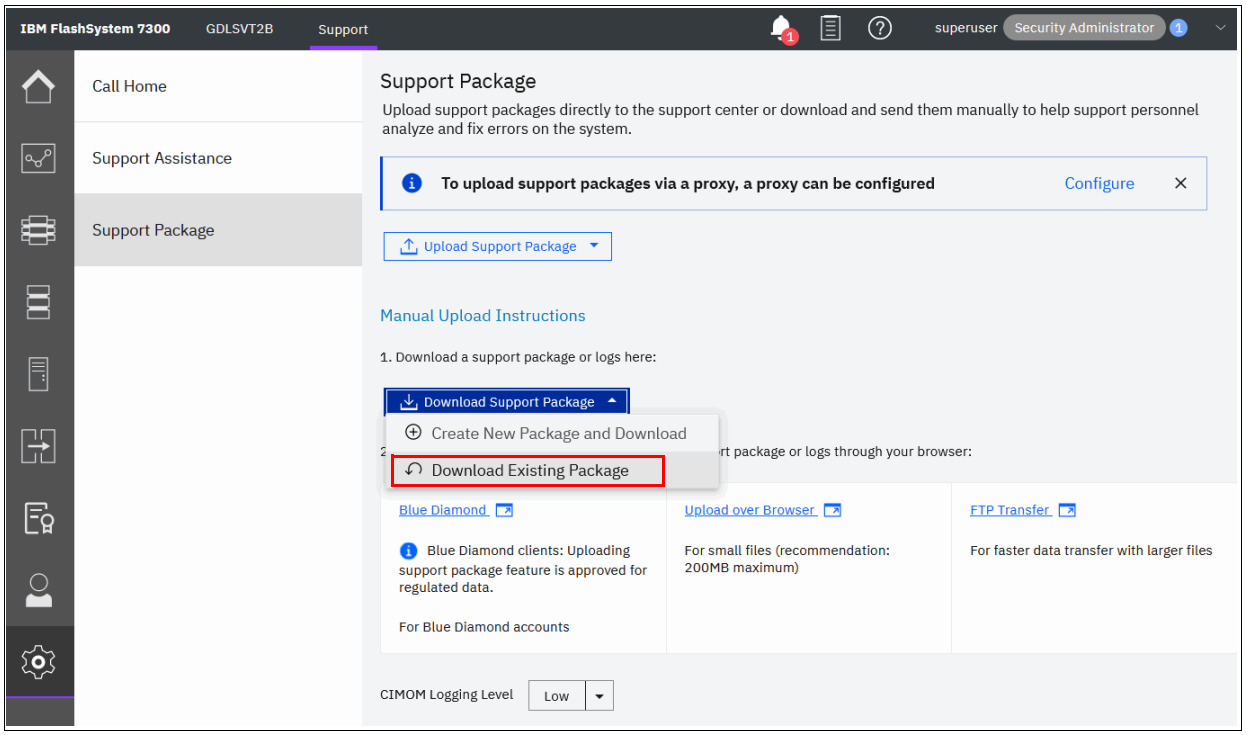


Figure 11-14 Download Existing Package

The Support Package selection window opens, as shown in Figure 11-15 on page 983.

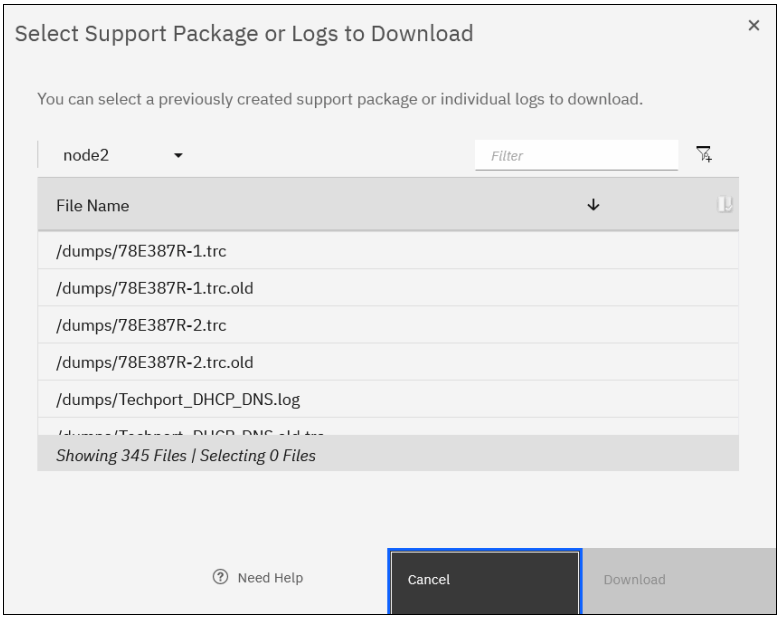


Figure 11-15 Support Package Selection

4. Filter the view by clicking the **Filter** box, entering `backup`, and pressing **Enter**, as shown in Figure 11-16

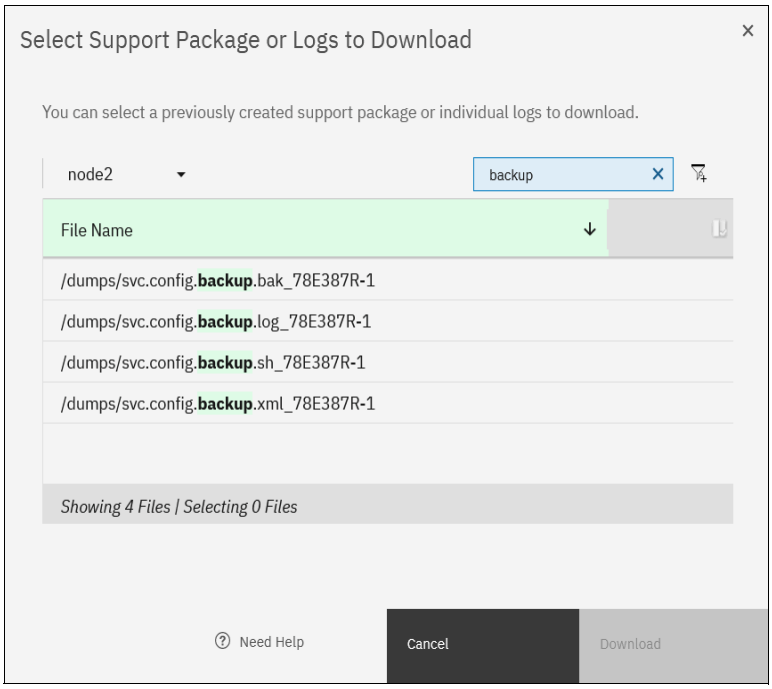


Figure 11-16 Filtering specific files for download

Note: You must select the configuration node in the upper-left drop-down menu because the backup files are stored there.

5. Select all the files to include in the compressed file, and click **Download**. Depending on your browser preferences, you might be prompted about where to save the file. Otherwise, the download is stored in your defined download directory.

11.5 Updating software

This section provides information about IBM Storage Virtualize software release numbering and describes the operations to update your system software.

To get current information about software updates, support bulletins, and other support-related information, log in to the IBM Support My Notifications page.

11.5.1 Storage Virtualize Upgrade Planning

When planning to update your Storage Virtualize system, you must first determine to which version you want to update. Storage Virtualize software is provided in two branches:

- Long Term Support (LTS):
  - Contains new functionality in addition to all functionality introduced in previous LTS and Non-LTS releases.
  - Plan for quarterly fix packs (PTFs) but no new functionality will be added. This can deliver maximum stability.
  - Plan for one LTS release every 1–2 years.
  - Identified by zero as the third digit of the release number, for example 8.5.0.8.
  - IBM expects most clients to run the LTS branch.
  - Select this branch when stability is a priority.
- Non-Long Term Support (Non-LTS)
  - Contains new functionality in addition to all functionality introduced in previous LTS and Non-LTS releases.
  - Do not receive regular fix packs. Fixes for issues introduced in a Non-LTS release will be delivered in a subsequent Non-LTS or LTS release.
  - Plan for multiple Non-LTS releases per year.
  - Identified by a nonzero as the third digit of the release number (for example 8.5.4.0).
  - To upgrade to a Non-LTS, a system must already be running a previous release (LTS or Non-LTS) on the same version. For example, to upgrade to 8.4.2 the system must be running 8.4.0.x or 8.4.1. To upgrade from 8.4.2 to 8.5.2, the system must first be upgraded to 8.5.0.x.
  - If new functionality is the priority, select this branch.

Under normal circumstances, always update to the latest recommended LTS version. If more recent code has a required feature, then a NON-LTS version can be installed.

For more information of recommended code levels see IBM Storage Virtualize Family of Products Upgrade Planning.

Note: IBM maintains and updates the newest code streams as well as the previous releases. As of this writing, you can get updated versions of the code streams for versions 8.6, 8.5, 8.4 and 8.3.

The latest code for each property record might not always be the best one to use. Review the release notes for your intended update-version and determine if the fixes are relevant for your Storage Virtualize system.

11.5.2 Precautions before the update

This section describes the precautions that you should take before you attempt an update.

Important: Before you attempt any code update, read and understand the concurrent compatibility and code cross-reference matrix for your system. For more information, see Concurrent Compatibility and Code Cross Reference for IBM Storage Virtualize and click **Latest system code**.

During the update, each node in the IBM FlashSystem/IBM SAN Volume Controller clustered system is automatically shut down and restarted by the update process. Because each node in an I/O group provides an alternative path to volumes, ensure suitable zoning and cabling is in place to so that node outage does not remove all paths to a host.

If you do not perform this check, certain hosts might lose connectivity to their volumes and experience I/O errors when the IBM FlashSystem node that provides that access is shut down during the update process.

11.5.3 IBM FlashSystem upgrade test utility

The software upgrade test utility is an IBM FlashSystem Software utility that checks for known issues that can cause problems during a software update. For more information, see Software Upgrade Test Utility. Download the software update utility from IBM Fix Central where you can also download the firmware. This procedure ensures that you receive the current version of this utility.

The Software Upgrade Test Utility and the firmware file can be uploaded in advance of the update process. Alternatively, when you use the update wizard, it can be uploaded and run during the software update.

You can run the utility multiple times on the same system to perform a readiness check as preparation for a software update. Run this utility a final time immediately before you apply the software update but always use the latest version of the utility.

The installation and use of this utility is nondisruptive, and it does not require a restart of any IBM FlashSystem nodes. Therefore, there is no interruption to host I/O. The utility is installed only in the current configuration node.

Example 11-10 shows how to upload the FlashSystem 9500 firmware files in advance by using the putty PSCP tool from a Windows command prompt. Uploading this way can be much faster than uploading by using the GUI, so in case of network issues PSCP is preferred over the GUI. The update itself can be done by using the GUI at a later time.

Example 11-10 Example of using PSCP to download update files (output shortened for clarity)

```
C:\Temp\FS9500 8600>dir
```

```
 Volume in drive C is SYSTEM
 Volume Serial Number is 1071-1234

 Directory of C:\Temp\FS9500 8600

21-06-2023  14:00 895.074.151 IBM_FlashSystem9x00_INSTALL_8.6.0.0
21-06-2023  14:00 414.677 IBM_INSTALL_FROM_8.4_AND_EARLIER_upgradetest_40.0
21-06-2023  14:00 421.565 IBM_INSTALL_FROM_8.5_AND_LATER_upgradetest_40.0
06-10-2016  13:48 359.336 pscp.exe
               4 File(s)    896.269.729 bytes
               2 Dir(s)  19.475.484.672 bytes free

C:\Temp\FS9500 8600>pscp IBM_FlashSystem9x00_INSTALL_8.6.0.0 superuser@172.15.130.140:/home/admin/upgrade
Using keyboard-interactive authentication.
Password:
IBM_FlashSystem9x00_INSTA | 874095 kB | 39731.6 kB/s | ETA: 00:00:00 | 100%

C:\Temp\FS9500 8600>pscp IBM_INSTALL_FROM_8.5_AND_LATER_upgradetest_40.0 superuser@172.15.130.140:/home/admin/upgrade
Using keyboard-interactive authentication.
Password:
IBM_INSTALL_FROM_8.5_AND_ | 411 kB | 411.7 kB/s | ETA: 00:00:00 | 100%

C:\Temp\FS9500 8600>
```

After the Upgrade Test Utility and firmware file are uploaded to the system to be updated, the firmware update wizard fields are pre-filled with the above shown filenames as can be seen in Figure 11-19 on page 988.

Tip: Currently two versions of the Upgrade Test Utility exists. One is for versions 8.4 and earlier and one is for versions 8.5 and higher. If your systems are to be updated from 8.4 and earlier to 8.5 or higher you will need both test utility files to update your system and drives. That is, if drives are updated AFTER the controller update then the system will be 8.5 or higher when the drives are to be updated, so both update test utilities are needed.

System administrators must continue to check whether the version of code that they plan to install is the latest version. For more information, see Concurrent Compatibility and Code Cross Reference for IBM Storage Virtualize.

The Upgrade Test Utility is intended to supplement rather than duplicate the tests that are performed by the IBM Storage Virtualize update procedure, for example, checking for unfixed errors in the error log.

A concurrent software update of all components is supported through the standard Ethernet management interfaces. However, most of the configuration tasks are restricted during the update process.

Note: IBM Storage Virtualize systems prevents you from upgrading if you use an Upgrade Test Utility that is older than one month. In this case, you must download a current Upgrade Test Utility from IBM Fix Central.

11.5.4 Updating your IBM FlashSystem to Version 8.6.0.0

In this example a FlashSystem 9110 is updated from version 8.5.0.8 to version 8.6.0.0. To update the IBM Storage Virtualize Software to Version 8.6.0.0, complete the following steps:

1. Log in by using superuser credentials or a user with similar permissions. The management home window opens. Hover over **Settings** and click **System** (Figure 11-17).

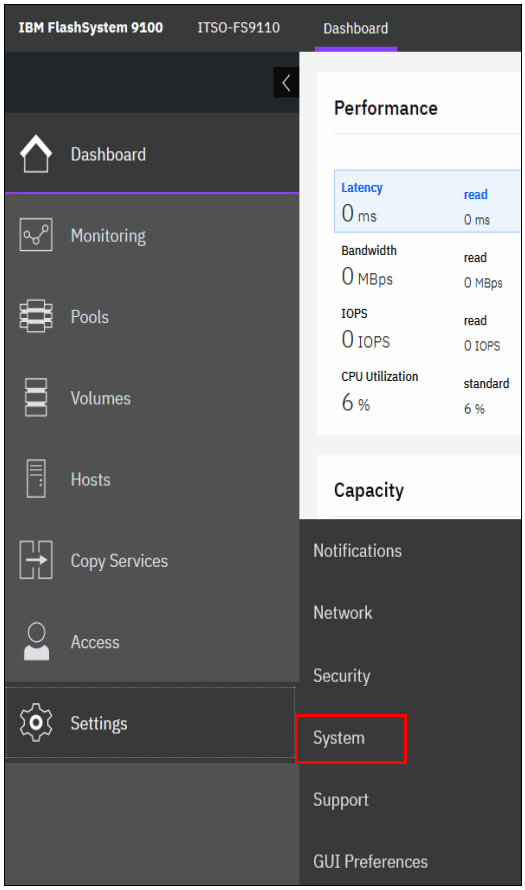


Figure 11-17 Settings menu

2. In the **System** menu, click **Update System**. The Update System window opens (see Figure 11-18).

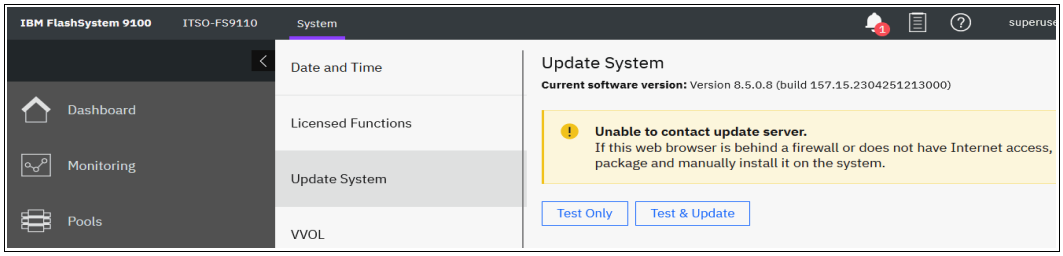


Figure 11-18 Update System window

Tip: You have the option to just do an update check without actually performing the update. This can be helpful because you can then run the Upgrade Test Utility before the update service window. The results of the Upgrade Test Utility indicate how long time the update will take. Updating FCM-drives takes approximately six minutes for each drive. As an example, a FlashSystem 9500 with 48 FCM-drives will take approximately 1 hour for the controllers to update and 5 hours for the drives, which is a total of 6 hours update time.

3. From this window, you can select to run the Upgrade Test Utility and continue with the code update or just run the Upgrade Test Utility. For this example, we click **Test and Update**.

   My Notifications: Use the My Notifications tool to receive notifications of new and updated support information to better maintain your system environment, especially in an environment where a direct internet connection is not possible.

   See My Notifications (an IBM account is required) to add your system to the notifications list to be advised of support information and to download the current code to your workstation for later upload.

4. Because you downloaded both files from the IBM Support Concurrent Compatibility and Code Cross Reference for IBM Storage Virtualize web page, you can click each folder, browse to the location where you saved the files, and upload them to the system. If the files are correct, the GUI detects and updates the target code level, as shown in Figure 11-19.

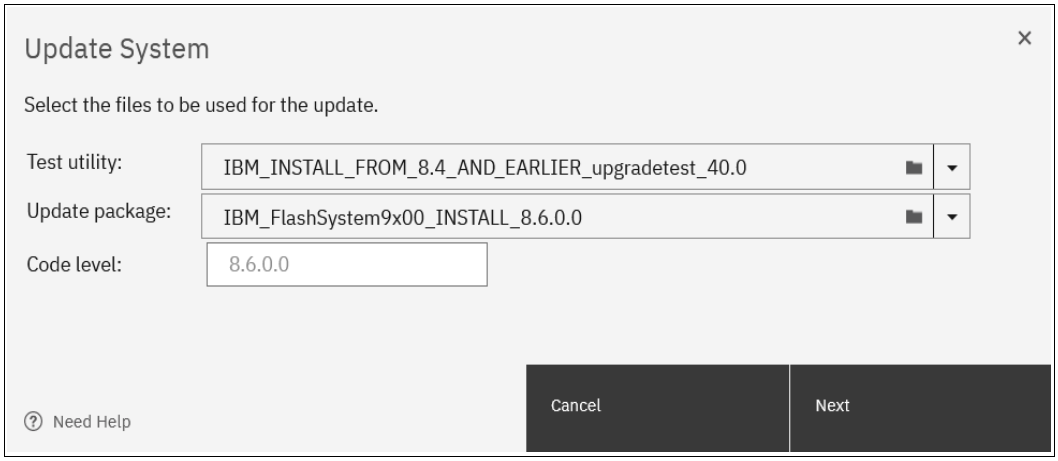


Figure 11-19 Upload option for both the Upgrade Test Utility and update package

5. Select the type of update you want to perform, as shown in Figure 11-20 on page 989. Select **Automatic update** unless IBM Support suggests **Service Assistant Manual update**. The manual update might be preferable in cases where a problem with host multipathing is known to cause a loss of access.

   Click **Next** to begin the update package upload process.

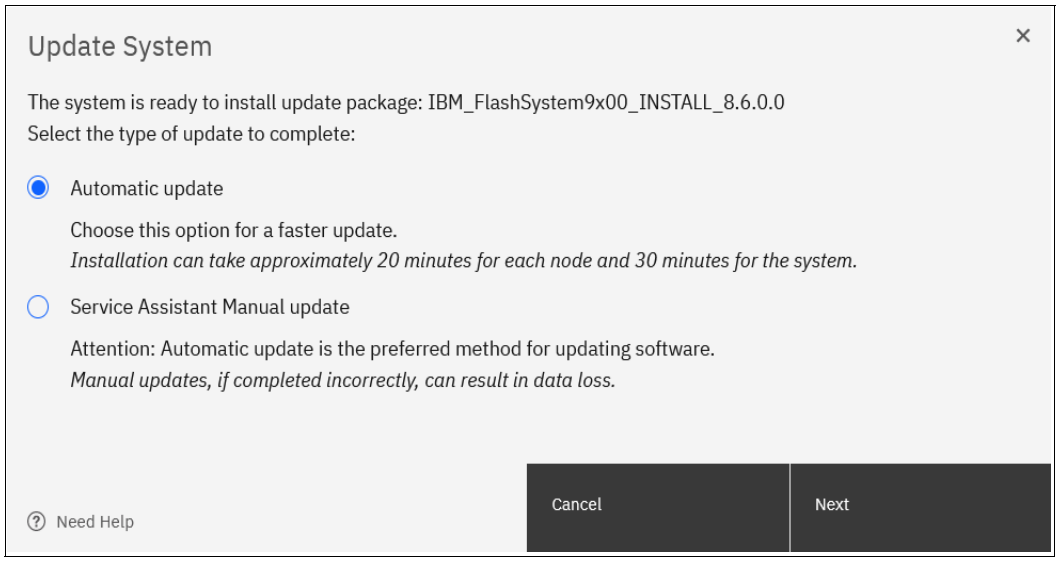


Figure 11-20 The update type selection

A new window opens, in which you can choose a fully automated update, one that pauses when half the nodes complete the update, or one that pauses after each node update, as shown in Figure 11-21. The pause option requires that you click **Resume** to continue the update after each pause. Click **Finish**.

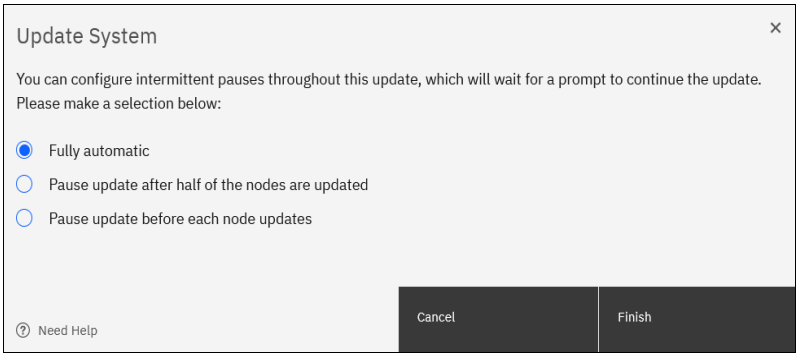


Figure 11-21 Update pause options

6. After the update packages upload, the Upgrade Test Utility looks for any known issues that might affect a concurrent update of your system. Click **Read more** (Figure 11-22).

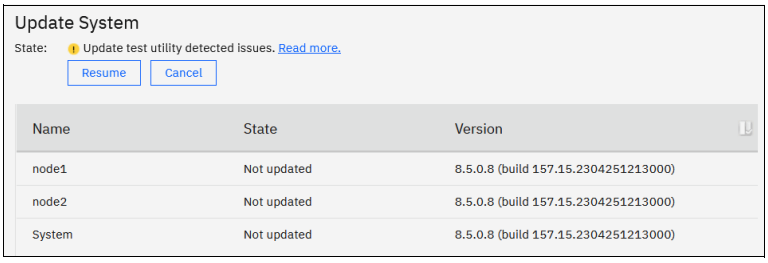


Figure 11-22 Issues that are detected by the Upgrade Test Utility

The results window opens and shows you what issues were detected (see Figure 11-23).

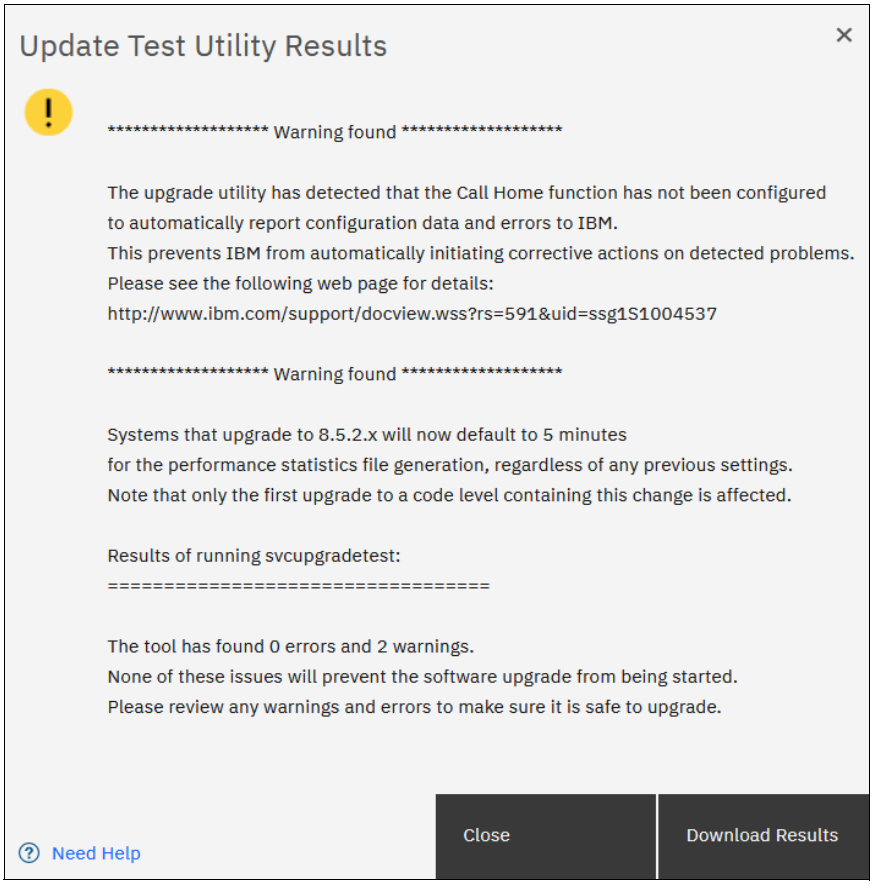


Figure 11-23 Description of the warning from the Upgrade Test Utility

In this example, the system identified only warnings about Call Home not enabled, and information about changes in performance statistics collection. Although these issues are not recommended conditions, it does not prevent the system update from running.

Therefore, click **Close** and proceed with the update. However, you might need to contact IBM Support to help resolve more serious issues before continuing.

In other cases you might see a warning that one or more drives in the system are running microcode with a known issue. Drive firmware can usually be updated AFTER the controller code update, but for some older versions the FCM-drive firmware must be updated before the controller code update can be started. A discussion of how drives are updated is in 11.5.5, “Updating the IBM FlashSystem drive code” on page 993.

7. Click **Resume** in the Update System window and the update proceeds, as shown in Figure 11-24.

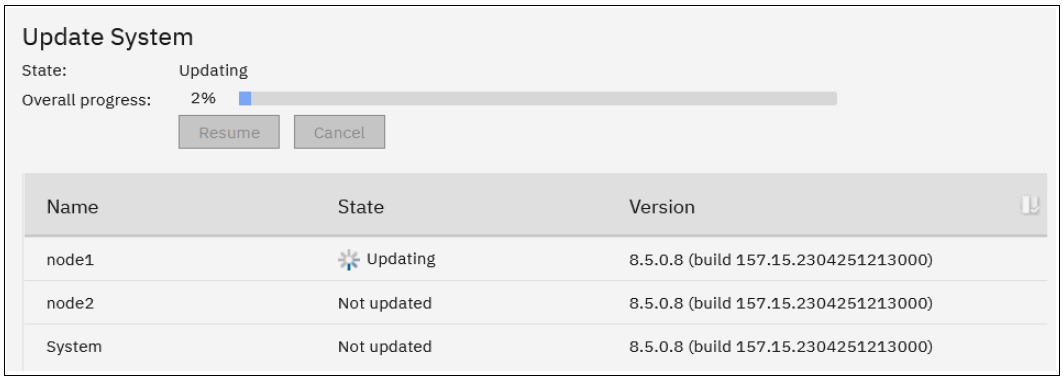


Figure 11-24 Resuming the update

Note: Because the utility detects issues, another warning appears to ensure that you investigated them and are certain that you want to proceed. When you are ready to proceed, click **Yes**.

8. The system begins updating the IBM Storage Virtualize Software by taking one node offline and installing the new code. This process takes approximately 20 minutes for each canister. After the node returns from the update, it is listed as complete, as shown in Figure 11-25.

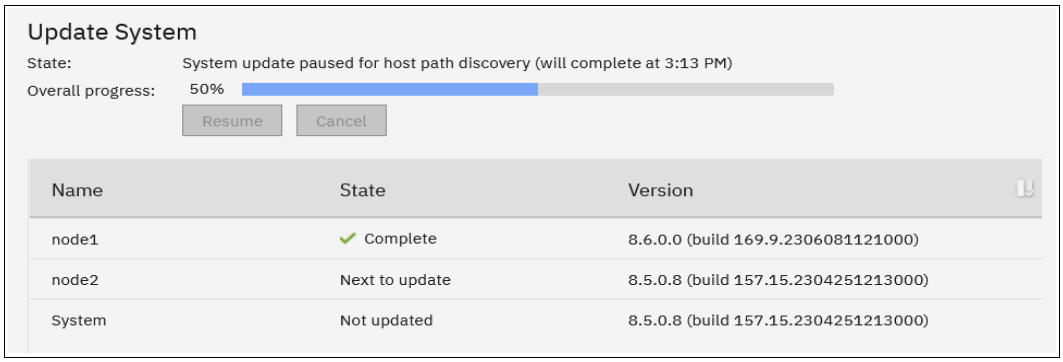


Figure 11-25 Update process paused for host path recovery

9. After a 30-minute pause at the half-way point to ensure that multipathing recovered on all attached hosts, a node failover occurs and you temporarily lose connection to the GUI. A warning window opens and prompts you to refresh the current session, as shown in Figure 11-26.

Tip: If you are updating from Version 7.8 or later, the 30-minute wait period can be adjusted by running **`applysoftware -delay (mins)`** parameter to begin the update instead of using the GUI.

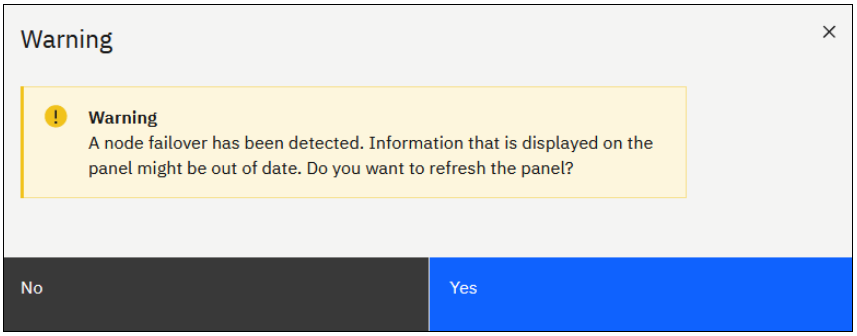


Figure 11-26 Node failover

Tip: If the above warning appears repeatedly during an update, then open your browser settings and clear all cookies.

The update process completes when all nodes and the system unit are committed. The final status indicates the new level of code that is installed in the system as shown in Figure 11-27

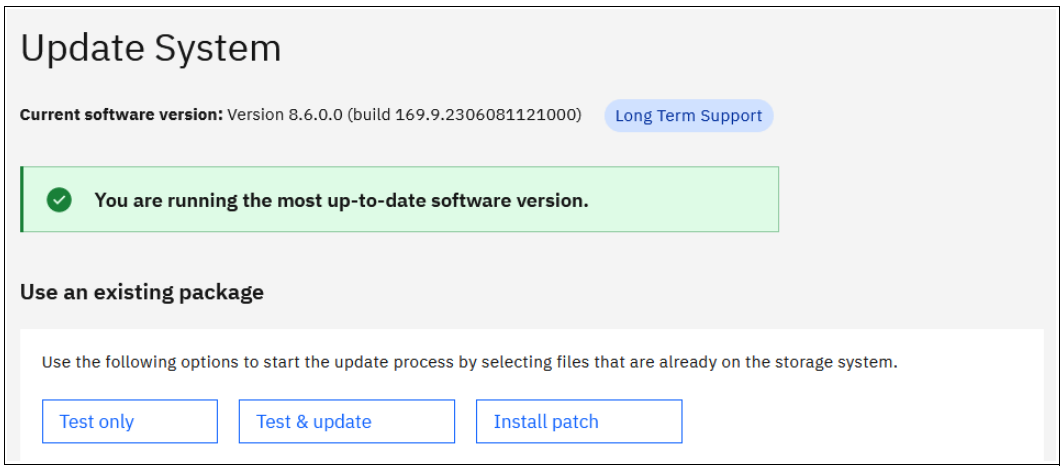


Figure 11-27 The update has finished

In Storage Virtualize version 8.6 and later, the Update System window also includes an Install Patch option. This is for updates to the system which do not require a complete code update. Install Patch is only when instructed by IBM Support personnel.

11.5.5 Updating the IBM FlashSystem drive code

Note: This section does not pertain to IBM SAN Volume Controller because no data drives are supported on IBM SAN Volume Controller hardware.

After completing the software update as described in 11.5, “Updating software” on page 984, the firmware of the disk drives in the system might also need to be updated. The upgrade test utility might identify drives in the system that require an update, as shown in Figure 11-28. However, this warning does not stop the system software update.

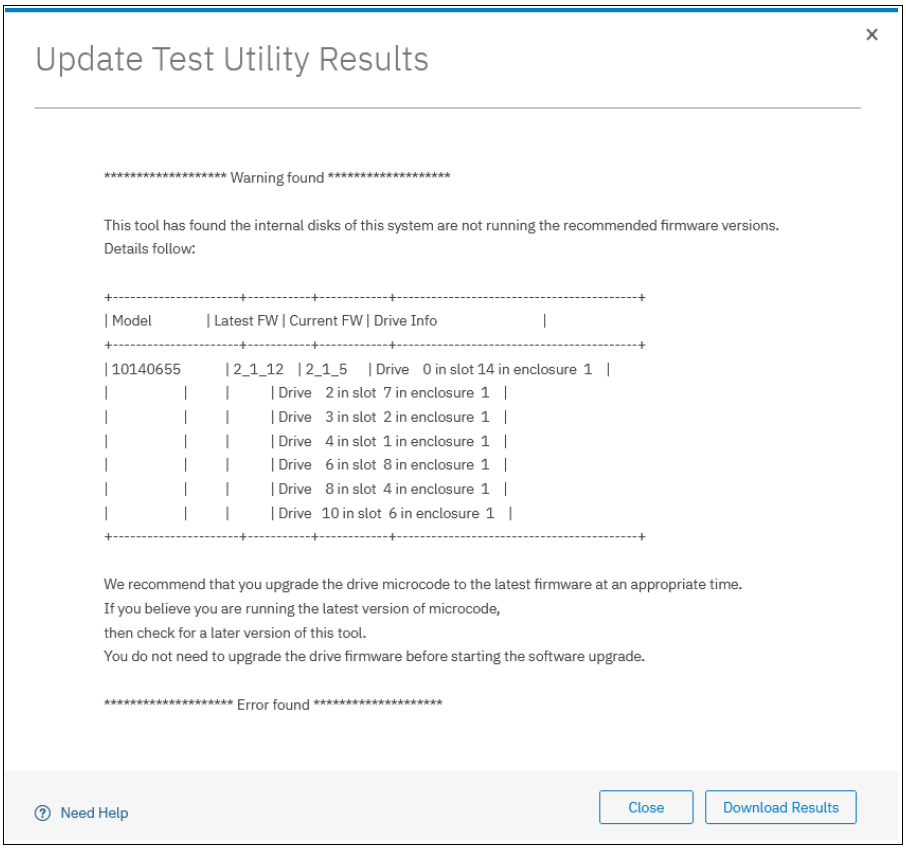


Figure 11-28 Upgrade Test Utility drive firmware warning

To update the drive code, complete the following steps:

1. Download the latest drive firmware package from this IBM Fix Central web page. Make sure that you select the correct product.
2. In the GUI, select **Pools** → **Internal Storage**.
3. Right-click the menu bar or click the icon as shown in Figure 11-29 on page 994. Then select **Firmware Level**.

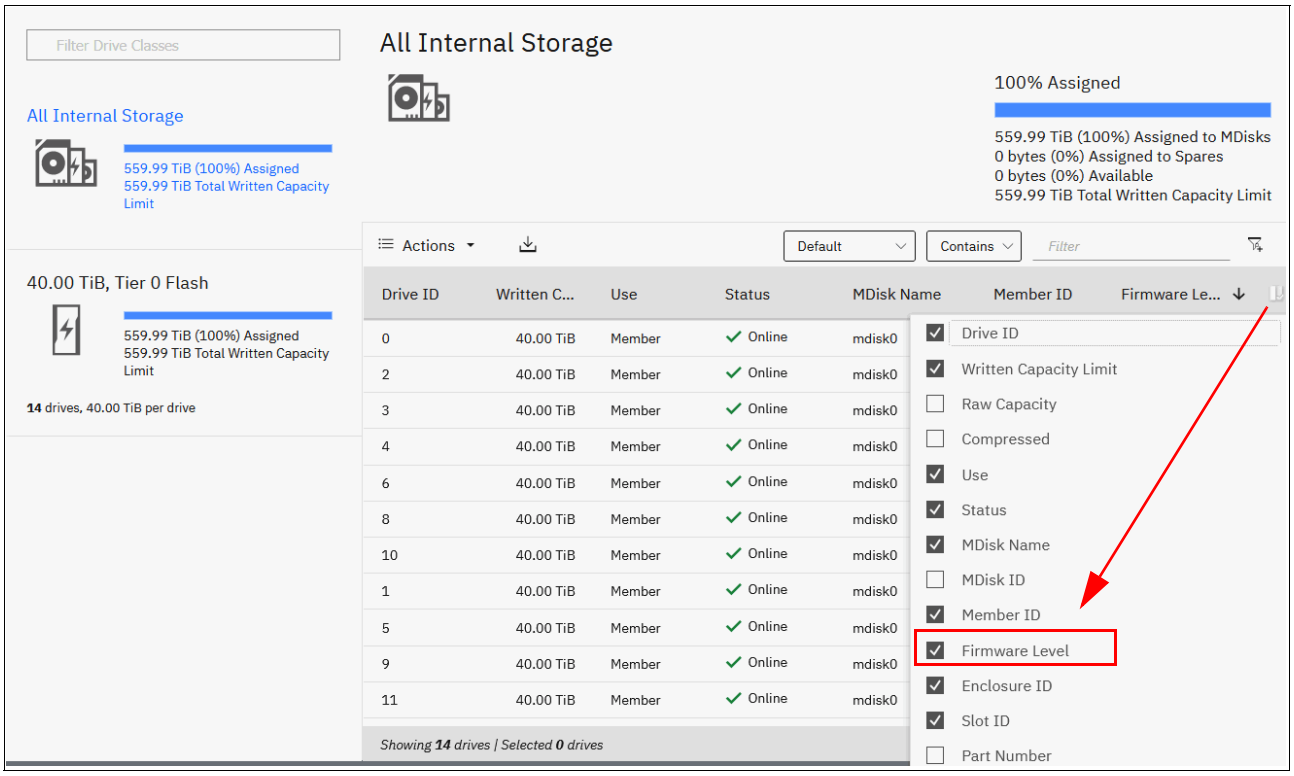


Figure 11-29 Internal Storage view versions

The GUI shows a system with 14 NVMe drives. Seven of these drives are on an older version of firmware and require an update.

4. Select the drives with downlevel code, right-click and select **Update** as shown in Figure 11-30.

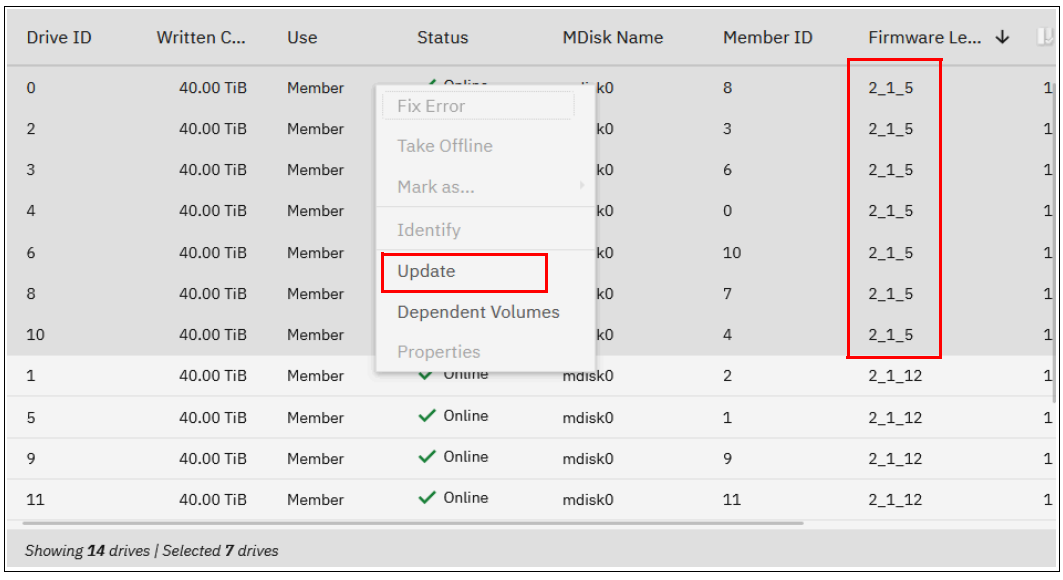


Figure 11-30 Select drives and update

Another option for drive update is to click the **Update all** function. This will select all drives and start the update process.

Tip: The **Update all** action is listed if you did not select any individual drive in the list. If you clicked an individual drive in the list, the action gives you individual drive actions; selecting **Update** upgrades only that drive's firmware. You can clear an individual drive by pressing Ctrl and clicking the drive again.

5. The Drive Update Checker window opens, as shown in Figure 11-31, in which you select the Upgrade Test Utility and the file containing the new drive firmware. Click **Next**.

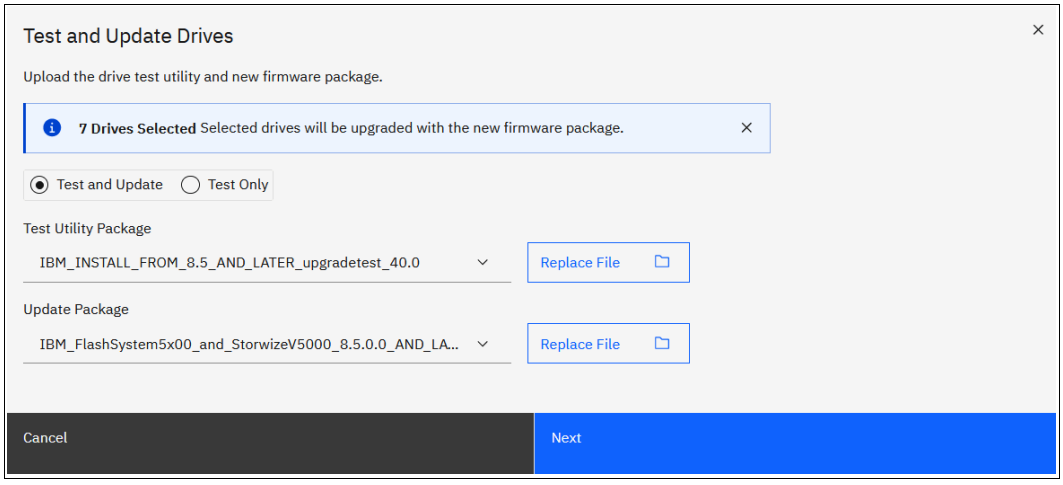


Figure 11-31 Selecting files on Drive Update Checker

6. The CLI commands for file upload and installation shown are shown in Figure 11-32. Click **Close** on the window.

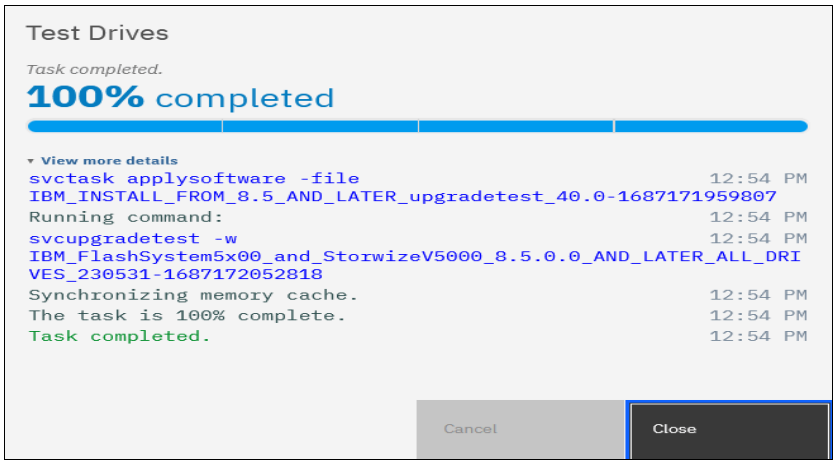


Figure 11-32 CLI command shows

7. If all drives are safe to update the drive update checker shows the Drive Update test Success window as shown in Figure 11-33. Click **Continue Update**.

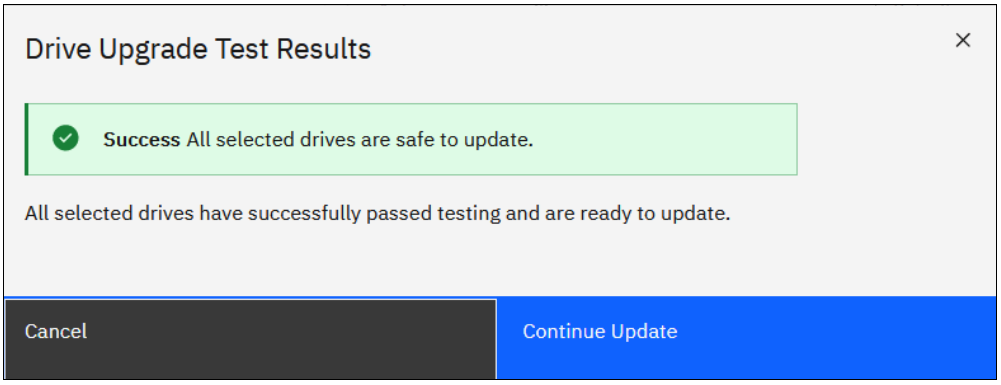


Figure 11-33 All drives safe to update

8. The CLI command for drive update start shows as in Figure 11-34. Click **Close** to proceed.

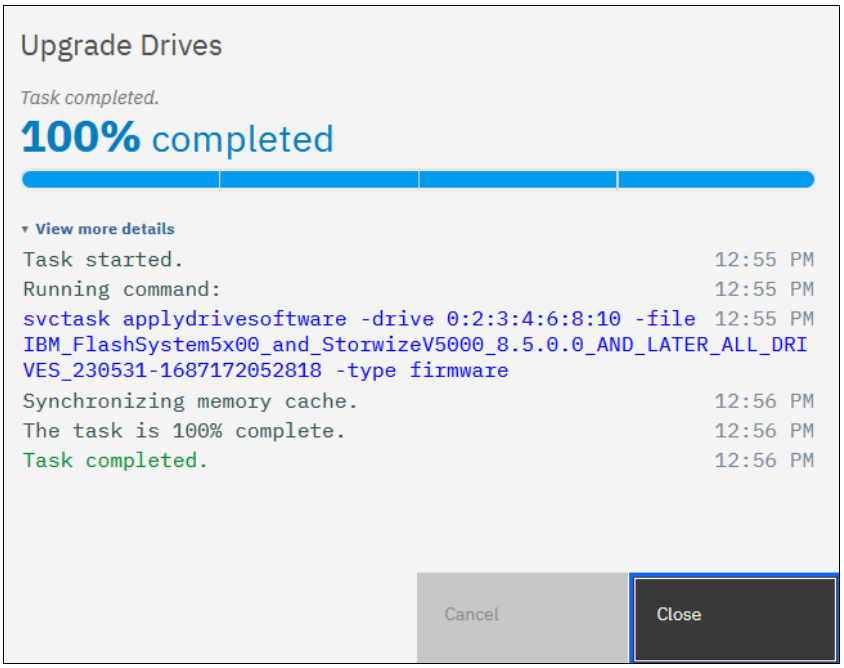


Figure 11-34 CLI command shows for updating drives

9. Figure 11-35 shows the Drive Update Started window. Click **Close** to complete the drive update start procedure.

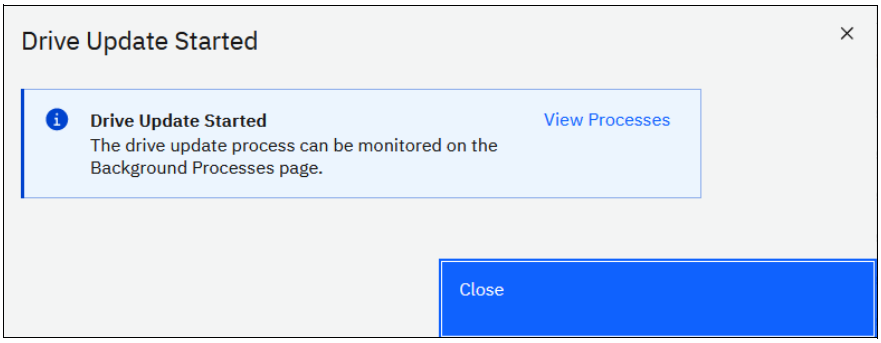


Figure 11-35 Drive update started

Note: The system upgrades each drive one at a time. Although the firmware upgrades are concurrent, they do cause a brief reset to the drive. However, the redundant array of independent disks (RAID) technology enables the system to continue after this brief interruption. After a drive completes its update, a calculated wait time exists before the next drive updates to ensure that the previous drive is stable after upgrading and can vary on system load.

10. With the drive upgrades running, you can view the progress by clicking the **Tasks** icon and clicking **View** for the Drive Upgrade running task, as shown in Figure 11-36.

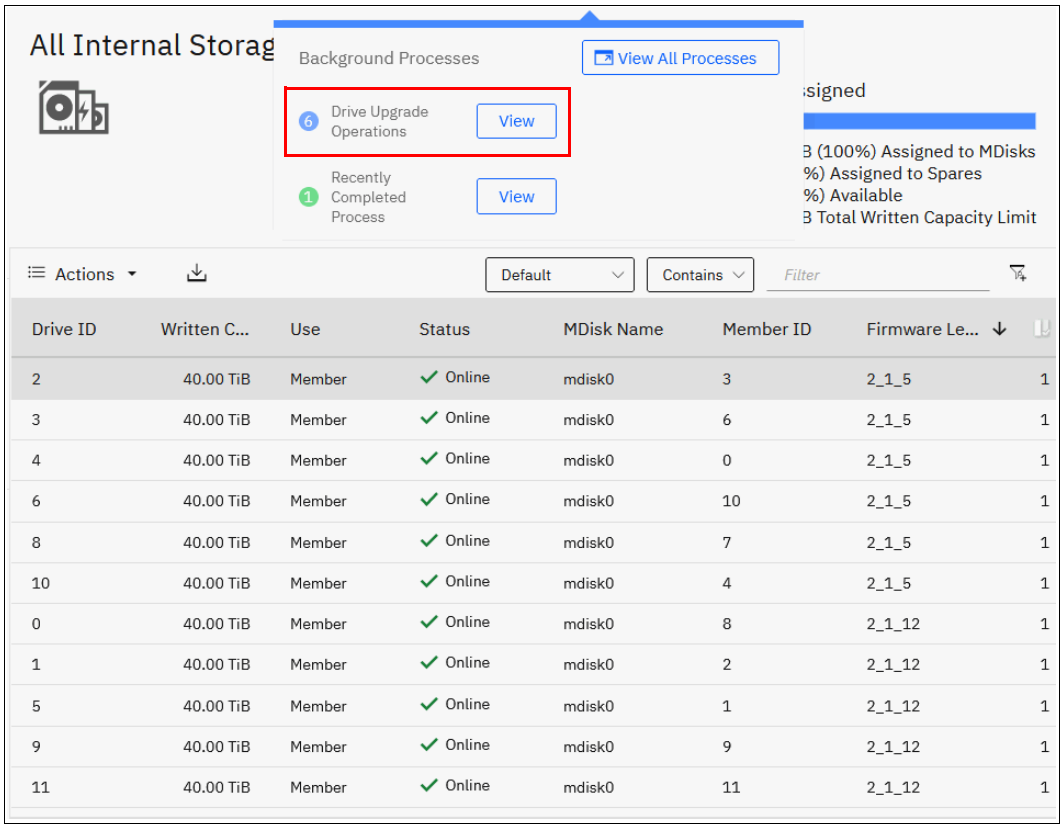


Figure 11-36 Selecting Drive Upgrade running task view

The Drive upgrade task window opens. The drives that are pending upgrade and an estimated time of completion are visible, as shown in Figure 11-37.

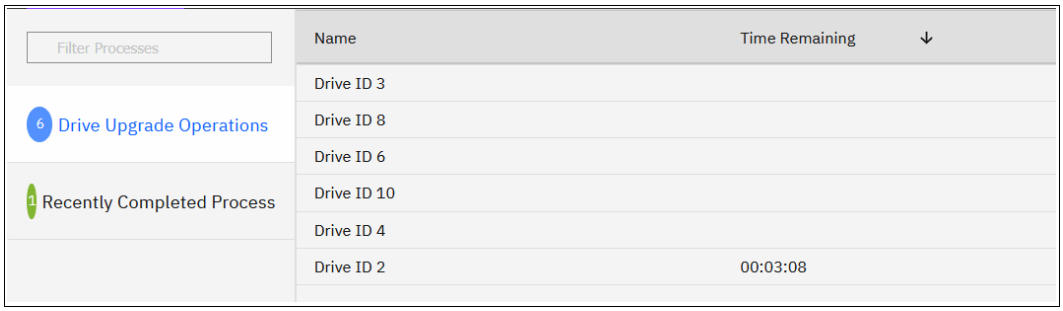


Figure 11-37 Drive upgrade progress for a single drive upgrade

When all drives are updated, the new firmware version will show in the GUI as shown in Figure 11-38

Drive ID	Written C...	Use	Status	MDisk Name	Member ID	Firmware
0	40.00 TiB	Member	Online	mdisk0	8	2_1_12
1	40.00 TiB	Member	Online	mdisk0	2	2_1_12
2	40.00 TiB	Member	Online	mdisk0	3	2_1_12
3	40.00 TiB	Member	Online	mdisk0	6	2_1_12
4	40.00 TiB	Member	Online	mdisk0	0	2_1_12
5	40.00 TiB	Member	Online	mdisk0	1	2_1_12
6	40.00 TiB	Member	Online	mdisk0	10	2_1_12
8	40.00 TiB	Member	Online	mdisk0	7	2_1_12
9	40.00 TiB	Member	Online	mdisk0	9	2_1_12
10	40.00 TiB	Member	Online	mdisk0	4	2_1_12
11	40.00 TiB	Member	Online	mdisk0	11	2_1_12

*Showing **14** drives | Selected **1** drive*

Figure 11-38 All drives updated

Tip: An FCM drive takes approximately six minutes to update. Drives are updated one at a time, so updating 24 drives in a system takes approximately 2.5 hours for the drives to update.

11.5.6 Manually updating the system

This example assumes that you have an 8-node canister cluster, as shown in Table 11-13.

Table 11-13 The iogrp setup

iogrp (0)	iogrp (1)	iogrp (2)	iogrp (3)
Node 1 (config node)	Node 3	Node 5	Node 7
Node 2	Node 4	Node 6	Node 8

After uploading the update utility test and software update package to the cluster by using PSCP and running the utility test, complete the following steps:

1. Start by removing node 2, which is the partner node of the configuration node in iogrp 0, by using the cluster GUI or CLI.
2. Log in to the service GUI to verify that the removed node is in the `candidate` status.
3. Select the candidate node and click **Update Manually** from the left pane.
4. Browse and find the code that you downloaded and saved to your PC.
5. Upload the code and click **Update**.

   When the update completes, a message caption indicating software update completion displays. The node then restarts, and appears again in the service GUI (after approximately 20 - 25 minutes) in the `candidate` status.
6. Select the node and verify that it is updated to the new code.
7. Add the node back by using the cluster GUI or the CLI.
8. Select **node 3** from `iogrp1`.
9. Repeat steps 1 - 7 to remove node 3, update it manually, verify the code, and add it back to the cluster.
10. Proceed to **node 5** in `iogrp 2`.
11. Repeat steps 1 - 7 to remove node 5, update it manually, verify the code, and add it back to the cluster.
12. Move on to **node 7** in `iogrp 3`.
13. Repeat steps 1 - 7 to remove node 5, update it manually, verify the code, and add it back to the cluster.

> **Note:** The update is 50% complete. You now have one node from each iogrp that is updated with the new code manually. Always leave the configuration node for last during a manual software update.

14. Select **node 4** from `iogrp 1`.
15. Repeat steps 1 - 7 to remove node 4, update it manually, verify the code, and add it back to the cluster.
16. Select **node 6** from `iogrp 2`.
17. Repeat steps 1 - 7 to remove node 6, update it manually, verify the code, and add it back to the cluster.
18. Select **node 8** in `iogrp 3`.

19. Repeat steps 1 - 7 to remove node 8, update it manually, verify the code, and add it back to the cluster.
20. Select and remove **node 1**, which is the configuration node in .

> **Note:** A partner node becomes the configuration node because the original configuration node is removed from the cluster, which keeps the cluster manageable.

The removed configuration node becomes a candidate, and you do not have to apply the code update manually. Add the node back to the cluster. It automatically updates itself and then adds itself back to the cluster with the new code.

21. After all the nodes are updated, you must confirm the update to complete the process. The confirmation restarts each node in order, which takes about 30 minutes to complete.

The update is complete.

11.6 Health checker feature

The IBM Storage Control health checker feature runs in IBM Cloud. Based on the weekly Call Home Connect Cloud inventory reporting, the health checker proactively creates recommendations. These recommendations are provided at IBM Call Home Connect Cloud web, which is found at Call Home Connect Cloud (login required).

For a video guide about how to set up and use IBM Call Home Connect Cloud web, see Introducing IBM Call Home Web.

Another feature is the *Critical Fix Notification* function, which enables IBM to warn users that a critical issue exists in the level of code that they are using. The system notifies users when they log on to the GUI by using a web browser that is connected to the internet.

Consider the following information about this function:

- It warns users only about critical fixes, and does not warn them that they are running a previous version of the software.
- It works only if the browser also has access to the internet. The system itself does not need to be connected to the internet.
- The function cannot be disabled. Each time that it displays a warning, it must be acknowledged (with the option to not warn the user again for that issue).

The decision about what is a *critical* fix is subjective and requires judgment, which is exercised by the development team. As a result, clients might still encounter bugs in code that were not deemed critical. They continue to review information about new code levels to determine whether they must update, even without a critical fix notification.

> **Important:** Inventory notification must be enabled and operational for these features to work. It is a best practice to enable Call Home Connect Cloud and Inventory reporting on your IBM Storage Virtualize clusters.

11.7 Troubleshooting and fixing procedures

The management GUI of IBM FlashSystem is a browser-based GUI for configuring and managing all aspects of your system. It provides extensive facilities to help troubleshoot and correct problems. This section explains how to effectively use its features to avoid service disruption of your system.

Figure 11-39 shows the Monitoring menu icon for System Hardware, Easy Tier Reports, viewing events, or seeing real-time performance statistics.

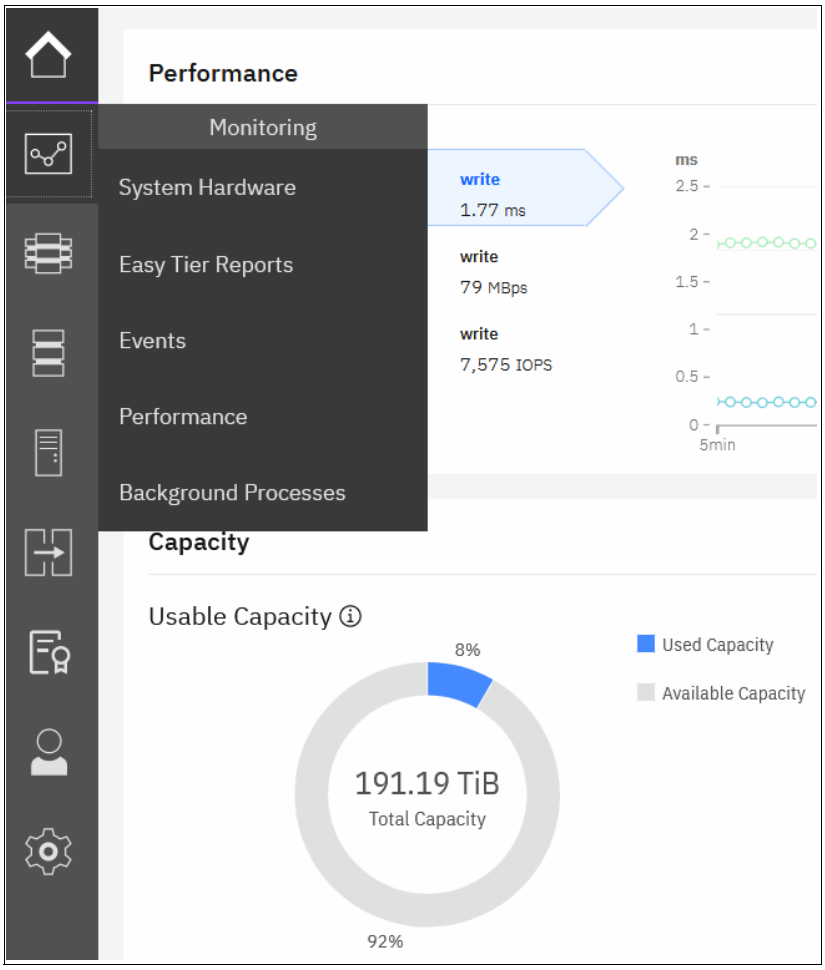


Figure 11-39 Monitoring options

Use the management GUI to manage and service your system. Select **Monitoring** → **Events** to list events that should be addressed and maintenance procedures that walk you through the process of correcting problems. Information in the Events window can be filtered four ways:

- Recommended Actions

  Shows only the alerts that require attention. Alerts are listed in priority order and should be resolved sequentially by using the available fix procedures. For each problem that is selected, you can perform the following tasks:

  - Run a fix procedure
  - View the properties

- Unfixed Alerts

Displays only the alerts that are not fixed. For each entry that is selected, you can perform the following tasks:

- Run a fix procedure
- Mark an event as fixed
- Filter the entries to show them by specific minutes, hours, or dates
- Reset the date filter
- View the properties

- Unfixed Messages and Alerts

  Displays only the alerts and messages that are not fixed. For each entry that is selected, you can perform the following tasks:

  - Run a fix procedure
  - Mark an event as fixed
  - Filter the entries to show them by specific minutes, hours, or dates
  - Reset the date filter
  - View the properties

- Show All

  Displays all event types whether they are fixed or unfixed. For each entry that is selected, you can perform the following tasks:

  - Run a fix procedure
  - Mark an event as fixed
  - Filter the entries to show them by specific minutes, hours, or dates
  - Reset the date filter
  - View the properties

Some events require a certain number of occurrences in 25 hours before they are displayed as unfixed. If they do not reach this threshold in 25 hours, they are flagged as *expired*. Monitoring events are below the coalesce threshold, and are transient.

Important: The management GUI is the primary tool that is used to *operate* and *service* your system. Real-time *monitoring* should be established by using SNMP traps, email notifications, or syslog messaging in an automatic manner.

11.7.1 Managing the event log

Regularly check the status of the system by using the management GUI. If you suspect a problem, first use the management GUI to diagnose and resolve the problem.

Use the views that are available in the management GUI to verify the status of the system, the hardware devices, the physical storage, and the available volumes by completing the following steps:

1. Select **Monitoring** → **Events** to see all problems that exist on the system (see Figure 11-40 on page 1003).

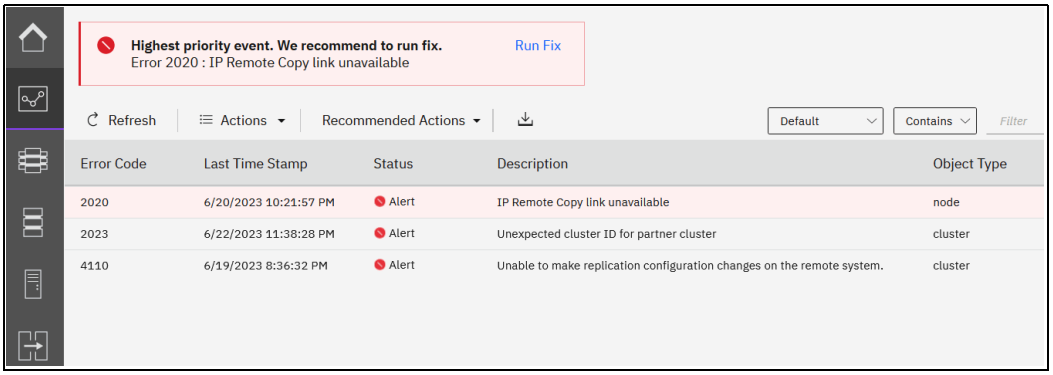


Figure 11-40 Messages in the event log

2. If not already chosen, select **Recommended Actions** from the drop-down list to display the most important events to be resolved (see Figure 11-41). The **Recommended Actions** tab shows the highest priority maintenance procedure that must be run. Use the troubleshooting wizard so that the system can determine the proper order of maintenance procedures.

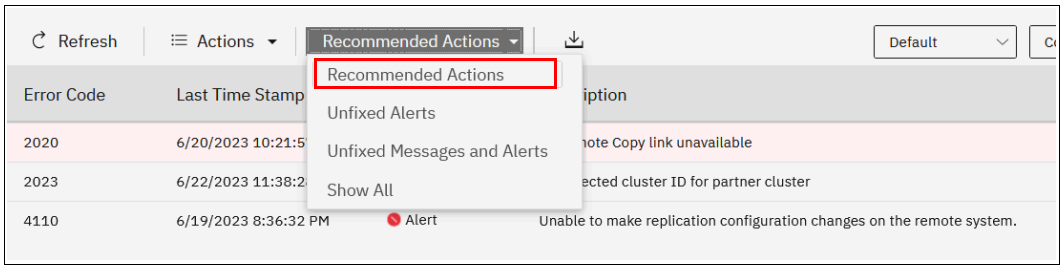


Figure 11-41 Recommended Actions

In this example, the highest priority message is for Remote Copy which has a link unavailable (service error code 2020). At any time and from any GUI window, you can directly go to this menu by clicking the **Status Alerts** icon at the top of the GUI (Figure 11-42).

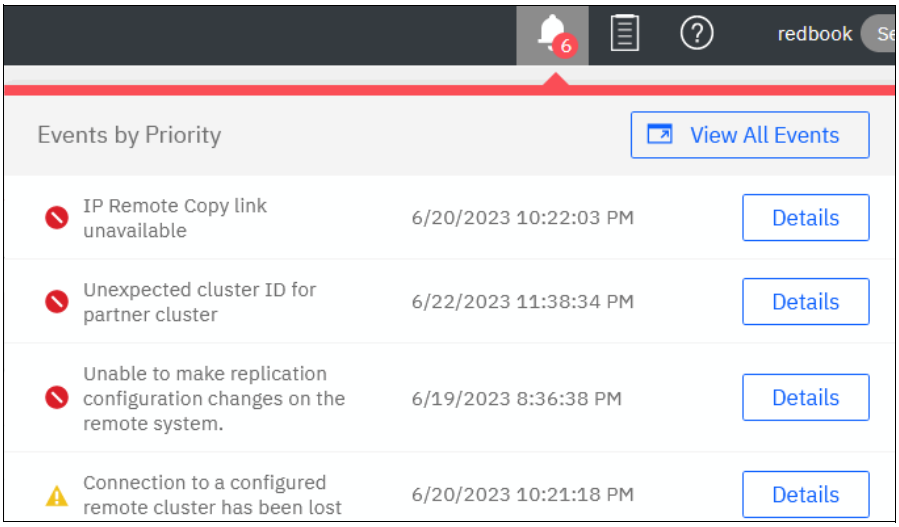


Figure 11-42 Status alerts

11.7.2 Running a fix procedure

If an error code exists for the alert, run the fix procedure to help you resolve the problem. These fix procedures analyze the system and provide more information about the problem. They suggest actions to take and walk you through the actions that automatically manage the system where necessary while ensuring availability. Finally, they verify that the problem is resolved.

If an error is reported, always use the fix procedures from the management GUI to resolve the problem for both software configuration problems and hardware failures. The fix procedures analyze the system to ensure that the required changes do not cause volumes to become inaccessible to the hosts. The fix procedures automatically perform configuration changes that are required to return the system to its optimum state.

The fix procedure displays information that is relevant to the problem, and it provides various options to correct the problem. Where possible, the fix procedure runs the commands that are required to reconfigure the system.

The fix procedure also checks that any other existing problems do not result in volume access being lost. For example, if a Power Supply Unit (PSU) in a node enclosure must be replaced, the fix procedure checks and warns you whether the integrated battery in the other PSU is not sufficiently charged to protect the system.

Hint: Always use **Run Fix**, which resolves the most serious issues first. Often, other alerts are corrected automatically because they were the result of a more serious issue.

Resolving alerts in a timely manner

To minimize any impact to your host systems, always perform the recommended actions as quickly as possible after a problem is reported. Your system is resilient to most single hardware failures.

However, if it operates for any period with a hardware failure, the possibility increases that a second hardware failure can result in some volume data unavailability. If several unfixed alerts exist, fixing any one alert might become more difficult because of the effects of the others.

11.7.3 Event log details

Multiple views of the events and recommended actions are available (see Figure 11-43 on page 1005). When you click the column icon at the right end of the table heading, or right-click the items bar, a menu for the column choices opens.

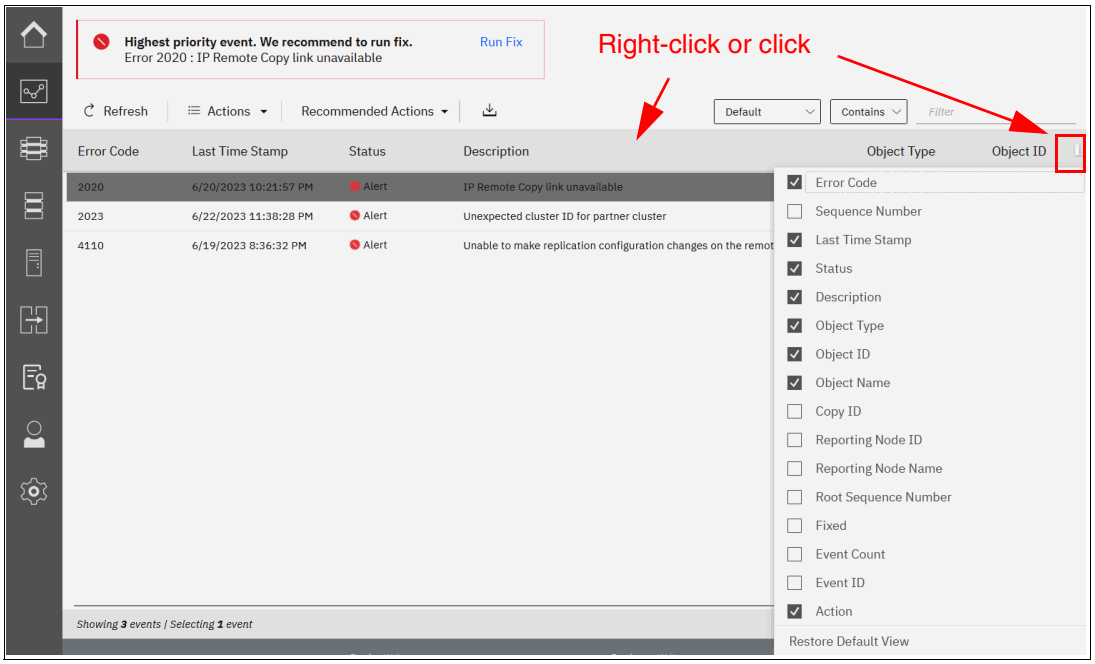


Figure 11-43 Grid options of the event log

Select or remove columns as needed. You can also extend or shrink the width of columns to fit your window resolution and size. This method is relevant for most windows in the management GUI of an IBM FlashSystem system.

Every field of the event log is available as a column in the event log grid. Several fields are useful when you work with IBM Support. The preferred method in this case is to use the Show All filter, with events sorted by time stamp. All fields have the sequence number, event count, and the fixed state. Clicking **Restore Default View** sets the grid back to the defaults.

You might want to see more details about each critical event. Some details are not shown in the main grid. To access the properties and sense data of a specific event, double-click the specific event anywhere in its row.

The properties window opens (see Figure 11-44) with all the relevant sense data. This data includes the first and last time of an event occurrence, number of times the event occurred, worldwide port name (WWPN), worldwide node name (WWNN), enabled or disabled automatic fix, and other information.

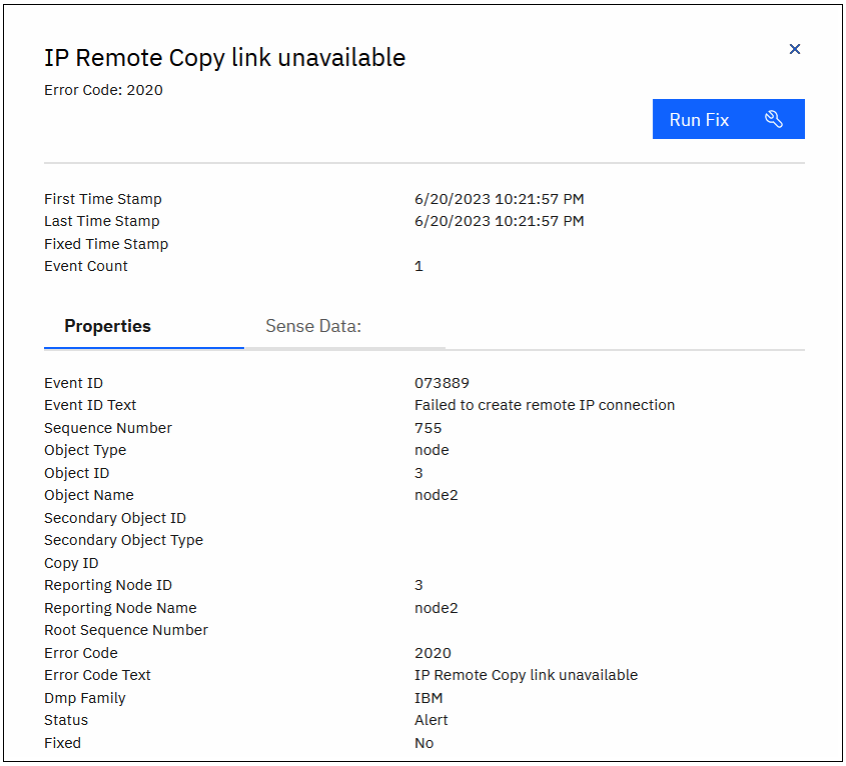


Figure 11-44 Event sense data and properties

For more information about troubleshooting options, search for “Troubleshooting” at IBM Documentation.

11.8 Monitoring and Event Notification

An important step is to correct any issues that are reported by your system as soon as possible. Configure your system to send automatic notifications to a standard Call Home server or to the Call Home Connect Cloud server when a new event is reported. To avoid the need to monitor the management GUI for new events, select the type of event for which you want to be notified. For example, you can restrict notifications to only events that require action.

The following event notification mechanisms are available:

- Call Home

  An event notification can be sent to one or more email addresses. This mechanism notifies individuals of problems. Individuals can receive notifications wherever they have email access, including mobile devices.

- Call Home Connect Cloud

  Cloud services for Call Home is the optimal transmission method for error data because it ensures that notifications are delivered directly to the IBM Support Center.

- SNMP

  An SNMP traps report can be sent to a data center management system, such as IBM Systems Director, which consolidates SNMP reports from multiple systems. With this mechanism, you can monitor your data center from a single workstation.

- Syslog

  A syslog report can be sent to a data center management system that consolidates syslog reports from multiple systems. With this option, you can monitor your data center from a single location.

If your system is within warranty or if you have a hardware maintenance agreement, configure your IBM FlashSystem system to send email events directly to IBM if an issue that requires hardware replacement is detected. This mechanism is known as *Call Home*. When this event is received, IBM automatically opens a problem report and, if appropriate, contacts you to help resolve the reported problem.

Important: If you set up Call Home to IBM, ensure that the contact details that you configure are correct and kept updated. Personnel changes can cause delays in IBM making contact.

Call Home Connect Cloud is designed to work with new service teams and improves connectivity, which should improve customer support.

Note: If the customer does not want to open the firewall, Call Home Connect Cloud does not work and the customer can disable Call Home Connect Cloud. Call Home is used instead.

11.8.1 Email notifications and the Call Home function

The Call Home function of IBM FlashSystem uses the email notification that is sent to the specific IBM Support Center. Therefore, the configuration is like sending emails to the specific person or system owner.

The following procedure summarizes how to configure email notifications and emphasizes what is specific to Call Home:

1. Prepare your contact information that you want to use for the email notification and verify the accuracy of the data. From the GUI menu, select **Settings** → **Support** → **Call Home.**
2. Select **Call Home**, and then click **Enable Notifications** (see Figure 11-45). For more information, see Setting up call home withemail notifications.

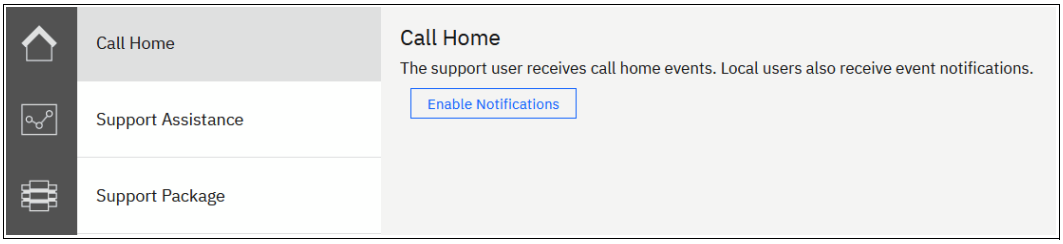


Figure 11-45 Configuring Call Home notifications

Figure 11-46 on page 1008 shows the Welcome screen of the Call Home enablement wizard. Click **Next** to proceed.

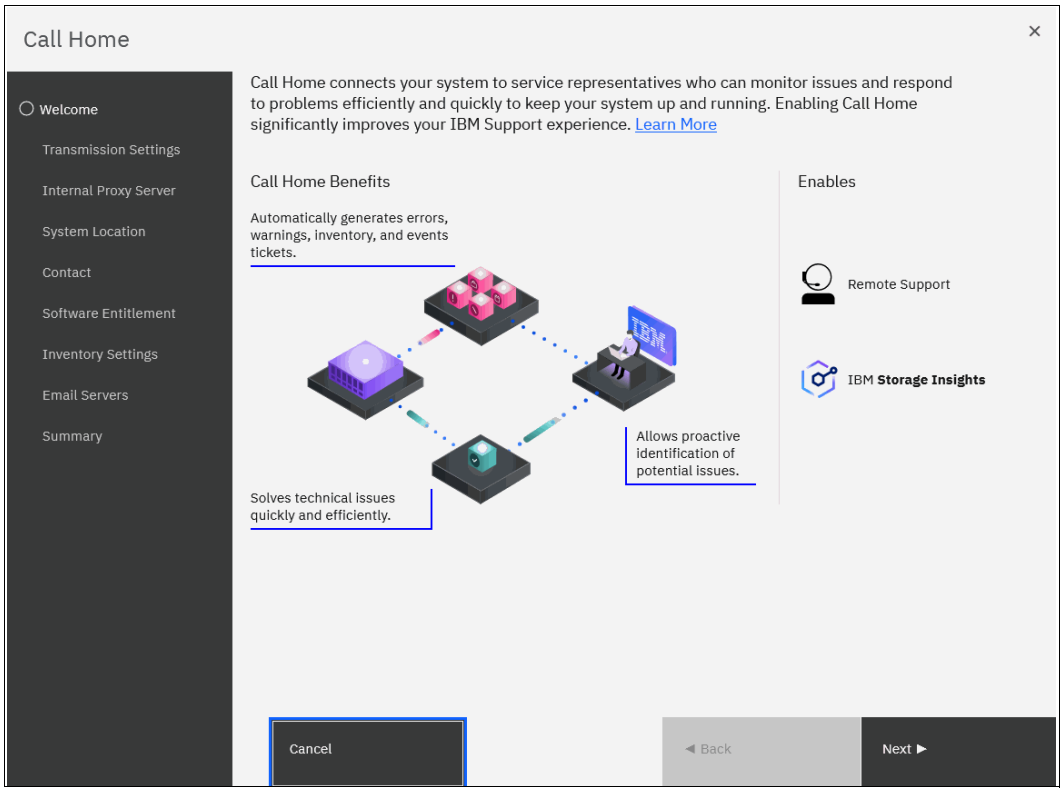


Figure 11-46 Call Home Connect Cloud service welcome screen

When you configure email notifications, ask your network administrator if Simple Mail Transfer Protocol (SMTP) is enabled on the management network and is not blocked by firewalls. Test the accessibility to the SMTP server by using the **`telnet`** command to any server in the same network segment. When you run the telnet command, use port 25 for a non-secured connection and port 465 for an SSL-encrypted communication.

The preferred method of transmission to IBM is using cloud services. Call home Connect Cloud is independent of email servers and sends notifications directly to support to improve error resolution. Call home Connect Cloud also and integrates with IBM Storage Insights for enhanced management of your systems.

Tip: Storage administrators might want to enable email notifications and cloud services. By enabling email notification local users can receive information, warning or error messages from the system.

3. Figure 11-47 shows how the transmission types are selected. Select one or both transmission types and click **Apply and next**.

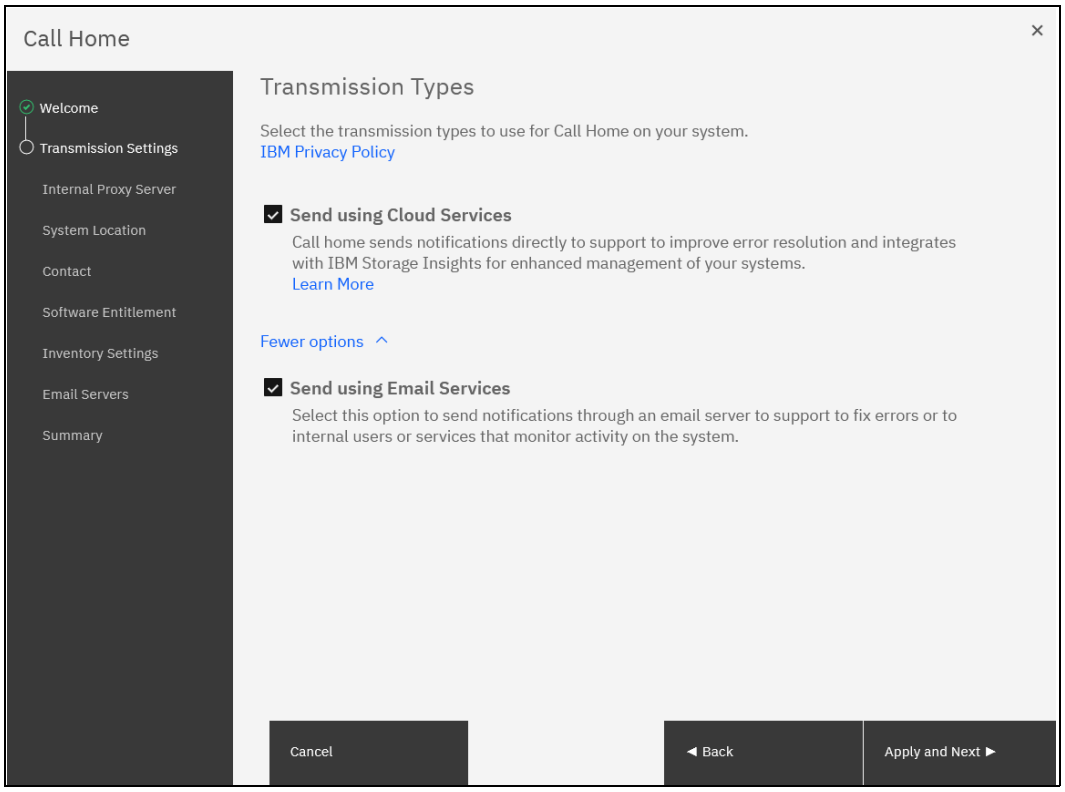


Figure 11-47 Transmission types

If using a proxy server select **Yes** at the next window. No proxy server is used in this example so **No** is selected. Click **Next** to continue as shown in Figure 11-48.

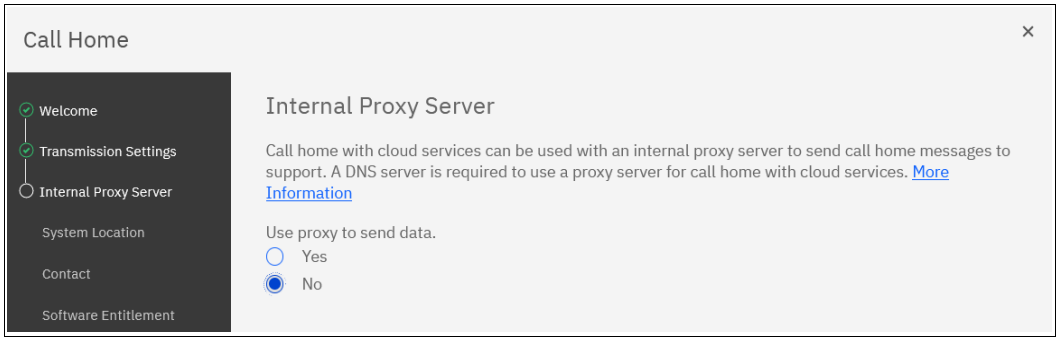


Figure 11-48 Proxy server

4. Enter the information about the location of the system (Figure 11-49) and contact information of the system administrator (Figure 11-50) to be contacted by IBM Support. *Always* keep this information current. Click **Apply and Next**.

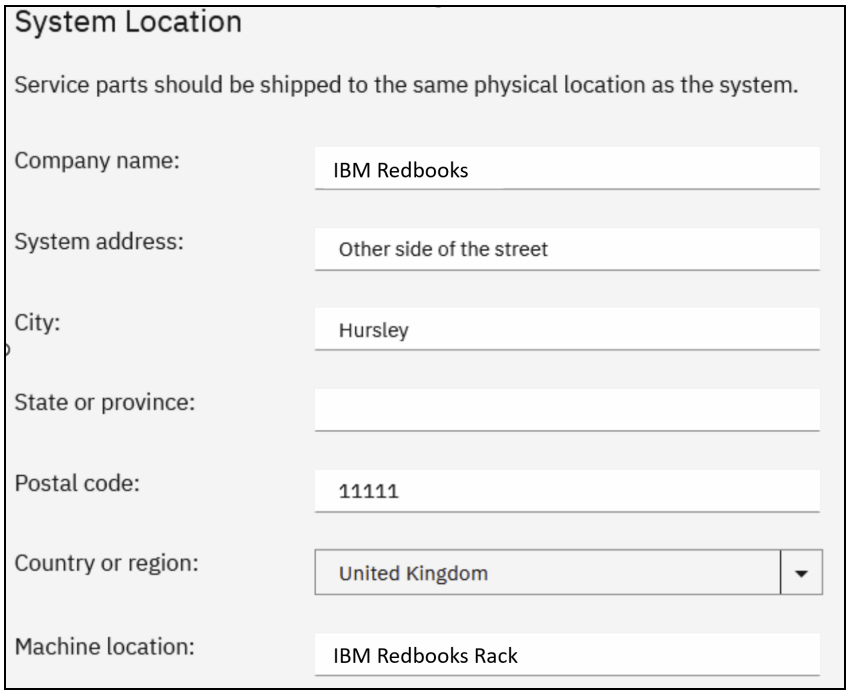


Figure 11-49 Location of the device

Next enter information of the contact person as shown in Figure 11-50. This is who IBM will contact in case of issues that have created a call home call. Click **Apply and Next**.

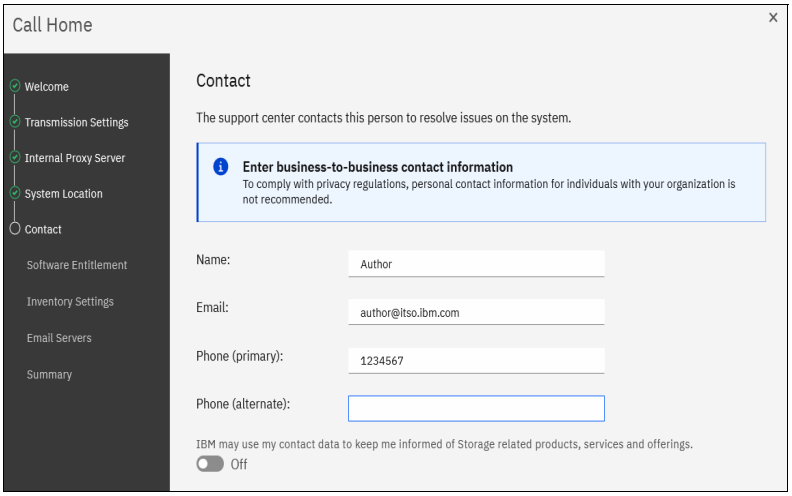


Figure 11-50 Contact information

Users can enter the account ID and the country code. Both can be found at IBM Passport Advantage®. If these values are unknown then leave the fields empty and click **Apply and next** as shown in Figure 11-51 on page 1011.

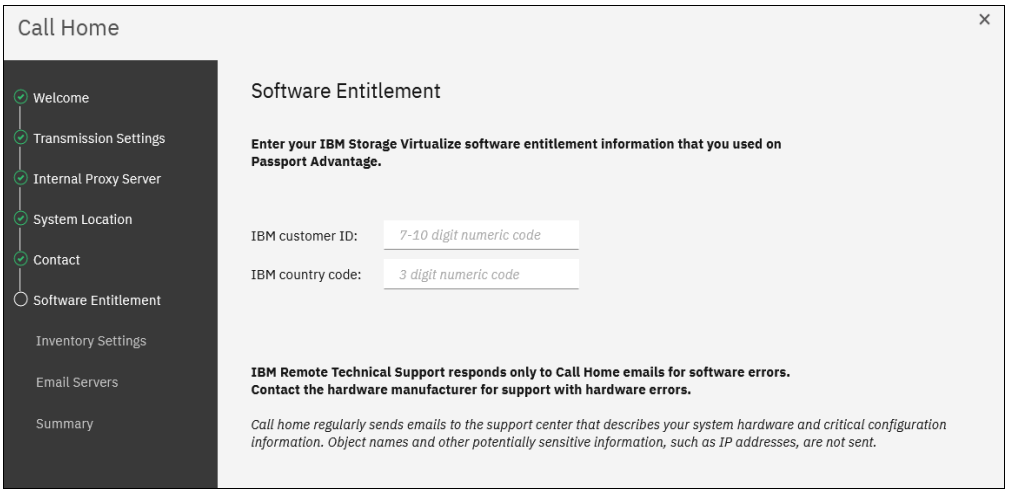


Figure 11-51 Software entitlement

Note: Software Entitlement is for applications with software maintenance only. This panel was introduced when support for SVC as Software was added. All new physical storage systems have warranty and hardware maintenance from IBM, which are entitled on the serial number directly. Software Entitlement is not necessary for enabling call home for storage hardware. The Software entitlement is present in the initial release of Storage Virtualize version 8.6.0.0 but is removed in later versions.

In the next window, you can enable Inventory Reporting and Configuration Reporting, as shown in Figure 11-52.

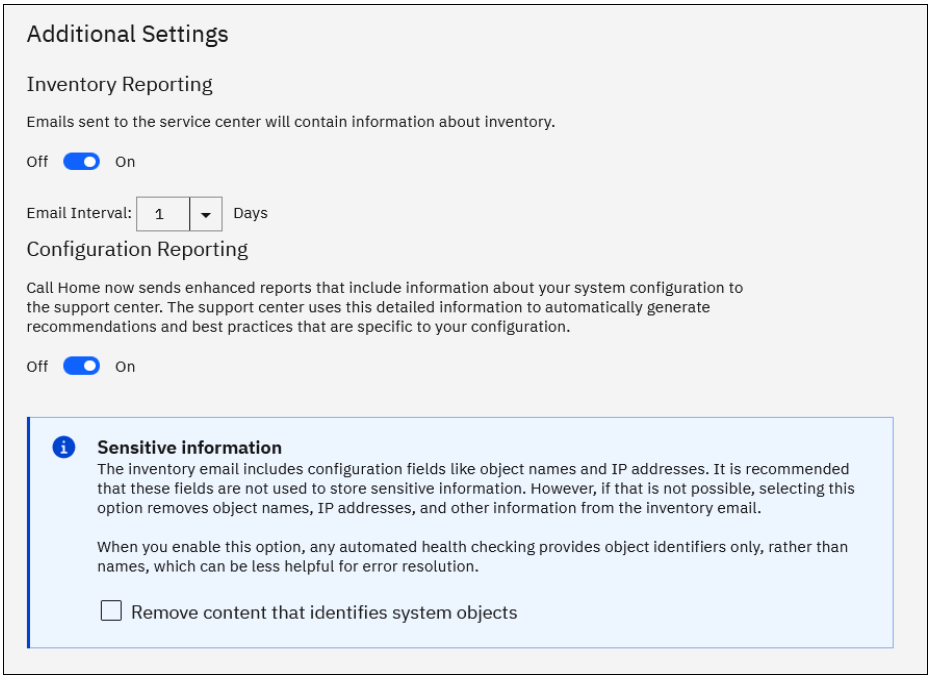


Figure 11-52 Inventory Reporting and Configuration Reporting

5. Configure the SMTP server according to the instructions that are shown in Figure 11-53. When the correct SMTP server is provided, you can test the connectivity by clicking **Ping** to verify that it can be contacted. If DNS has been configured on the storage system, then the name of the SMTP server can be entered. Otherwise enter the IP address. Then click **Apply and Next**.

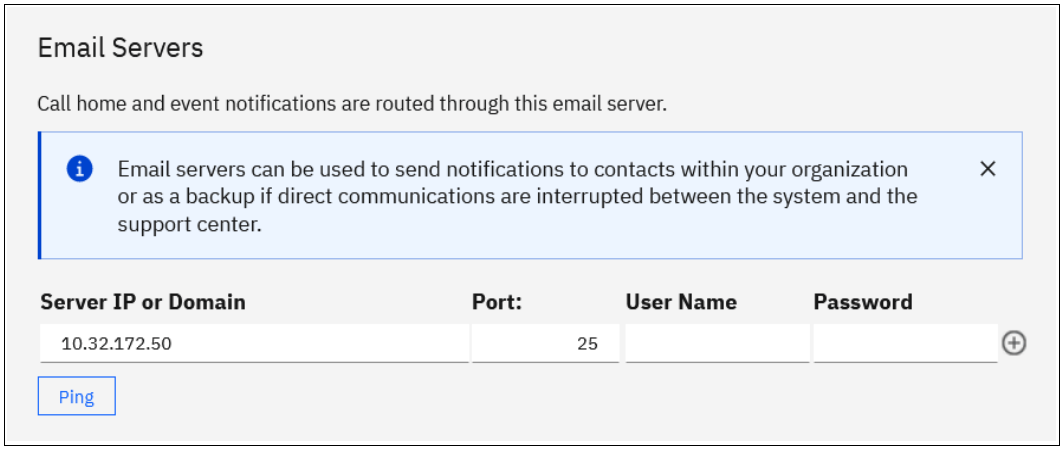


Figure 11-53 Configuring email servers and inventory reporting

6. A summary window opens. Verify all the information and click **Finish**. You are returned to the Call Home settings window, where you can verify the email addresses of IBM Support, which is `callhome1@de.ibm.com`, and you can optionally add local users who also need to receive notifications (see Figure 11-54).

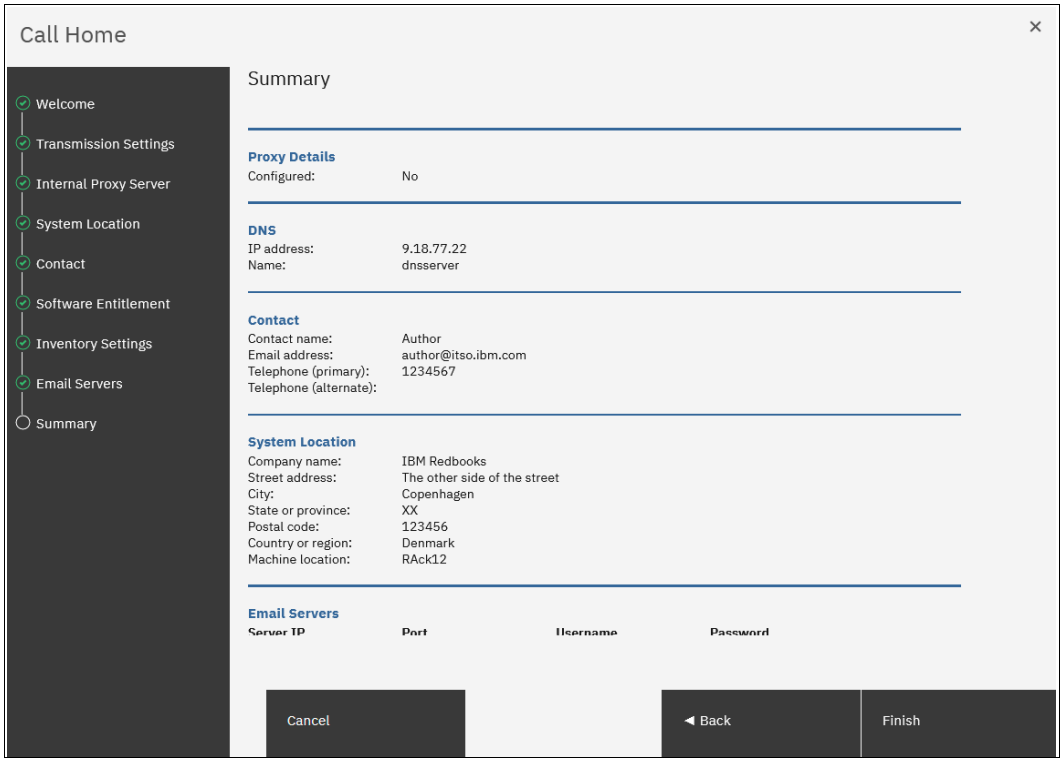


Figure 11-54 Call home settings summary

Note: The default support email address `callhome1@de.ibm.com` is predefined by the system to receive Error Events and Inventory messages. Do not change these settings or disable the 7-day reporting interval at the bottom of the Settings window.

You can modify or add local users by using Edit mode after the initial configuration is saved.

The **Inventory Reporting** function is enabled by default for Call Home. Rather than reporting a problem, an email is sent to IBM and other configured mail-receivers that describes your system hardware and critical configuration information. Object names and other information, such as IP addresses, are *not* included. By default, the inventory email is sent weekly, which allows an IBM Cloud service to analyze the inventory email and inform you whether the hardware or software that you are using requires an update because of any known issue, as described in 11.6, “Health checker feature” on page 1000.

Figure 11-54 on page 1012 shows the configured email notification and Call Home settings.

7. After completing the configuration wizard, test the email function. To do so, enter Edit mode, as shown in Figure 11-55. In the same window, you can define more email recipients or alter any contact and location details as needed.

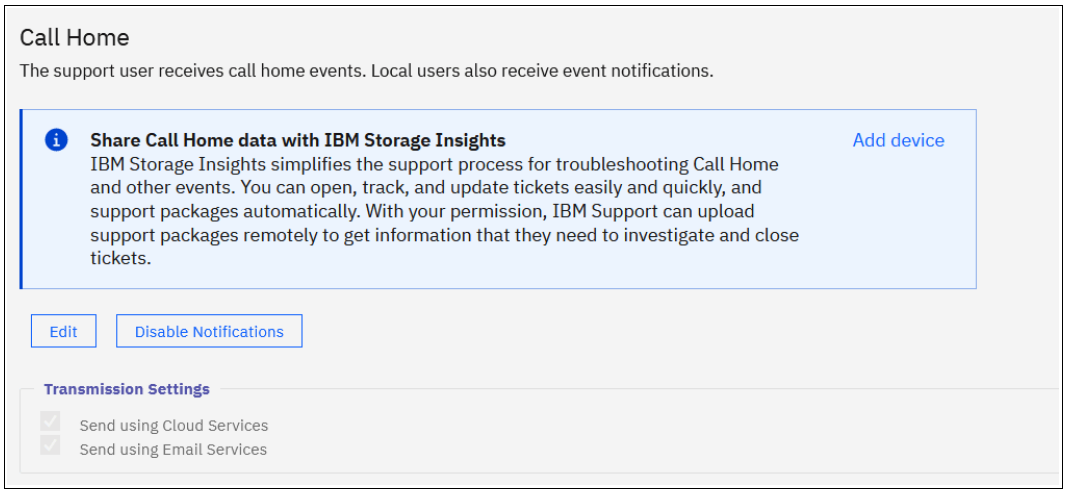


Figure 11-55 Entering Edit mode

8. In Edit mode, you can change any of the previously configured settings. After you are finished editing these parameters, adding more recipients, or testing the connection, save the configuration so that the changes take effect (see Figure 11-56).

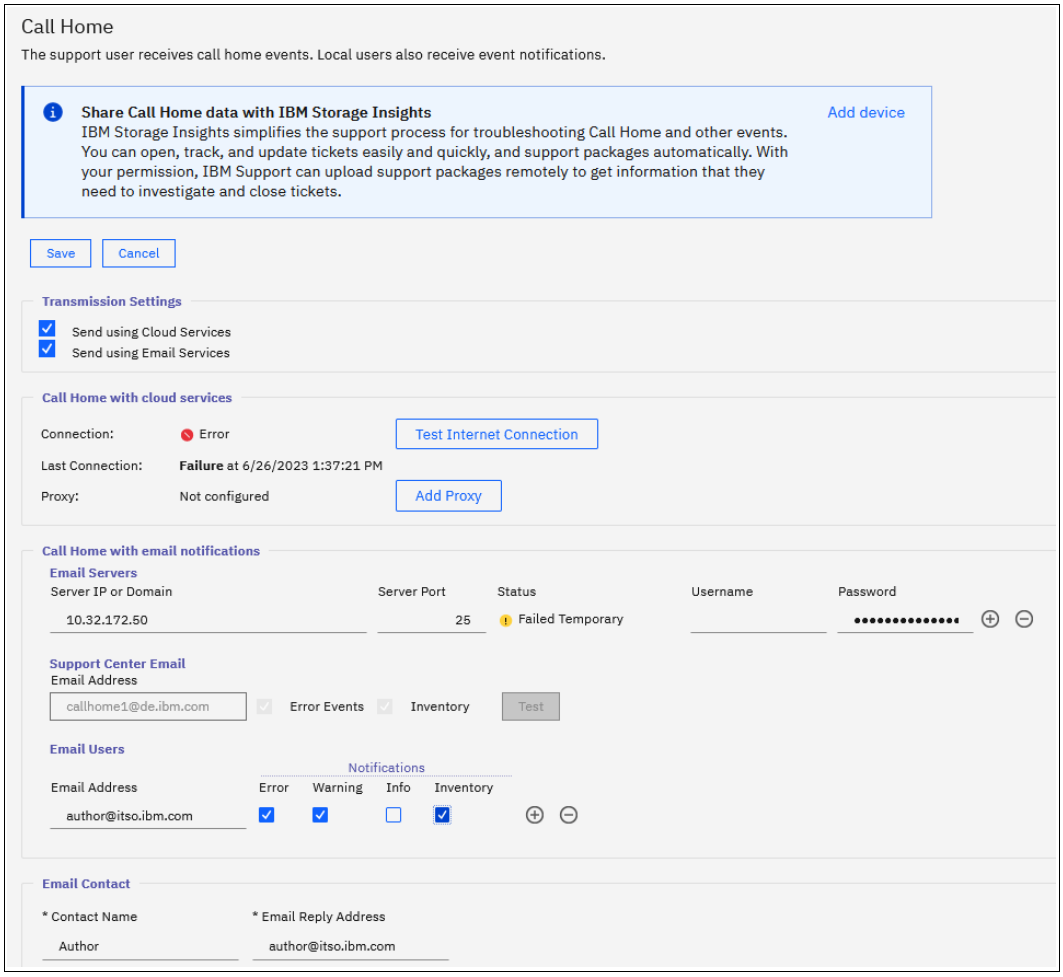


Figure 11-56 Saving a modified configuration

Note: The Test button appears for new email users after first saving and then editing again.

Disabling and enabling notifications

At any time, you can temporarily or permanently disable email notifications, as shown in Figure 11-57 on page 1015.

This best practice is useful when performing activities in your environment that might generate errors on IBM Storage Virtualize, such as SAN reconfiguration or replacement activities. After the planned activities, remember to re-enable the email notification function. The same results can be achieved by running the commands **`svctask stopemail`** and **`svctask startemail`**.

Demonstration video: For a demonstration video, see IBM Storage Spectrum Virtualize 8.6 GUI, including Volume Group Snapshots (with Safeguarded Copy) and Policy Based Replication.

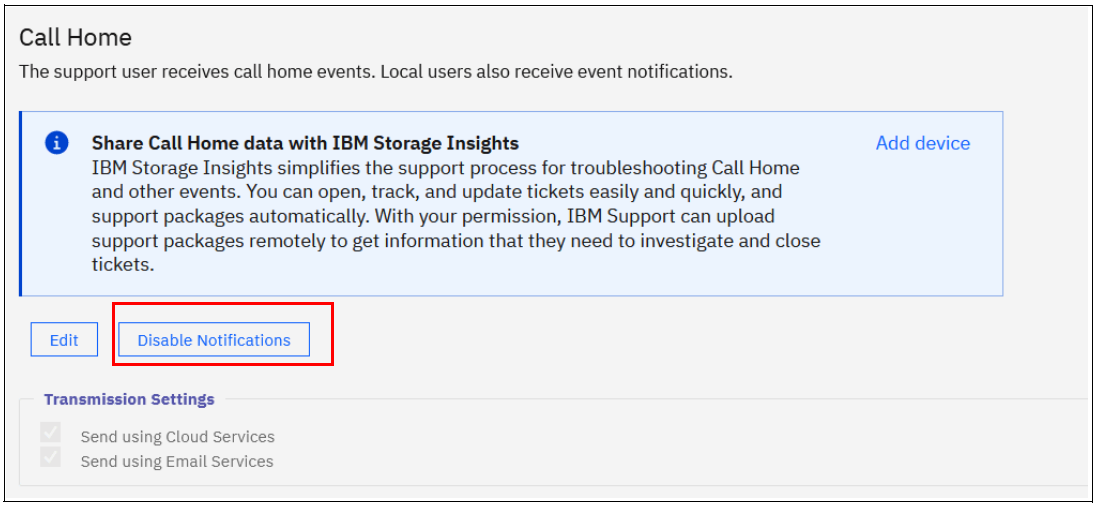


Figure 11-57 Disabling or enabling email notifications

11.8.2 Remote Support Assistance

For IBM Storage Virtualize versions 8.1 and later, Remote Support Assistance enables IBM Support to remotely connect to an IBM FlashSystem system through a secure tunnel to perform analysis, log collection, and software updates. The tunnel can be enabled as needed by the client or as a permanent connection.

Note: Customers who purchased Enterprise Class Support (ECS) server and IBM Storage Expert Care are entitled to IBM Support by using Remote Support Assistance to quickly connect and diagnose problems. However, IBM Support might choose to use this feature on non-ECS systems at their discretion. Therefore, configure and test the connection on all systems.

If you are enabling Remote Support Assistance, ensure that the following prerequisites are met:

- Call Home Connect Cloud or a valid email server are configured (Call Home Connect Cloud is used as the primary method to transfer the token when you initiate a session, with email as backup).
- A valid service IP address is configured on each node in the system.
- If your IBM FlashSystem system is behind a firewall or if you want to route traffic from multiple storage systems to the same place, you must configure a Remote Support Proxy server. Before you configure Remote Support Assistance, the proxy server must be installed and configured separately. During the setup for Support Assistance, specify the IP address and the port number for the proxy server on the Remote Support Centers window.
- If you do not have firewall restrictions and the nodes are directly connected to the internet, request your network administrator to allow connections to `170.225.126.11`, `170.225.126.12`, `170.225.127.11` and `170.225.127.12` on Port 22 from each and all service IPs of the node canisters.
- Uploading support packages and downloading software have direct connections to the internet. A DNS must be defined on your system for both of these functions to work.

- IBM Call Home IP Addresses changed after 1 March 2024. To ensure that support packages are uploaded correctly, configure the firewall to allow connections to the following IP addresses on port 443: `129.42.21.70`, `129.42.18.70` and `129.42.19.70`
- To ensure that software is downloaded correctly, configure the firewall to allow connections to the following IP addresses on port 22: `170.225.15.105`, `170.225.15.104`, `170.225.15.107`, `129.35.224.105`, `129.35.224.104`, and `129.35.224.107`.

See the following URL For more detailed information of call home servers at IBM, see IBM Call Home IP Addresses Are Changing on 1 March 2024 (formerly 30 May & 19 September 2023) - Action Required To Ensure Continuing Functionality

For more information about the **`downloadsoftware`** command to download selected code bundles from a Fix Central server, see downloadsoftware.

Figure 11-58 shows how to find Setup Remote Support Assistance if you closed the window.

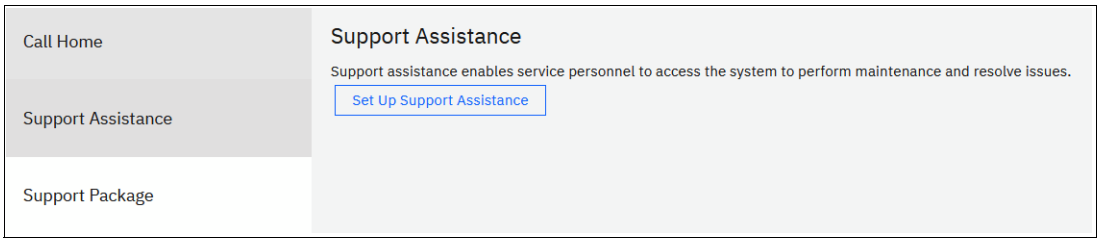


Figure 11-58 Support Assistance menu

Choosing to set up Support Assistance opens a wizard to guide you through the following configuration process:

1. Figure 11-59 on page 1017 shows the first wizard window. To keep remote assistance disabled, select **I want support personnel to work on-site only**. To enable remote assistance, select **I want support personnel to access my system both on-site and remotely**. Click **Next**.

   Note: Though you select **I want support personnel to work on-site only,** most maintenance contracts are for customer-replaceable unit (CRU) support, where IBM diagnoses your problem and sends a replacement component for you to install, if required.

   If you prefer to have IBM perform replacement tasks for you, contact your local sales representative to discuss an upgrade to your current maintenance contract.

2. Figure 11-60 on page 1017 shows the IBM Support Center IP addresses and Secure Shell (SSH) port that must be open in your firewall. You can also define a Remote Support Assistance Proxy if you have multiple systems in the data center, which allows for a firewall configuration being required only for the proxy server rather than every storage system. In this example, we do not have a proxy server and leave the field blank. Click **Next**.

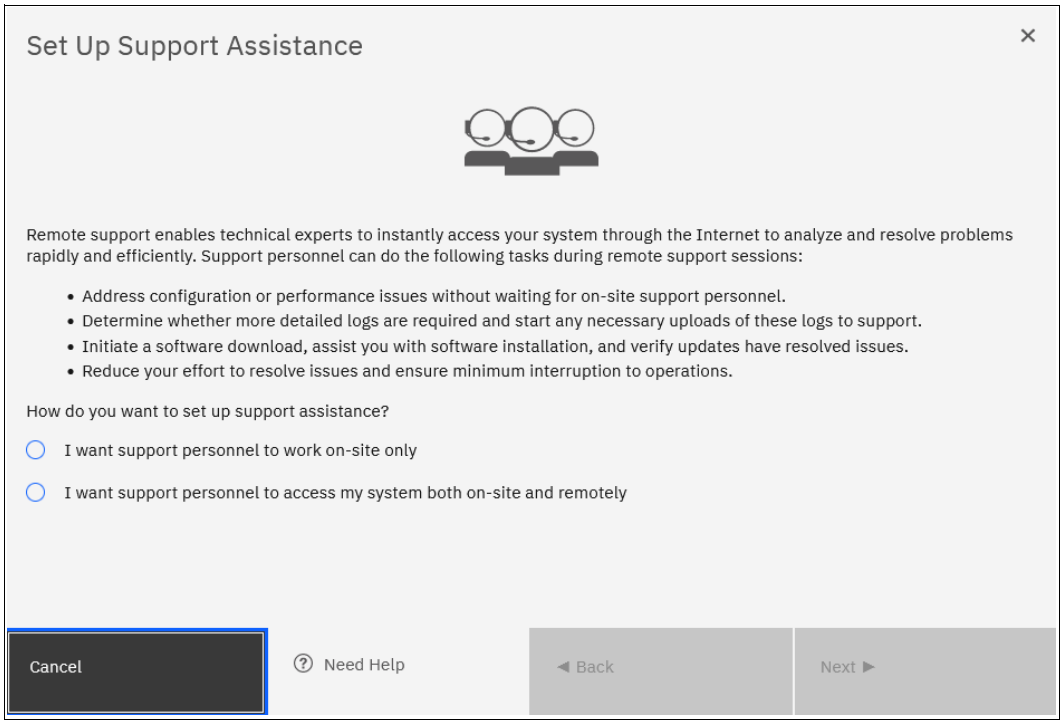


Figure 11-59 Enabling or disabling the support wizard

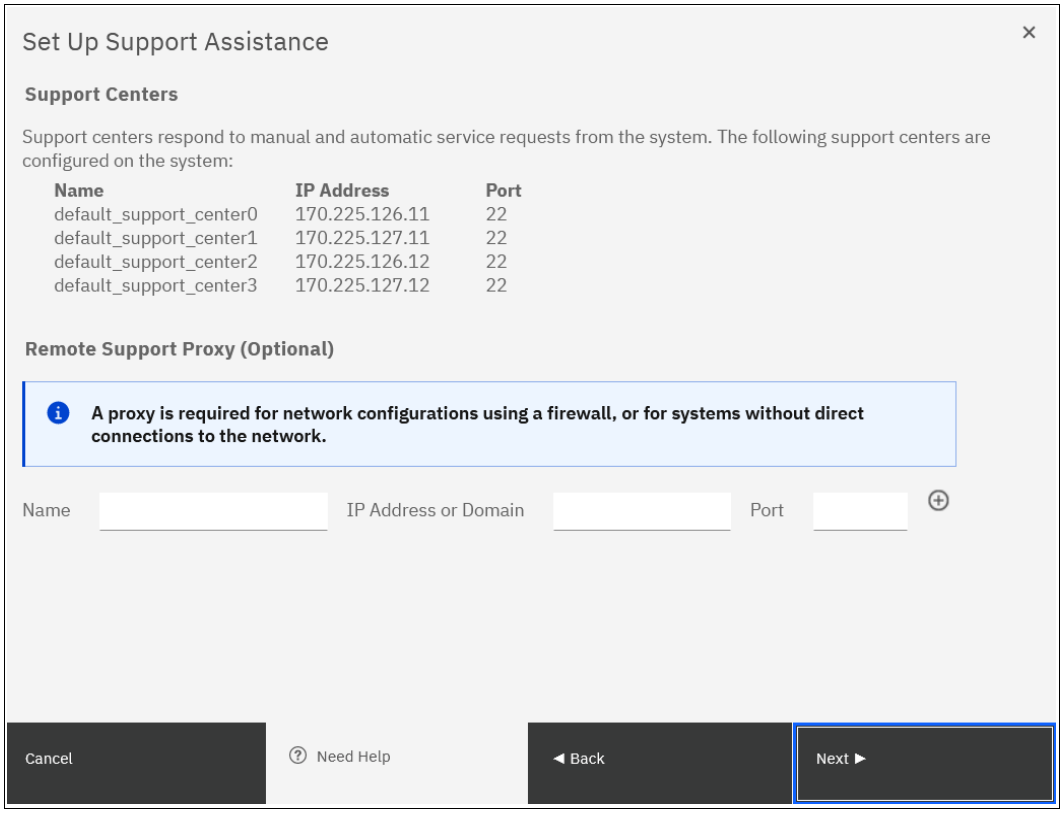


Figure 11-60 Support wizard proxy setup

3. The next window prompts you about whether you want to open a tunnel to IBM permanently, which allows IBM to connect to your system **At Any Time**, or **On**

Permission Only, as shown in Figure 11-61. **On Permission Only** requires a storage administrator to log on to the GUI and enable the tunnel when required. Click **Finish**.

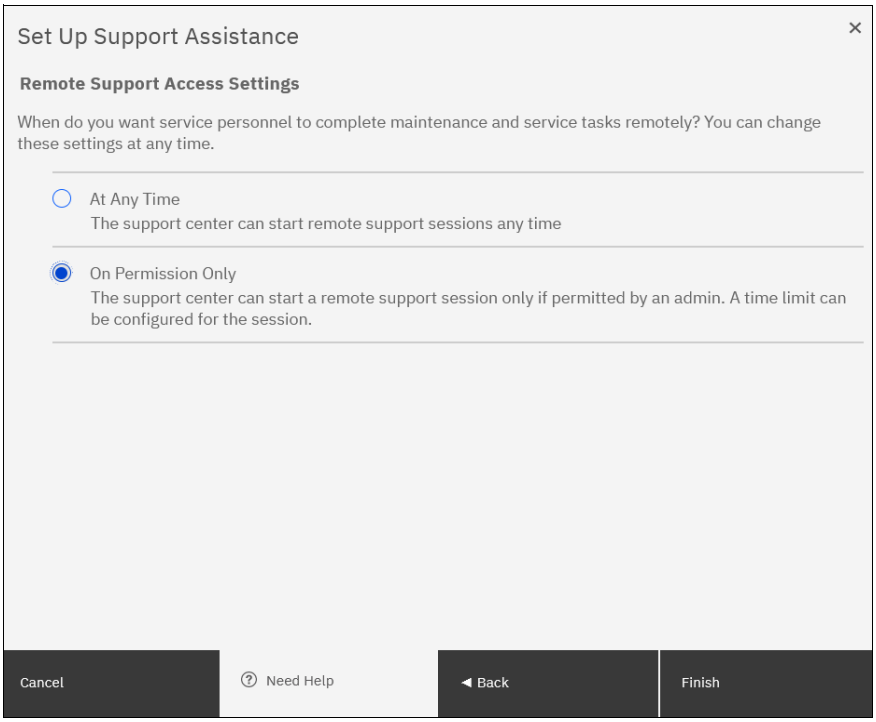


Figure 11-61 Support wizard access choice

4. After completing the remote support setup, you can view the status of any remote connection, start a session, test the connection to IBM, and reconfigure the setup. As shown in Figure 11-62, we successfully tested the connection. Click **Start New Session** to open a tunnel through which IBM Support can connect.

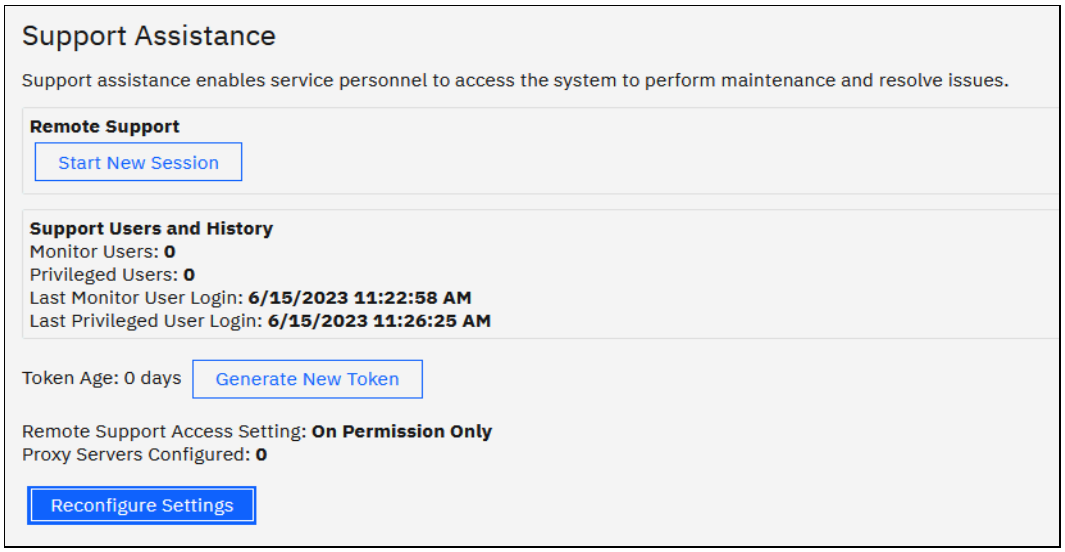


Figure 11-62 Support status and session management

5. A window prompts you for how long you want the tunnel to remain open if no activity occurs by setting a timeout value.

11.8.3 SNMP configuration

SNMP is a standard protocol for managing networks and exchanging messages. The system can send SNMP messages that notify personnel about an event. You can use an SNMP manager to view the SNMP messages that are sent by the system.

You can configure an SNMP server to receive various informational, error, or warning notifications by entering the following information (see Figure 11-63):

- IP Address

  The address for the SNMP server.

- Server Port

  The remote port (RPORT) number for the SNMP server. The RPORT number must be a value of 1 - 65535, where the default is port 162 for SNMP.

- Community

  The SNMP community is the name of the group to which devices and management stations that run SNMP belong. Typically, the default of `public` is used.

- Event Notifications:

  Consider the following points about event notifications:

  – Select **Error** if you want the user to receive messages about problems, such as hardware failures that require prompt action.

  > **Important:** Browse to **Recommended Actions** to run the fix procedures on these notifications.

  – Select **Warning** if you want the user to receive messages about problems and unexpected conditions. Investigate the cause immediately to determine any corrective action such as a space efficient volume running out of space.

  > **Important:** Go to **Recommended Actions** to run the fix procedures on these notifications.

  – Select **Info** if you want the user to receive messages about expected events. No action is required for these events.

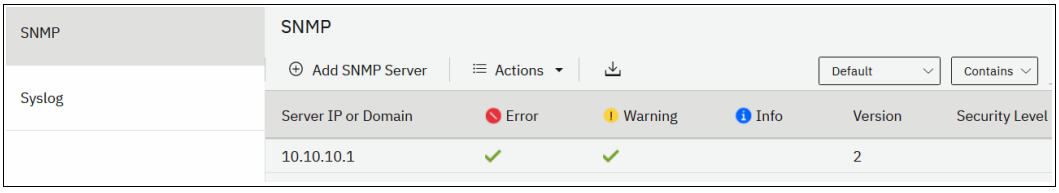


Figure 11-63 SNMP configuration

To add an SNMP server, select **Actions** → **Add** and complete the Add SNMP Server window, as shown in Figure 11-64. To remove an SNMP server, click the line with the server that you want to remove, and select **Actions** → **Remove**.

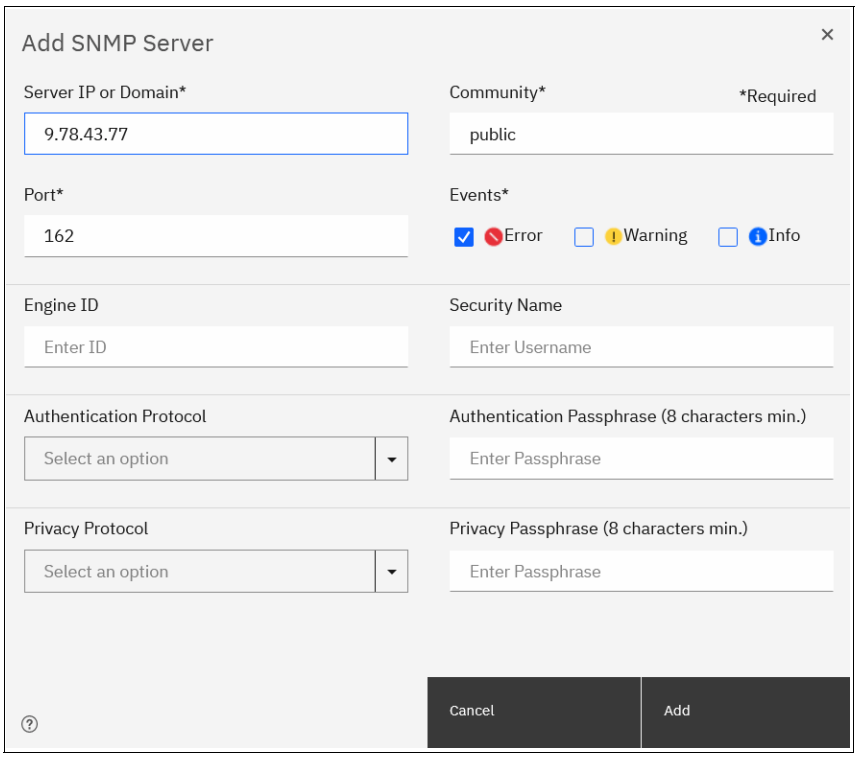


Figure 11-64 Add SNMP Server

Note: The following properties are either required or optional depending of the SNMP version:

- Engine ID

  Indicates the unique identifier (UID) in hexadecimal that identifies the SNMP server.
- Security Name

  Indicates which security controls are configured for the SNMP server. Supported security controls are none, authentication, or authentication and privacy.
- Authentication Protocol

  Indicates the authentication protocol that is used to verify the system to the SNMP server.
- Privacy Protocol

  Indicates the encryption protocol that is used to encrypt data between the system and the SNMP server.
- Privacy Passphrase

  Indicates the user-defined passphrase that is used to verify encryption between the system and SNMP server.

11.8.4 Syslog notifications

The syslog protocol is a standard protocol for forwarding log messages from a sender to a receiver on an IP network. The IP network can be IPv4 or IPv6. The system can send syslog messages that notify personnel about an event.

You can configure a syslog server to receive log messages from various systems and store them in a central repository by selecting **Settings** → **Notifications** → **Syslog**, as shown in Figure 11-65.

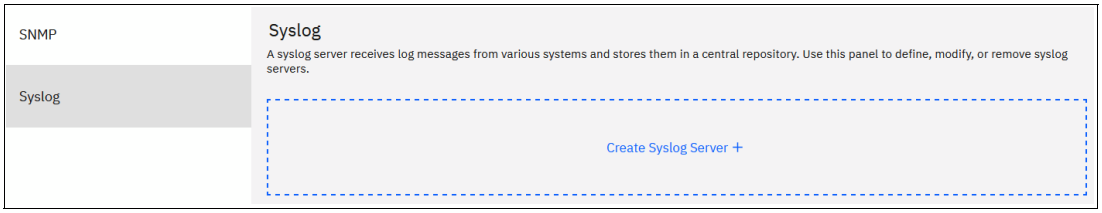


Figure 11-65 Syslog Servers menu

Enter the following information, as shown in Figure 11-66:

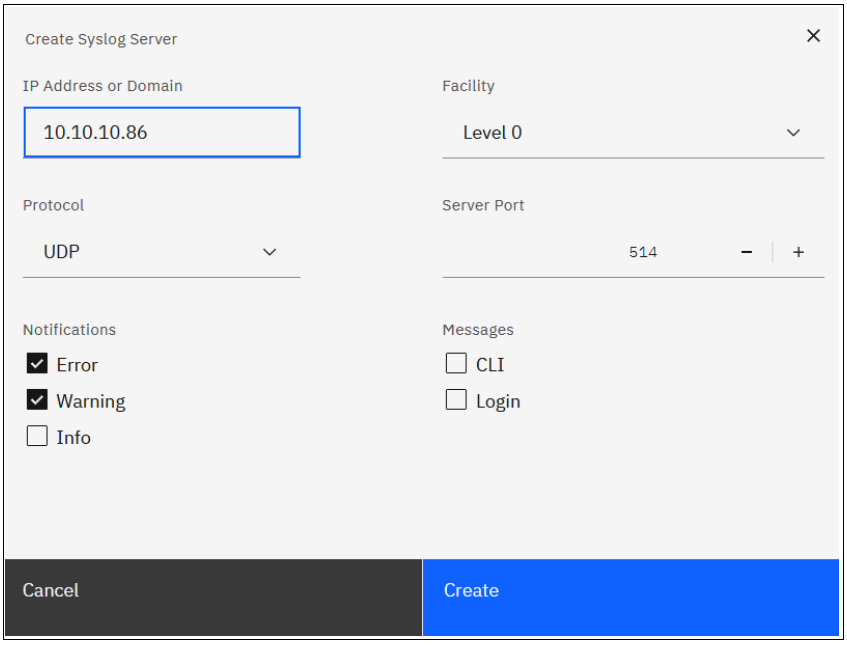


Figure 11-66 Syslog configuration

- IP Address

  The IP address for the syslog server.

- Facility

  The facility determines the format for the syslog messages. The facility can be used to determine the source of the message.

- Protocol

  The protocol to be used (UDP or TCP).

- Server Port

  The port to communicate with the syslog server.

- Notifications

  Choose one of the following items for event notifications:

  - Select **Error** if you want the user to receive messages about problems, such as hardware failures that must be resolved immediately.

    Important: Go to **Recommended Actions** to run the fix procedures on these notifications.

  - Select **Warning** if you want the user to receive messages about problems and unexpected conditions. Investigate the cause immediately to determine whether any corrective action is necessary.

    Important: Go to **Recommended Actions** to run the fix procedures on these notifications.

  - Select **Info** if you want the user to receive messages about expected events. No action is required for these events.

- Messages

  Choose one of the following items for messages:

  - **CLI**

    Select this option to include any CLI or management GUI operations on the specified syslog servers.

  - **Login**

    Select this option to send successful and failed authentication attempts to the specified syslog servers.

11.9 Audit log

The audit log is useful when analyzing past configuration events, especially when trying to determine, for example, how a volume ended up being shared by two hosts, or why the volume was overwritten. The audit log is also included in the `svc_snap` support data to aid in problem determination.

The audit log tracks action commands that are issued through an SSH session, management GUI, or Remote Support Assistance. It provides the following entries:

- Identity of the user who ran the action command.
- Name of the actionable command.
- Timestamp of when the actionable command ran on the configuration node.
- Parameters that ran with the actionable command.

The following items are not documented in the audit log:

- Commands that fail are not logged.
- A result code of 0 (success) or 1 (success in progress) is not logged.
- Result object ID of node type (for the **`addnode`** command) is not logged.
- Views are not logged.

Several specific service commands are not included in the audit log:

- `dumpconfig`
- `cpdumps`
- `cleardumps`
- `finderr`
- `dumperrlog`
- `dumpintervallog`
- `svcservicetak dumperrlog`
- `svcservicetask finderr`

Figure 11-67 shows the access to the audit log. Click **Audit Log** in the left menu to see which configuration CLI commands were run on the system.

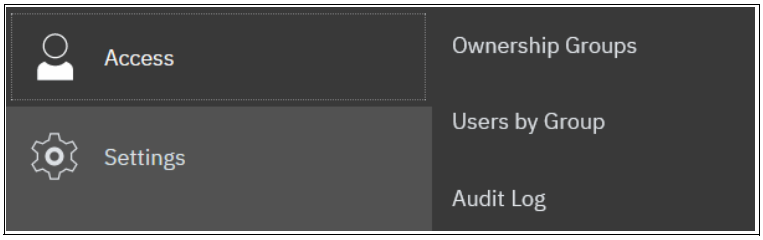


Figure 11-67 Audit Log from the Access menu

An example of the audit log after a volume is created and mapped to a host is shown in Figure 11-68.

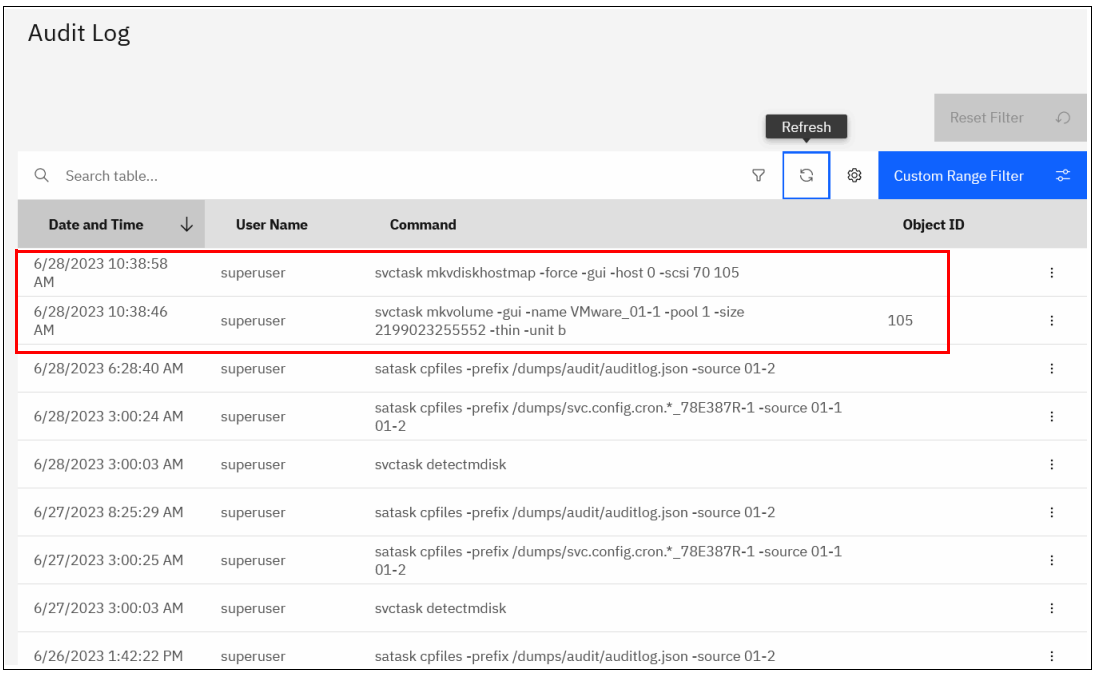


Figure 11-68 Audit log

Changing the view of the Audit Log grid is possible by clicking the Settings button in the column headings line (see Figure 11-69). The grid layout and sorting is under the user's control, so you can view everything in the audit log, sort different columns, and reset the default grid preferences.

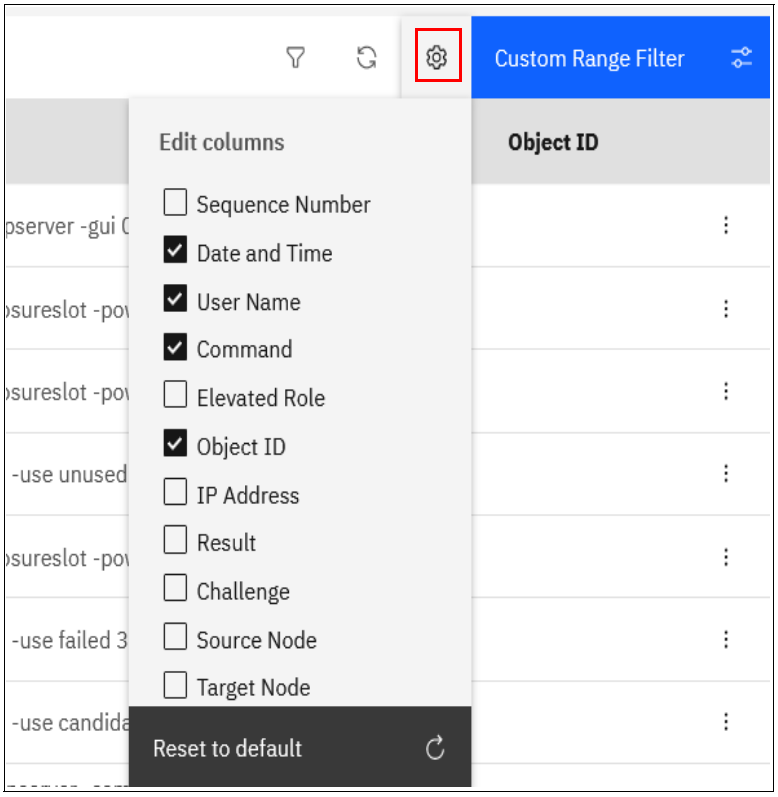


Figure 11-69 How to change audit log column headings

Also extensive filtering options are available either by clicking the **Filter** button, or clicking the **Custom Range Filter** menu.

11.10 Collecting support information by using the GUI, CLI, and USB

If you encounter a problem and contact the IBM Support Center, you will be asked to provide a support package. You can collect and upload this package by selecting **Settings** → **Support** menu.

11.10.1 Collecting information by using the GUI

You can use two methods to collect and upload the support package from the GUI of your Storage Virtualize system:

1. Upload Support Package. Use this feature if your system is connected to the internet to upload the Support Package directly from the storage system.
2. Download Support Package. Use this feature if your system is not connected to the internet to upload the Support Package manually by using IBM Support Open a Case. For more details, see Let's troubleshoot, or you can connect to EcuRep.

In the following example, the system is connected to the internet. To collect and upload the Support Package by using the GUI, complete the following steps:

1. Select **Settings** → **Support** and then the **Support Package** tab (Figure 11-70).

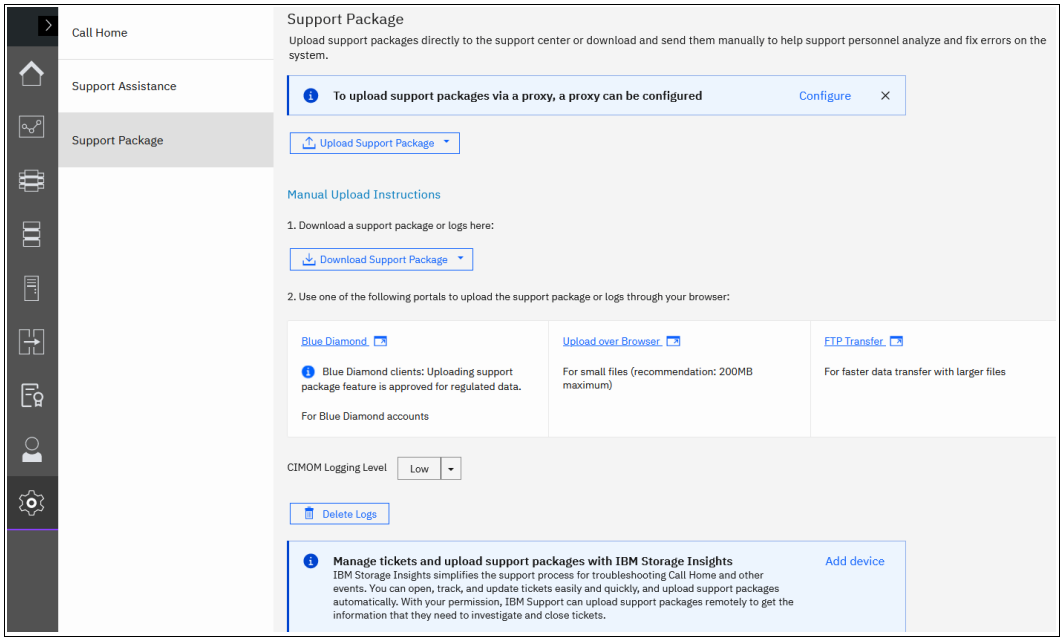


Figure 11-70 Support Package window

2. Click **Upload Support Package** and then **Create New Package and Upload**.

   Assuming that the problem that was encountered was an unexpected node restart that logged a 2030 error, collect the default logs and the most recent statesave from each node to capture the most relevant data for support.

   Note: When a node unexpectedly restarts, it first dumps its current statesave information before it restarts to recover from an error condition. This statesave is critical for IBM Support to analyze what occurred. Collecting a snap type 4 creates statesaves at the time of the collection, which is not useful for understanding the restart event.

3. The Upload Support Package window provides four options for data collection. If you are contacted by IBM Support because your system called home or you manually opened a call with IBM Support, you receive a case number in the format of TSxxxxxxxxx. Enter that number into the **Case Number** field and select the snap type (often referred to as an *option 1, 2, 3, 4 snap*) as requested by IBM Support (see Figure 11-71 on page 1026). In our example, we entered our case number, selected **snap type 3 (option 3)** because this option automatically collects the statesaves that were created at the time that the node restarted, and clicked **Upload**.

   Tip: To open a service request online, see Let's troubleshoot.

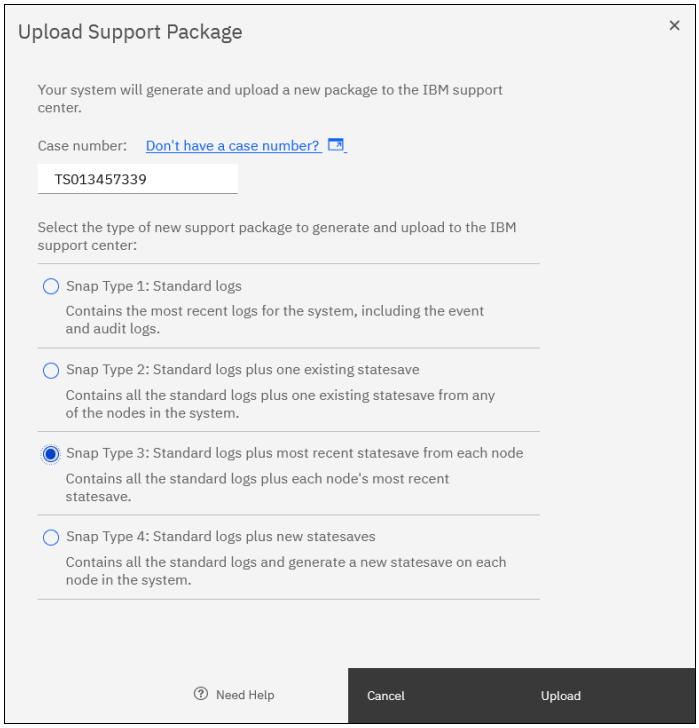


Figure 11-71 Upload Support Package window

4. The procedure to generate the snap on the system begins and includes the most recent statesave from each node canister. This process might take several minutes (see Figure 11-72).

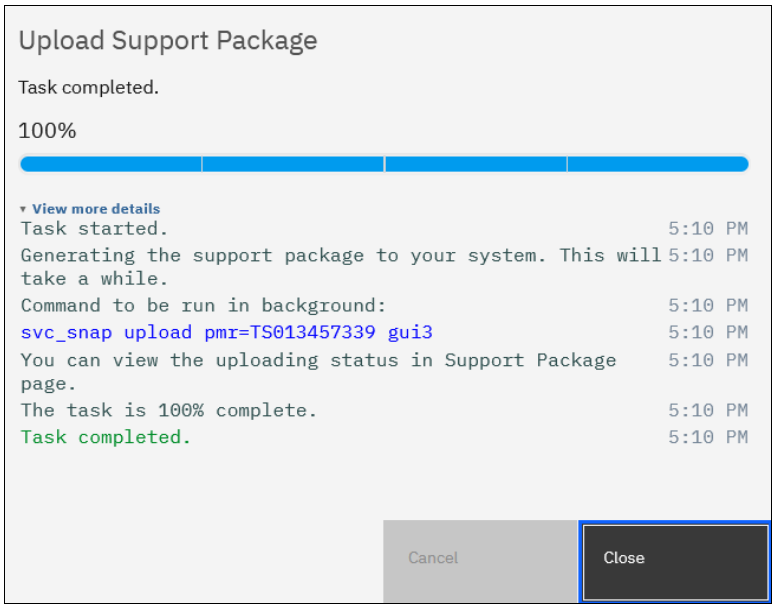


Figure 11-72 Task detail window

The support package is collected. The process can take several minutes depending on the type of snap that was requested. When collection has finished, the upload to IBM will start as seen in Figure 11-73 on page 1027.

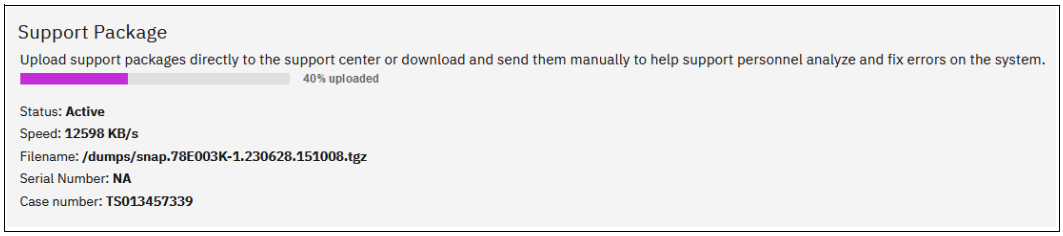


Figure 11-73 Support package upload to IBM

After the upload completes the GUI will show either failure to upload or success as shown in Figure 11-74.

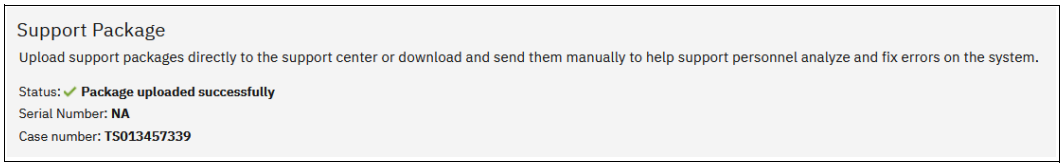


Figure 11-74 Support package upload to IBM - success

The time that it takes to generate the snap and the size of the file that is generated depends mainly on two things: the snap option that you selected and the size of your system. An option 1 snap takes much less time than an option 4 snap because the required data has already been collected. An option 4 snap requires the system to collect new statesaves from each node. In an 8-node cluster, this task can take quite some time, so you should always collect the snap option that IBM Support recommends.

Table 11-14 lists the approximate file sizes for each SNAP option.

Table 11-14 Types of snaps

Option	Description	Approximate size (one I/O group, 30 volumes)	Approximate size (four I/O groups, 250 volumes)
1	Standard logs	10 MB	340 MB
2	Standard logs plus one existing statesave	50 MB	520 MB
3	Standard logs plus the most recent statesave from each node	90 MB	790 MB
4	Standard logs plus new statesaves	90 MB	790 MB

11.10.2 Collecting logs by using the CLI

The CLI can be used to collect and upload a support package as requested by IBM Support by performing the following steps:

1. Log in to the CLI and run the **svc_snap** command that matches the type of snap that is requested by IBM Support:
   - Standard logs (type 1):

```
svc_snap upload pmr=TSxxxxxxxxx gui1
```

– Standard logs plus one existing statesave (type 2):

```
svc_snap upload pmr=TSxxxxxxxxx gui2
```

– Standard logs plus most recent statesave from each node (type 3):

```
svc_snap upload pmr=TSxxxxxxxxx gui3
```

– Standard logs plus new statesaves (type 4):

```
svc_livedump -nodes all -yes
svc_snap upload pmr=TSxxxxxxxxx gui3
```

The example shows the collection of a type 3 (option 3) package and uploads it automatically uploaded to the PMR number that is provided by IBM Support, as shown in Example 11-11.

Example 11-11 The svc_snap command

```
IBM_FlashSystem:FS7300:superuser>svc_snap upload pmr=TS013457339 gui3
Collecting data
Packaging files
Snap data collected in /dumps/snap.78E003K-1.230628.140925.tgz
IBM_FlashSystem:FS7300:superuser>
```

If you do *not* want to automatically upload the snap to IBM, do not specify the **upload pmr=TSxxxxxxxxx** part of the commands. When the snap creation completes, it creates a file name that uses the following format:

`/dumps/snap.<panel_id>.YYMMDD.hhmmss.tgz`

It takes a few minutes for the snap file to complete. It can take longer if statesaves are included.

The generated file can then be retrieved from the GUI by selecting **Settings** → **Support** → **Manual Upload Instructions** → **Download Support Package**, and then clicking **Download Existing Package**, as shown in Figure 11-75.

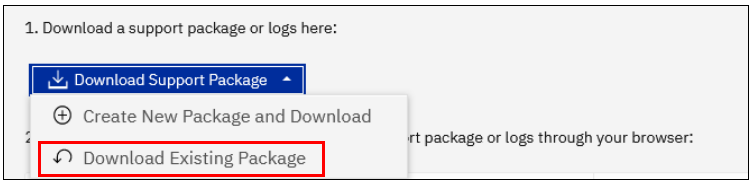


Figure 11-75 Downloaded Existing Package

2. Click in the **Filter** box and enter `snap` to see a list of snap files, as shown in Figure 11-76 on page 1029. Find the exact name of the snap that was generated by running the **svc_snap** command that was run earlier. Select that file, and click **Download**.

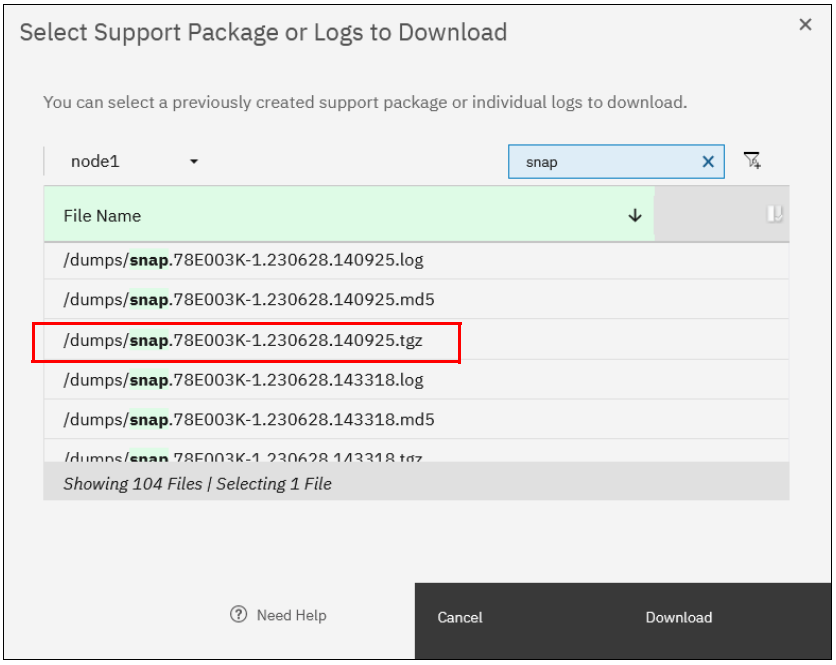


Figure 11-76 Filtering on snap to download

3. Save the file to a folder of your choice on your workstation.

11.10.3 Collecting logs by using a USB flash drive

As a backup, in case there is no connectivity to CLI and GUI (rare), it is possible to get a snap from a node from the USB ports on the rear.

Note: This procedure collects a single snap from the node canister, not a cluster snap. It is useful for determining the state of the node canister.

When a USB flash drive is plugged into a node canister, the canister code searches for a text file that is named `satask.txt` in the root directory. If the code finds the file, it attempts to run a command that is specified in the file. When the command completes, a file that is called `satask_result.html` is written to the root directory of the USB flash drive. If this file does not exist, it is created. If it exists, the data is inserted at the start of the file. The file contains the details and results of the command that was run and the status and the configuration information from the node canister. The status and configuration information matches the detail that is shown on the service assistant home page windows.

To collect a snap, complete the following steps:

1. Ensure that your USB drive is formatted with an FAT32 file system on its first partition.
2. Create a text (`.txt`) file on the USB flash drive in the root directory called `satask.txt` (case-sensitive).
3. In the `satask.txt` file, write **`satask snap`** and save the file.
4. Put the USB into one of the USB ports on the rear of the canister and wait for a short time. The fault LED on the node canister flashes when the USB service action is being completed. When the fault LED stops flashing, it is safe to remove the USB flash drive.

5. Unplug the USB drive from the system and plug it into your workstation. If the procedure was successful, the `satask.txt` file was deleted and you have a `satask_result.html` file and a single snap from the node canister. This snap can be uploaded to the IBM Support Center, as shown in 11.10.4, “Uploading files to the IBM Support Center” on page 1030.

> **Note:** If there was a problem with the procedure, the HTML file is still generated, and reasons why the procedure did not work are listed in it.

11.10.4 Uploading files to the IBM Support Center

If you chose to not have the system upload the support package automatically, it can still be uploaded for analysis from the Enhanced Customer Data Repository (ECuRep). Any uploads should be associated with a specific IBM Support ticket number. This is also known as a *service request* and is a mandatory requirement when uploading.

To upload the information, complete the following steps:

1. Using a web browser, go to Enhanced Customer Data Repository (ECuRep) (see Figure 11-77).

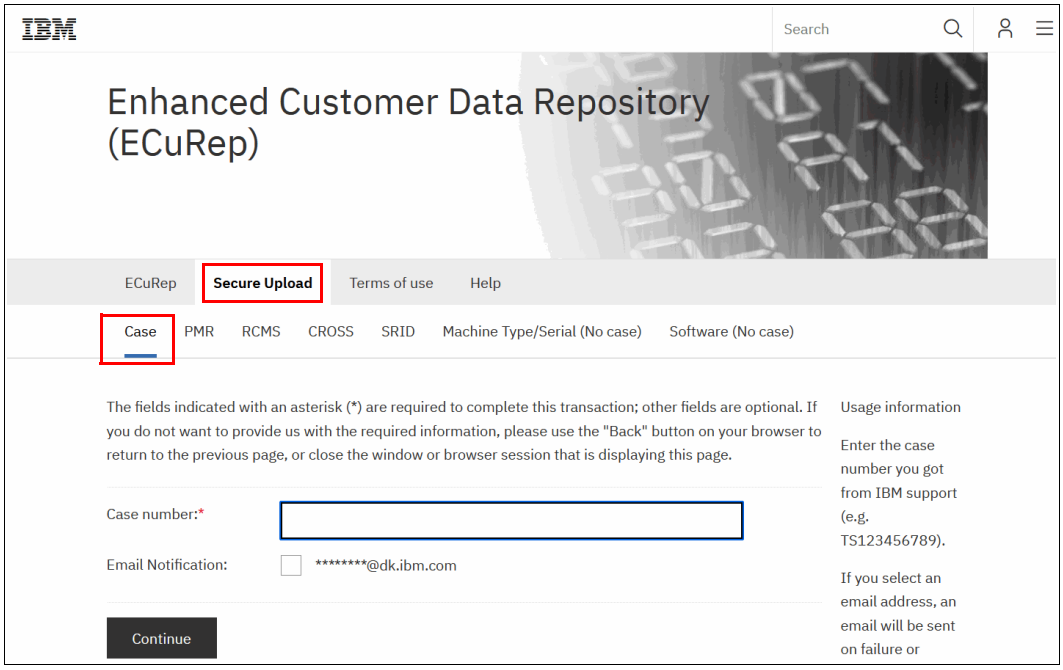


Figure 11-77 ECuRep details

2. Complete the required fields:
   - If not already on the correct page select **Secure Upload** and **Case**.
   - Fill in the **Case number** that is provided by IBM Support for your specific case. This number uses the format of `TSxxxxxxxxx; for example TS013457339.`
   - **Upload is for**: Select **Hardware** from the drop-down menu.

   Although the **Email address** field is not mandatory, it is a best practice to enter your email address so that you are automatically notified of a successful or unsuccessful upload.

3. When the form is completed, click **Continue** to open the input window (see Figure 11-78).

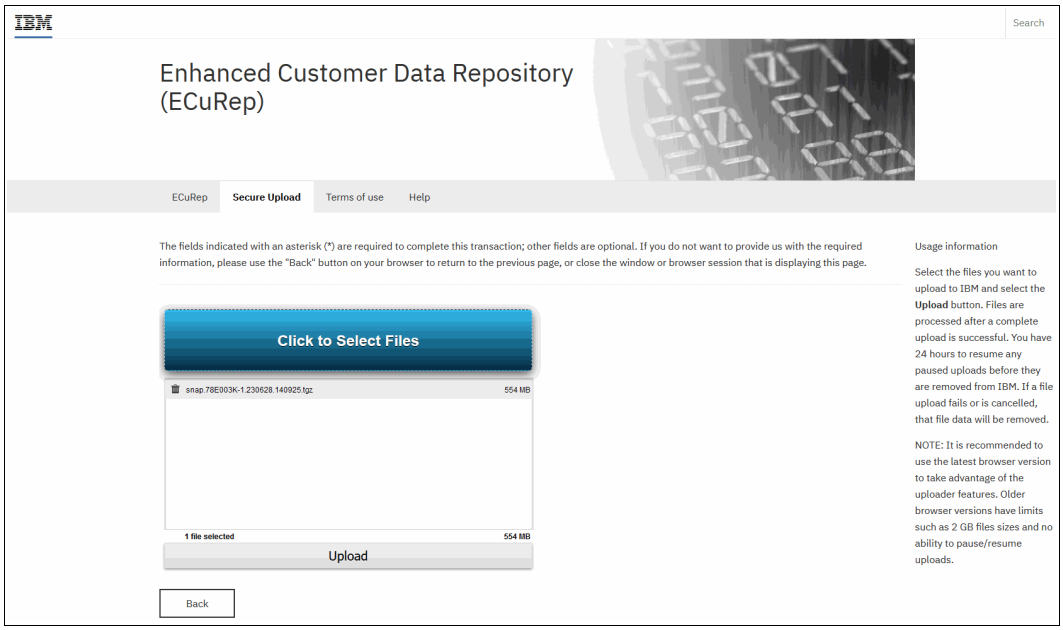


Figure 11-78 ECuRep File upload

4. Select one or more files, click **Upload** to continue, and follow the directions.

11.11 Service Assistant Tool

The Service Assistant Tool (SAT) is a web-based GUI that is used to service individual nodes, primarily when a node has a fault and is in a service state. A node is not an active part of a clustered system while it is in service state.

Typically, the system is configured with the following IP addresses:

- One service IP address for each of the nodes/node canisters.
- One cluster management IP address, which is set when the cluster is created.

The SAT is available even when the management GUI is not accessible. The following information and tasks can be accomplished with the SAT:

- Status information about the connections and the node canister
- Basic configuration information, such as configuring IP addresses
- Service tasks, such as restarting the Common Information Model Object Manager (CIMOM) and updating the WWNN
- Details about node error codes
- Details about the hardware, such as IP addresses and Media Access Control (MAC) addresses

The SAT GUI is available by using a service assistant IP address that is configured on each node/node canister. It can also be accessed through the cluster IP addresses by appending `/service` to the cluster management IP.

It is also possible to access the SAT GUI of the config node if you enter the Uniform Resource Locator (URL) of the service IP address of the config node into any web browser and click **Service Assistant Tool** (see Figure 11-79).

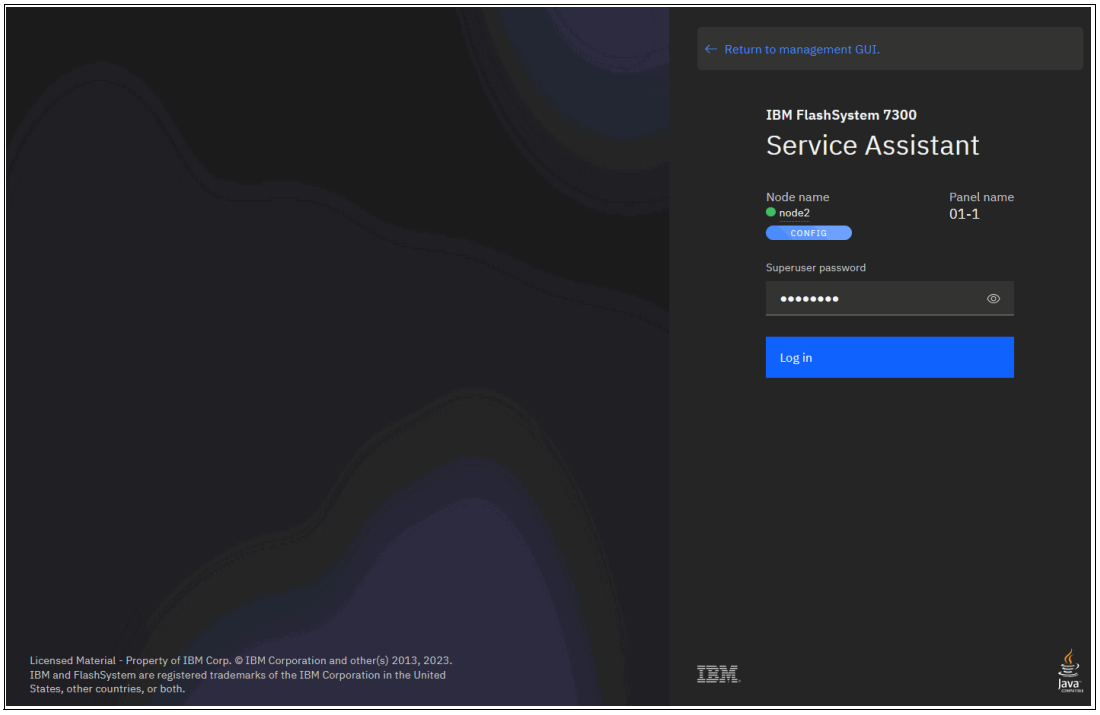


Figure 11-79 Service Assistant Tool login

If the clustered system is down, the only method of communicating with the node canisters is through the SAT IP address directly. Each node can have a single service IP address on Ethernet port 1, which should be configured on all nodes of the cluster.

To open the SAT GUI, enter one of the following URLs into a web browser:

- Enter `http(s)://<cluster IP address of your cluster>/service`.
- Enter `http(s)://<service IP address of a node>/service`.
- Enter `http(s)://<service IP address of config node>` and click **Service Assistant Tool**.

To access the SAT, complete the following steps:

1. If you are accessing SAT by using `cluster IP address/service`, the configuration node canister SAT GUI login window opens. Enter the Superuser Password, as shown in Figure 11-80 on page 1033.

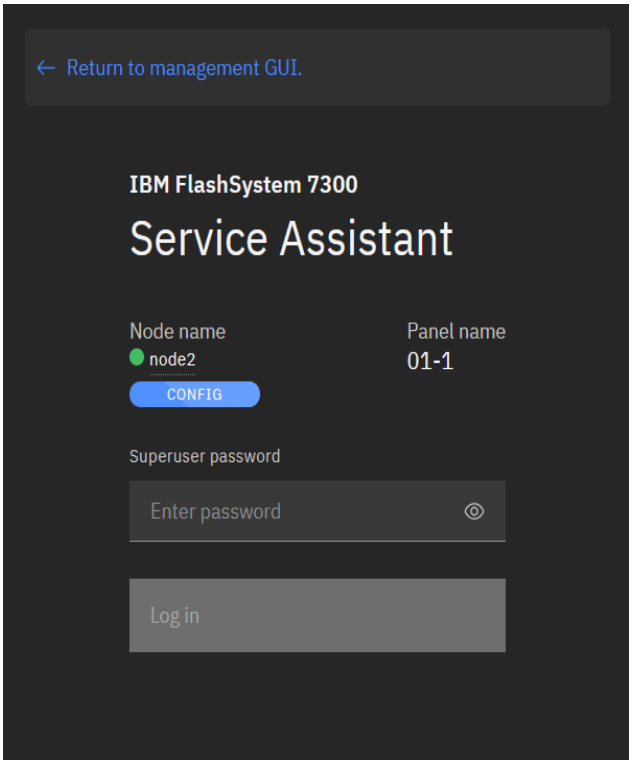


Figure 11-80 Service Assistant Tool Login GUI

2. After you are logged in, you see the Service Assistant Home window, as shown in Figure 11-81. The SAT can view the status and run service actions on other nodes in addition to the node to which the user is logged in.

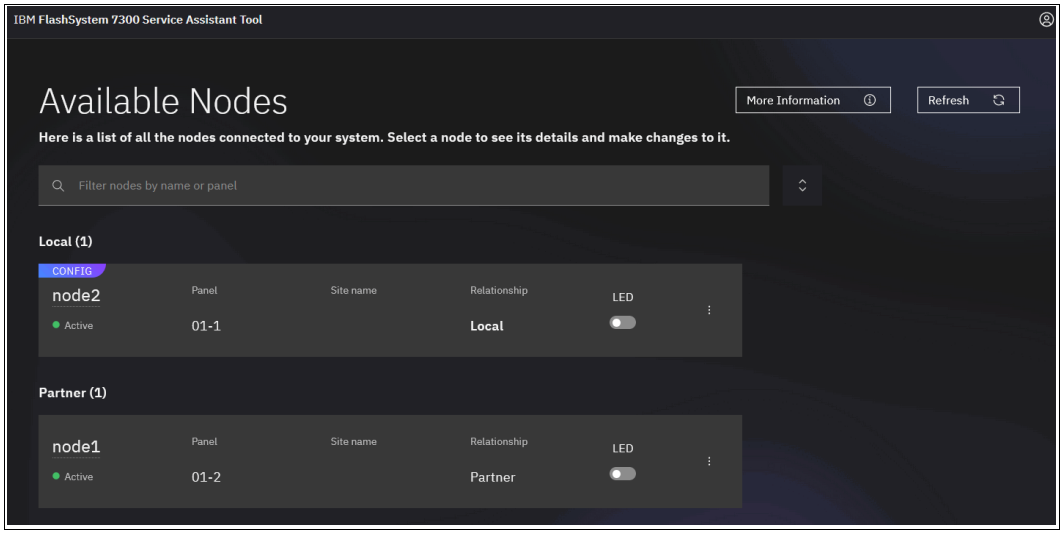


Figure 11-81 Service Assistant Tool GUI

3. In the Service Assistant Home window, you have an overview of which nodes are in the cluster, and which may be available for adding. Also, you might see nodes that are in service state or in error. In this window you can select a node and perform node actions such as Power off, Reboot, Restart and Enter or Exit service state as shown in Figure 11-82 on page 1034.

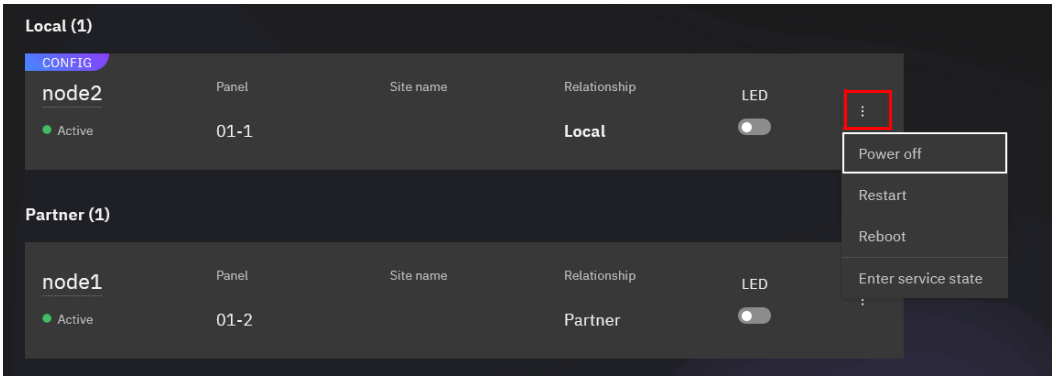


Figure 11-82 SAT home window options

4. From SAT home, you can also click one of the nodes to get additional details of the node as shown in Figure 11-83.

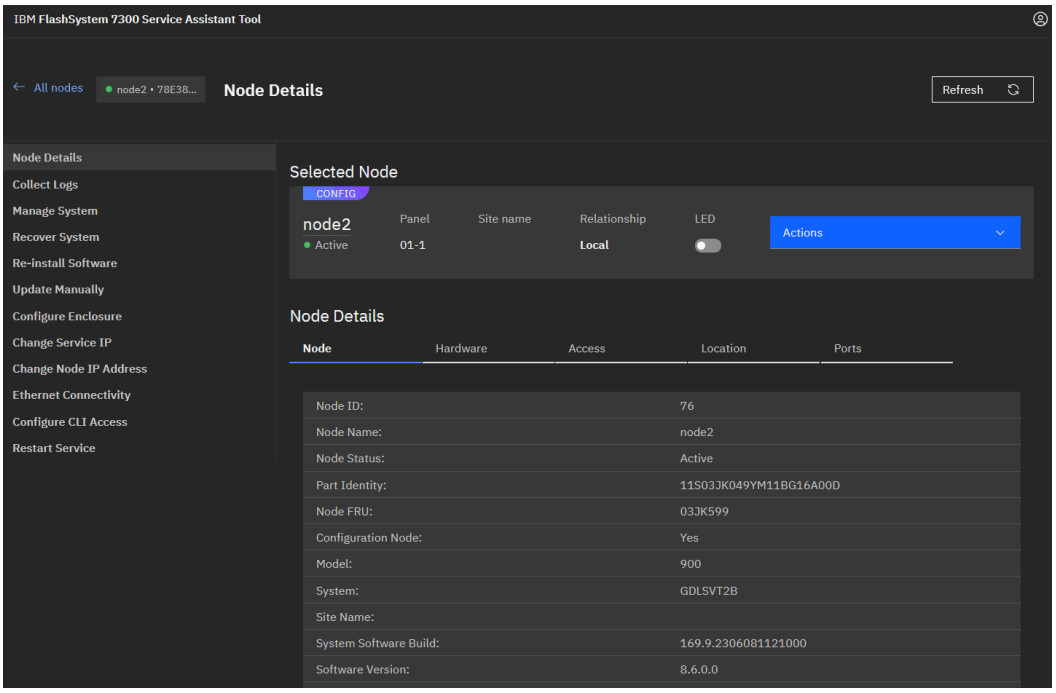


Figure 11-83 SAT node details

When you view the Node Details window you have many options for node management and log collection. As in the SAT home window, nodes can be powered off, rebooted, restarted and can Enter or Exit service state by clicking the **Actions** button. The node can also be removed completely or it can be updated manually to any code version needed. Never operate the SAT tool without guidance from IBM support.

Note: The SAT GUI provides access to service procedures and shows the status of the node canisters. These procedures should be carried out only if you are directed to do so by IBM Support.

For more information about how to use the SAT, see Accessing the service assistant.

11.12 IBM Storage Insights monitoring

With IBM Storage Insights, you can optimize your storage infrastructure by using this cloud-based storage management and support platform with predictive analytics.

The monitoring capabilities that IBM Storage Insights provides are useful for things like capacity planning, workload optimization, and managing support tickets for ongoing issues.

For a live demo of IBM Storage Insights, see Storage Insights Demo (requires login).

Demonstration videos: To view videos about Storage Insights, see What's New in IBM Storage Insights. The videos include new features and enhancements of IBM Storage Insights.

After you add your systems to IBM Storage Insights, you see the Dashboard, where you can select a system that you want to see the overview for, as shown in Figure 11-84.

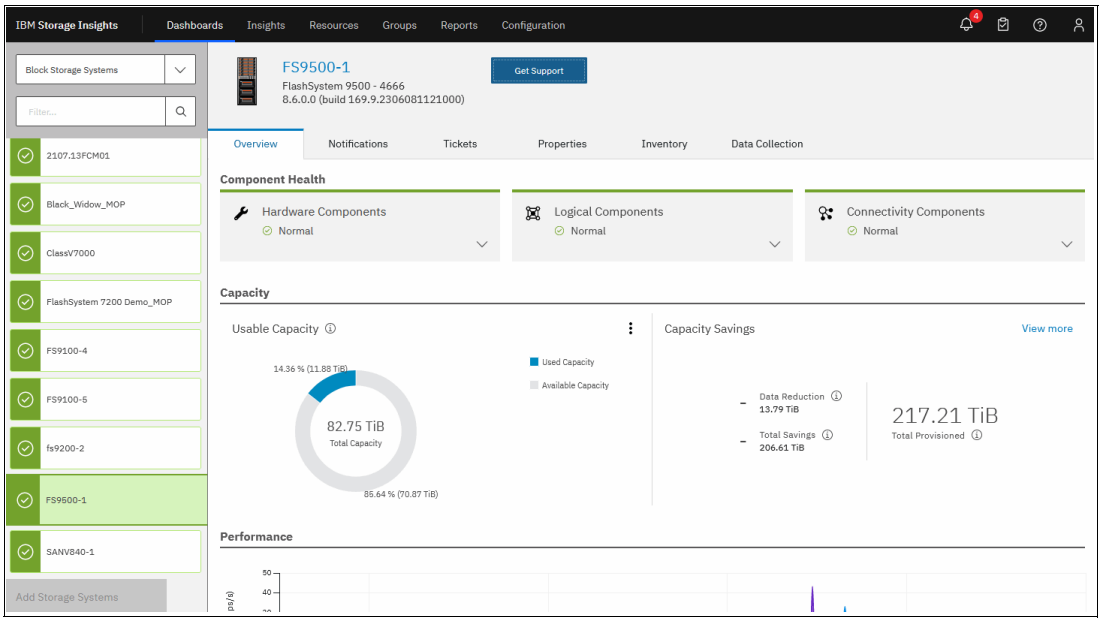


Figure 11-84 IBM Storage Insights System overview

Component health is shown at the upper center of the window. If there is a problem with one of the Hardware, Logical or Connectivity components, errors are shown here, as shown in Figure 11-85.

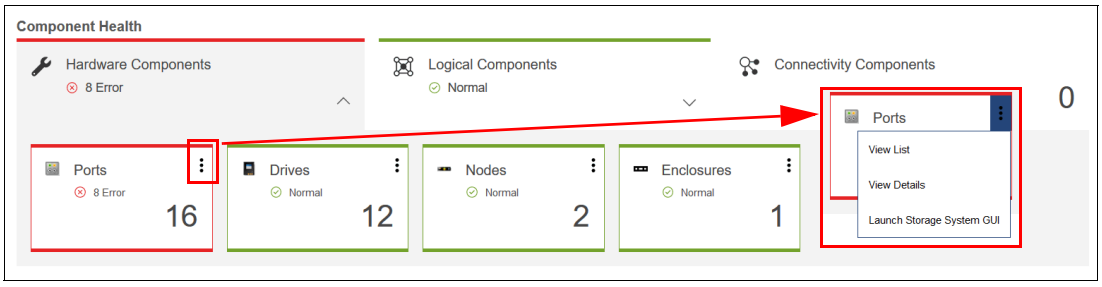


Figure 11-85 Component Health overview

The error entries can be expanded to obtain more details by selecting the three dots at the upper right corner of the component that has an error and then selecting **View Details**. The relevant part of the more detailed System View opens, and what you see depends on which component has the error, as shown in Figure 11-86.

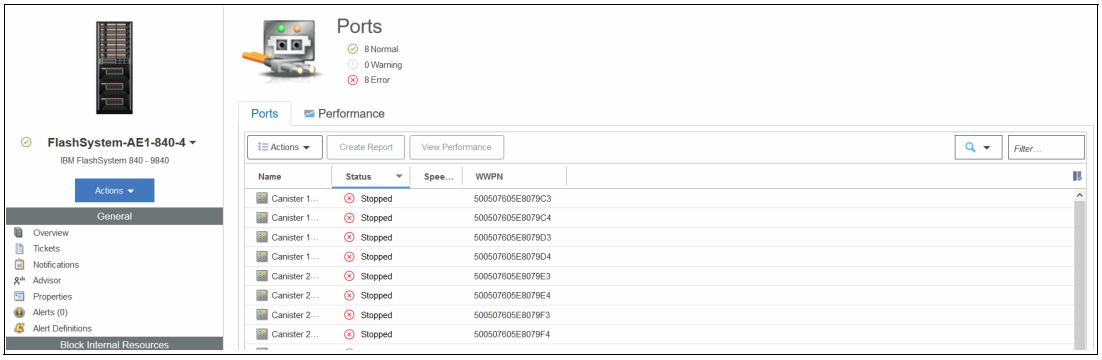


Figure 11-86 Ports in error

From here, it is now obvious which components have the problem and exactly what is wrong with them, so now you can log a support ticket with IBM if necessary.

11.12.1 Capacity monitoring

You can see key statistics such as Usable and Provisioned Capacity and Capacity Savings as shown in Figure 11-87. Capacity can be viewed by volume or pool and the **View More** button shows a trend curve.

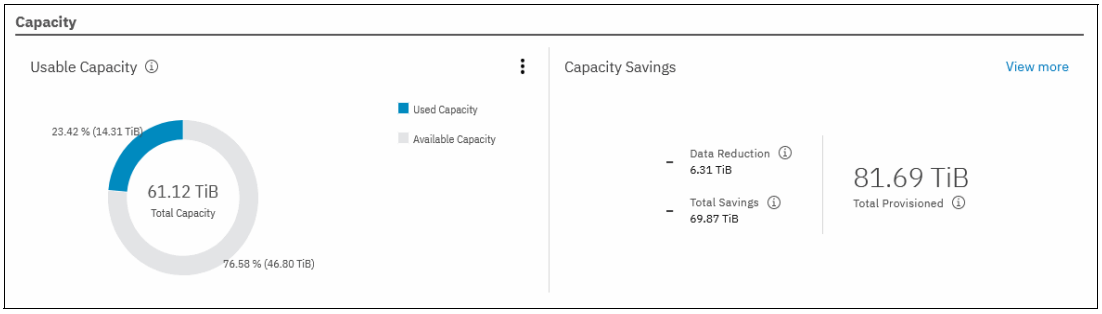


Figure 11-87 Capacity area of the IBM Storage Insights system overview

In the Capacity view, the user can click **View Pools**, **View Compress Volumes**, **View Deduplicated Volumes**, and **View Thin-Provisioned Volumes**. Clicking any of these items takes the user to the detailed system view for the selection option. From there, you can click **Capacity** to get a historical view of how the system capacity changed over time, as shown in Figure 11-88 on page 1037. At any time, the user can select the timescale, resources, and metrics to be displayed on the graph by clicking any options around the graph.

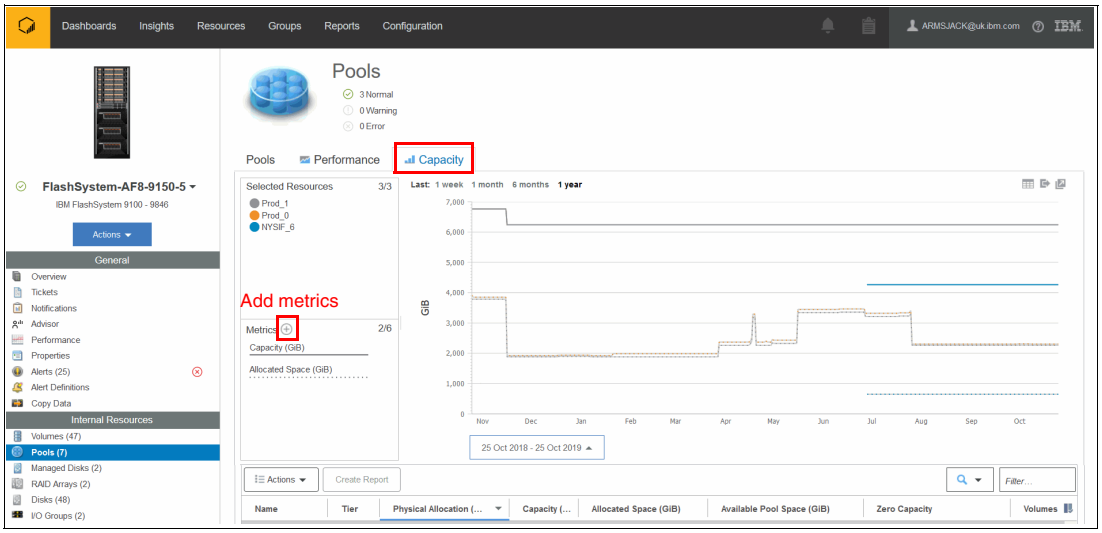
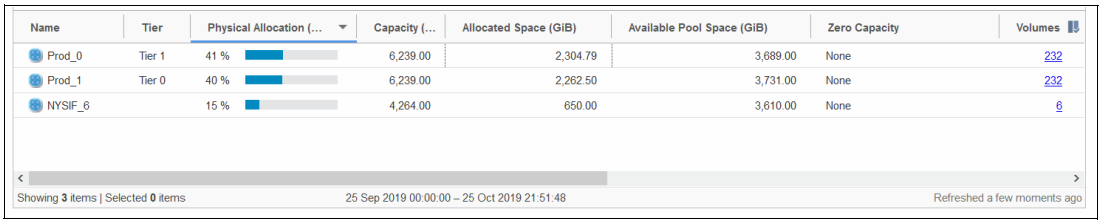


Figure 11-88 IBM Storage Insights Capacity view

If you scroll down below the graph, you find a list view of the selected option. In this example, we selected **View Pools** so the configured pools are shown with the relevant key capacity metrics, as shown in Figure 11-89. Double-clicking a pool in the table display the properties for it.



Name	Tier	Physical Allocation (...	Capacity (...	Allocated Space (GiB)	Available Pool Space (GiB)	Zero Capacity	Volumes
Prod_0	Tier 1	41 %	6,239.00	2,304.79	3,689.00	None	232
Prod_1	Tier 0	40 %	6,239.00	2,262.50	3,731.00	None	232
NYSIF_6		15 %	4,264.00	650.00	3,610.00	None	6

Showing 3 items | Selected 0 items
25 Sep 2019 00:00:00 – 25 Oct 2019 21:51:48
Refreshed a few moments ago

Figure 11-89 Pools list view

11.12.2 Performance monitoring

From the system overview, you can scroll down and see the three key performance statistics for your system, as shown in Figure 11-90. For the Performance overview, these statistics are aggregated across the whole system, and you cannot drill down by Pool, Volume, or other items.

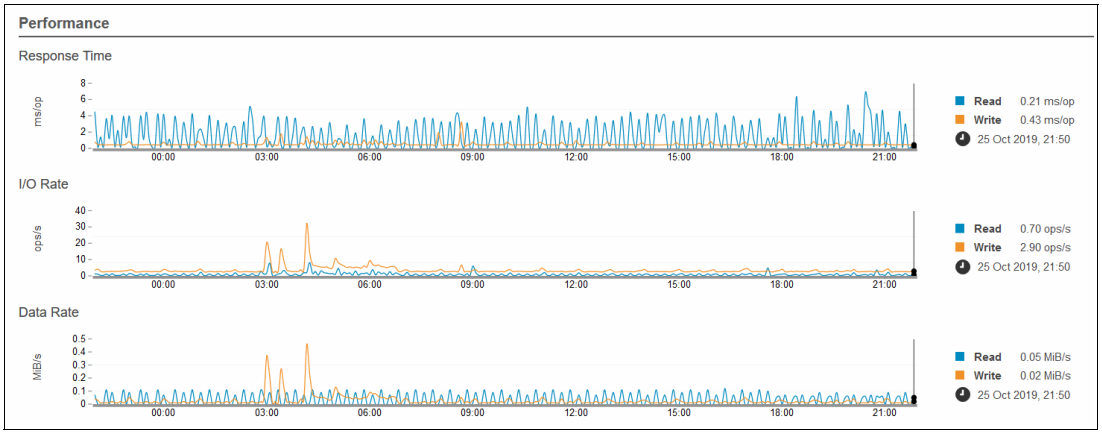


Figure 11-90 System overview: Performance

To view more detailed performance statistics, enter the system view again, as described in 11.12.1, “Capacity monitoring” on page 1036.

For this performance example, we select **View Pools**, and then select **Performance** from the System View pane, as shown in Figure 11-91.

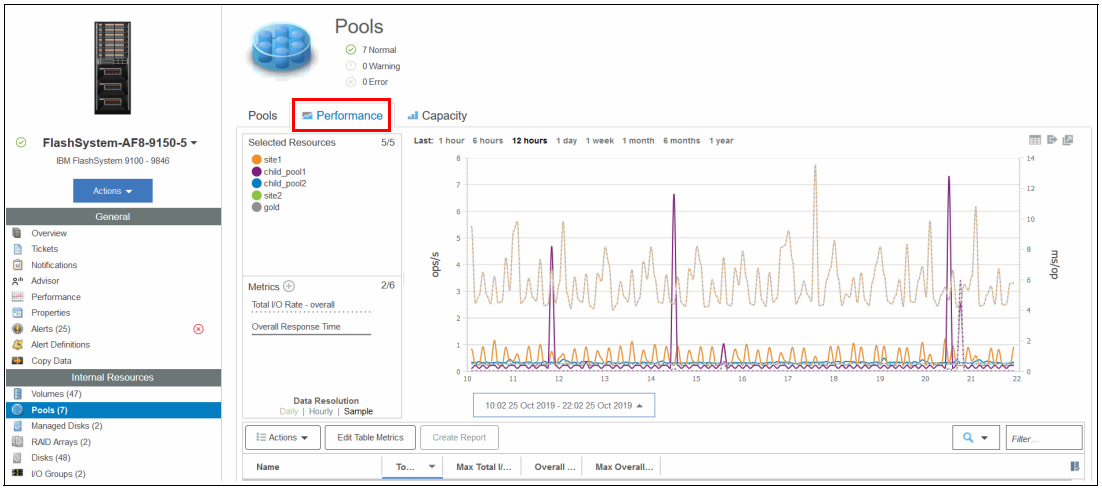


Figure 11-91 IBM Storage Insights: Performance view

It is possible to customize what can be seen on the graph by selecting the metrics and resources. In Figure 11-92, the Overall Response Time for one pool over a 12-hour period is displayed.

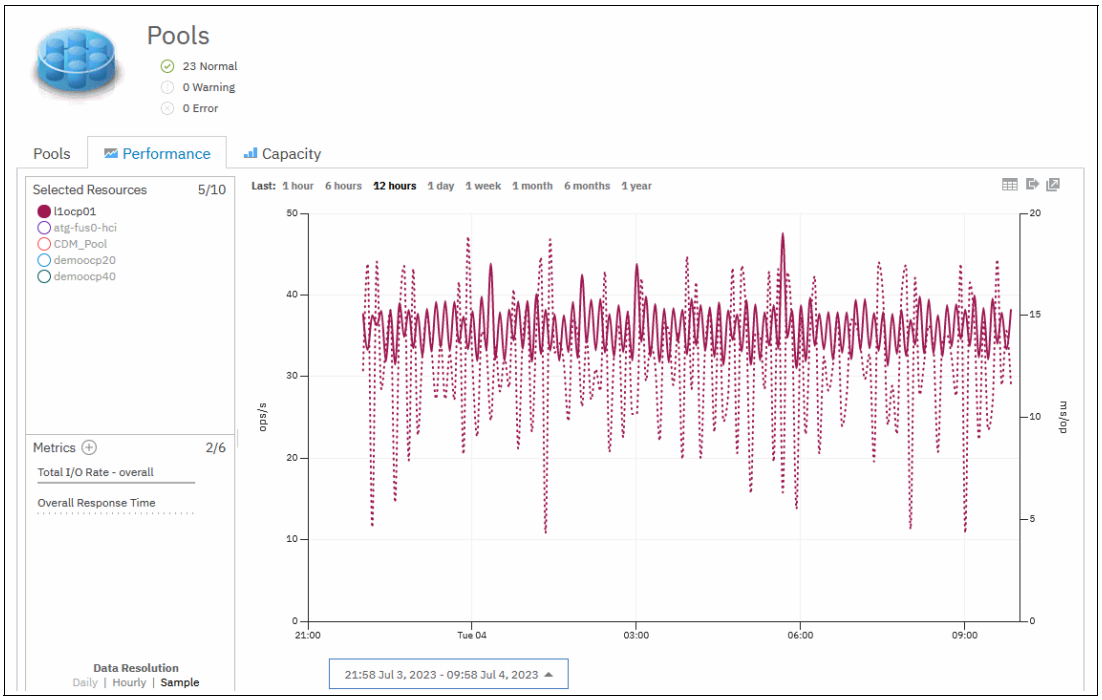


Figure 11-92 Filtered performance graph

Scrolling down the graph, the Performance List view is visible, as shown in Figure 11-93 on page 1039. Metrics can be selected by clicking the filter button at the right of the column headers. If you select a row, the graph is filtered for that selection only. Multiple rows can be selected by holding down the Shift or Ctrl keys.

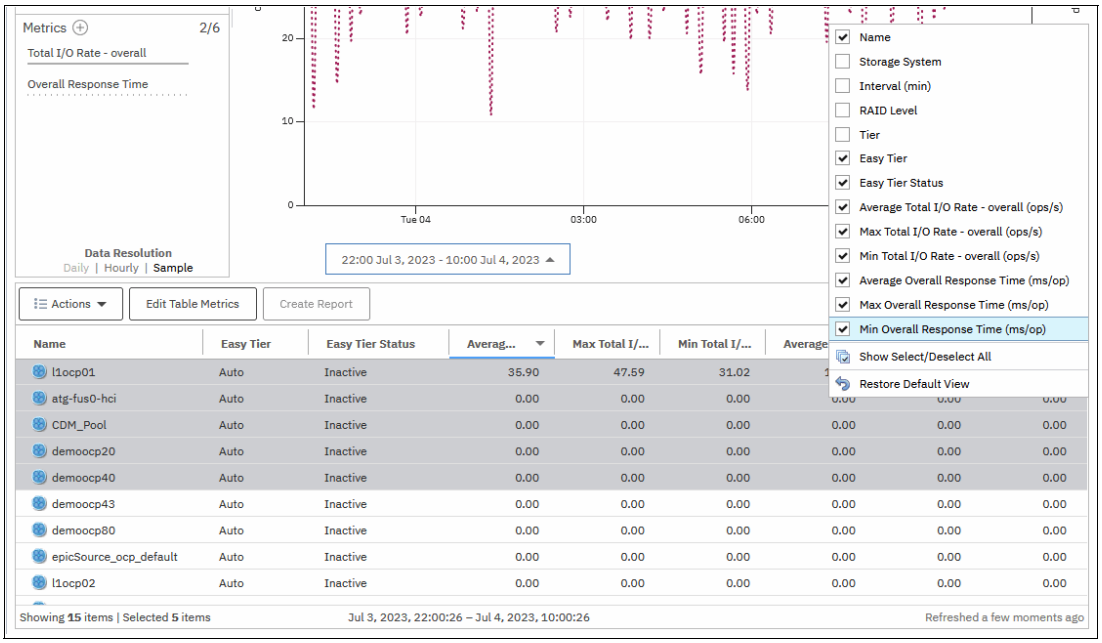


Figure 11-93 Performance List View

11.12.3 Logging support tickets by using IBM Storage Insights

With IBM Storage Insights, you can log existing support tickets that greatly complement the enhanced monitoring opportunities that the software provides. When an issue is detected and you want to engage IBM Support, complete the following steps:

1. Select the system to open the System Overview and click **Get Support**, as shown in Figure 11-94.

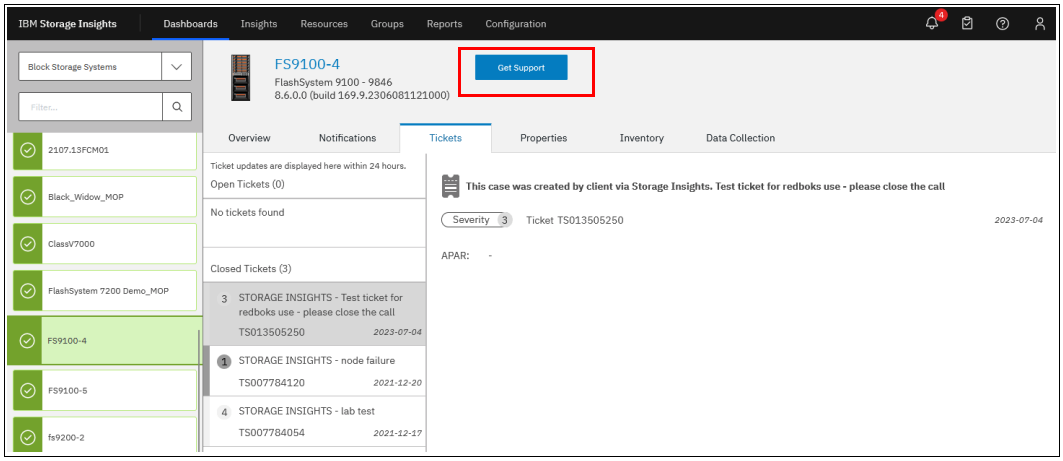


Figure 11-94 Get Support

A window opens where you can create a ticket or update an existing ticket, as shown in Figure 11-95.

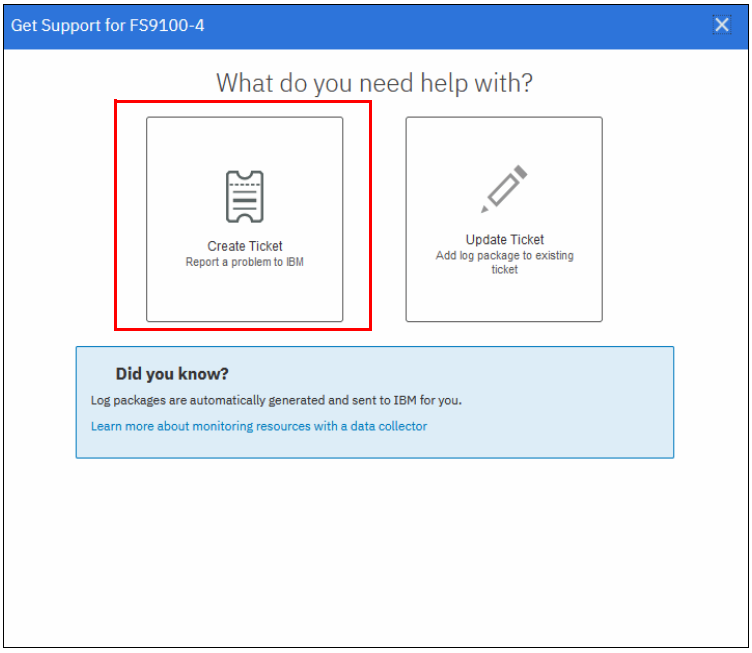


Figure 11-95 Get Support window

2. Select **Create Ticket**, and the ticket creation wizard opens. Details of the system are automatically populated, including the customer number, as shown in Figure 11-96. Select **Next**.

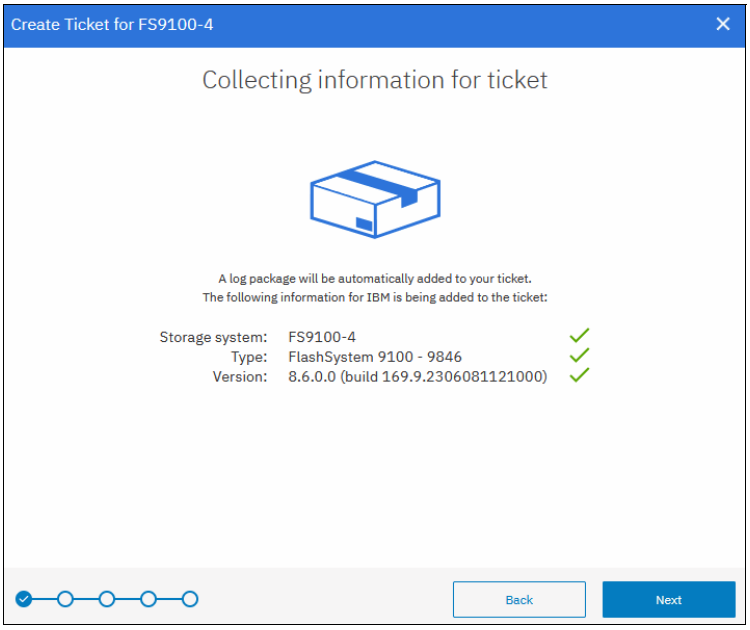


Figure 11-96 Create Ticket wizard

3. You can enter relevant details about your problem to the ticket, as shown in Figure 11-97. It is also possible to attach images or files to the ticket, such as PuTTY logs and screen captures. Once done, select **Next**.

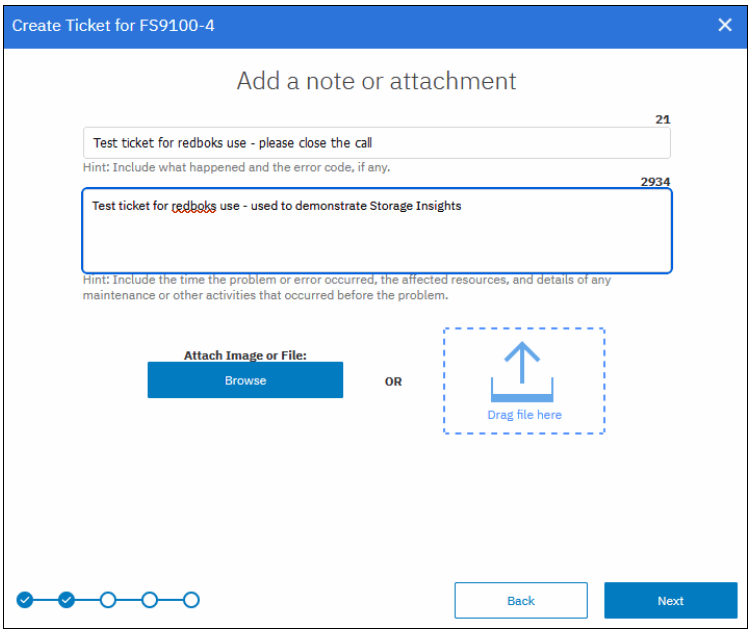


Figure 11-97 Add a note or attachment window

4. You can select a severity for the ticket. Examples of what severity you should select are shown in Figure 11-98. Because in our example there are storage ports offline with no impact, we select **severity 2** because we lost redundancy.

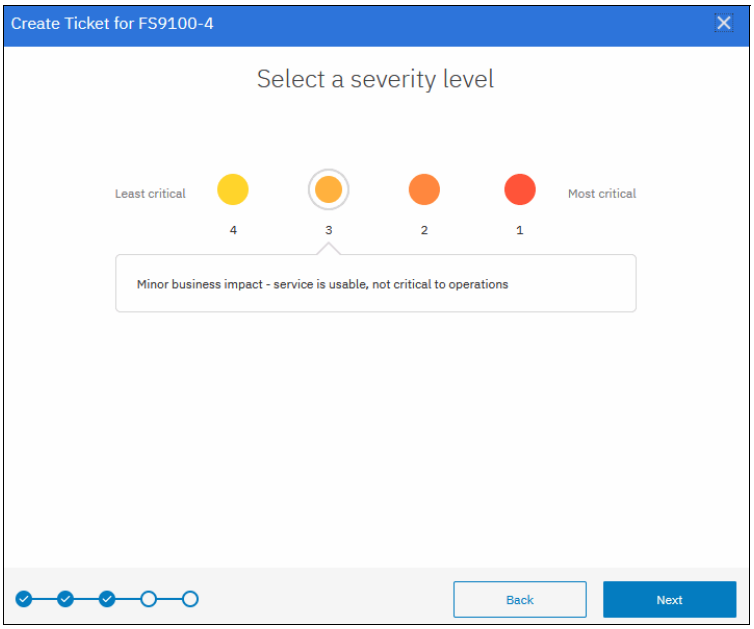


Figure 11-98 Selecting a Severity Level window

5. Choose whether this is a hardware or a software problem. Select the relevant option (for this example, the offline ports are likely caused by a physical layer hardware problem). Once done, click **Next**.

6. Review the details of the ticket that will be logged with IBM, as shown in Figure 11-99. Contact details must be entered so that IBM Support can respond to the correct person. You must also choose which type of logs should be attached to the ticket. For more information about the types of snap, see Table 11-14 on page 1027. Click **Create Ticket**.

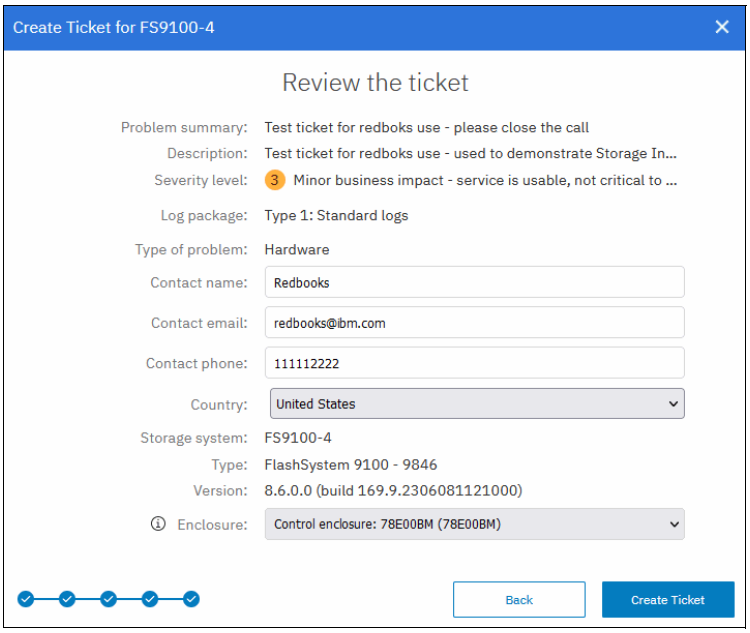


Figure 11-99 Review the ticket window

7. Once done, select **Create Ticket**. A confirmation window opens, as shown in Figure 11-100, and IBM Storage Insights automatically uploads the snap to the ticket when it is collected. Click **Close**.

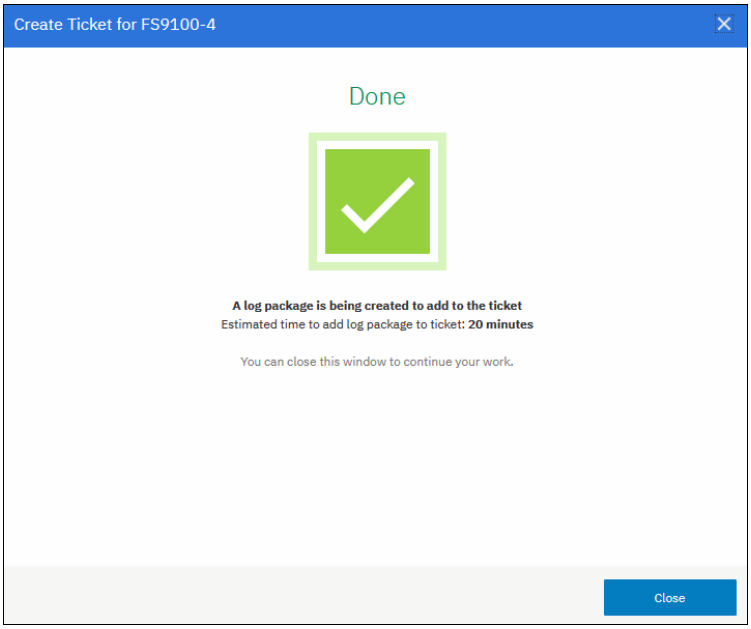


Figure 11-100 Ticket Creation confirmation window

11.12.4 Managing existing support tickets by using IBM Storage Insights

With IBM Storage Insights, you can track existing support tickets and upload logs to them. To do so, complete the following steps:

1. From the System Overview window, select **Tickets**, as shown in Figure 11-101.

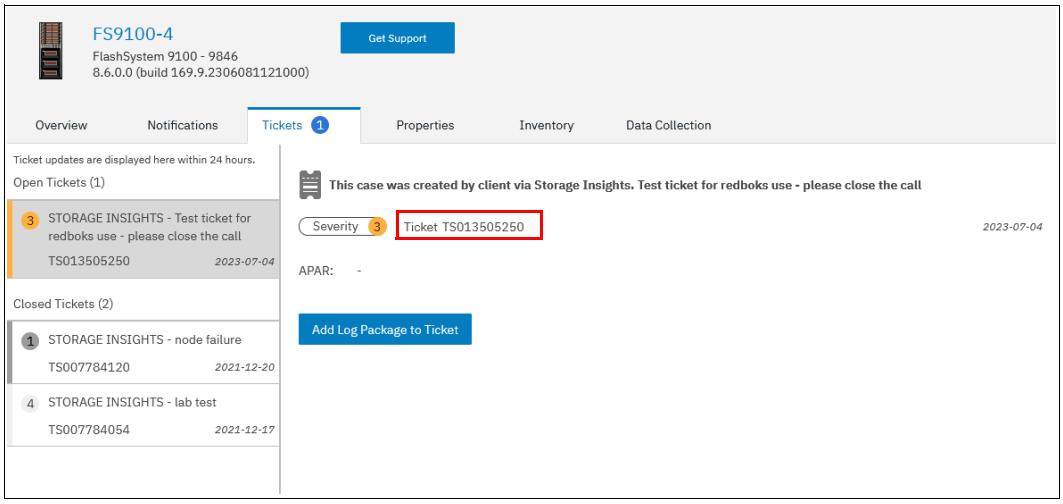


Figure 11-101 View Tickets

This window shows the newly created ticket number and a history of support tickets that were logged through IBM Storage Insights for the system. Tickets that are not currently open are listed under **Closed Tickets,** and currently open tickets are listed under **Open Tickets**.

2. To quickly add logs to a ticket without having to browse to the system GUI or use IBM ECuRep, click **Get Support** and **Add Log Package to Ticket**. A window opens that guides you through the process, as shown in Figure 11-102. After entering the support ticket number, you can select which type of log package you want and add a note to the ticket with the logs.

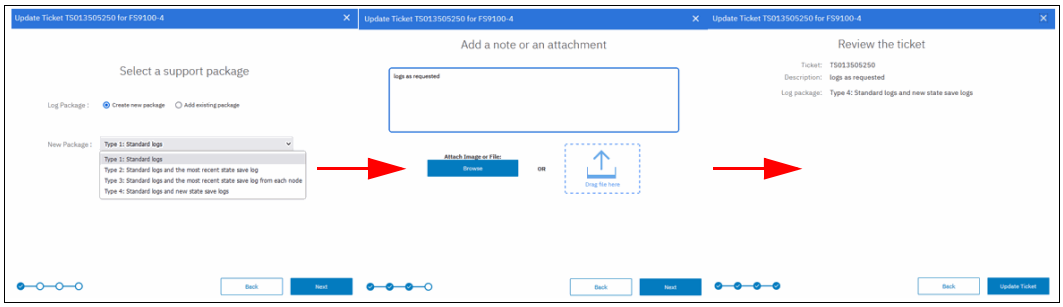


Figure 11-102 Adding a log package to the ticket

3. After clicking **Update Ticket**, a confirmation opens, as shown in Figure 11-103. You can exit the wizard. IBM Storage Insights runs in the background to gather the logs and upload them to the ticket.

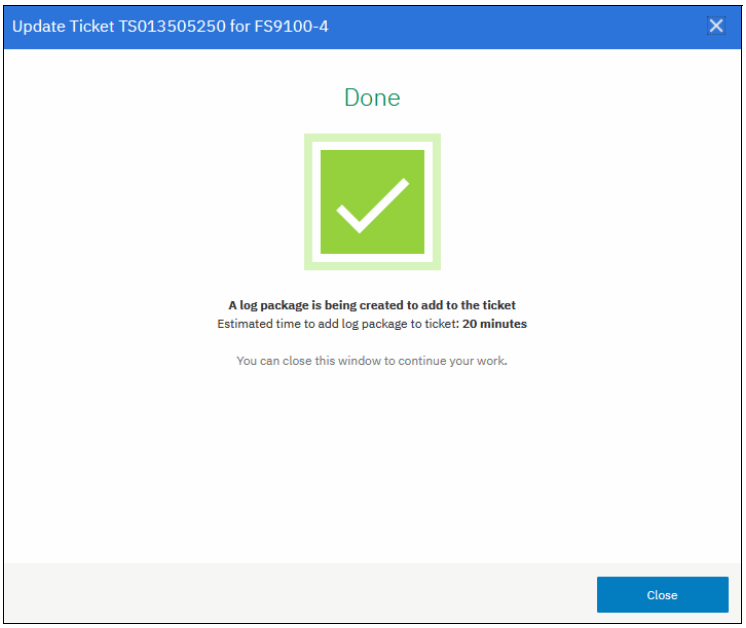


Figure 11-103 Confirming the log upload

12

Security and encryption

IBM Storage Virtualize is a storage platform that implements various security-related features in terms of system-level access controls and data-level security features.

This chapter discusses system-level security and data-level security in terms of encryption.

In addition, you may want to review *IBM Storage Virtualize, IBM Storage FlashSystem, and IBM SAN Volume Controller Security Feature Checklist*, REDP-5716.

This IBM Blueprint can help you and your organization's security team to understand the security features and plan to implement the same into your environment.

This chapter includes the following topics:

- 12.1, “New security features in IBM Storage Virtualize V8.6” on page 1046
- 12.2, “Introduction to IBM Storage Virtualize security” on page 1046
- 12.3, “Configuring users and password policy” on page 1052
- 12.4, “Multifactor authentication” on page 1066
- 12.5, “Ownership groups principles of operations” on page 1083
- 12.6, “Two Person Integrity (TPI)” on page 1093
- 12.7, “Encryption” on page 1098
- 12.8, “Activating and enabling encryption” on page 1103
- 12.9, “Using encryption” on page 1141

© Copyright IBM Corp. 2024.

12.1 New security features in IBM Storage Virtualize V8.6

Several new security-related features were introduced with IBM Storage Virtualize version 8.6.0:

- Support for non-superuser ability to manage the system
- SMTP Authentication
- Support for TLS 1.3 and RSA keys with a length of 4096 bits

12.1.1 Support for non-superuser ability to manage the system

Prior to Storage Virtualize version 8.6, only users with the superuser role could run certain system management commands through the Command Line Interface (CLI). Starting with version 8.6, these commands can now be run by users with either the 'Security Administrator' or 'Administrator' role. The applicable commands are described in the following list:

- **`svcinfo lsnodestatus`**, which is equivalent to **`sainfo lsservicestatus`**
- **`satask restartservice`**, which is equivalent to **`satask restartservice`**
- **`svctask stopsystem -enterservicestate`**, which is equivalent to **`satask startservice`**
- **`svctask startsystem`**, which is equivalent to **`satask stopservice`**
- **`svcinfo traceroute`**, which is equivalent to **`sainfo traceroute`**
- **`svctask chnodeserviceip`**, which is equivalent to **`satask chserviceip`**

12.2 Introduction to IBM Storage Virtualize security

IBM Storage Virtualize features multiple levels of security to protect against threats and to keep the attack vector as small as possible:

- Strict verification features that prevent unauthorized users from using login interfaces and gaining access to the system and its configuration
- Least privilege features that restrict the environment and limit any effect if a malicious actor managed to gain access to a system configuration interface
- Minimal, locked down mode to prevent damage spreading to the kernel and the rest of the operating system
- Protection of data at rest that is stored on the system from theft, loss, and malicious or accidental data corruption

The topics that are discussed in this chapter can be broadly divided into these two categories:

1. System Security

   System security includes features that prevent unauthorized access to the system, protect the logical configuration of the storage system, and restrict what actions users can perform. System security also ensures visibility and reporting of system-level events that can be used by a Security Information and Event Management (SIEM) solution, such as IBM QRadar®.

   These features include, but are not limited to, multifactor authentication (MFA), role-based access control (RBAC), object-based access control (OBAC), or disabling access to the command line interface (CLI), graphical user interface (GUI), and Representational State Transfer interface (REST) API.

2. Data Security

   Features in the fourth line of defense that protect the data that is stored on the system against theft, loss, or attack. These features include, but are not limited to, encrypting data at rest (EDAR) and IBM Safeguarded Copy (SGC).

12.2.1 Configuring System TLS Certificates

During system initialization, a Transport Layer Security (TLS) certificate is automatically generated and signed by the system-internal root certificate authority (CA), which was introduced with IBM Storage Virtualize V8.5.3. IBM Storage Virtualize systems that were updated from earlier Storage Virtualize releases, by default have a self signed TLS certificate in place, unless they had been configured with a signed certificate supplied by a trusted CA. If a self-signed certificate does expire, or it is renewed before its expiration date, it is also upgraded to a signed certificate.

Note: TLS has succeeded the deprecated Secure Sockets Layer (SSL) protocol. However, TLS certificates are commonly still referred to as SSL certificates. Therefore, the terms TLS and SSL can be used interchangeably.

The system TLS certificate serves to secure various features:

- Encryption of communication between browser and web-based GUI
- Communication with Encryption Key Servers
- VASA Provider (vVols)
- IP Replication over IPsec
- IP quorum
- Multi-Factor Authentication (MFA)

Secure communication between browser and web-based GUI

The web-based GUI is the central point of management of an IBM Storage Virtualize system. Communication between a web browser running on a workstation and the GUI is always encrypted using TLS.

The web browser used to access the Storage Virtualize management GUI raises a security warning concerning self-signed certificates or a certificate that is signed by an untrusted CA. Also, using self-signed certificates might not comply with organizational security guidelines.

Signed TLS certificates are issued by a CA, which can be either an internal one, for example part of your organization's own Public Key Infrastructure (PKI), or a public trusted CA such as GlobalSign, Let's Encrypt, and others. Web browsers maintain a list of trusted CAs that are identified by their root certificate. The root certificate must be included in this list for the signed certificate to be trusted, so no security warning will be raised.

Organization-internal Root CAs are usually configured as trusted authorities across your company.

The purpose of the internal Root CA of an IBM Storage Virtualize system is to add its certificate to truststores on systems communicating through TLS-encrypted paths. This can include administrator workstations that are used to access the management GUI, Encryption Key Servers or other Storage Virtualize systems in a Remote Copy partnership. The advantage of applying the internal root CA certificate over using the system's certificate is that the communication can seamlessly continue between these systems, even if the certificate was renewed after it had expired.

To see the details of your system's certificate in the GUI, select **Settings** → **Security** and then click **System Certificates**, as shown in Figure 12-1. In the CLI, you can use the command **`lssystemcert`**.

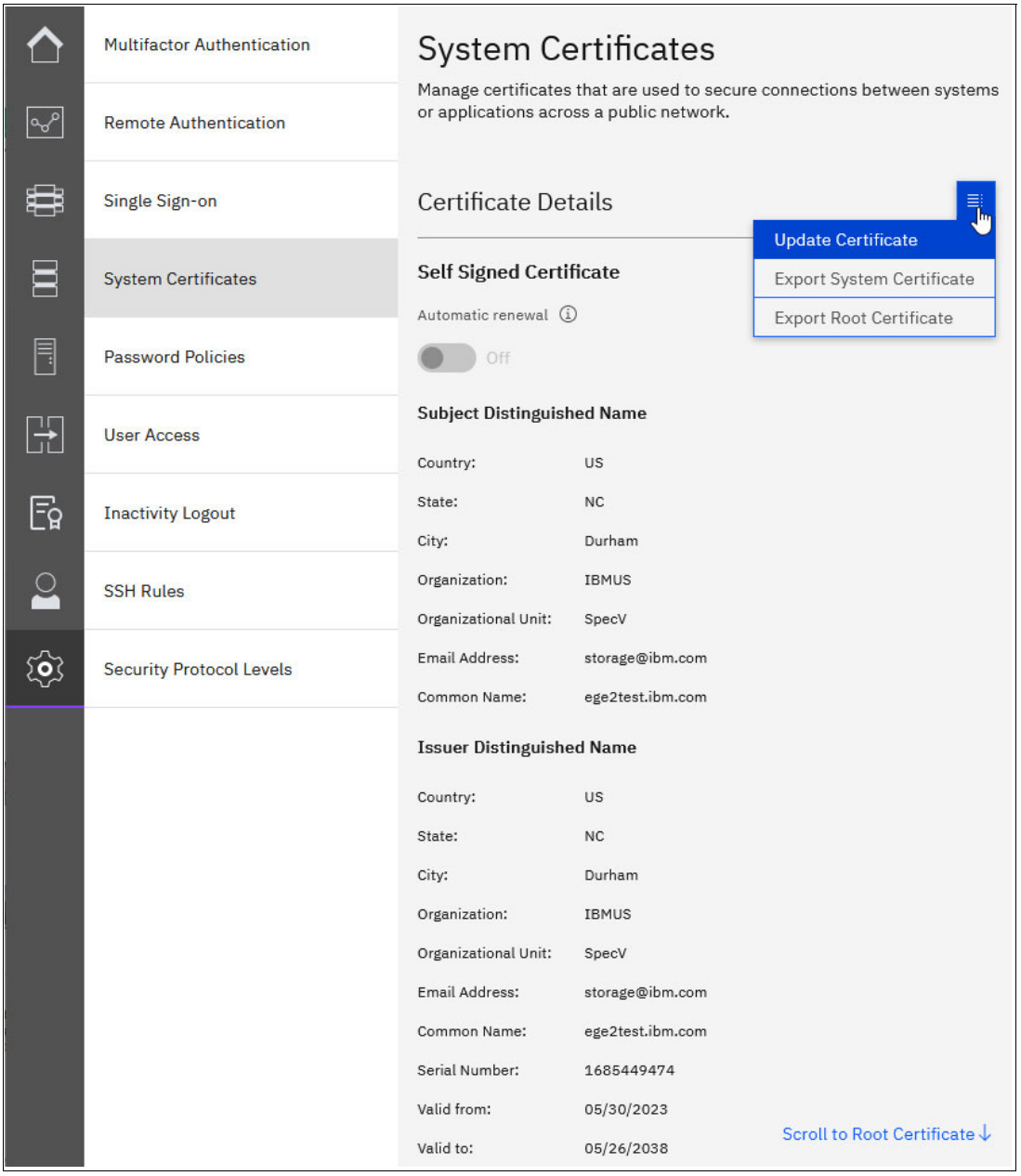


Figure 12-1 Accessing the System Certificates window

Depending on your organization's security policies and needs, you can create a new certificate, which is a self-signed by the system. Another option is to create a Certificate Signing Request (CSR), which can be submitted to be signed by a CA within your own organization or by an external, trusted CA. Also, beginning with Storage Virtualize V8.5.3, the option to use a system-internal root CA was added, which can be used to sign certificate signing requests created for this system.

Generating an Internally Signed Certificate

If the system certificate has expired or is about to expire soon, you can generate a new, internally signed one using the GUI as shown in Figure 12-1 on page 1048.

Before creating an internally signed certificate, ensure that the key type that you are going to use is supported by your web browser and is in compliance with your organization's security policy. For Storage Virtualize V8.6 and later, support for key type RSA-4096 was added.

1. Select **Update Certificate** on the **System Certificates** window.
2. Select **Internally Signed Certificate** and enter the details for the new certificate as shown in Figure 12-2 and click **Update**.

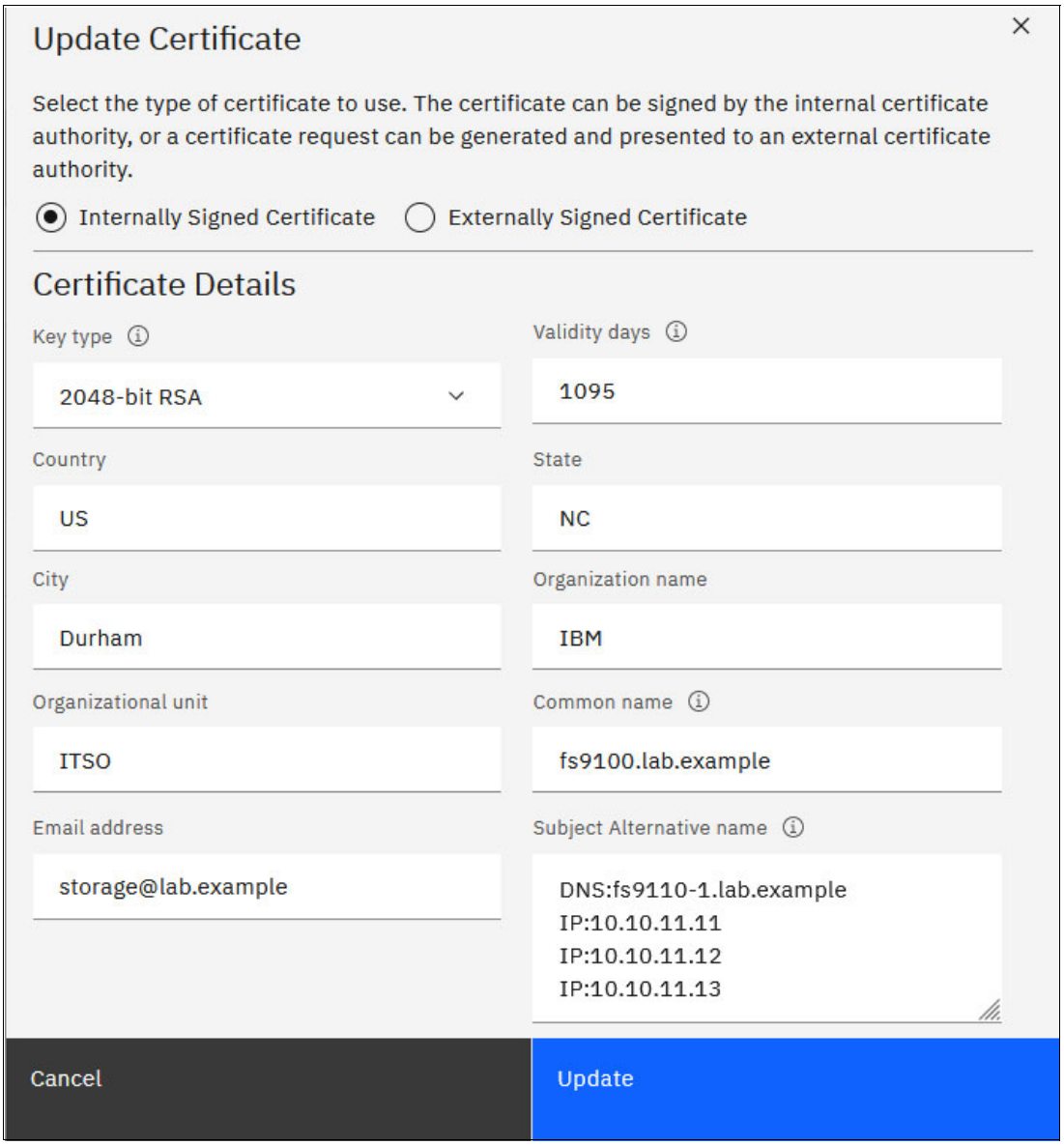


Figure 12-2 Creating a self-signed system certificate

3. You are prompted to confirm the action. Click **Yes** to proceed. Close the browser. Wait approximately 2 minutes and reconnect to the management GUI.

Certain web browsers, such as those that are based on Google Chrome, require the certificate to have a Subject Alternative Name (subjectAltName) in addition to the Common Name. The **Subject Alternative Name** field can be up to 512 characters. The syntax is in the form of **`key:value`** pairs as shown in the following examples:

- `DNS:fs9110-1.lab.example`

  This value contains the fully qualified domain name (FQDN). It was assigned within your infrastructure's DNS environment. It is not supposed to contain a DNS server's IP address.

- `IP:10.10.11.11`

  If you access your system also by using an IP address instead of its DNS name, add all configured *management* and *service IP addresses* to avoid browser certificate warnings.

- `email:storage@lab.example`

The CLI equivalent of the GUI example is shown in Example 12-1:

Example 12-1 Generating a self-signed certificate

```
IBM_FlashSystem:FS9110:superuser>chsystemcert -commonname fs9100.lab.example
-country US -email storage@lab.example -gui -keytype rsa4096 -locality Durham
-mksystemsigned -org IBM -orgunit ITSO -state NC -subjectalternativename
"DNS:fs9110-1.lab.example IP:10.10.11.11 IP:10.10.11.12 IP:10.10.11.13" -validity
1095
```

With the addition of the Internal Root CA functionality, the **`chsystemcert`** CLI command parameter **`-mkselfsigned`** was succeeded by **`-mksystemsigned`**. The difference between both is that the generated system certificate is no longer a self-signed one, it is signed by the internal root CA. However, the certificate of the Internal Root CA is still a self-signed one and is therefore not a trusted CA. You can offload both the root and the system certificate by selecting **Settings** → **Security** → **System Certificates** → **Options menu** → **Export** and add it to the store of trusted authorities of your browser or your workstation OS respectively:

An internally signed system certificate can be renewed automatically 30 days before its expiration date.

Applying an Externally Signed certificate

If your company's security policy requests certificates to be signed by a trusted certificate authority (CA), complete the following steps to configure a signed certificate:

1. Select **Update Certificate** in the System Certificates window.
2. Select **Externally Signed Certificate** and enter the details for the new certificate signing request, as shown in Figure 12-3 on page 1051. For the Country field, use a two-letter country code according to ISO 3166-1 alpha-2.
3. Click **Generate Request**.

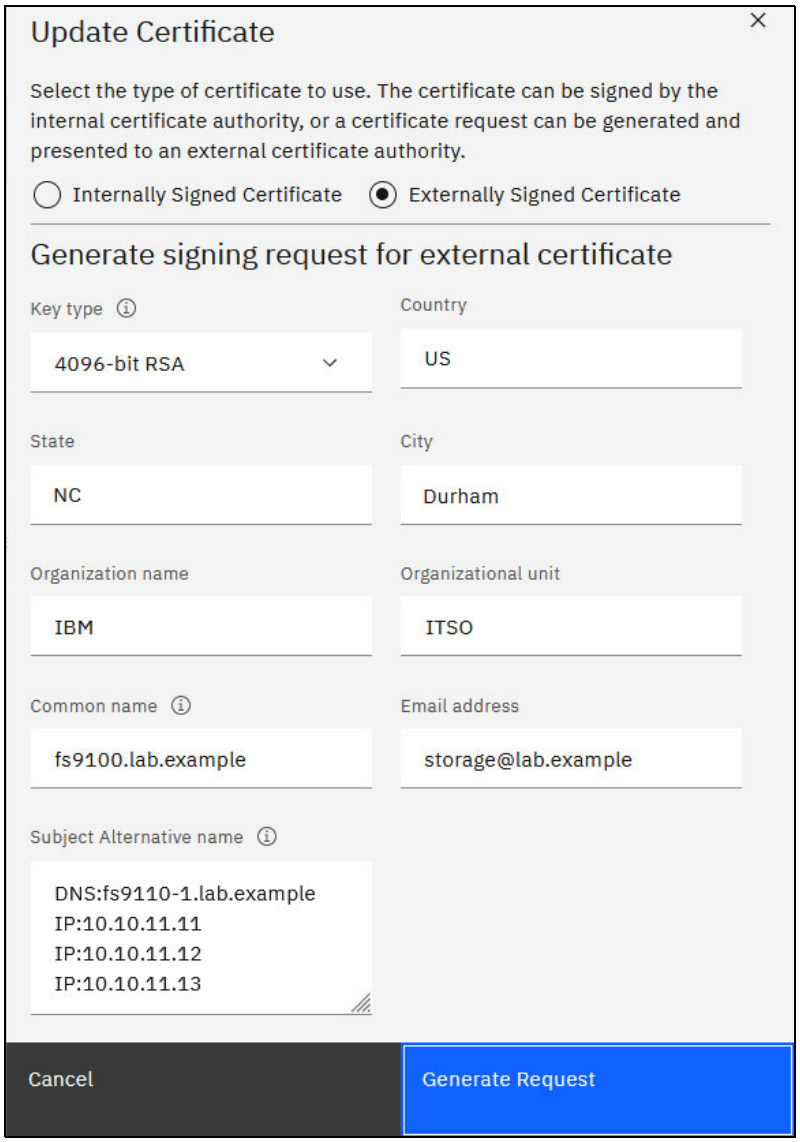


Figure 12-3 Generating a certificate signing request

4. When prompted, save the `certificate.csr` file that contains the certificate signing request. If you must download the CSR another time, you can find the CSR file `certificate.csr` in the `/dumps` directory of the configuration node. Until the signed certificate is installed, the **System Certificates** window shows that an outstanding signing request exists.
5. Submit the request to the appropriate CA to have the CSR signed. A signed certificate is generated from the CSR. The certificate must be in PEM format. The file name does not matter, though common filename extensions are `.pem`, `.crt` or `.cer`. Example 12-2 shows an excerpt of a base64-encoded certificate in PEM format.

Example 12-2 TLS Certificate in PEM format

```
-----BEGIN CERTIFICATE-----
MIIFODCCBCCgAwIBAgIIRNDAZIllU8OwDQYJKoZIhvcNAQELBQAwgZgxCzAJBgNV
BAYTAkFUMQOwCwYDVQQIEwRXaWVuMQOwCwYDVQQHEwRXaWVuMQOwCwYDVQQKEwRC
[...][truncated][...]
IvY/T3hzJz15ZdxoowCDIsHjlURk9RDSV/ccgw/E3Ig5EOIE69EOOErnMhBvKBG8
```

```
1U3o4sO+Tb712Wd9HZT/xXvzibKO23cryRwGVsGtBcGg+9xloIaObe5VQFw=
-----END CERTIFICATE-----
```

6. When you receive the certificate signed by the CA, select **Install Signed Certificate** in the **System Certificates** window, then **Add file**.
7. Navigate to and select the signed certificate file, click on **Install**.
8. You are prompted to confirm the action. Click **Continue** to proceed. While the certificate is being installed, there can be no further GUI response until the session eventually times out and a connection error is shown. Allow approximately two minutes before you reconnect your browser to the management GUI.
9. After you have reconnected successfully, navigate to **Settings** → **Security** → **System Certificates**. The Certificate Details window shows that you are using an Externally Signed Certificate, as shown in Figure 12-4.

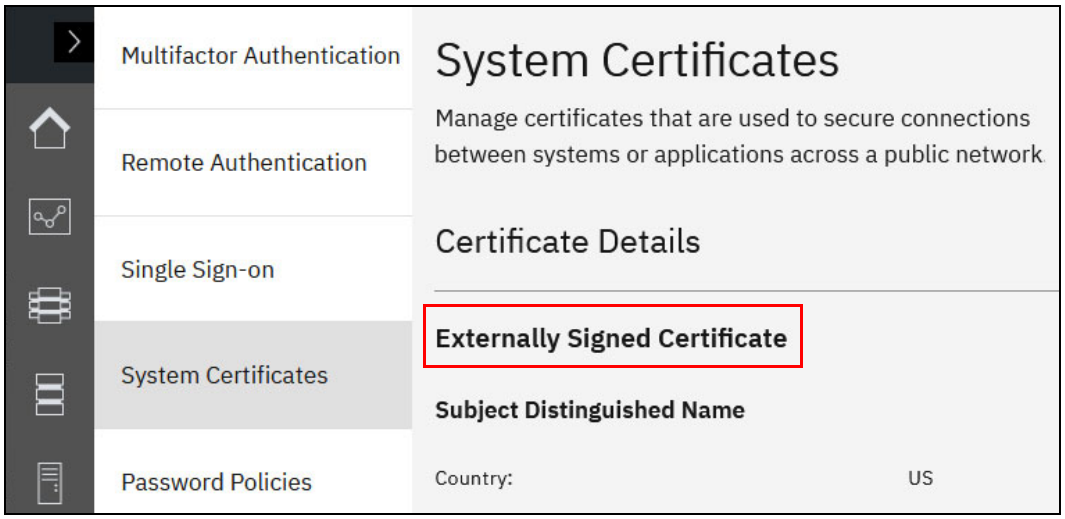


Figure 12-4 Signed certificate installed

12.3 Configuring users and password policy

To administer, configure, and monitor the system, an authenticated user is required. The system supports local and remote users. Local users are defined on the system itself and managed internally by the system. This must be done on each individual Storage Virtualize system. Remote users are defined and maintained in an authentication service external to the Storage Virtualize system, for example in your organization's LDAP directory.

Each defined user belongs to a user group, by which its access level privileges are defined.

12.3.1 Local users

A local user is a user whose account is managed entirely within the system itself. A local user can belong to one user group only. It must have a password, an SSH public key, or both. Each user has a username, which must be unique across all users in a system.

Usernames can contain up to 256 printable characters defined in the American Standard Code for Information Interchange (ASCII). The following characters cannot be used:

- Single quotation mark (')
- Colon (:)
- Percent symbol (%)
- Asterisk (*)
- Comma (,)
- Double quotation marks (“).

Also, a username cannot begin or end with a blank space.

Passwords for local users can be up to 64 printable ASCII characters, but cannot begin or end with a space.

When you connect to the CLI by using an SSH client, SSH public key authentication is attempted first, with the username and password combination available as an alternative. The SSH key authentication method is available for CLI access and file transfer using Secure copy protocol (SCP). For GUI access, only the combination of username and password is used. The system supports up to 200 or 400 local users, depending on your platform. Locally administered users can coexist with remote authentication.

12.3.2 Remote authentication

Remote authentication can simplify the burden of user administration, as it provides a single point of administration instead of maintaining user ids on each individual system. This applies particularly in large environments that are composed of more than a few systems, but even in smaller environments it is advisable to consider the implementation of remote authentication.

Remote users are administered in the organization’s LDAP directory infrastructure, where the actual authentication of remote users is done. The LDAP infrastructure can be a single, dedicated LDAP server, or part of a more complex environment, such as Microsoft Active Directory Services (MSAD) or Red Hat Directory Server.

Communication between the IBM Storage Virtualize system and an LDAP server is realized through the Internet Protocol (IP). Up to six LDAP servers can be configured. Optionally, the communication between the system and the LDAP server can be configured for transport encryption using TLS or LDAPS.

Remote users and user groups

User groups for remote users must be configured with the same name in the Storage Virtualize system and in the LDAP directory The group name is case sensitive. The role associated with each individual user group defines the access level privileges or permissions of a remote user, who can be a member of one or more user groups in the directory. The membership of the group with the highest privileges defines the user’s access level privileges.

User groups used for remote authentication must be flagged accordingly, either during creation or later.

Users that are authenticated by an LDAP server can log in to the management GUI and the CLI. These users do not need to be configured locally for CLI access, and they do not need an SSH key that is configured to log in by using the CLI.

If multiple LDAP servers are available, you can configure up to six LDAP servers to improve resiliency. If more than one LDAP server is configured, one ore more of these can be marked as preferred. Authentication requests are processed by those LDAP servers, unless the

connection fails or a user is not found. Requests are distributed across all preferred servers for load balancing in a round-robin fashion.

Note: All LDAP servers that are configured within a system must be of the same type.

12.3.3 Default user

During initial configuration, the system defines a single local Security Administrator user that is called a *superuser*. For newly created systems, the default superuser password must be changed at first login. This user cannot be deleted or configured for remote authentication, but it can be disabled for higher security levels.

During initial setup, define the required users for the system. After this definition is complete, it is a best practice to disable the default superuser. It is advisable not to use the superuser account for daily administration tasks. The superuser account can be reenabled with only physical access to the system by using the technician port, by another user with the Security Administrator role, or by an IBM remote support engineer by using remote support assistance (RSA).

If a security information and event management (SIEM) solution is used, consider forwarding the system audit log so that the SIEM system can be triggered upon this action.

Access to the Service Assistant GUI (SA GUI) or performing certain service procedures through Service Assistant CLI commands, which are prefixed by either **`satask`** or **`sainfo`**, require the superuser login. If the superuser account is disabled, it might be necessary to temporarily re-enable the account in certain situations.

If you chose not to disable the superuser account, ensure that the highest level of protection is afforded to the superuser account. For example, use a strong password, limit knowledge of the password, and configure multifactor authentication.

The password for the superuser is set during the system setup. The superuser password cannot be reset to its default value of `passw0rd`. However, it can be reset by using a procedure that is described in this IBM Documentation article: Reset service IP address and superuser password command.

12.3.4 User groups and roles

A local user can be assigned to a single user group on the system. The system supports up to 256 user groups. Each user group is assigned a role that is associated with a set of privileges The user group privileges determine which configuration commands can be run by the users that are members of this group. The available roles are described next.

Monitor

Users with this role can view all objects but cannot manage the system or its resources. Support personnel can be assigned this role to monitor the system and to determine the causes of problems. This role is suitable for use by automation tools, such as the *Storage Insights* data collector for collecting status about the system.

Copy Operator

Users with this role have monitor role privileges and can create, change, and manage all Copy Services functions (Remote Copy and FlashCopy) but cannot create consistency groups or modify host mappings.

FlashCopy Administrator

Users can create, change, and delete all of the FlashCopy mappings and consistency groups and create and delete host mappings.

Administrator

Users with this role can access all functions on the system, except those that deal with managing users, user groups, and authentication. This standard role is assigned to users who administer the system and perform tasks, such as provisioning storage.

Security Administrator

Users with this role can access all functions on the system, including managing users, user groups, all aspects of security, and user authentication.

Service

These users can delete dump files, add and delete nodes, apply service, and shut down the system. Users can also perform the same tasks as users in the monitor role.

Restricted Administrator

Users with this role can perform the same tasks as the Security Administrator role, but are restricted from deleting specific objects. Support personnel can be assigned this role to solve problems.

3-Site Administrator

Users with this role can configure, manage, and monitor 3-site replication configurations through specific command operations that are available only on the 3-site orchestrator. This role is the only role that is intended to be used with the 3-site orchestrator.

VASA Provider

Users with this role can manage virtual volumes (vVOLs) that are used by VMware vSphere and managed through IBM Storage Connect software and IBM Storage Virtualize Plugin for vSphere.

12.3.5 Configuring remote authentication

Complete the following steps in the GUI to configure an LDAP-based remote authentication service:

1. Navigate to **Settings** → **Security** → **Remote Authentication**.
2. Select the **Configure Remote Authentication** button.

   The system displays the Configure Remote Authentication window, as shown in Figure 12-5 on page 1056.

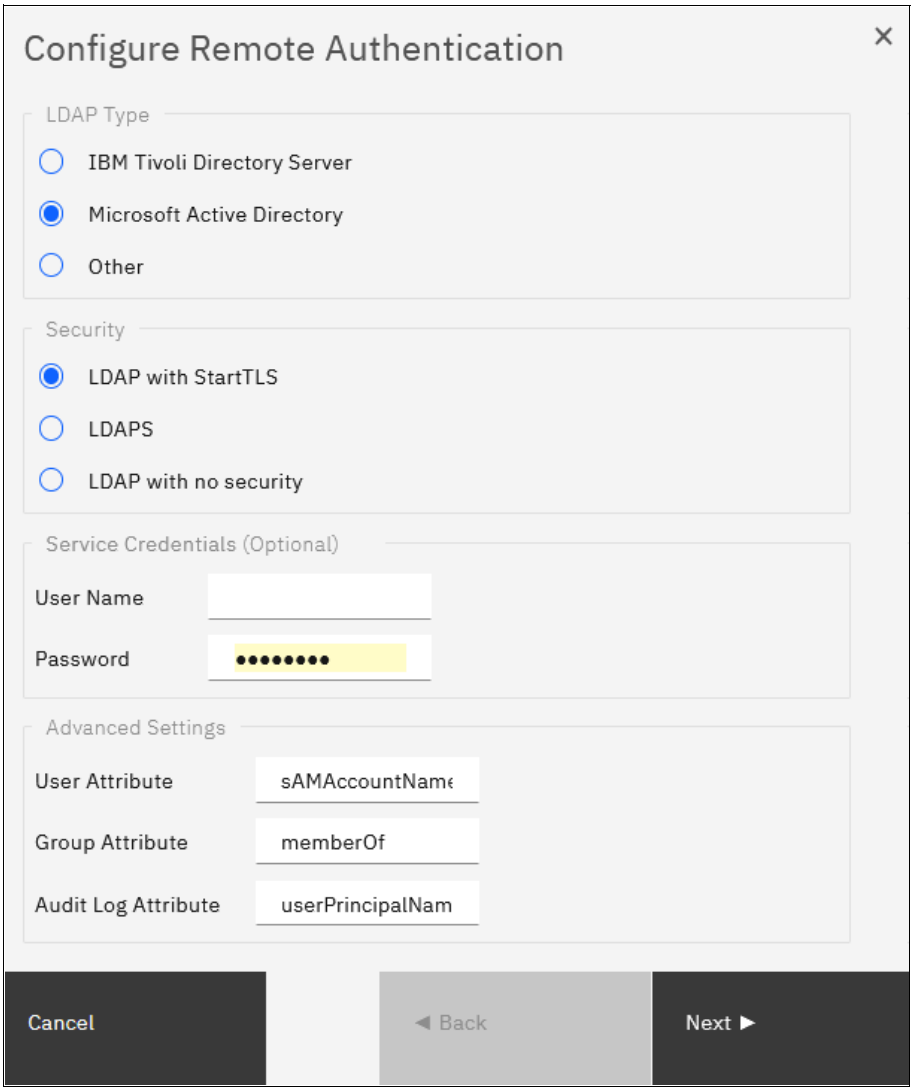


Figure 12-5 Configure remote authentication window with Advanced Settings expanded

3. Select the type of LDAP server and the security settings. Refer to your LDAP server documentation or administrator for details.

   The following settings are available:

   – LDAP type:

   - IBM Security Directory Server 6.4.0.
   - Microsoft Active Directory (MSAD), see Active Directory documentation.
   - Other LDAP v3-capable directory servers, such as Open LDAP.

   – Security:

   - LDAP with StartTLS

     Select this option to use the StartTLS extension according to RFC 2830. StartTLS establishes an unencrypted connection with an LDAP server through TCP port 389, and then performs a TLS handshake over this connection.

   - LDAPS

     Select to use LDAP over SSL and establish secure connections through default secure TCP port 636.

- LDAP with no security

  Select to transport data in clear text format without encryption.

  Note: It is highly recommended to configure security so that the communication between the IBM Storage Virtualize system and the LDAP server is encrypted. Otherwise, user and service credentials would traverse the network in clear text. However, to keep things simple, initially you might want to set up remote authentication with no security until everything is working as desired. LDAP with StartTLS or LDAPS respectively can be configured afterwards.

– Service Credentials

  Enter the credentials of an existing directory for an administrative bind. This user needs permission to query the directory. If your organization's security guidelines permit, the same service user might be used across multiple IBM Storage Virtualize systems. To avoid LDAP authentication failures, consider having this service user configured with no password expiration. Most of the LDAP implementations do not allow anonymous binds, so configuring Service Credentials is usually a requirement. The username usually has to be in User Principal Name (UPN) format, for example `user@lab.example`.

– Advanced settings

  Verify with the administrator of the LDAP server to ensure that these fields have the correct information. These attributes are defined in the LDAP directory's schema, which defines the directory structure. Each attribute has a name and can hold one or more values, which are returned in the response of an LDAP server upon an authentication request:

  - User Attribute

    Determines the username of remote users. The attribute must exist in your LDAP schema and must be unique for each user. The default User Attribute is `sAMAaccountName` for MSAD and `uid` for IBM Tivoli Directory Server and other LDAP servers.

  - Group Attribute

    Determines a user's membership in one or more groups. The Group Attribute defaults to `memberof` for MSAD and Other, and to `ibm-allGroups` for IBM Security Directory Server. For Other LDAP type implementations, you might need to configure the memberOf overlay, if it is not in place.

  - Audit Log Attribute

    Determines the identity of remote users. When an LDAP user performs an audited configuration action, this identity is recorded in the system audit log. This advanced setting defaults to userPrincipalName for MSAD and to uid for IBM Security Directory Server and Other. Under certain circumstances, a user might not have a User Principal Name set in its directory record. The LDAP server would then send a response without a value for this attribute, which leads to an error being logged. In MSAD environments it might be helpful thenthen to configure this value to `sAMAaccountName` instead.

4. When ready, click **Next** and the system displays the LDAP server details window.
5. Enter the details for one or more LDAP servers.

   Enter the IP Address or, if a DNS server is configured, the Domain name for the LDAP server and the Base Distinguished Name (Base DN). You can add the TLS or SSL certificate that is used to verify the LDAP server's identity if security was configured to either LDAP with StartTLS or LDAPS. The following settings are available:

- Preferred

  One or more configured LDAP servers can be marked as Preferred. Requests are distributed among these servers. LDAP servers not marked as Preferred are used only if all the preferred servers are not reachable.

- IP address or Domain

  Enter the IP Address or, if a DNS server is configured, the Domain name of the LDAP server

- Base DN

  The distinguished name in the LDIF format of the starting point in the directory to search for users, for example `ou=storageadmins,dc=infra,dc=lab,dc=example`. Although it is listed as optional, most LDAP servers require a Base DN to be configured. A Base DN limits the search for users to a certain scope. In the example, `ou=storageadmins,dc=infra,dc=lab,dc=example`, the Organizational Unit (OU) `storageadmins` reduces the search time in the directory because the search for the given user id does not need to traverse the entire directory structure.

- SSL Certificate

6. The SSL or TLS certificate is used to verify the LDAP server's identity, if security was configured to either LDAP with StartTLS or LDAPS.

   Note: Transport encryption is effective only if two conditions are met.

   1. Security is configured to anything other than LDAP with no security.
   2. A valid SSL certificate has been applied.

   If your organization uses a tiered CA hierarchy, obtain and install the certificate of the Root CA in your CA hierarchy. This is also helpful if multiple LDAP servers, such as MSAD domain controllers, are located behind a load balancer but can be reached through a single IP address or domain name.

3. If you have more than one LDAP server, add the appropriate entries for each server.

   If you set a certificate and you want to remove it, click the red cross that is next to **Configured**.

   Click the plus (+) and minus (-) signs to add or remove LDAP server records. You can define up to six servers.

4. When ready, click **Finish** to complete the remote authentication setup.

To verify the configuration is working as expected, conduct these two tests from the **Settings** → **Security** → **Remote Authentication** panel:

1. Test LDAP Connections
   - Connectivity to all configured LDAP servers are tested.
   - A bind request using the configured Service Credentials are attempted.
2. If the test was successful, then test LDAP authentication:
   - Connectivity and bind tests are run again.
   - An authentication attempt with the provided user credentials is conducted.

If both test are successful, log in via CLI or GUI. If you log in by using the GUI, log out of the current session before you log in again or use a new private browser window.

12.3.6 Adding locally administered users

Complete the following steps to add new local users by using the GUI:

1. Select **Access** → **Users by Group**.
2. Click the **Create User** button on the resulting page.
3. The Create User window opens, as shown in Figure 12-6.

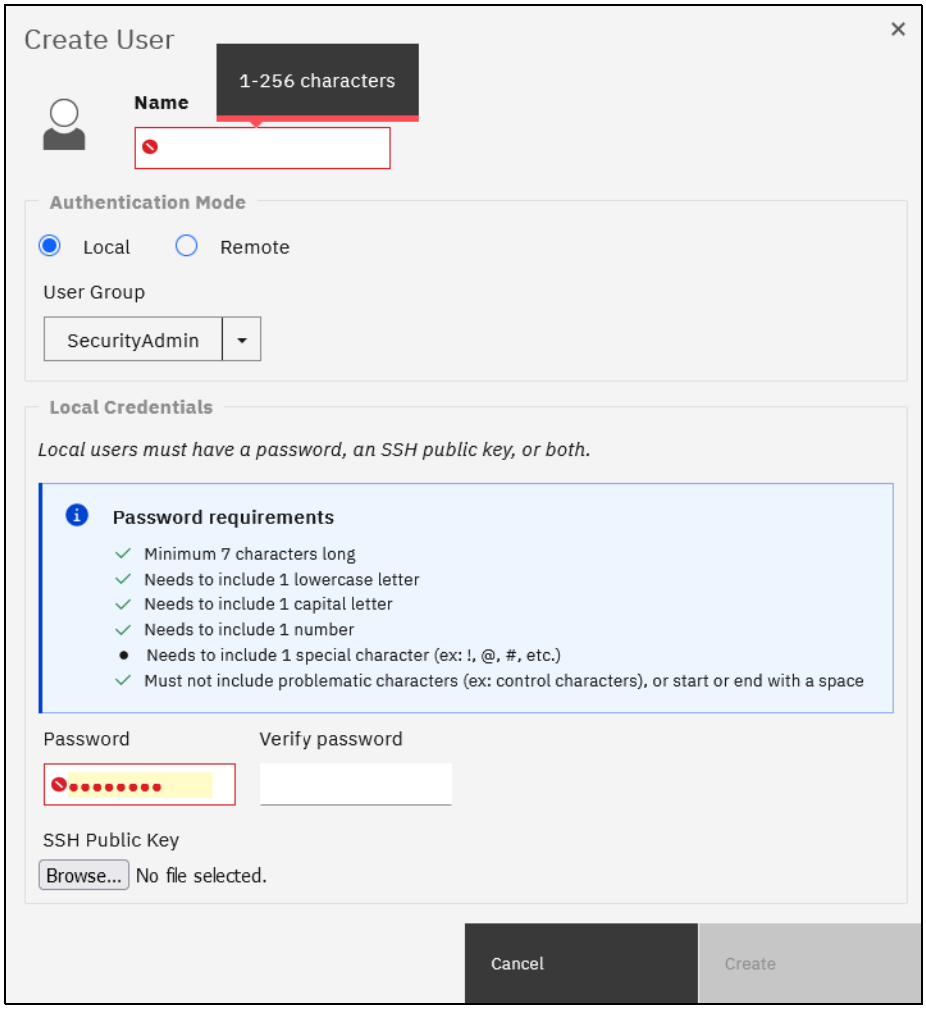


Figure 12-6 Create User window

4. Enter the **Name** and select **Local** or **Remote**. For remote users the name must exist in the remote LDAP directory.

> **Note:** The username is case sensitive.

Note: The concept of a local user with the 'remote' flag can be initially confusing. However, it serves a purpose in specific scenarios. For instance, if you require CLI access via SSH public key, then setting the 'remote' flag eliminates the need for a password at login, which is ideal for automated tasks.

For remote user roles and group memberships, a remote user inherits its role and group memberships from the directory groups to which it belongs.

5. Select the suitable user group or role from the drop-down list.
6. When you have at least one user group that is enabled for remote authentication. If you have set up your user group on the LDAP server correctly, then the following conditions are true:
   - The name of the user group on the LDAP server matches the one that you modified or created on the storage system.

     Note: The user group name is case-sensitive.

   - Each user that you want to authenticate remotely is a member of the LDAP user group that is configured for the intended system role.
7. To test the user authentication, select **Settings** → **Security** → **Remote Authentication** and then select **Global Actions** → **Test LDAP Authentication**. Enter the user credentials of a user that is defined on the LDAP server and click **Test**. A successful test returns the message: `CMMVC70751 The LDAP task completed successfully.`

   A user can log in by using their short name without the domain component or by using the fully qualified username in the UPN format, `user@domain`.
8. For local users, add the initial password for the user. The defined password requirements as described in 12.3.7, “Defining a password policy” on page 1060 are shown in the **Create User** panel.
9. Users that are configured with a password and a public ssh key can use either method to login to the CLI. A login to the GUI requires a password. To raise the security level, the ssh key can be generated with a passphrase.
10. Click **Create** to complete creation of the user.

Note: If LDAP authentication does not work, only local users can log in.

12.3.7 Defining a password policy

Users that are members of the Security Administrator group can define a set of password policies that apply to all users on the system. This set of attributes can be tailored to match your organizational mandates regarding password rules.

The following attributes can be defined:

- Minimum password length: 6–64 characters
- Minimum number of:
  - Uppercase characters: 1–3
  - Lowercase characters: 1–3
  - Special characters: 1–3
  - Digits: 1–3
- History check (0–10) before password reuse

- Password expiry: 0–365 days
- Password expiry warning (0–30 days) which is displayed on CLI at login only
- Password age (1–365 days) which is the minimum age before a password can change
- Force change on next login which is a one-time option by Security Admin

Note: The force password change on next login can be used to keep a newly created user account deactivated until the user initially logs in.

Complete the following steps to set the password policy in the GUI:

1. Select **Settings** → **Security**.
2. Select the **Password Policies** tab and expand the **Password creation** section.
3. Complete all attributes as required (Figure 12-7) and click **Save**.
4. Before you click **Confirm** in the Confirm policy changes window, you can select to have all existing passwords expire immediately, so all users are forced to create a new password at their next login.

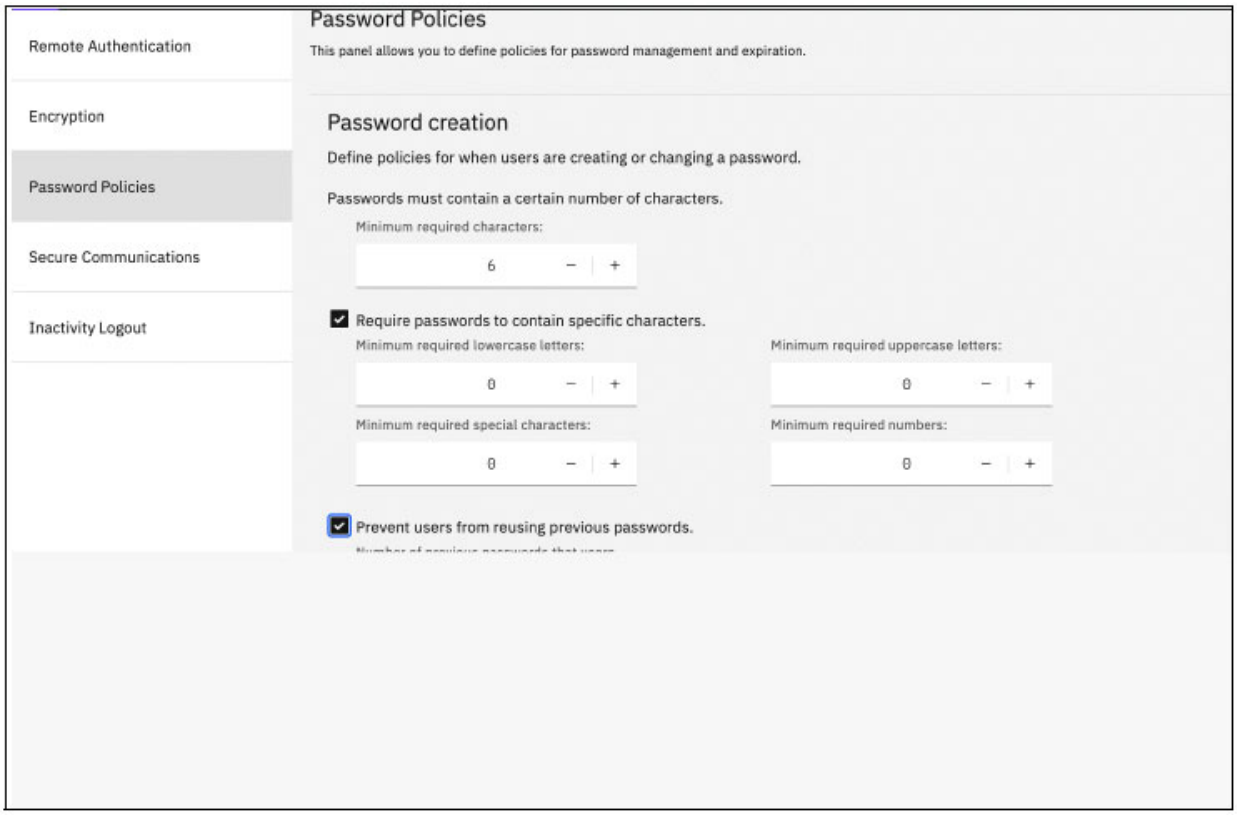


Figure 12-7 Password policies - Password creation

12.3.8 Setting password expiration and account locking

The following options can be used to apply password expiry to passwords:

- Passwords can be set to expire after 0–365 days.
- All existing passwords are set to expire in n days when the setting is first enabled.
- A user with an expired password can log in to the system, but cannot run any **`svctask`** commands until they change their password.

- An expiry warning of 0–30 days can be defined, which warns the user upon login that their password will expire in *n* days.

The security administrator can force a user to change their password at any time. The password expires immediately. If you use the CLI, you can expire individual users. If you use the GUI, you can reset all user passwords.

Typically, password expiry is used when creating a user, and to require a password change on first login. Another example use case is to expire all user passwords when the password policy settings are changed.

Account locking can be managed by using one of the following methods:

- Manually

  The security administrator can manually lock and unlock user accounts by using the CLI, as shown in Example 12-3.

  Example 12-3 Manually lock and unlock a user account

```
IBM_FlashSystem:FS9110:secadmin>chuser -lock Alice
IBM_FlashSystem:FS9110:secadmin>chuser -unlock Bob
```

  Note: A locked account cannot log in to the system.

- Automatically

  Accounts can be locked automatically by using the following parameters:

  – By setting the maximum number of failed login attempts from 0 – 10.

    Note: The counter is reset on a successful login.

  – By setting the length of time of 0–10080 minutes that a user is locked out of the system. The setting of 10080 minutes equals 7 days. A setting of 0 means that the lockout is indefinite.

Locking the superuser Account

As an additional measure to increase the security level and to minimize the attack surface, the superuser account can be locked.

Disabling the superuser account and session timeouts is available only on platforms with a dedicated Technician port.

Note: This feature is not available on IBM FlashSystem 5015 because it does not have a dedicated Technician port.

Complete the following steps to disable the superuser account by using the GUI or CLI:

1. Use a specific option to enable superuser locking, as shown in Example 12-4.

   Example 12-4 Manually enable superuser account locking option

```
IBM_FlashSystem:FS9110:secadmin>chsecurity -superuserlocking enable
Changing the system security settings could result in a loss of access to the
system via SSH or the management GUI. Refer to the Command Line Interface help
```

```
for more information about the risks associated with each parameter. Are you
sure you wish to continue? (y/yes to confirm) yes
```

2. Lock the superuser, as shown in Example 12-5.

Example 12-5 Manually lock the superuser account

```
IBM_FlashSystem:FS9110:superuser>chuser -lock superuser
IBM_FlashSystem:FS9110:superuser>
```

A good use case is assuming that some enterprises include policies that all systems must use remote authentication. Therefore, configure remote authentication, create a remote security administrator, and disable the superuser. Now, no local accounts can log in to the system.

Note: The superuser account is still required for **satask** actions and recovery actions, such as a T3 recovery. It can be unlocked by another user with SecurityAdmin privileges from the GUI or CLI, by connecting to the system's Technician Port, using a USB key, or by Remote Support Assistance (RSA) if RSA is configured. For more information, see Locking user accounts.

Setting password expiration and account lockout by using the GUI

In the Security settings window, complete the following steps to configure password expiration and account lockout settings:

1. Select **Settings** → **Security**.
2. Select the **Password policy** tab and expand the **Password expiration and account lockout** section (see Figure 12-8 on page 1064).
3. Configure all settings as needed and click **Save**.

Note: In this window you can select **Allow locking of the superuser**.

In this section, a button is available that can be used to expire all passwords. Clicking this button forces every local user to change their password when they next login.

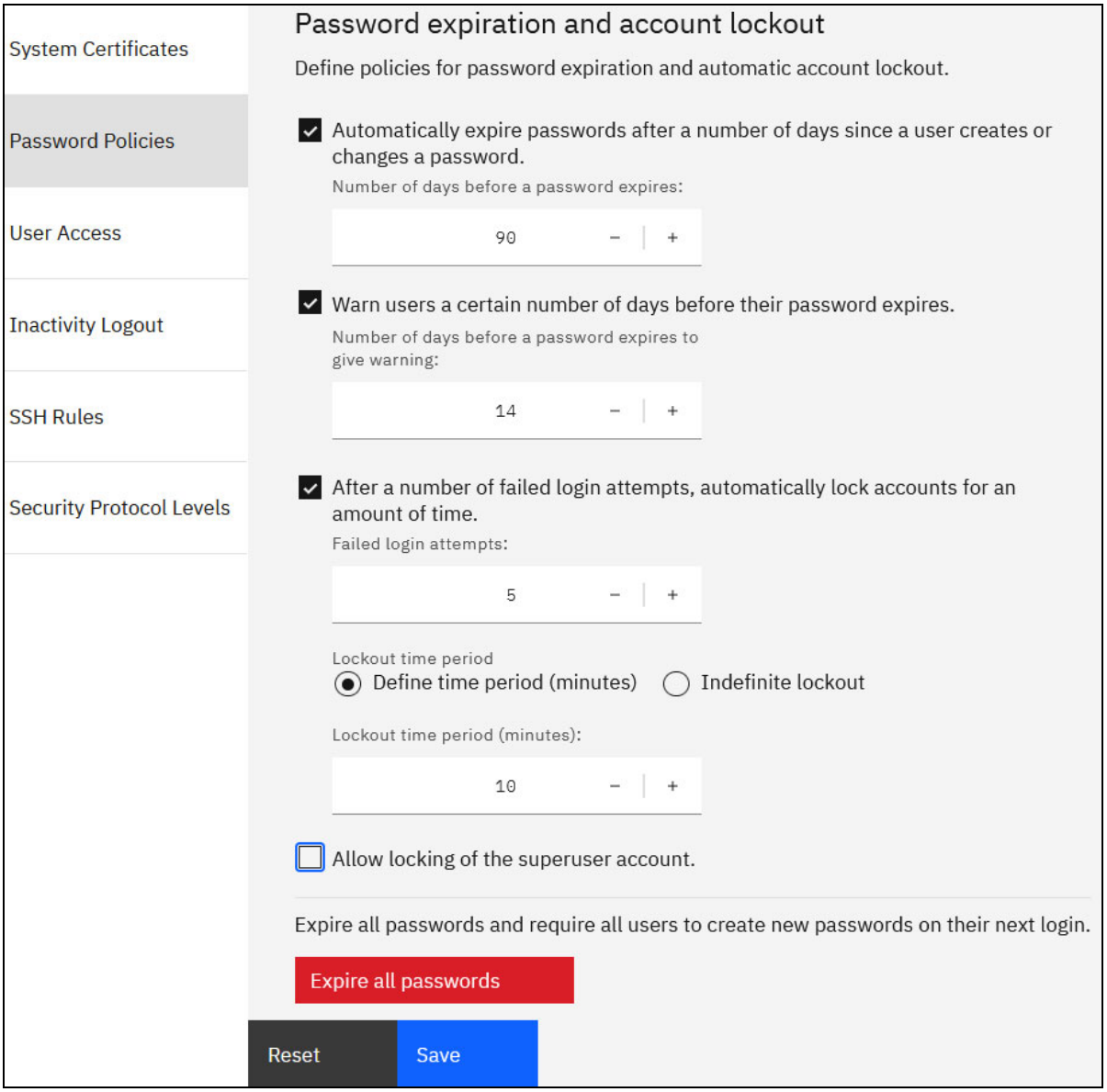


Figure 12-8 Password expiration and account lockout

4. Click **Allow locking of the superuser account** (Figure 12-9).

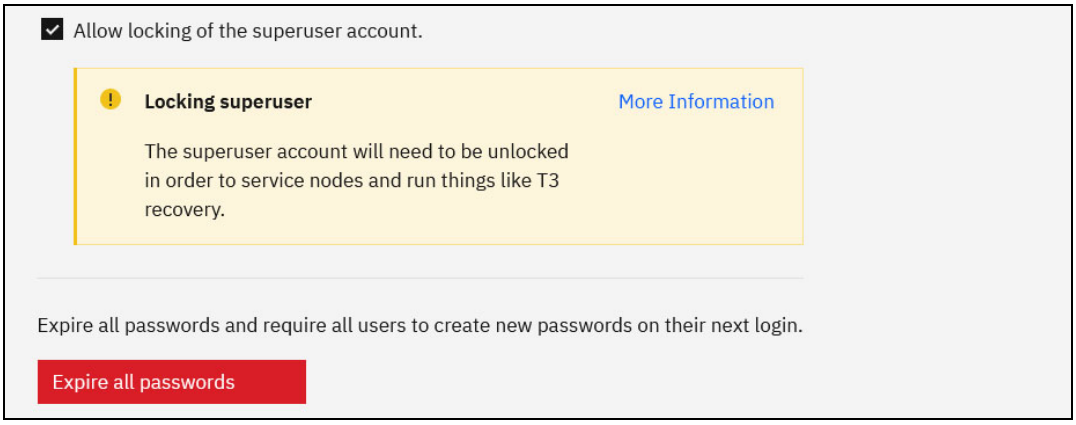


Figure 12-9 Allow locking of the superuser account

5. Click **Save** to complete the setup.

12.3.9 Changing the default session timeouts

The **Inactivity Logout** settings also can be changed from the **Security** settings window. You can define independent settings for the CLI and GUI. The following settings define the time in minutes that a user is automatically logged off the respective interface. Figure 12-10 shows the settings including the allowed values:

- A configurable CLI timeout of 5–240 minutes
- A configurable GUI timeout of 5–240 minutes
- A REST API token timeout of 10–120 minutes

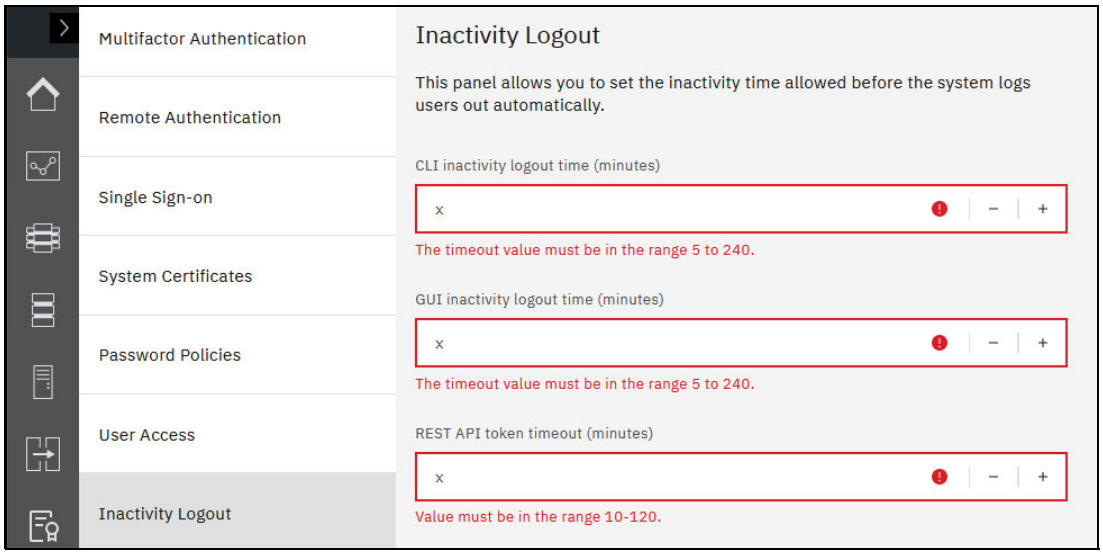


Figure 12-10 Inactivity Logout settings window

12.4 Multifactor authentication

The system allows you to use a second factor authentication mechanism to provide an improved user access security.

In addition to username and password, which is something a user *knows*, a second factor is required to log in when MFA is enabled. The second factor can be, for example, an application or a device generating a one-time passcode (OTP) The second factor can also be a biometrical characteristics such as a fingerprint or a retinal scan.

MFA can be configured for both local and remote (LDAP) users. MFA works for GUI and CLI access.

For IBM Storage Virtualize version 8.5.0 and later, a multifactor authentication provider is required to enable multifactor authentication (MFA). Cloud-based authentication providers IBM Security Verify or Duo Security by Cisco can be used as an authentication service that can interface with the most commonly used second factor authentication methods. Examples of a second factor include a fingerprint that is used in the IBM Security Verify application installed on a smartphone, an OTP sent to the user via email or a text message, or a Timed One-Time Password (TOTP) capable authentication app.

Demonstration video: For a demonstration video, see IBM Storage Virtualize V8.6 Cisco DUO MFA.

A schematic flow of authentication for remote (LDAP) users is shown in Figure 12-11:

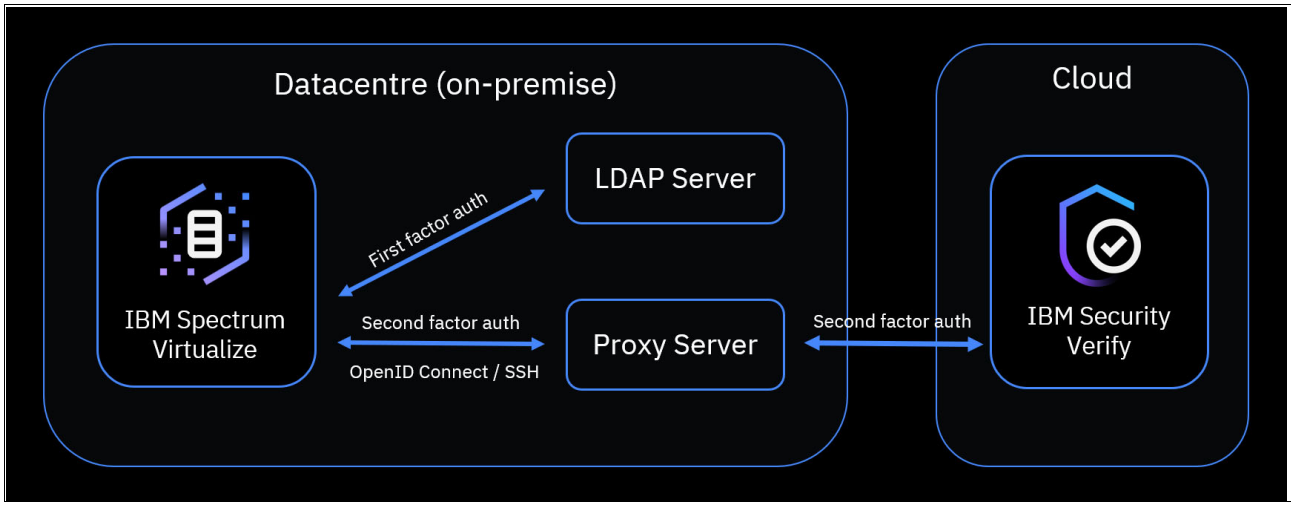


Figure 12-11 General flow of authentication

As shown in Figure 12-11, the initial authentication can be done locally at the local site by means of remote authentication like an on-premises LDAP directory infrastructure. The second factor authentication is done by using an outbound connection to IBM Security Verify cloud service, which connects to the configured user authenticator application by using the industry-standard OpenID Connect protocol.

12.4.1 Configuring Multifactor authentication

Multifactor Authentication requires DNS name resolution to be configured. If DNS is not configured, complete the following steps before configuring MFA:

1. Configure a DNS server to enable MFA. If a DNS is not yet configured, click the **Configure** link when you select the MFA window (Figure 12-12).

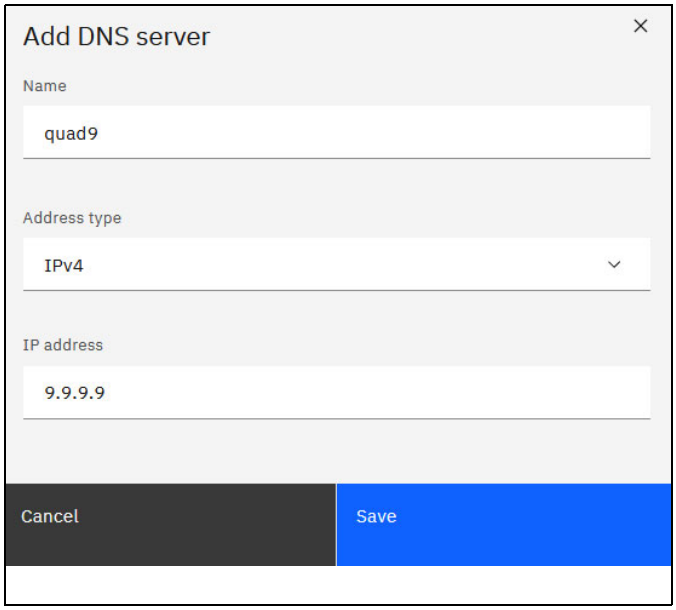


Figure 12-12 Add DNS server

2. If you require a proxy server to reach IP addresses external to your organization, configure the suitable proxy settings (Figure 12-13).

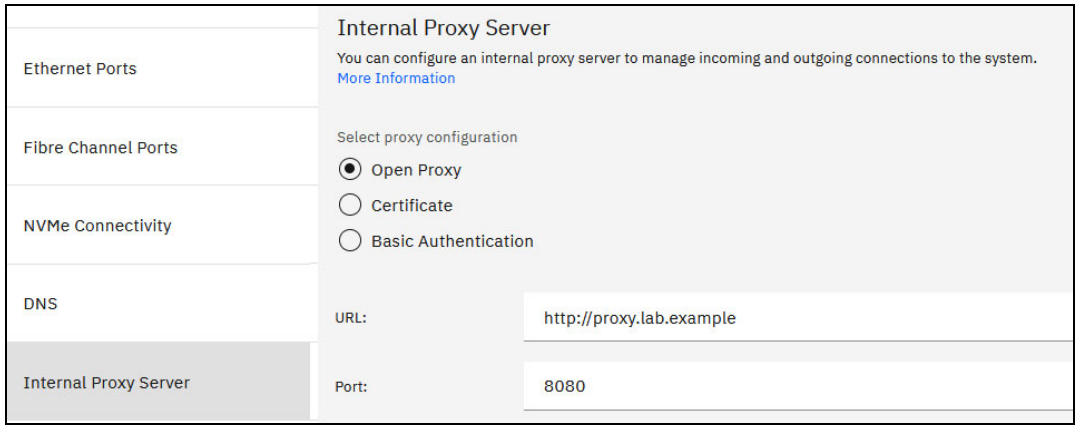


Figure 12-13 Internal Proxy Server configuration

It is assumed that you have purchased and set up your IBM Security Verify (ISV) instance (Figure 12-14 on page 1068). For more information, see IBM Security Verify (SaaS). A free trial is available at IBM Security Verify to test the service.

After these prerequisite steps are completed, configure MFA by completing the following steps:

1. Select **Settings** → **Security** → **Multifactor Authentication**, (Figure 12-14).

2. Export the system's TLS certificate by clicking the button and save the resulting file. Note down or copy the shown certificate name. You will need it later in step 5. The certificate name equals the system's 16-digit system ID. If needed, you can retrieve the system ID anytime from the GUI by selecting **Monitoring** → **System Hardware** → **System Actions** → **Properties** or by using the CLI command `lssystem`.

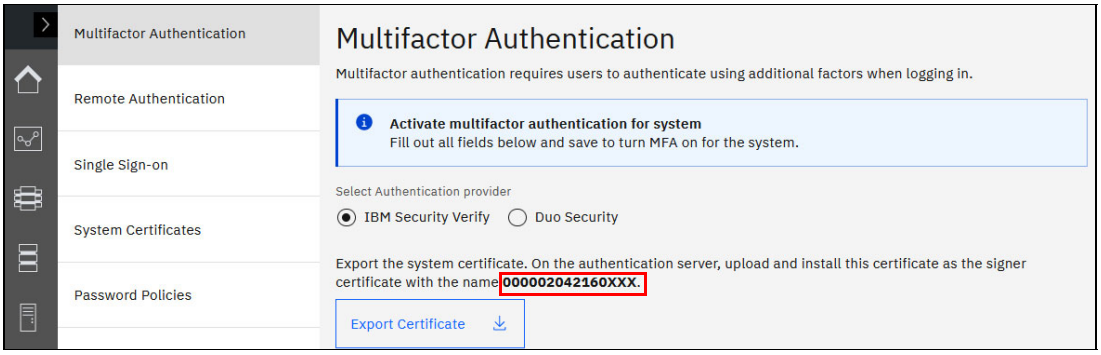


Figure 12-14 Exporting the unique certificate

3. Log on to your instance of IBM Security Verify and select **Security** → **Certificates** (Figure 12-15).

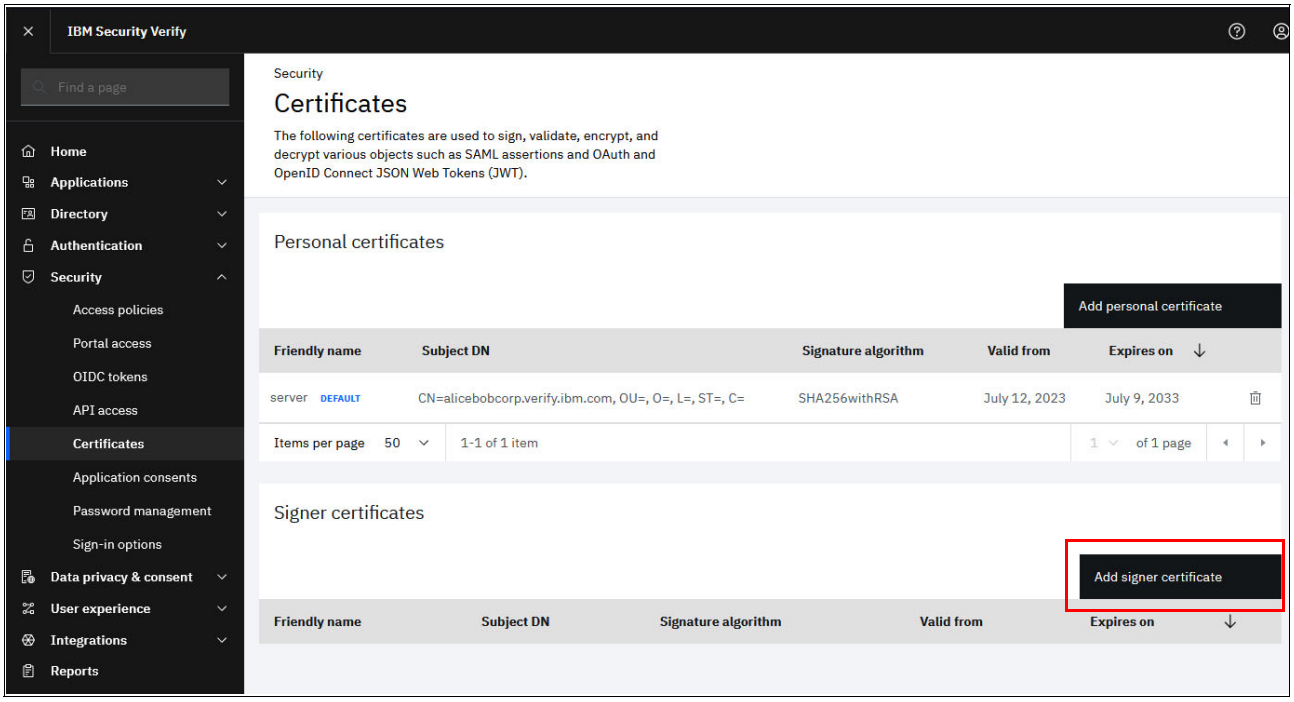


Figure 12-15 ISV Certificates

4. Select **Add Signer Certificate** (Figure 12-15).
5. Add the certificate file and enter the **Friendly name** (Figure 12-16 on page 1069), which is the system ID that was saved in Step 2. Otherwise, the MFA fails because of a signature verification error.

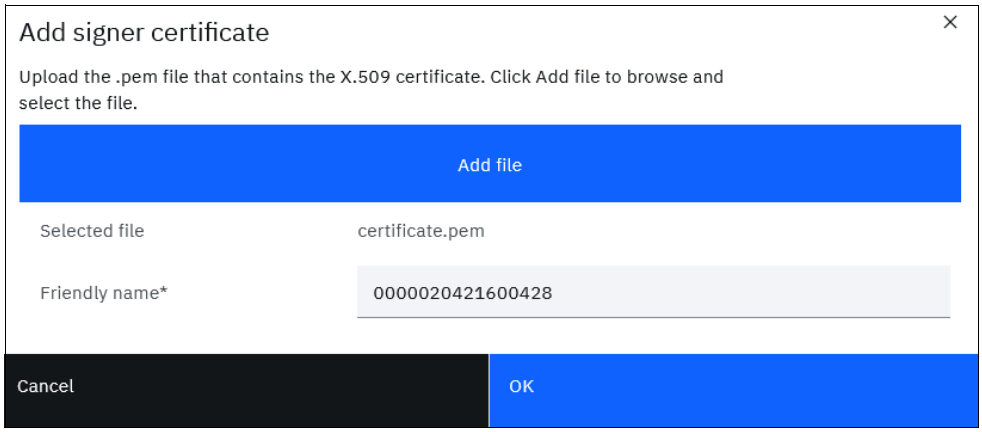


Figure 12-16 Add signer certificate

6. Return to the IBM Storage Virtualize GUI and fill in the fields as displayed (see Figure 12-17 on page 1070).

> **Note:** OpenID and API Client Credentials are generated when you configure the IBM Security Verify instance. Your security team should be able to provide the suitable Client ID and Client secret information.

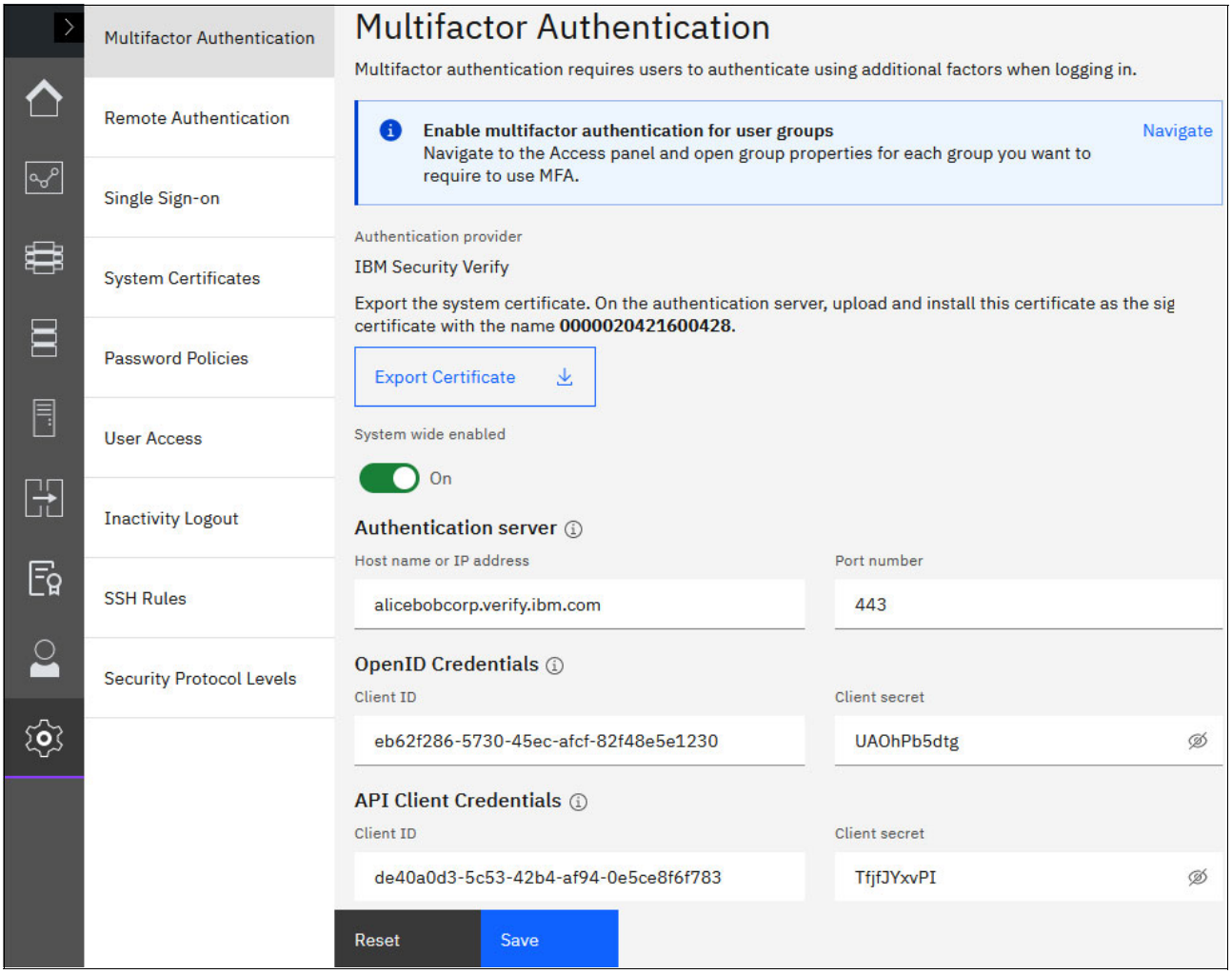


Figure 12-17 MFA Configuration Details showing sample credentials

7. Click **Save** and **Confirm** to enable MFA (Figure 12-18).

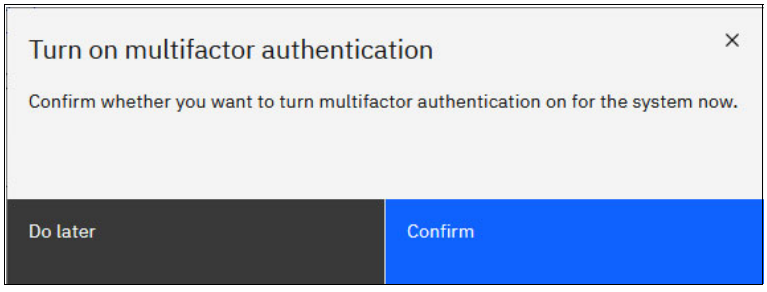


Figure 12-18 Confirm Turn on multifactor authentication window

Return to **Settings** → **Security** → **MFA** to verify that MFA was enabled system-wide.

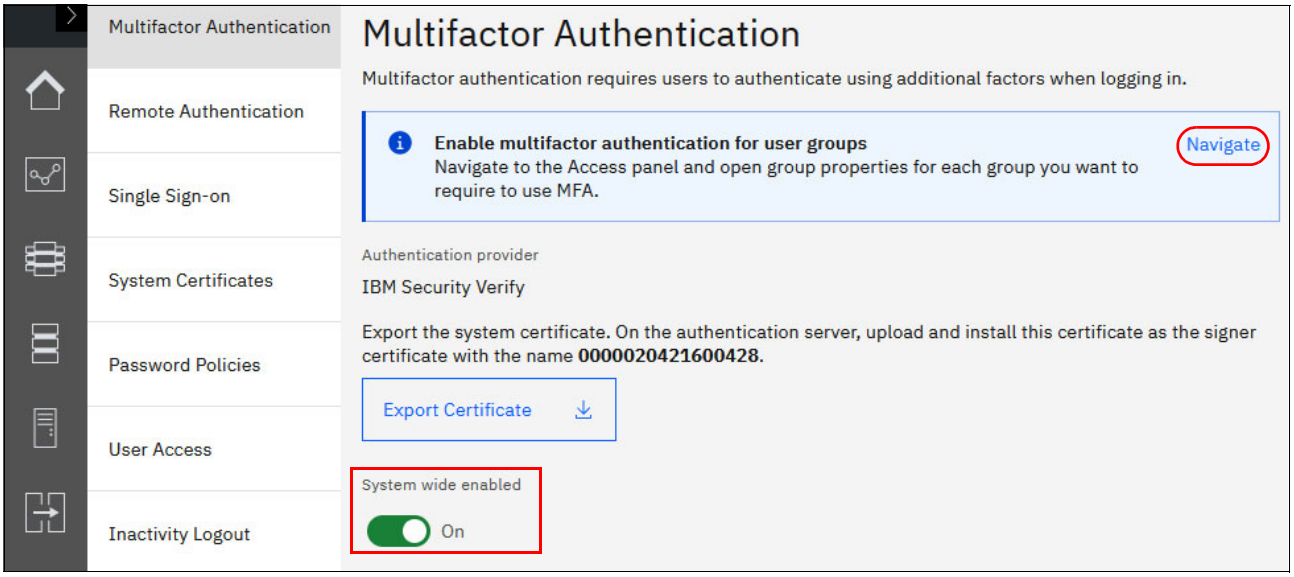


Figure 12-19 MFA enabled system wide, navigate to user groups settings

8. The final configuration step requires you to create suitable users and user groups that feature MFA enabled. A link is provided in the message area. Click **Navigate** to take you directly to the configure user groups window (Figure 12-19).

   In the **Access** → **Users by Group** window, select **Create User Group**.

9. In the **Create User Group** window, select **On** in the Multifactor Authentication section to enable MFA for the user group (Figure 12-20).

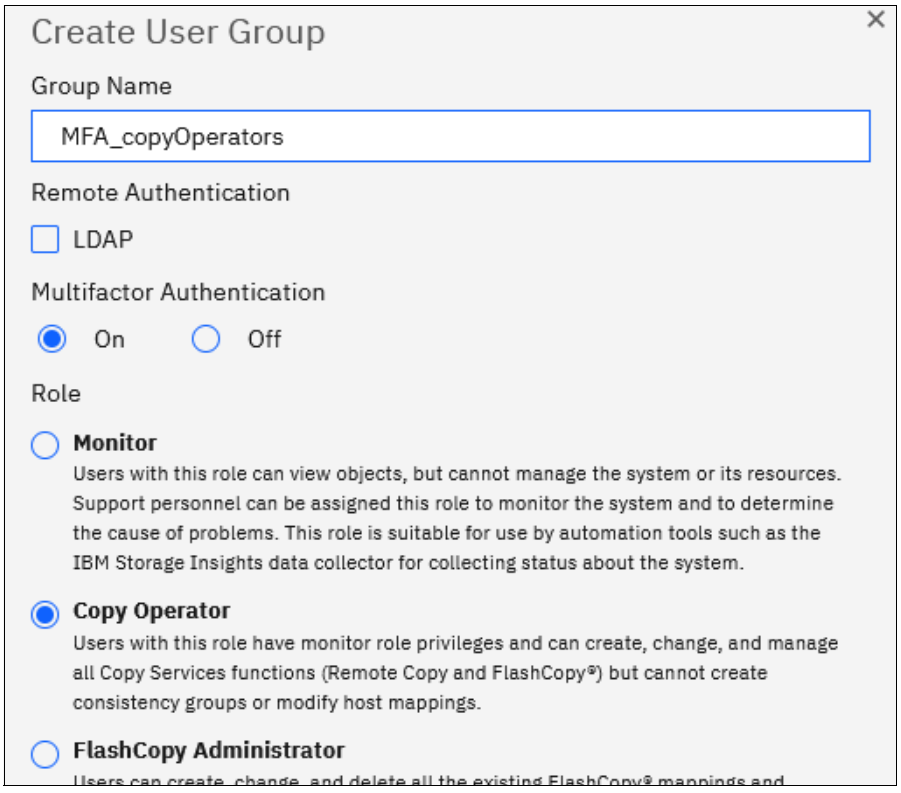


Figure 12-20 Create User Group window

10. To add a user to this MFA enabled group, select **Create User** and assign the user to the MFA enabled group that you created in Step 9 (Figure 12-21).

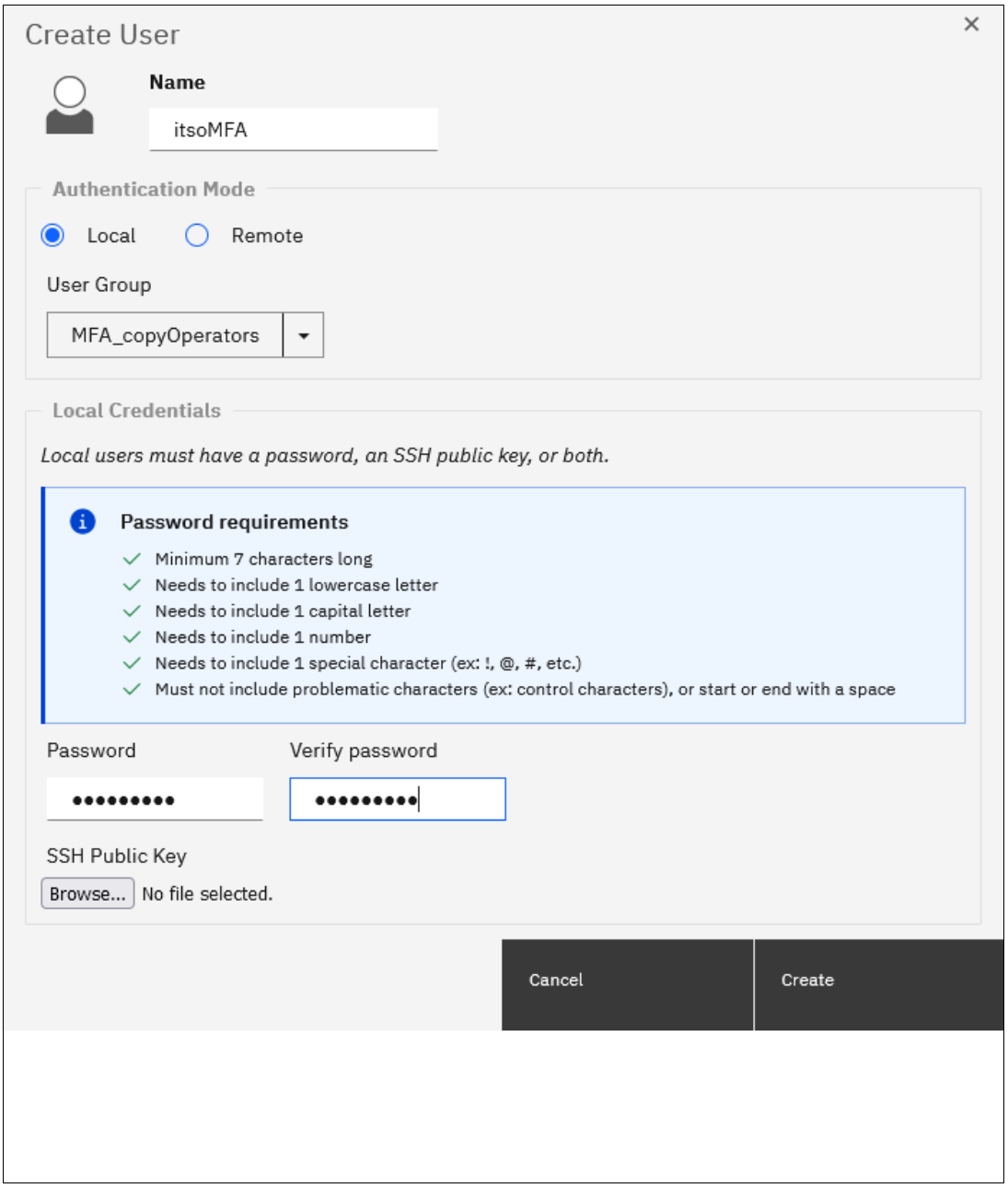


Figure 12-21 Create User window

Example of logging in by using MFA

The login process for a user that is member of an MFA-enabled includes a prompt for the second factor in both the GUI and CLI interfaces. The login flow by using the GUI is described next. The IBM FlashSystem login window is shown in Figure 12-22.

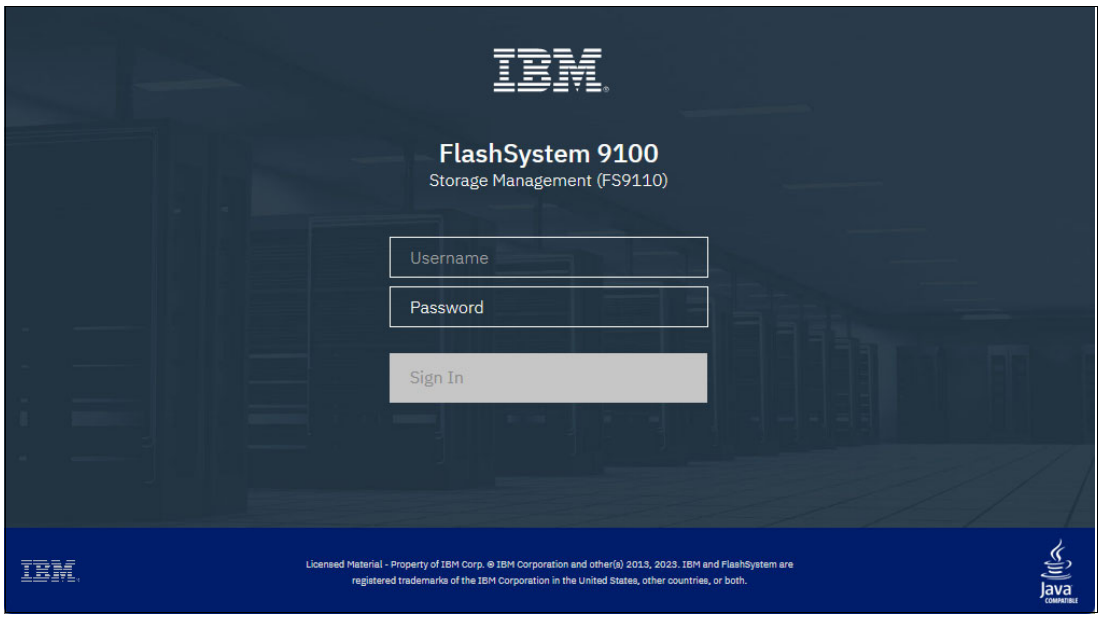


Figure 12-22 IBM FlashSystem GUI login screen

Complete the following steps to log in by using MFA:

1. Sign in with your user MFA-enabled useridID and password (see Figure 12-23).

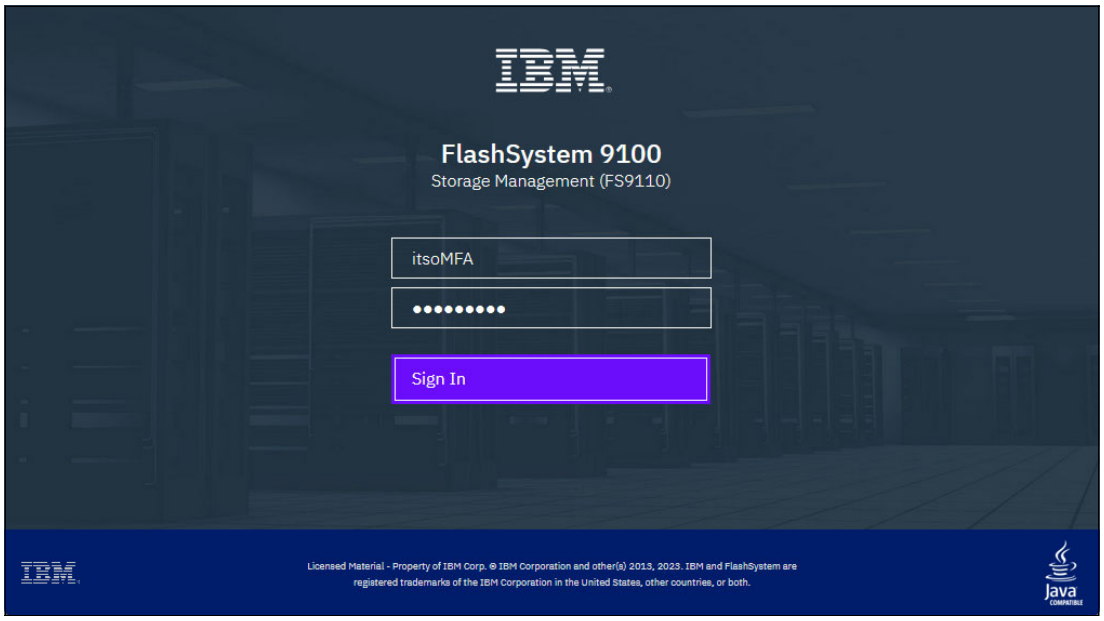


Figure 12-23 Signing in by using user ID and password

2. In the next window, you are prompted to choose a method for authentication. For information how to manage second factors see Managing your IBM Verify Authenticators. In this example, we select the **IBM Verify app Touch Approval** method. Click **Send Push** when you are done (Figure 12-24).

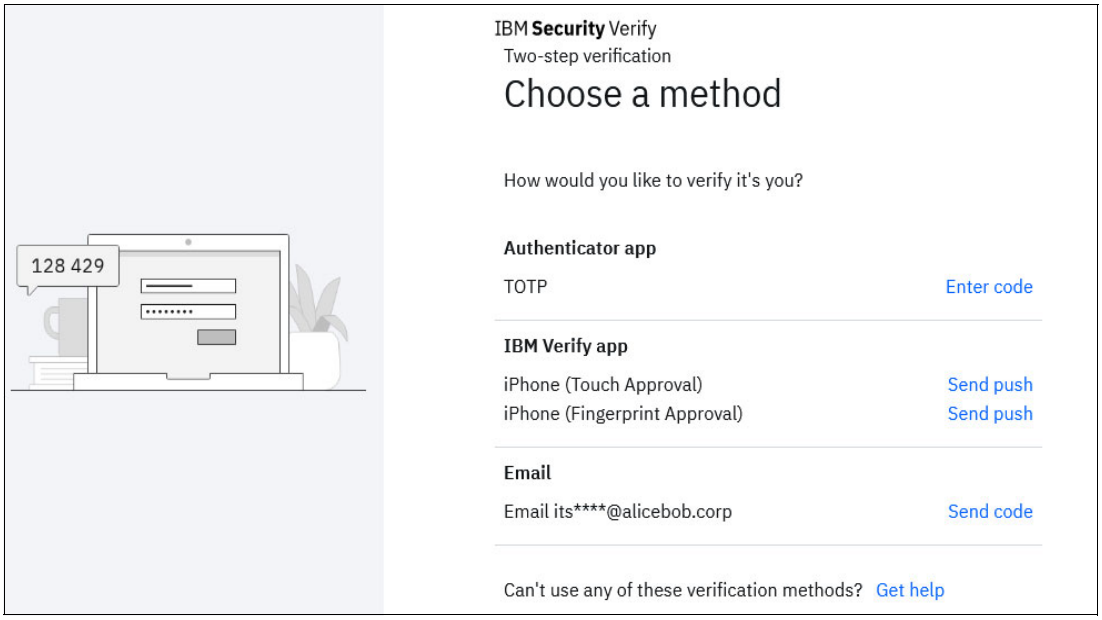


Figure 12-24 Choose a two-step verification method for authentication

The pending authentication message that is shown in Figure 12-25 is displayed.

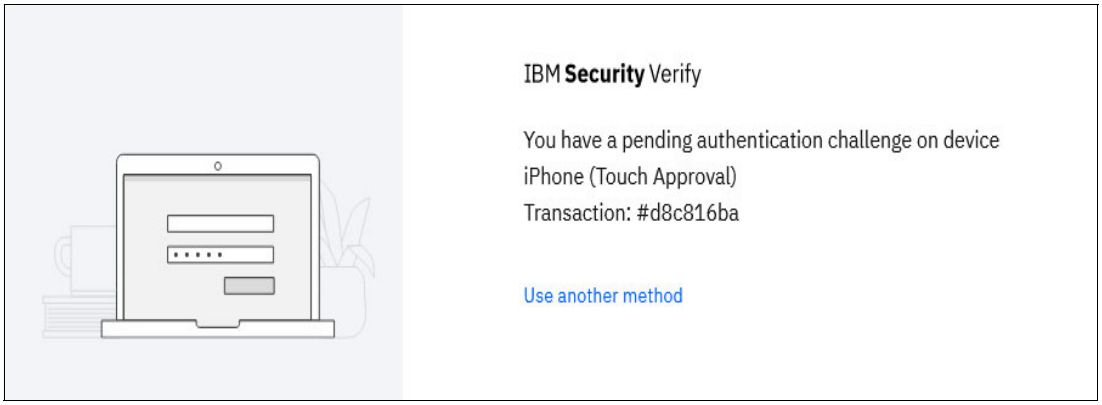


Figure 12-25 Touch Approval pending authentication challenge message

Figure 12-26 shows the verification question. Click **Approve**.

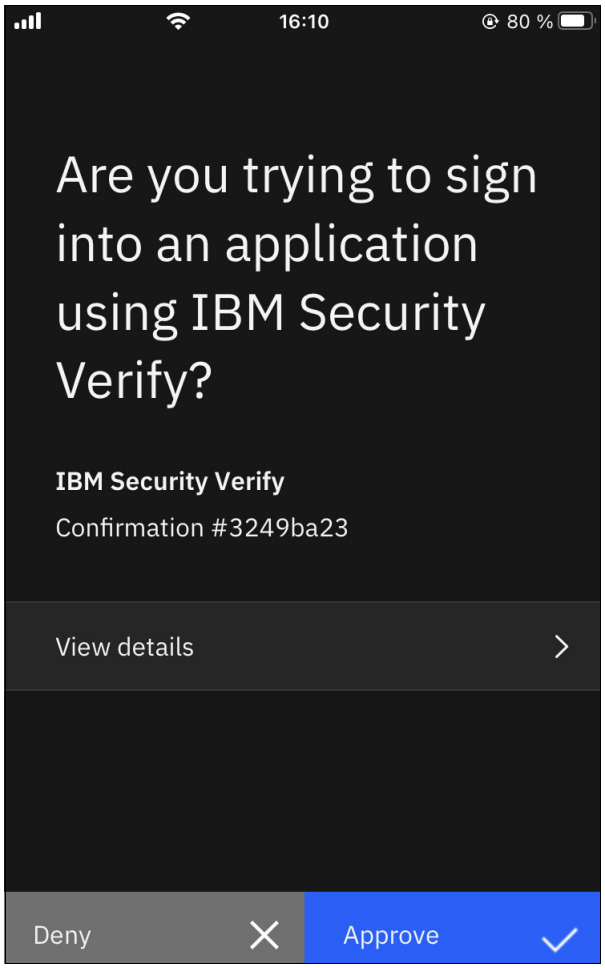


Figure 12-26 Verification request in IBM Verify app on mobile phone

The message that is displayed that indicates the request is verified is shown in Figure 12-27.

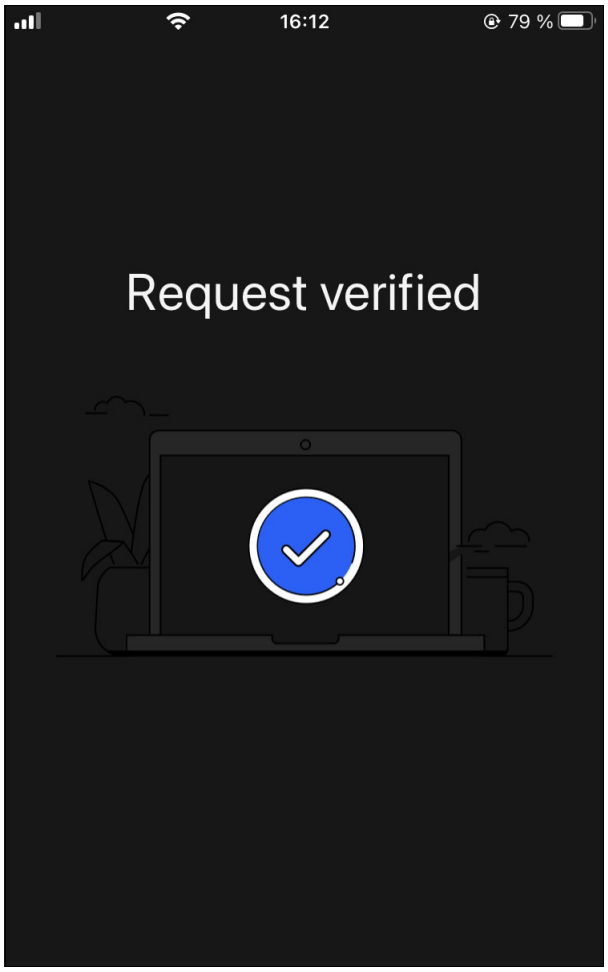


Figure 12-27 Request verified in IBM Verify app

On the CLI, you are prompted to select the second factor device or method. An example after successful completion is shown in Example 12-6:

Example 12-6 MFA prompt at CLI login

```
login as: itsoMFA
Keyboard-interactive authentication prompts from server:
| Password: (*********)
| 1) TOTP
| 2) itsomfa@alicebob.corp
| ? 1
| Enter OTP -123456
End of keyboard-interactive prompts from server
IBM_FlashSystem:FS9110:itsoMFA>
```

12.4.2 User authentication using Single Sign-on

Single Sign-on (SSO) authentication enables users to log in with a single, centrally managed user ID to various, independent systems. After a user has authenticated with the SSO provider, the user can login to other SSO-enabled IBM Storage Virtualize systems without entering their credentials again. Depending on the security policies and the effective configuration in the SSO provider, it might be required to enter a second factor each time the user logs in into another device.

SSO Authentication was introduced first with IBM Storage Virtualize version 8.5.0. Several SSO service providers are supported, such as IBM Security Verify and Microsoft Active Directory Federation Services (ADFS). The list of supported SSO services is updated regularly and documented in IBM Documentation article Supported authentication providers.

As with MFA, SSO requires DNS name resolution to be configured and working. If no DNS server is configured, click the **Configure** link in the blue box to navigate to the DNS server configuration panel (Figure 12-28). DNS server configuration is described in chapter 12.4.1, “Configuring Multifactor authentication” on page 1067.

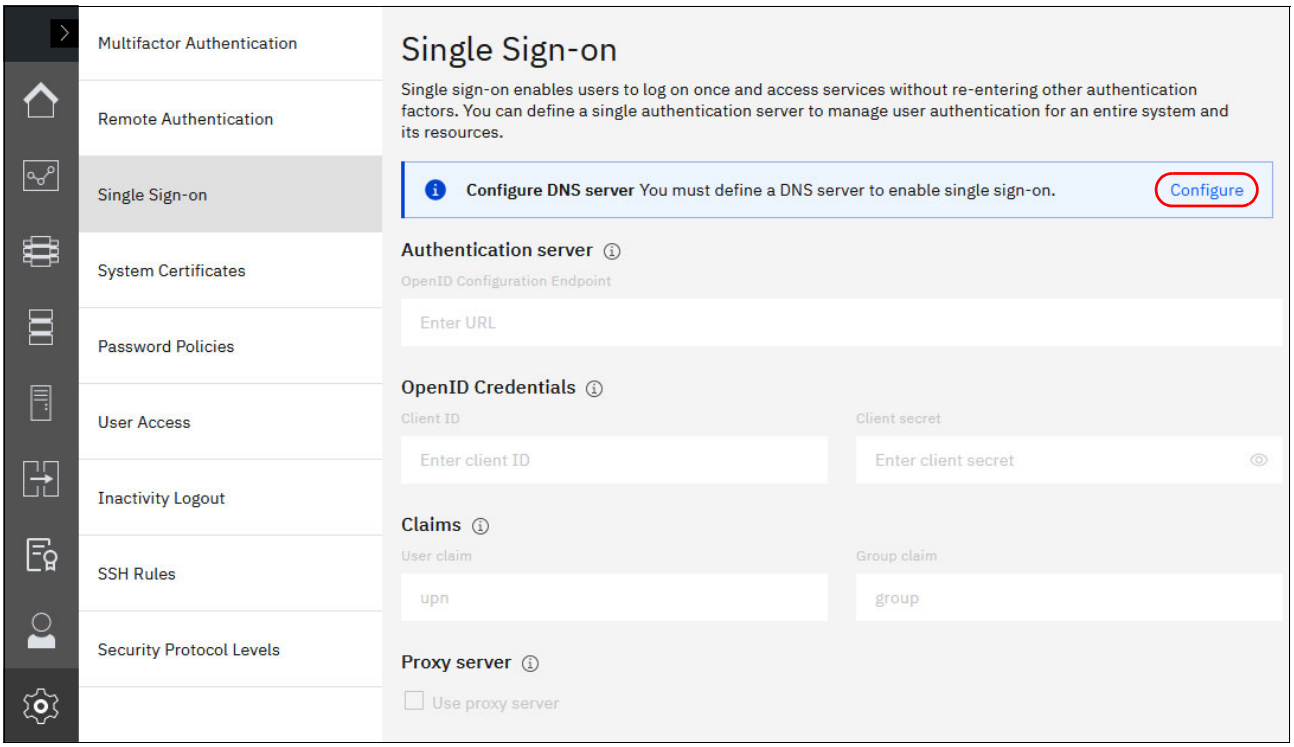


Figure 12-28 Single Sign-on configuration

Similarly, you might need to define an Internal Proxy Server. Select the **Internal Proxy Server** tab from the **Settings** → **Network** window and enter your proxy information.

It is assumed you purchased and set up your IBM Security Verify instance. For more information, see IBM Security Verify (SaaS).

Complete the following steps to configure IBM Security Verify for SSO. You can find further details on these steps in the GUI-internal help pages:

1. Navigate to **Settings** → **Security** → **Single Sign-on**.
2. Enter data in all the fields as shown in Figure 12-29 on page 1078:

– **Authentication server** → **OpenID Configuration Endpoint**. The endpoint's URL according to your SSO provider, for ISV this can be, for example, `https://<your_tenant_name>.verify.ibm.com/oidc/endpoint/default/.well-known/openid-configuration`
– OpenID Credentials. OpenID **Client ID** and **Client secret** were generated when you configured the Application object in IBM Security Verify for the Storage Virtualize system.
– Claims:
  • **User claim**. The username attribute according to your SSO provider. For ISV set it to `preferred_username`.
  • **Group claim**. The group attribute according to your SSO provider. For ISV set it to `groupIds`.
– Select the **Proxy server** control box if access to external addresses goes through a proxy server in your environment.

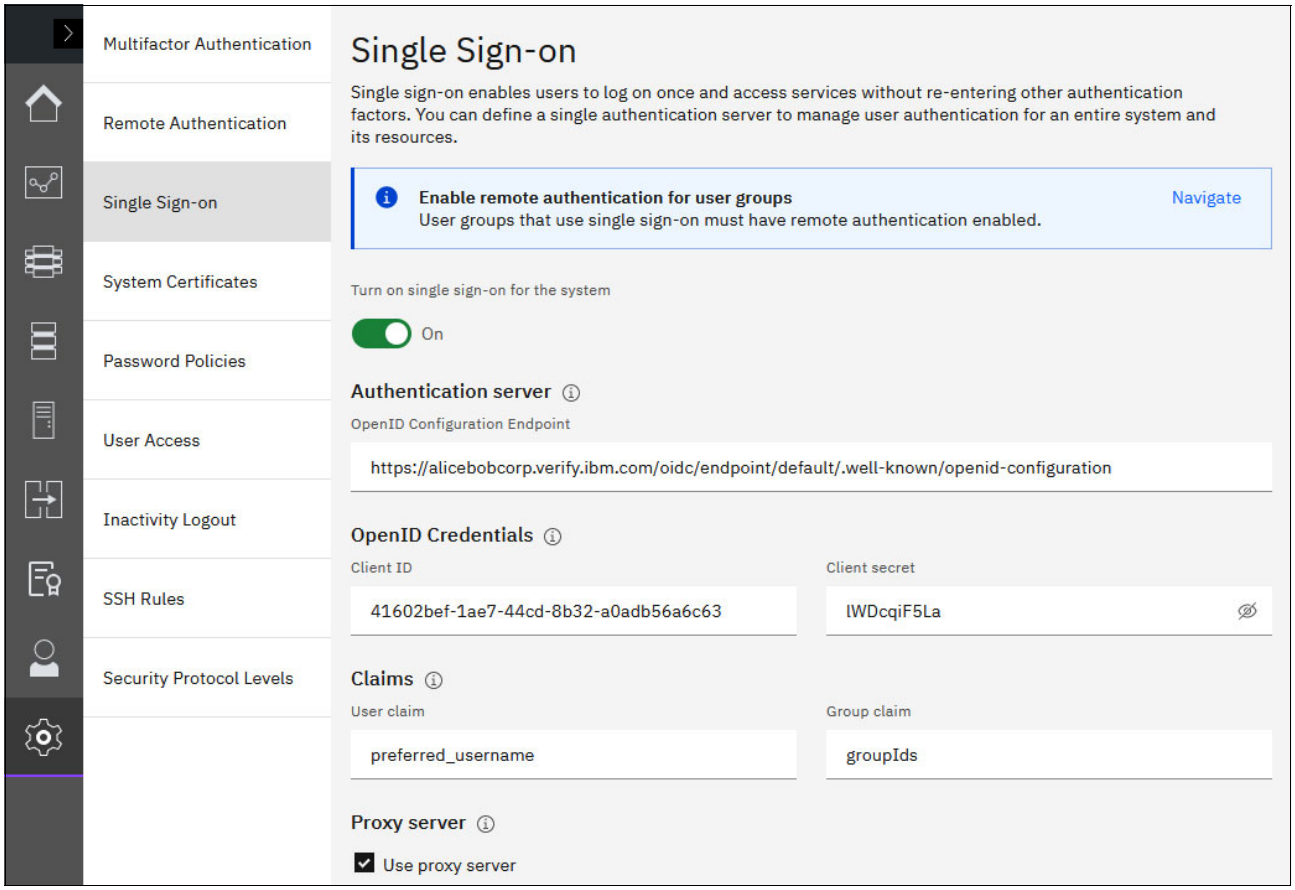


Figure 12-29 Complete all fields as displayed

3. Click **Save**. You are prompted to enable SSO (see Figure 12-30).



Figure 12-30 Enable single sign-on

When enabled, you see the system-wide enabled status. However, to complete the configuration, you must create user groups that are configured to use SSO. The SSO window shows a link to the suitable window.

Complete the following steps:

1. Select **Access** → **Users by Group** and then click **Create User Group**.
2. Complete the Create User Group window fields and then select the enable SSO option, as shown in Figure 12-31.

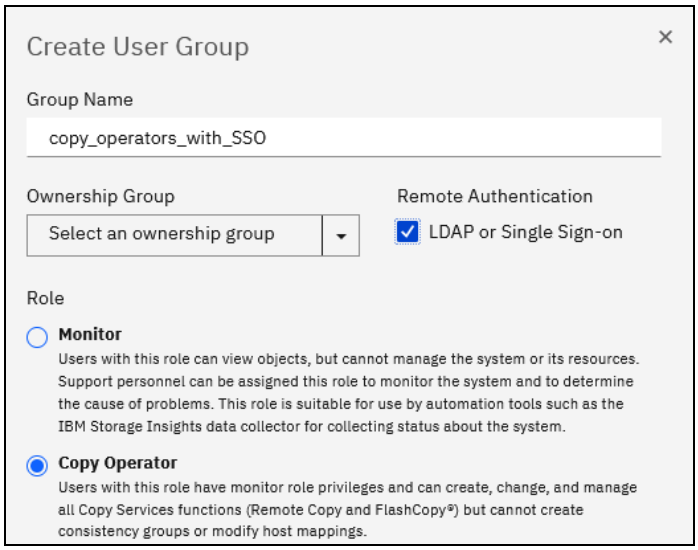


Figure 12-31 Create User Group for SSO login

SSO passes all authentication factors to the remote authentication service, so you do not need to create any local users on the IBM Storage Virtualize system. The remaining setup is performed in IBM Security Verify.

3. Log in to your IBM Security Verify instance and select the Directory window (Figure 12-32).

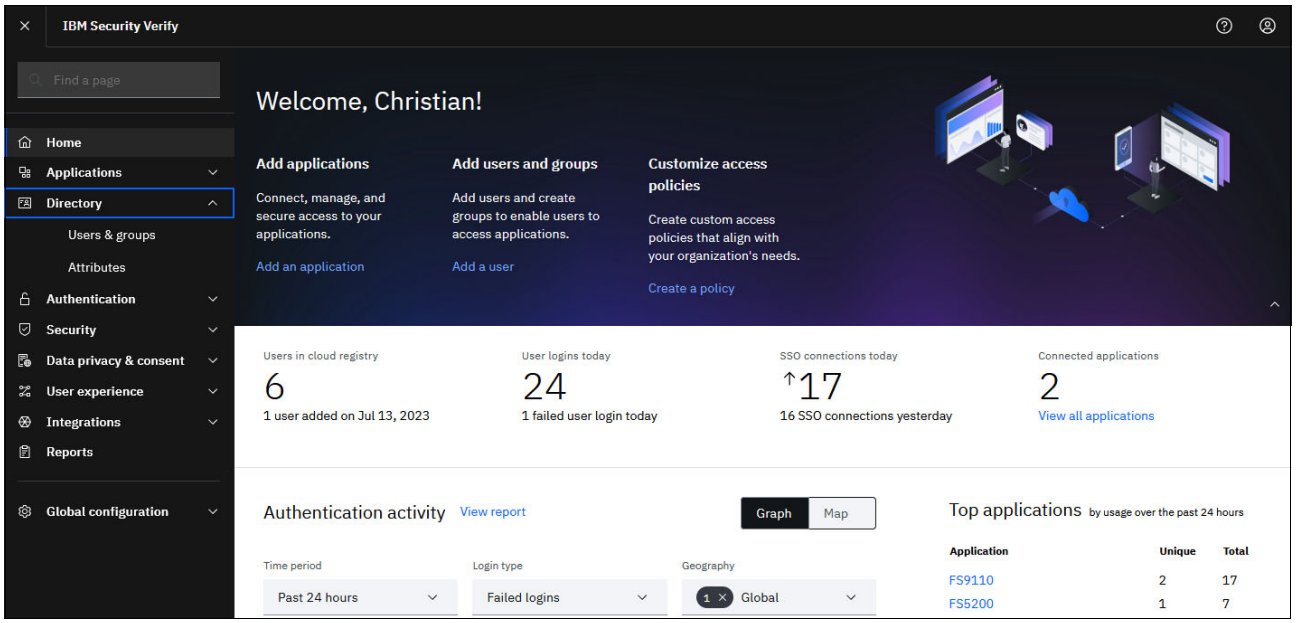


Figure 12-32 Select User Directory

4. Select the **Users and Groups** window (Figure 12-33).

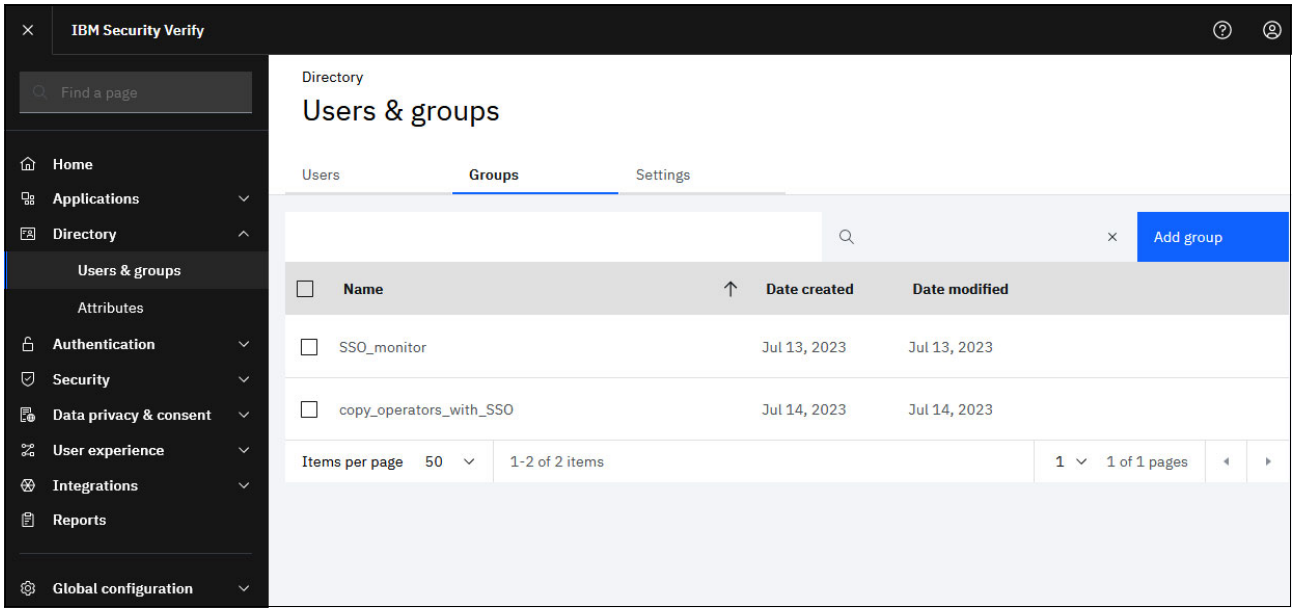


Figure 12-33 Users and groups window

5. Click **Add Group** and enter exactly the same User group name that was created in the Storage Virtualize GUI. Enter a short description for future reference.
6. Add one or more users by searching for their name. You also can select whether you want to email the user to notify them of the access you granted (Figure 12-34 on page 1081).

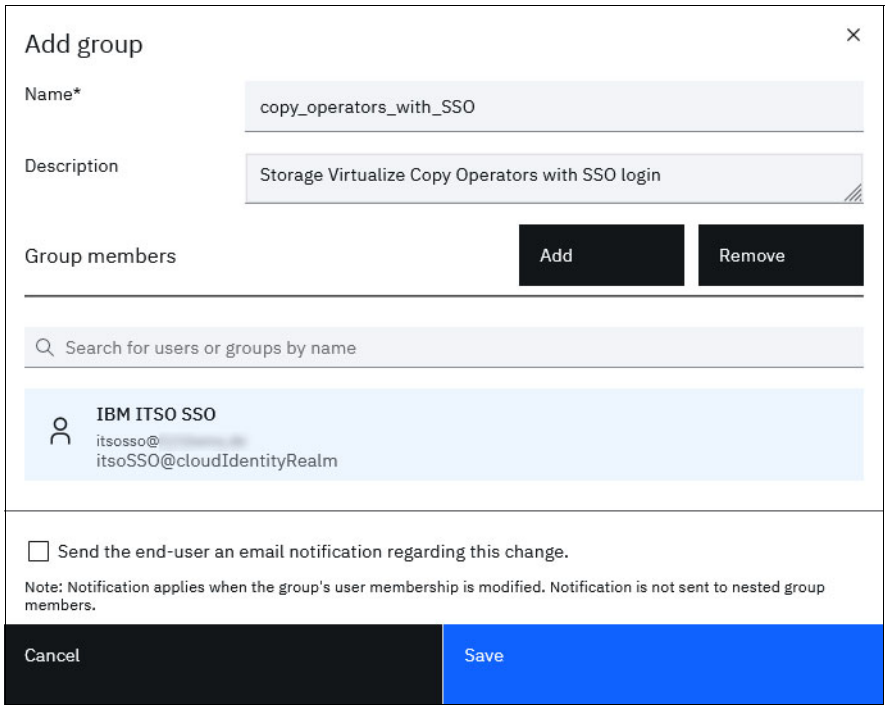


Figure 12-34 ISV Add group window

Example logging in by using single sign-on

After the configuration process is complete, the login window for the IBM Storage Virtualize GUI shows the **Sign In with SSO** button (see Figure 12-35):



Figure 12-35 Sign In with SSO

You will be redirected to your SSO provider to login with your credentials and supply a second factor if required. Figure 12-36 on page 1082 and Figure 12-37 on page 1082 show verification by using ISV.

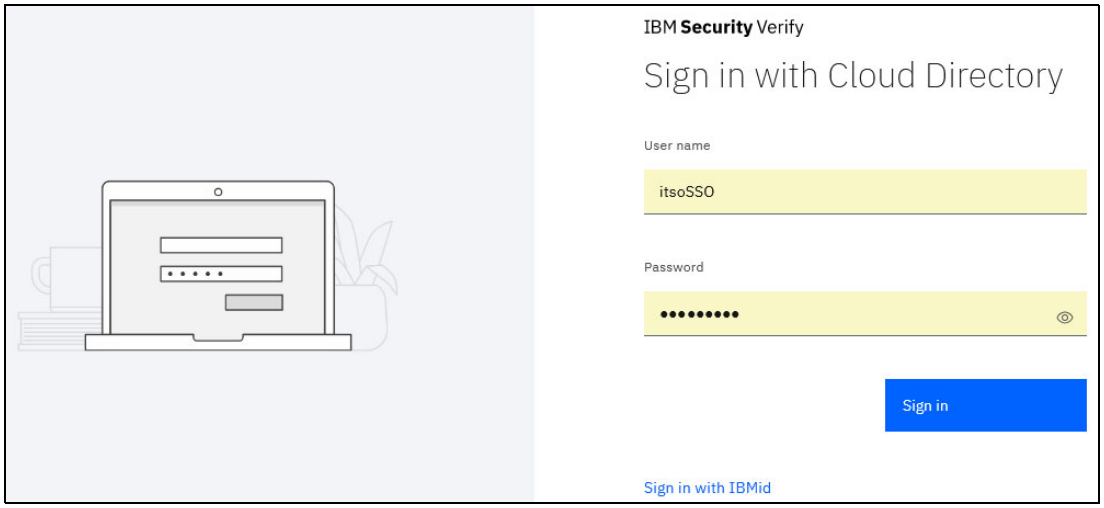


Figure 12-36 SSO provider login screen

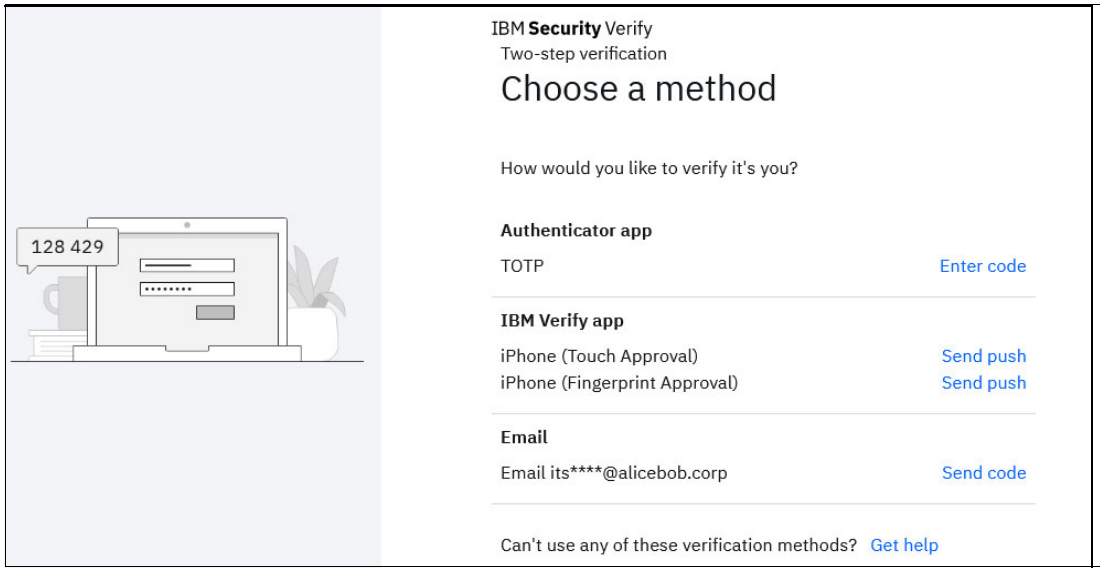


Figure 12-37 Choosing a second factor verification method

In this example, the **Authenticator app** method with Time-based One-Time Password (**TOTP**) is selected as shown in Figure 12-38.



Figure 12-38 Enter the TOTP

After a successful login, the authenticated user's name along with his effective group membership will be displayed in the GUI as usual (Figure 12-39).

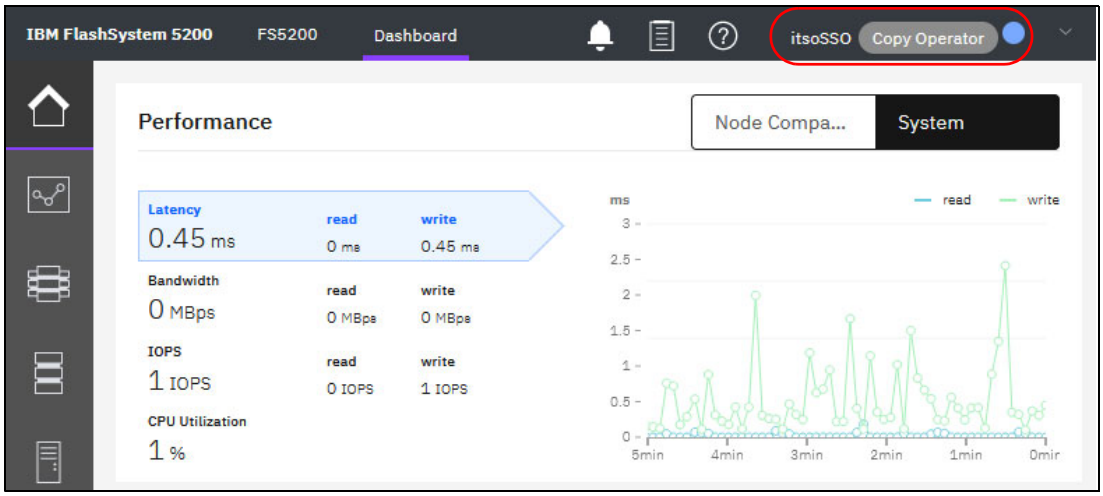


Figure 12-39 SSO login successfully completed

12.5 Ownership groups principles of operations

Ownership groups enable the allocation of storage resources to several independent tenants with the assurance that one tenant cannot access resources that are associated with another tenant.

Ownership groups restrict access for users in the ownership group to only those objects that are defined within that ownership group. An owned object can belong to one ownership group.

Users in an ownership group are restricted to viewing and managing objects within their ownership group only. Users that are not in an ownership group can continue to view or manage all the objects on the system based on their defined user role, including objects within ownership groups.

Only users with Security Administrator roles can configure and manage ownership groups.

The system supports several resources that you assign to ownership groups:

- Child pools
- Volumes
- Volume groups
- Hosts
- Host clusters
- Host mappings
- FlashCopy mappings
- FlashCopy consistency groups
- User groups
- Portsets

An owned object can belong to only one ownership group. An owner is a user with an ownership group that can view and manipulate objects within that group.

Before you create ownership groups and assign resources and users, review the following guidelines:

- Users can be in only one ownership group at a time. This applies to local and remotely authenticated users.
- Objects can be within at most one ownership group.
- Global resources, such as drives, enclosures, and arrays, cannot be assigned to ownership groups.
- If a user role permits it, a global user that does not belong to an ownership group can view and manage any resources on the system, including the resources that belong to an ownership group, and including users within an ownership group.
- Users within an ownership group cannot have the Security Administrator role. All Security Administrator role users are global users.
- Users within an ownership group can view or change resources within the ownership group in which they belong.
- Users within an ownership group cannot change any objects that are outside of their ownership group. This restriction includes global resources that are related to resources within the ownership group. For example, a user can change a volume in the ownership group, but not the drive that provides the storage for that volume.
- Users within an ownership group cannot view or change resources if those resources are assigned to another ownership group or are not assigned to any ownership group. However, users within ownership groups can view and display global resources. For example, users can display information about drives on the system because drives are a global resource that cannot be assigned to any ownership group.

When a user group is assigned to an ownership group, the users in that user group retain their role, but are restricted to only those resources that belong to the same ownership group. The role that is associated with a user group can define the permitted operations on the system, and the ownership group can further limit access to individual resources.

For example, you can configure a user group with the Copy Operator role, which limits user access to FlashCopy operations. Access to individual resources, such as a specific FlashCopy consistency group, can be further restricted by assigning it to an ownership group.

A child pool is a key requirement for the ownership groups feature. By defining a child pool and assigning it to an ownership group, the system administrator provides capacity for volumes that ownership group users can create or manage.

Depending on the type of resource, the owning group for the resource can be defined specifically or inherited from defined objects. For example, a child pool needs an ownership group parameter to be set by a system administrator; however, volumes that are created in that child pool automatically inherit the ownership group from a child pool.

For more information about ownership inheritance, see “Ownership groups.

When the user logs on to the management GUI or command line interface (CLI), only resources that they can access through the ownership group are available. Also, only events and commands that are related to the ownership group in which a user belongs are viewable by those users.

12.5.1 Implementing ownership groups on a new system

This section describes ownership group implementation process for a new system that has no volumes and users that must be migrated to ownership groups.

Creating an ownership group

To create the first ownership group, select **Access** → **Ownership Groups**, as shown in Figure 12-40. Enter a name for the ownership group, and the, click **Create Ownership Group**.

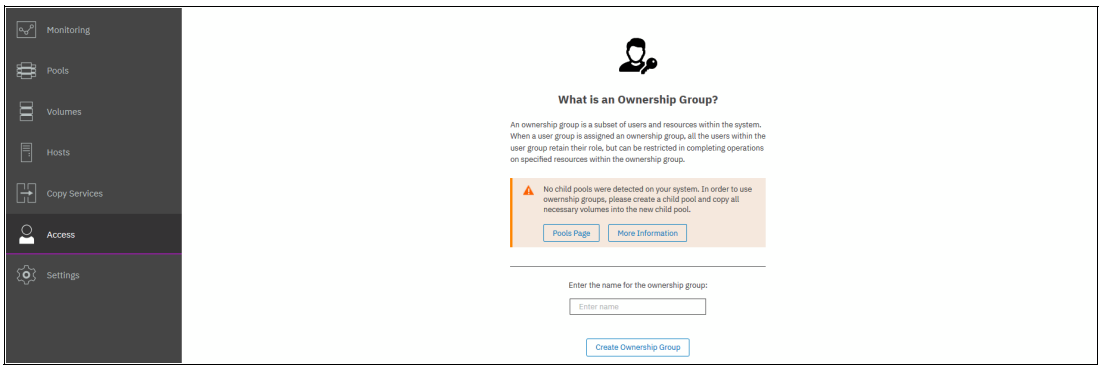


Figure 12-40 Creating the first ownership group

After the first group is created, the window changes to ownership group mode, as shown in Figure 12-41. The new ownership group has no user groups and no resources that are assigned to it.

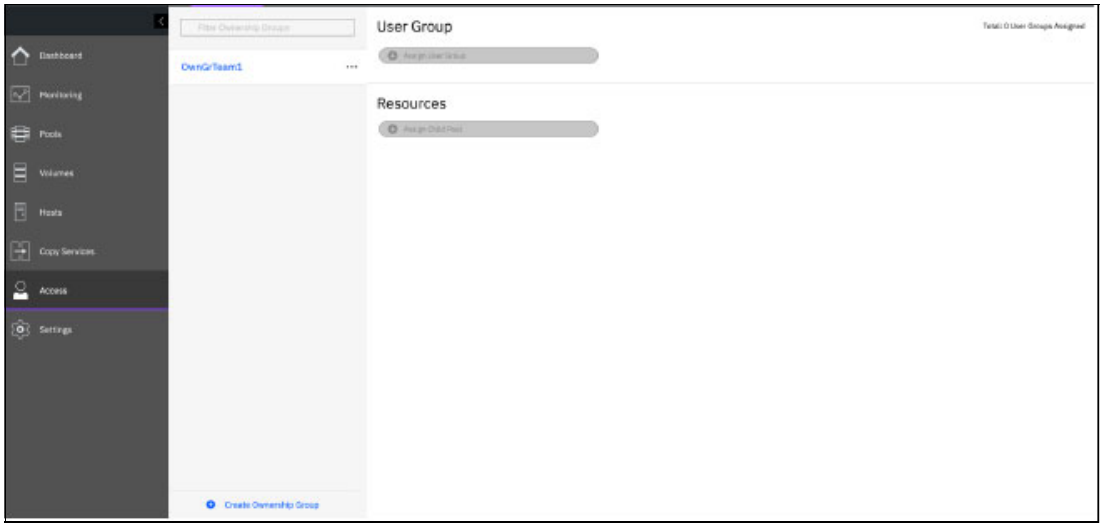


Figure 12-41 Ownership groups management window

To create more ownership groups, click **Create Ownership Group.**

Assigning users to an ownership group

To create accounts for users that use ownership group resources, the user group must be created and assigned to the ownership group.

To create a user group, select **Access** → **Users by Group** and then click **Create User Group**. The Create User Group window opens, as shown in Figure 12-42 on page 1086.

Specify the User Group name, select an ownership group to tie this user group to, and select a role for the users in this group.

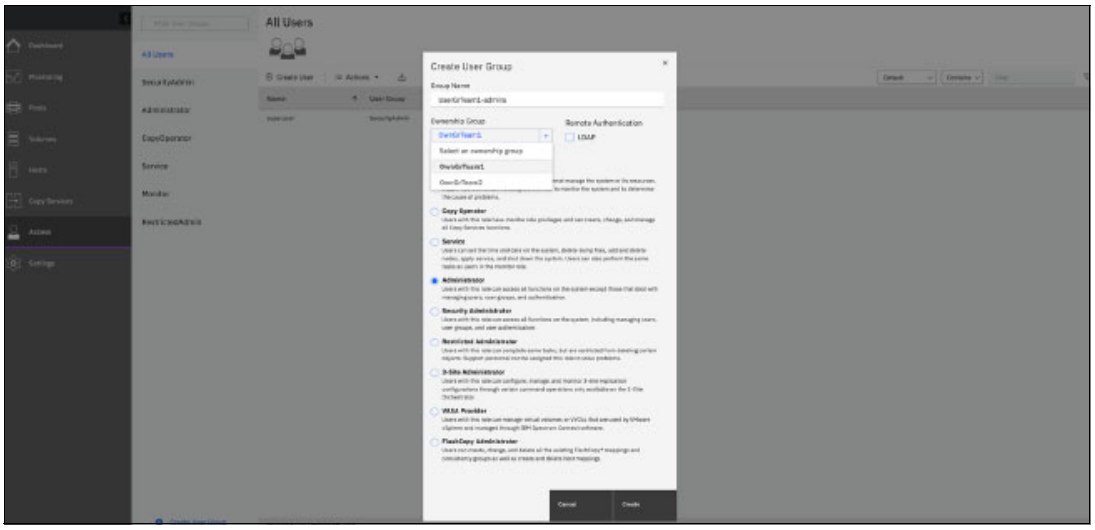


Figure 12-42 Creating and assigning a user group

For more information about user roles, see Ownership groups.

To create volume, host, and other objects in an ownership group, users must have an Administrator or Restricted Administrator role. Users with the Security Administrator role cannot be assigned to an ownership group.

You can also set up a user group to use remote authentication, if it is enabled. To do so, select the **Lightweight Directory Access Protocol (LDAP)** option.

Note: Users that use LDAP can belong to multiple user groups, but belong to only one ownership group that is associated with one of the user groups.

If remote authentication is not configured, you must create a user (or users) and assign it to a created user group, as shown in Figure 12-43.

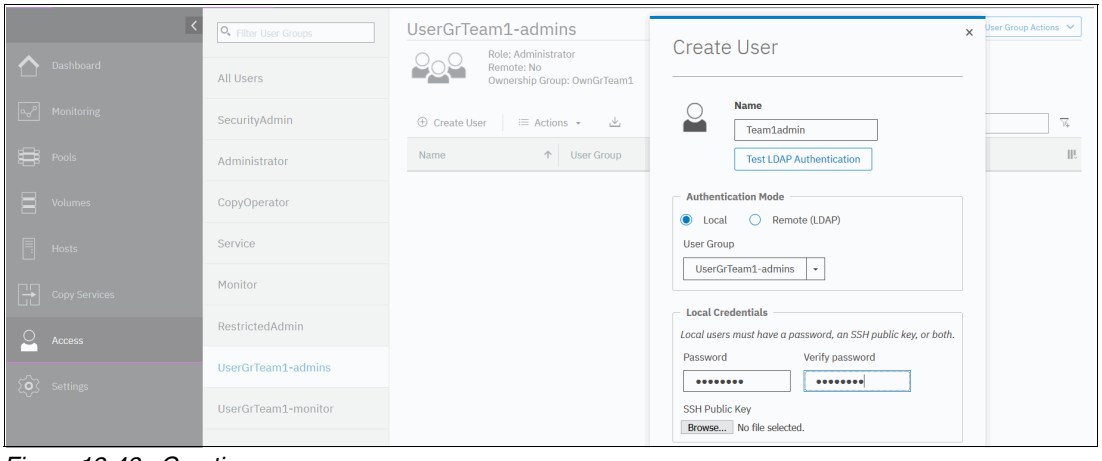


Figure 12-43 Creating a user

You can manage user groups that are assigned to an ownership group by selecting **Access** → **Ownership Groups**, as shown in Figure 12-44. To assign a user group that exists but is not assigned to any ownership group, click **Assign User Group**. To unassign user group, click the **"..."** icon that is next to the assigned user group name.

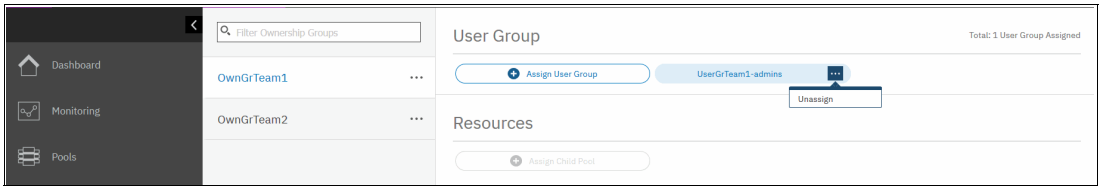


Figure 12-44 Unassigning user groups

Multiple user groups with different user roles can be assigned to one ownership group. For example, you can create and assign a user group with the Monitor role in addition to a group with the Administrator role to have two sets of users with different privilege levels accessing an ownership group's resources.

Creating ownership group resources

To create ownership group volumes and other resources, a child pool must be created and assigned to the ownership group.

To create a child pool, select **Pools** → **Pools**, right-click a parent pool that is designated as a container for child pools and then click **Create Child Pool**, as shown in Figure 12-45.

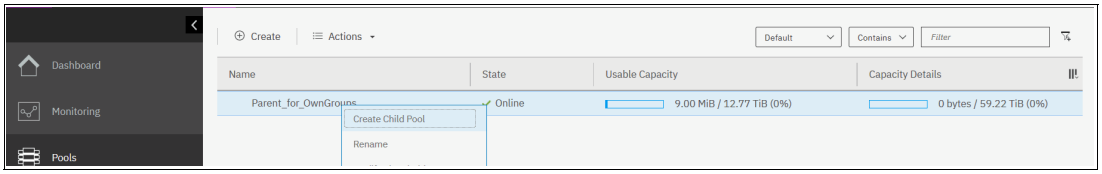


Figure 12-45 Creating a child pool

When creating a child pool, specify an ownership group for it and assign a part of the parent's pool capacity, as shown in Figure 12-46. Ownership group objects can use only capacity that is provisioned for them with the child pool.

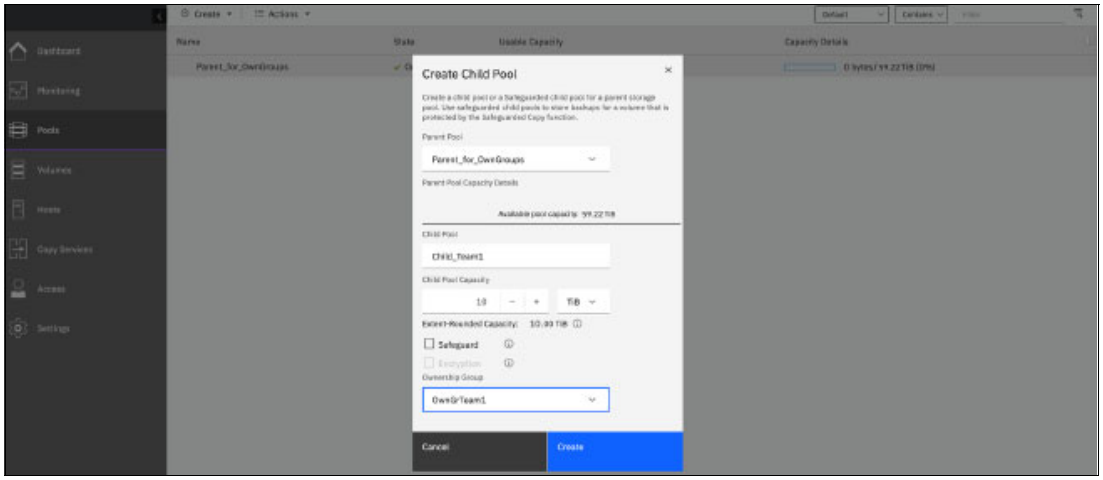


Figure 12-46 Creating a child pool and assigning it to an ownership group

Multiple child pools that are created from the same or different parent pools can be assigned to a single ownership group.

After a child pool is created and assigned, the ownership group management window (which you open by selecting **Access** → **Ownership Groups**) changes to show the assigned and available resources, as shown in Figure 12-47.

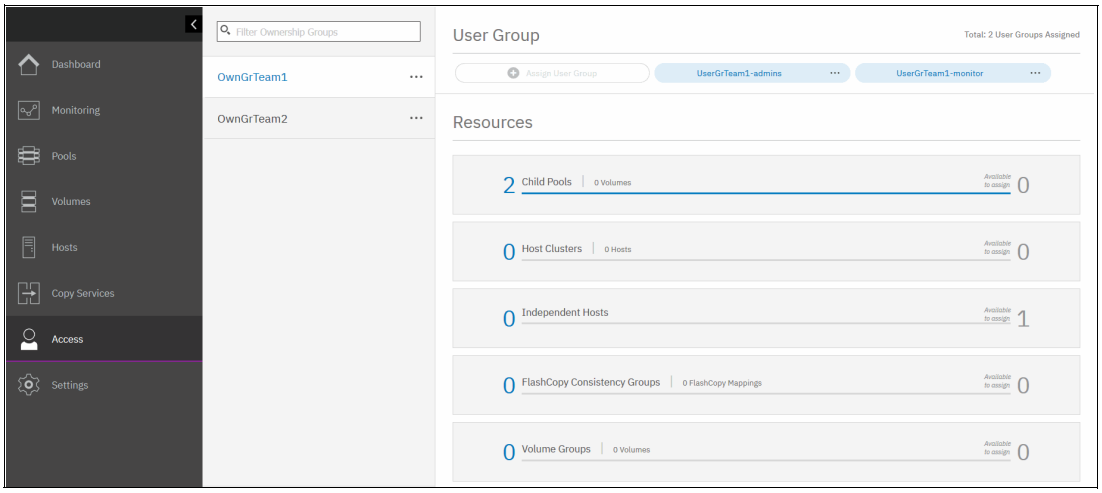


Figure 12-47 Ownership group management window

Any volumes that are created on a child pool that is assigned to an ownership group inherits ownership from the child pool.

After a child pool and user group are assigned to an ownership group, ownership group administrators can log in with their credentials and start creating volumes, host and host clusters, or FlashCopy mappings.

For more information about creating those objects, see:

- Chapter 6, “Volumes” on page 419
- Chapter 7, “Hosts” on page 517
- Chapter 10, “Advanced Copy Services” on page 717

Although an ownership group administrator can create objects only within the resources that are assigned to them, the system administrator can create, monitor, and assign objects for any ownership group.

Listing ownership group resources

By default, the Ownership Group attribute is not enabled in the GUI windows that list volumes and other objects that can be owned. For convenience, the system administrator can enable this attribute, as shown in Figure 12-48.

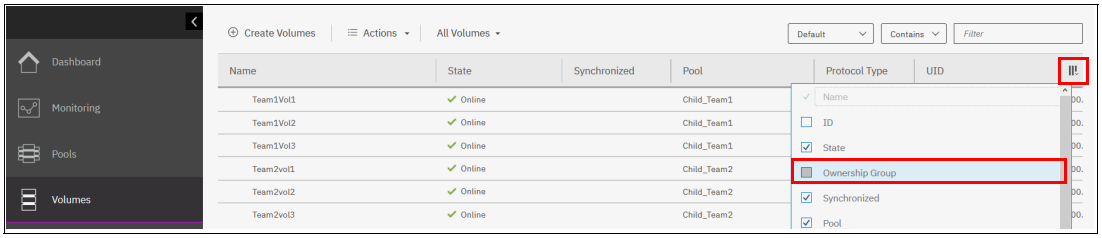


Figure 12-48 Enabling the ownership group attribute display

For example, the volume listing for a system administrator looks the listing that is shown in Figure 12-49.

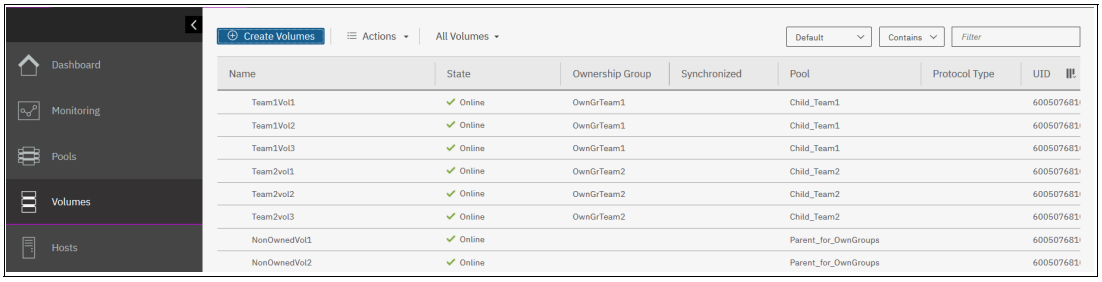


Figure 12-49 Listing volumes for all ownership groups

The global system administrator can see and manage the resources of all ownership groups and resources that are not assigned to any groups.

When the ownership group user logs in, they can see and manage only resources that are assigned to their group. Figure 12-50 shows the initial login window for an ownership group user with the Administrator role.

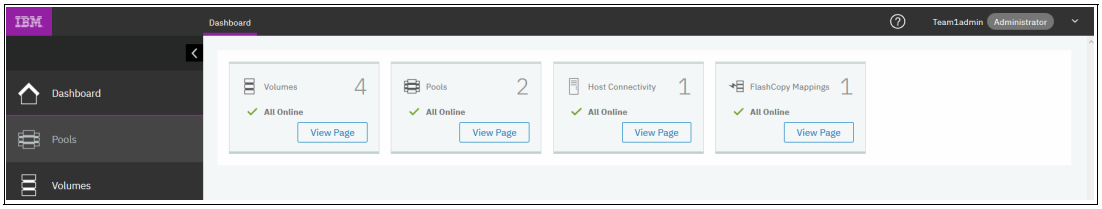


Figure 12-50 Ownership group administrator view

This user does not see a dashboard with global system performance and capacity parameters, but instead can see only tiles for their ownership group resources. For example, of the eight volumes that are configured on a system and shown in Figure 12-49, they can see and manage only three volumes that belong to the group.

The ownership group user can use the GUI to browse, create, and delete (depending on their user role) resources that are assigned to their group. To see information about the global resources (for example, list-managed disks [MDisks] or arrays on the pool), they must use the CLI. Ownership group users cannot manage global resources; they can view them only.

Actions on ownership groups

The global system administrator can rename or remove an ownership group. This process is done by selecting **Access** → **Ownership Groups**, clicking the **"..."** icon that is next to the group name and then selecting the required task, as shown in Figure 12-51.

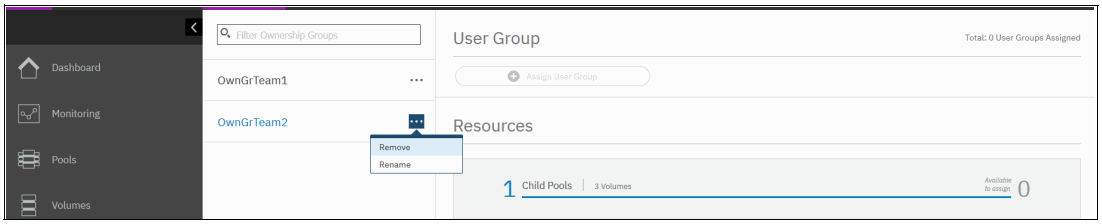


Figure 12-51 Renaming or removing an ownership group

When an ownership group is removed by using the GUI, all ownership assignment information for all the objects of the ownership group is removed; however, the objects remain configured. Only the system administrator can manage those resources later.

12.5.2 Migrating objects to ownership groups

If you want to use ownership groups for objects that are on your system, you must reconfigure specific resources if you want to configure ownership groups.

If child pools are on the system, you can define an ownership group to the child pool or child pools. Before you define an ownership group to child pools, determine other related objects that you want to migrate. Any volumes that are in the child pool inherit the ownership group that is defined for the child pool.

If no child pools are on the system, you must create child pools and move any volumes to those child pools before you can assign them to ownership groups. If volumes are in a parent pool, volume mirroring can be used to create copies of the volume within the child pool.

Alternatively, volume migration can be used to relocate a volume from a parent pool to a child pool within that parent pool without requiring copying.

To nondisruptively migrate a volume to become an ownership group object, complete the following steps:

1. Create a child pool. Do not assign it to an ownership group yet.
2. Migrate volumes that must be assigned to an ownership group to that child pool by using the volume migration or volume mirroring function.

   Figure 12-52 shows the `NonOwnedVol1` volume, which is in a parent pool and does not belong to any ownership group. This volume is mapped to a Small Computer System Interface (SCSI) host.

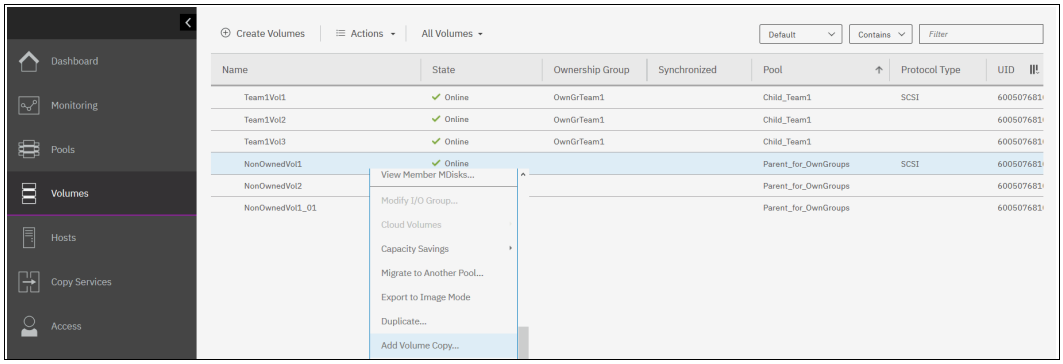


Figure 12-52 Adding a volume copy for migration

To start migration, complete the following steps:

a. Right-click the volume, and then click **Add Volume Copy**. You can also use **Migrate to Another Pool**. However, this method is suitable only if you are migrating from a pool with the same extent size (for example, from a parent pool to a child pool of the same pool), and provides less flexibility.

b. In the **Volume Copy** window, select the child pool that is to be assigned to an ownership group, as shown in Figure 12-53.

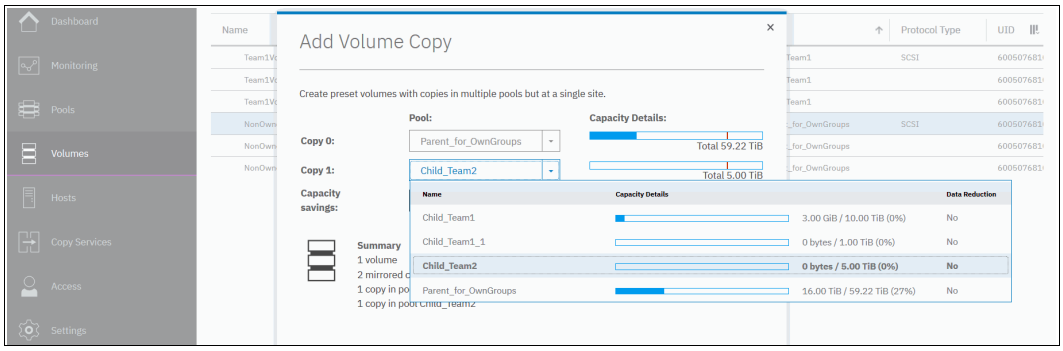


Figure 12-53 Migrating to a child pool

3. Repeat step 2 on page 1090 for all volumes that must belong to an ownership group and then remove the source copies.
4. Create an ownership group as described in “Creating an ownership group” on page 1085. Assign a user group to it, as described in “Assigning users to an ownership group” on page 1085.
5. As shown in Figure 12-54, in **Access** → **Ownership Groups**, select the wanted ownership group and then click **Assign Child Pool**.

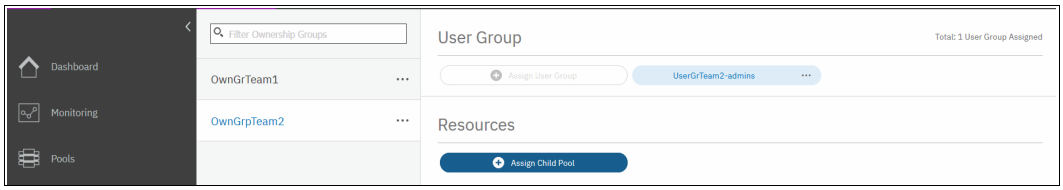


Figure 12-54 Assigning a child pool to an ownership group

6. Select a child pool to assign, as shown in Figure 12-55.

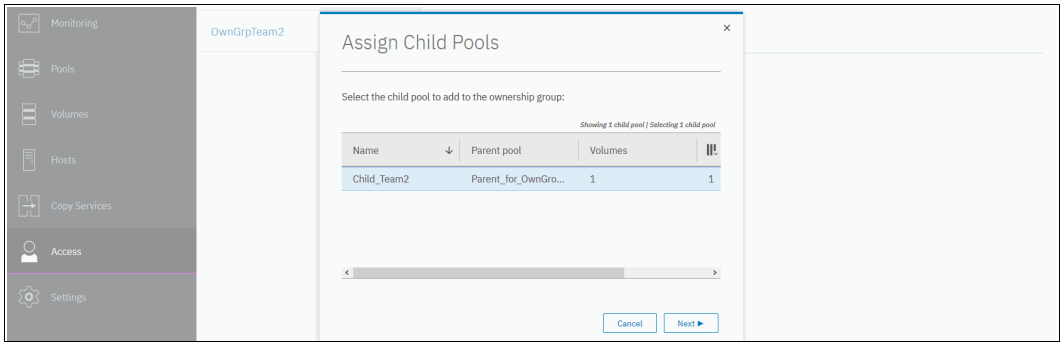


Figure 12-55 Selecting a child pool to assign

7. Click **Next**. The system notifies you that more resources are to inherit ownership from a volume. Also, because the volume is mapped to a host, the host becomes an ownership group object, as shown in Figure 12-56.

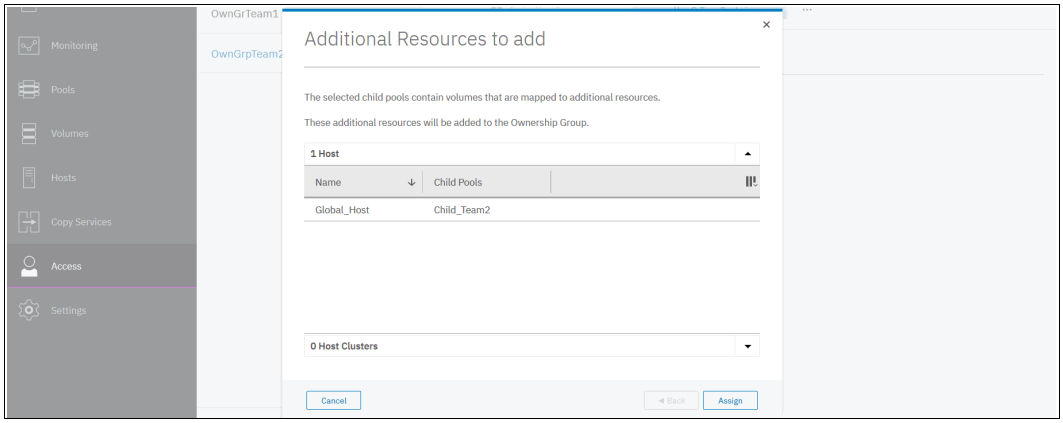


Figure 12-56 Additional Resources to add window

8. As shown in Figure 12-57, a volume and a host both belong to an ownership group. Because a host and volume are in a group, host mapping inherits ownership and becomes a part of an ownership group.

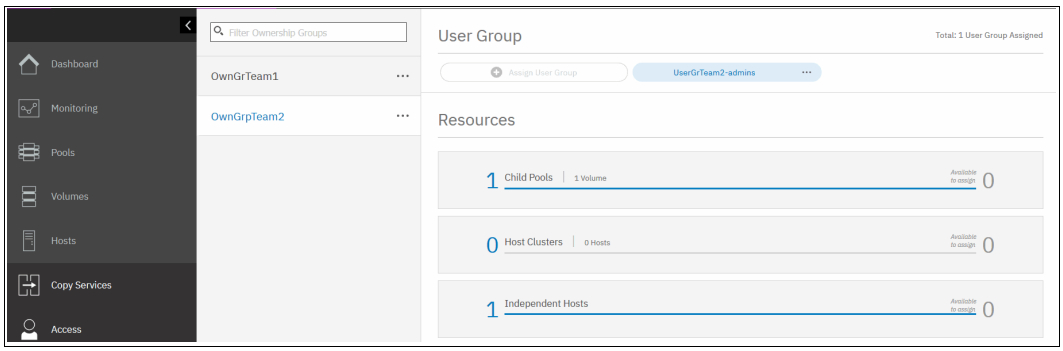


Figure 12-57 Resources of an ownership group

Now, a child pool is assigned to an ownership group.

If you must migrate more volumes to the child pool later, the same approach can be used. However, during migration, one volume copy is in an owned child pool, and the original copy remains in an unowned parent pool. Such a condition causes *inconsistent ownership*, as shown in Figure 12-58.

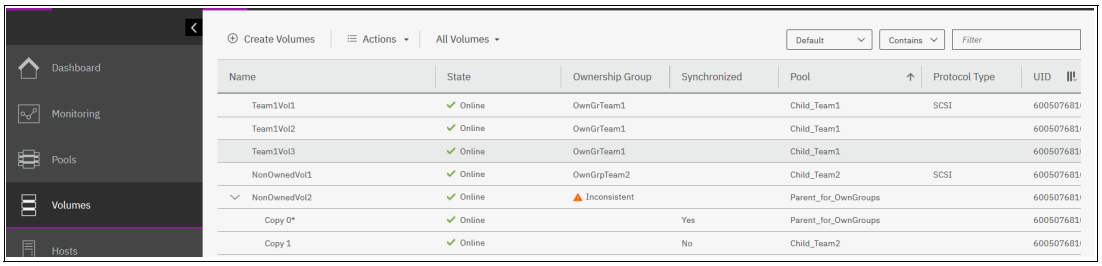


Figure 12-58 Example of inconsistent volume ownership

Until the inconsistent volume ownership is resolved, the volume does not belong to an ownership group and cannot be seen or managed by an ownership group administrator. To resolve it, delete one of the copies after both are synchronized.

12.6 Two Person Integrity (TPI)

Starting with Storage Virtualize release version 8.5.4, Two Person Integrity (TPI) is available. It is also known as the four-eyes-principle. It provides another level of security, by which a defined set administration tasks require agreement by another administrator, so they cannot be conducted by a single person, such as a rogue user with administration permissions.

TPI works in that way that the *superuser* account gets locked. Also, all users with the Security Administrator role are changed to the Restricted Security Administrator role, which is available only when TPI is enabled. It is not a configurable option when creating or changing user groups.

When a configuration task requires Security Administrator privileges, approval of a second user with Restricted Security Administrator role is required. This privilege elevation is limited to a maximum time allowed of 24 hours. A maximum of four elevated users at a time is permitted.

Demonstration video: To view a demonstration video, see Configuring Two Person Integrity with IBM Storage Virtualize V8.6.

Note: Only local users or users who are authenticated through LDAP Remote Authentication with Security Administrator privileges can request a role elevation. Users authenticated through SSO, even those with Security Administrator privileges, cannot do this.

12.6.1 Prerequisites

Enabling TPI is restricted to IBM Storage Virtualize systems featuring a Technician Port. This is to avoid users locking themselves out of the system because a locked superuser account can be unlocked from the Technician Port interface. TPI can be enabled by superuser or any other user with Security Administrator privileges

Further requirements are:

- Two local Security Administrator users (excluding superuser) need to be in place.
- Alternatively, a remote user group with Security Admin role is configured and Remote Authentication is enabled.

12.6.2 Tasks affected by TPI

The following are tasks affected by TPI:

- Restricted Security Administrators cannot create or change any users or user groups related to Security Administrator role.
- Remove and change Safeguarded backups and Safeguarded backup locations.
- Delete Safeguarded snapshots.
- Use a provisioning policy to define a set of rules that are applied when volumes are created within a storage pool or child pool.

- Change security settings related to LDAP, MFA, SSO.
- Change certificates.
- Remove the Safeguarded snapshot policy association from a volume group.
- Change the system time.

12.6.3 Enabling and Configuring TPI

TPI can be enabled in the GUI via **Settings** → **Security** → **User Access** → **Two Person Integrity (TPI)**, see Figure 12-59.

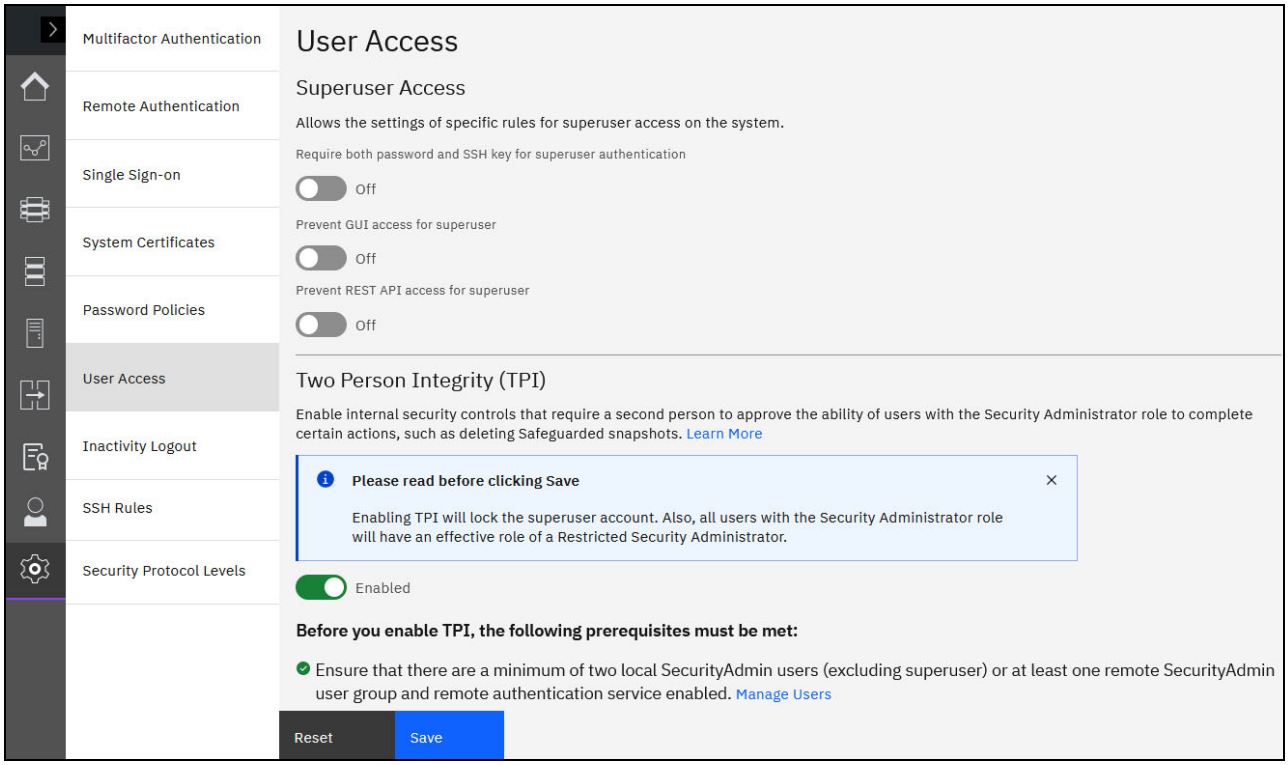


Figure 12-59 GUI TPI Configuration Page

After TPI is enabled and the configuration is saved, a GUI warning concerning changed permissions of the currently logged in user is displayed (see Figure 12-60).

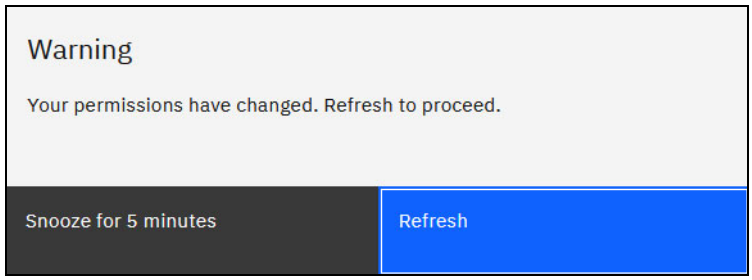


Figure 12-60 Warning - Permissions changed upon enabling TPI

After refreshing the GUI, the Restricted Security Administrator role will be shown for the current user, the selection of GUI tasks in **Settings** → **Security** is also reduced (Figure 12-61 on page 1095).

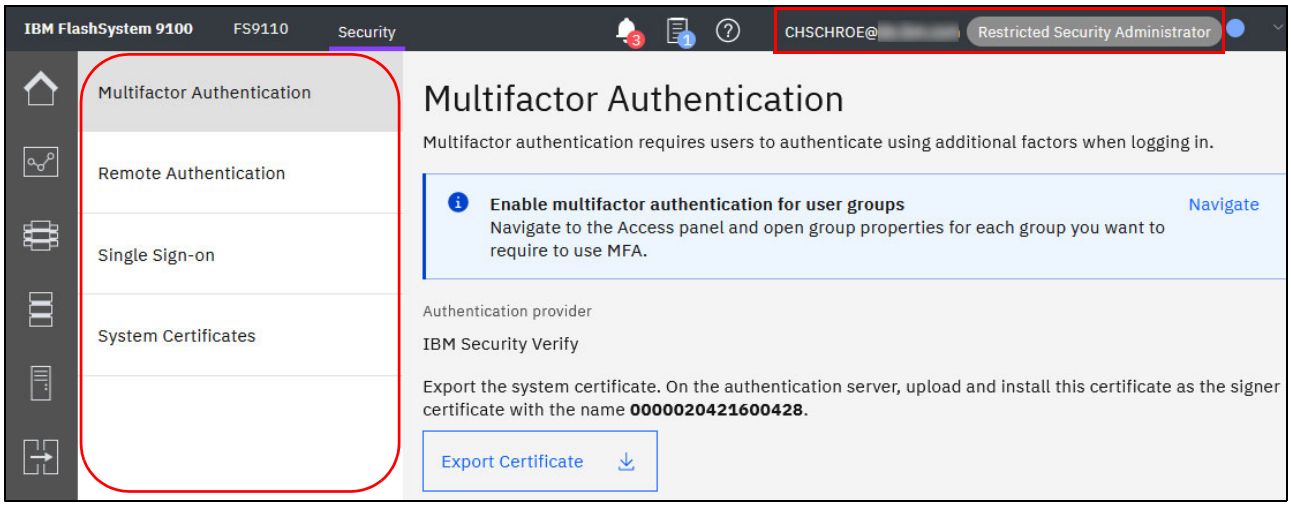


Figure 12-61 TPI enabled

An attempt to create a new local user account with Security Administrator role fails, as shown in Example 12-7.

Example 12-7 Creating Security Admin user fails with TPI enabled

```
IBM_FlashSystem:FS9110:secAdminLocal>mkuser -name TPIenabled -password s0Secret!
-usergrp 0
CMMVC8864E The current user is not allowed to create new users in SecurityAdmin
user groups when two person integrity is enabled.
```

12.6.4 Requesting a temporary Role Elevation

When a Restricted Security Administrator needs to perform a task, which requires Security Administrator privileges, they can request a time-based role elevation. Allowed values for the time limit are 10 minutes– 24 hours. The request can be made in the GUI by clicking the user account pod at the upper right corner in the GUI (Figure 12-62 and Figure 12-63 on page 1096).

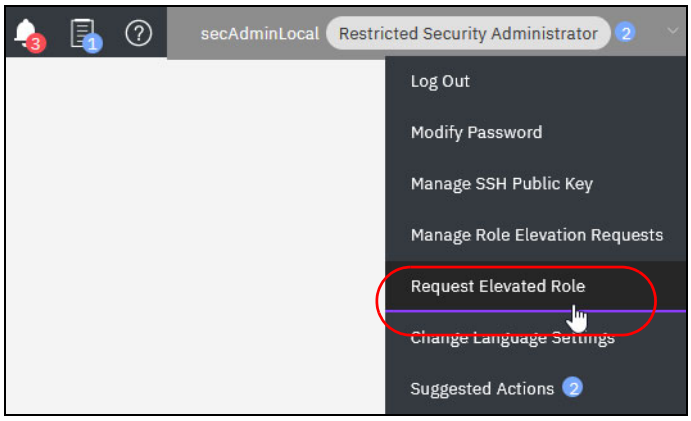


Figure 12-62 Request a time base role elevation

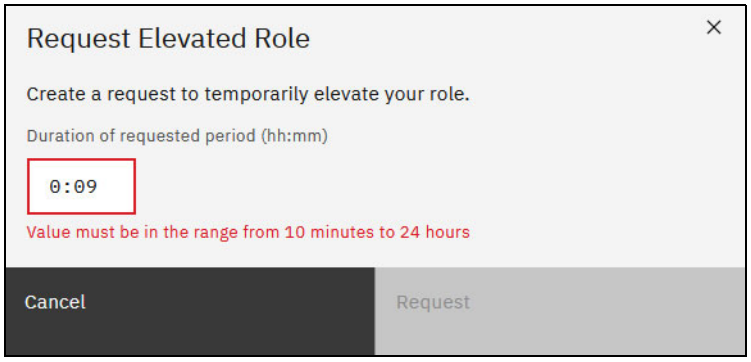


Figure 12-63 Duration of Role Elevation request

The Restricted Security Administrator user that initiated the role elevation request sees their pending request and where it can also be canceled (see Figure 12-64).

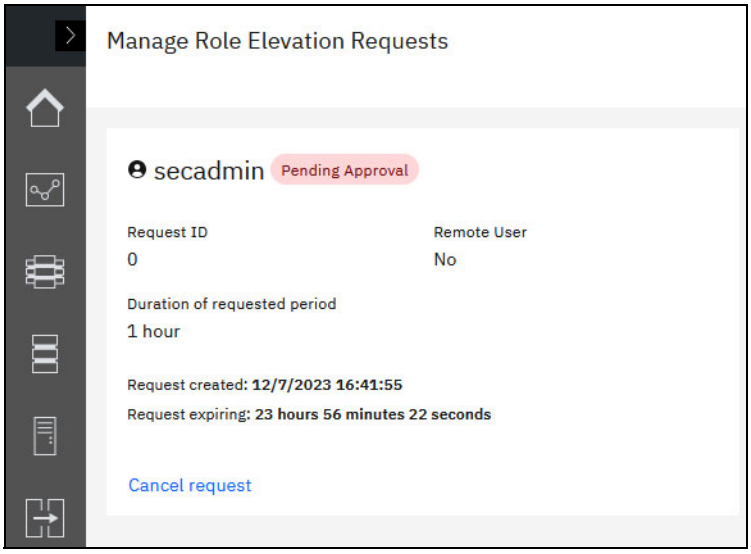


Figure 12-64 Pending Role Elevation Request.

Any other Restricted Security Administrator can navigate through the user account menu **Manage Role Elevation Requests** and either Approve or Deny the request (Figure 12-65 on page 1097). When approving the request, the approver can also change the duration of the grant period of the request.

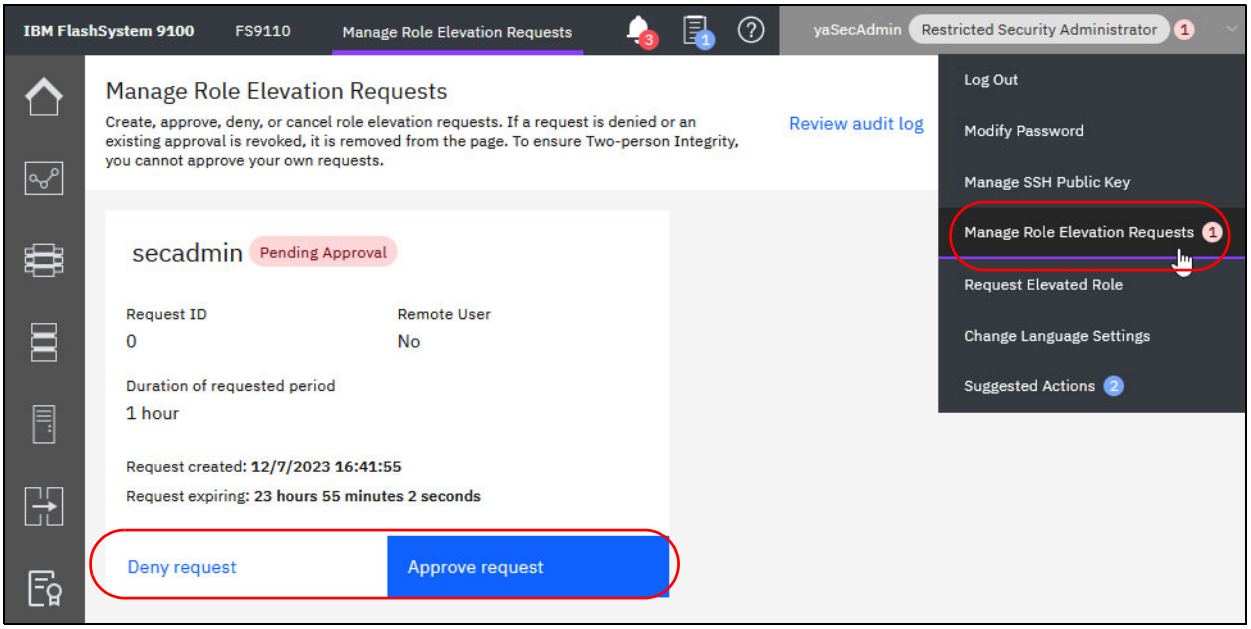


Figure 12-65 Approve or Deny a pending Role Elevation Request

Under **Manage Role Elevation Requests** a previously approved request can be revoked before it expires by any other Restricted Security Admin (Figure 12-66).

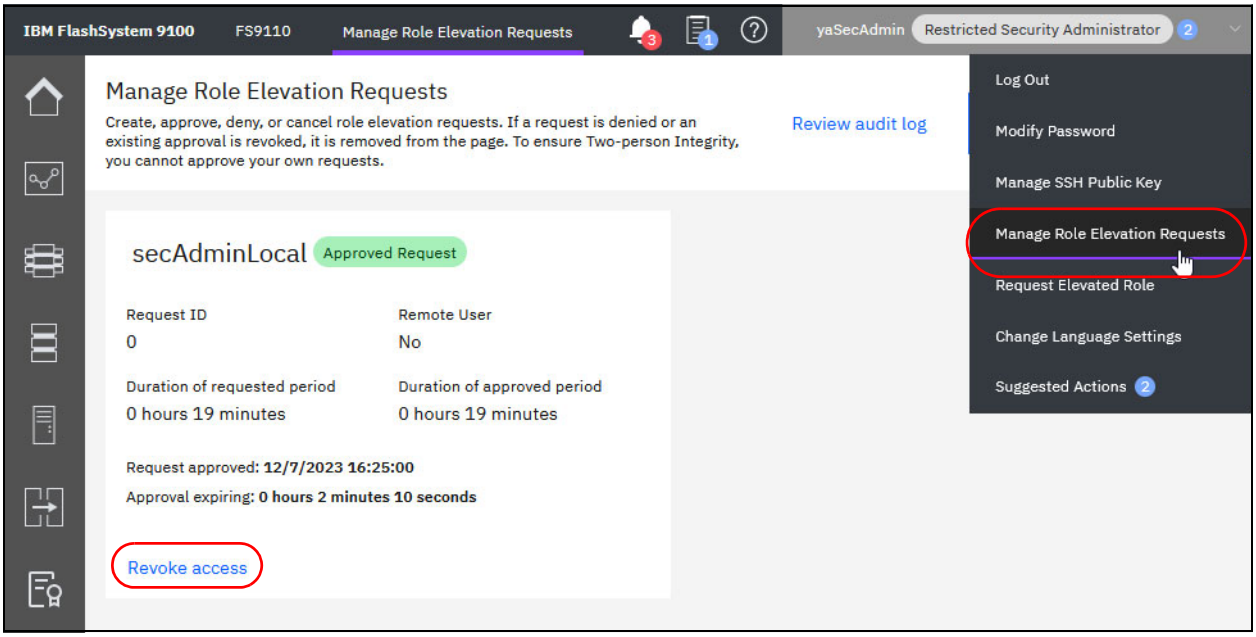


Figure 12-66 Approved Role Elevation Request

An issued role elevation request is logged to the system event log, as shown in Example 12-8.

Example 12-8 Eventlog entry 3375/009219

```
lseventlog <eventlog_sequence_number>
[...]
event_count 1
```

```
status alert
fixed no
auto_fixed no
notification_type warning
event_id 009219
event_id_text Two person integrity requests are pending approval
error_code 3375
error_code_text Two person integrity activity in progress
[...]
```

Note: Only local users or users authenticated through LDAP Remote Authentication with Restricted Security Admin privileges can request a role elevation. Users authenticated through SSO, even those with Security Admin privileges, cannot do this.

12.7 Encryption

Encryption protects against the potential exposure of valuable and sensitive user data that is stored on discarded, lost, or stolen storage devices. IBM SAN Volume Controller and other storage devices that are driven by IBM Storage Virtualize support optional encryption of data at-rest.

For IBM Storage Virtualize, IBM SAN Volume Controller, and IBM FlashSystem, the following different types of encryption are available:

- Externally virtualized storage
- Serial-attached SCSI internal storage
- Non-Volatile Memory Express internal storage

These types of encryption are described in the following sections.

Demonstration video: To view a demonstration video, see How to enable and activate encryption for IBM FlashSystem with IBM Storage Virtualize V8.6.

12.7.1 Externally virtualized storage

Data is decrypted and encrypted as read/write I/Os are issued to the external storage. An encryption key can be used per storage pool or per child storage pool. Migrating a volume between pools (by using volume copying) can be used as a technique for encrypting and decrypting the data.

The key per pool (and in particular, allowing different keys for child pools) supports some part of the multi-tenant use case (if you delete a pool, you delete the key and cryptoerase the data). However, all the keys are wrapped and protected by a single master key that is obtained from a USB stick or an external key server.

As a special case, it is possible to turn off encryption for individual MDisks within the storage pool; for example if the MDisk is self-encrypting. If an external storage controller supports encryption, you can choose to allow it to encrypt the data instead. Where the MDisk reports in its SCSI C2 inquiry page that it is self-encrypting, they system automatically marks the MDisk as self-encrypting.

12.7.2 Serial-attached SCSI internal storage

Data is decrypted and encrypted by the serial-attached Small Computer System Interface (SCSI) (SAS) controller. An encryption key is available per Redundant Array of Independent Disks (RAID). Normally, all arrays in a storage pool are encrypted to form an encrypted storage pool. Although you can create child storage pools, only one key is available per RAID array. Multitenancy is possible only if you have more than one array and storage pool, which often is not practical.

You can migrate volumes from a nonencrypted storage pool to an encrypted storage pool, or you can add an encrypted array to a storage pool and then delete the unencrypted array (which migrates all of the data automatically) as a way of encrypting data.

12.7.3 Non-Volatile Memory Express internal storage

Data is decrypted and encrypted by the Non-Volatile Memory Express (NVMe) drives. Each drive has a media encryption key, but this key is wrapped and protected by an encryption key per RAID array. Therefore, it has the same properties as SAS internal storage.

A storage pool can include a mixture of two or all three types of storage. In this case, the SAS and NVMe internal storage use a key per RAID array for encryption and the externally virtualized storage uses the pool level key.

Because it is almost impossible to control exactly what storage is used for each volume, from a security perspective, this lack of control is effectively a single key for the entire pool. Also, a cryptographic erase is possible only by deleting the entire storage pool and arrays.

12.7.4 Planning for encryption

Data-at-rest encryption is a powerful tool that can help organizations protect the confidentiality of sensitive information. However, as with any other tool, encryption must be used correctly to fulfill its purpose.

Multiple drivers exist for an organization to implement data-at-rest encryption. These drivers can be internal, such as protection of confidential company data, and ease of storage sanitization, or external, such as compliance with legal requirements or contractual obligations.

Therefore, before encryption is configured on the storage, the organization defines its needs and, if it is decided that data-at-rest encryption is required, includes it in the security policy. Without defining the purpose of the specific implementation of data-at-rest encryption, it is difficult or impossible to choose the best approach to implement encryption and verify whether the implementation meets the set of goals.

The following factors are worth considering during the design of a solution that includes data-at-rest encryption:

- Legal requirements
- Contractual obligations
- Organization's security policy
- Attack vectors
- Expected resources of an attacker
- Encryption key management
- Physical security

Multiple regulations mandate data-at-rest encryption, from processing of sensitive personal information to the guidelines of the payment card industry. If any regulatory or contractual obligations govern the data that is held on the storage system, they often provide a wide and detailed range of requirements and characteristics that must be realized by that system. Apart from mandating data-at-rest encryption, these documents might contain requirements concerning encryption key management.

Another document that should be consulted when planning data-at-rest encryption is the organization's security policy.

The outcome of a data-at-rest encryption planning session answers the following questions:

- What are the goals that the organization wants to realize by using data-at-rest encryption?
- How is data-at-rest encryption to be implemented?
- How can it be demonstrated that the proposed solution realizes the set of goals?

12.7.5 Defining encryption of data at-rest

Encryption is the process of encoding data so that only authorized parties can read it. Secret keys are used to encode the data according to well-known algorithms.

Encryption of data-at-rest as implemented in IBM Storage Virtualize is defined by the following characteristics:

- Data-at-rest means that the data is encrypted on the end device (drives or MDisks).
- The algorithm that is used is the Advanced Encryption Standard (AES) US government standard from 2001.
- Encryption of data at-rest complies with the Federal Information Processing Standard 140-2 (FIPS-140-2) standard.
- AES 256 is used for master access keys.
- XTS-AES 256 encryption is a FIPS 140-2 compliant algorithm and is used for data encryption.
- The algorithm is public; the only secrets are the keys.
- A symmetric key algorithm is used. The same key is used to encrypt and decrypt data.

The encryption of system data and metadata is not required; therefore, they are not encrypted.

12.7.6 Encryption methods

Two types of encryption are available on devices that are running IBM Storage Virtualize: hardware encryption and software encryption. Both types protect against the potential exposure of sensitive user data that is stored on discarded, lost, or stolen media. Both also can facilitate the warranty return or disposal of hardware.

Which method is used for encryption is chosen automatically by the system based on the placement of the data:

- Hardware encryption. Data is encrypted by using SAS hardware, or self-encrypting drives. For example, if FlashCore Modules (FCM) are present in the system, then hardware-based data compression and self-encryption is used. It is used only for internal storage drives.
- Software encryption. Data is encrypted by using nodes' CPU. Encryption code uses an AES-NI CPU instruction set. It is used only for external storage.

Both methods of encryption use the same encryption algorithm, key management infrastructure, and license.

Note: The design for encryption is based on the concept that a system is encrypted or not encrypted. Encryption implementation is intended to encourage solutions that contain only encrypted volumes or only unencrypted volumes. For example, after encryption is enabled on the system, all new objects, such as pools, are created as encrypted by default.

12.7.7 Encrypted data

IBM Storage Virtualize performs data-at-rest encryption, meaning that data is encrypted or decrypted when it is written to or read from internal drives (hardware encryption) or external storage systems (software encryption).

Therefore, data is encrypted when transferred across the storage area network (SAN) only between IBM Storage Virtualize systems and external storage. Data in transit is *not* encrypted when transferred on SAN interfaces under the following circumstances:

- Server-to-storage data transfer
- Remote Copy (RC); for example, Global Mirror (GM) or Metro Mirror (MM)
- Intracluster (Node-to-Node) communication

Note: Only data-at-rest is encrypted. Host-to-storage communication and data that is sent over links that are used for Remote Mirroring are not encrypted.

Figure 12-67 shows an encryption example. Unencrypted data paths are shown with red arrows. Software-encrypted disks and software-encrypted data paths are marked in blue. Hardware-encrypted disks and hardware-encrypted data paths are marked in green.

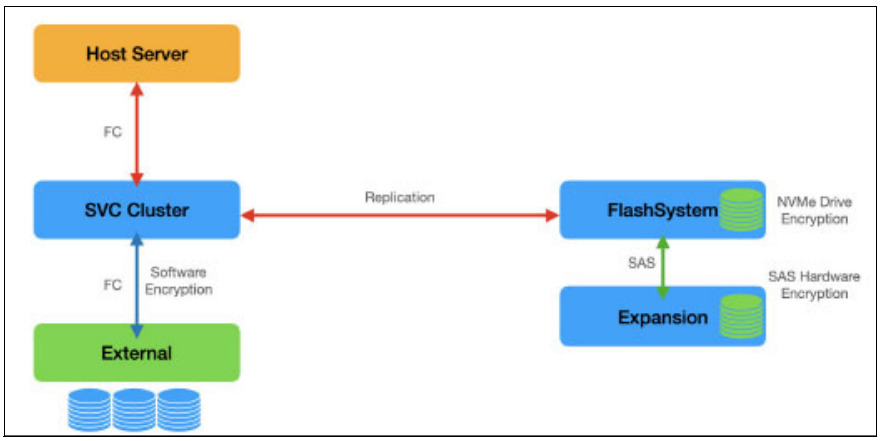


Figure 12-67 Encryption example

The example that is shown in Figure 12-67 uses all three types of encryption. The IBM SAN Volume Controller provides software encryption and data is encrypted before it is sent to the external storage controller. Thus, all data that is stored by the external controller is encrypted. This security feature provides a mechanism to enable data at rest encryption on an external storage controller that otherwise does not natively support encryption.

The IBM SAN Volume Controller sends the replication traffic in clear form (unencrypted) to the remote FlashSystem. This FlashSystem contains NVMe drives in the control enclosure

that natively support drive encryption. These drives handle the encryption internally and are hardware encrypting. In addition, the FlashSystem includes an expansion enclosure that is attached by using SAS. Here, the SAS adapter is performing hardware encryption. Data that is flowing over the SAS network is encrypted.

12.7.8 Data reduction and encryption

Data that is encrypted does not compress or de-duplicate well, if at all. With the hardware encryption that is provided by IBM Storage Virtualize platforms, the encryption is performed as the last operation. That is, if Data Reduction Pools (DRPs) are used, data is reduced before it reaches the hardware encryption components. If Flash-Core Module (FCM) compression is being used, data is first compressed by the FCM, then encrypted.

If external storage controllers are used, consider using software-based encryption. If you want to use data reduction technology, you must use DRPs to ensure that data is reduced first and then encrypted before sending to the external controller.

Note: When externally virtualized storage is encrypted by IBM Storage Virtualize, any data reduction technology that is provided by the external storage controller is rendered ineffective.

12.7.9 Encryption keys

Hardware and software encryption use the same encryption key infrastructure. The only difference is the object that is encrypted by using the keys. The following objects can be encrypted:

- Pools (software encryption)
- Child pools (software encryption)
- Arrays (hardware encryption)

Consider the following points regarding encryption keys:

- Keys are unique for each object, and they are created when the object is created.
- The following types of keys are defined in the system:
  - Master access key:
    - The master access key is created when encryption is enabled.
    - The master access key can be stored on USB flash drives, key servers, or both. One master access key is created for each enabled encryption key provider.
    - It can be copied or backed up as necessary.
    - It is *not* permanently stored anywhere in the system.
    - It is required at boot time to unlock access to encrypted data.
  - Data encryption keys (one for each encrypted object):
    - Data encryption keys are used to encrypt data. When an encrypted object (such as an array, pool, or child pool) is created, a new data encryption key is generated for this object.
    - MDisks that are not self-encrypting are automatically encrypted by using the data encryption key of the pool or child pool to which they belong.

- MDisks that are self-encrypting are not reencrypted by using the data encryption key of the pool or child pool they belong to by default. You can override this default by manually configuring the MDisk as not self-encrypting.
- Data encryption keys are stored in secure memory.
- During cluster internal communication, data encryption keys are encrypted with the master access key.
- Data encryption keys cannot be viewed or changed.
- When an encrypted object is deleted, its data encryption key is discarded (*secure erase*).

Important: Consider the following points:

- If all master access key copies are lost and the system must cold restart, all encrypted data is gone. No method exists (even for IBM) to decrypt the data without the keys. If encryption is enabled and the system cannot access the master access key, all SAS hardware is offline, including unencrypted arrays.
- A self-encrypting MDisk is an MDisk from an encrypted volume in an external storage system.

12.7.10 Encryption licenses

Encryption is a licensed feature that uses key-based licensing. A license must be present for each node in the system before you can enable encryption.

If you add a node to a system that in which encryption is enabled, the node must also be licensed.

No trial license for encryption exists on the basis that when the trial ends, the access to the data is lost. Therefore, you must purchase an encryption license before you activate encryption. Licenses are generated by IBM Data Storage Feature Activation (DSFA) based on the serial number (S/N) and the machine type and model (MTM) of the node.

You can activate an encryption license during the initial system setup (on the Encryption window of the initial setup wizard) or later on, in the running environment.

Contact your IBM marketing representative or IBM Business Partner to purchase an encryption license.

12.8 Activating and enabling encryption

In this section we cover activating and enabling encryption.

12.8.1 Activating encryption

Encryption is enabled at a system level and all of the following prerequisites must be met to use encryption:

- You must purchase an encryption license before you activate the function.

  If you did not purchase a license, contact an IBM marketing representative or IBM Business Partner to purchase an encryption license.

- At least three USB flash drives are required if you plan to *not* use a key management server. They are available as a feature code from IBM. When purchasing your IBM Storage Virtualize storage system with an encryption license, you *must* add the three USB flash drives to your configuration.

  Note: Three USB keys are required per FlashSystem when Encryption is enabled.

- You must activate the license that you purchased.
- Encryption must be enabled.

Activation of the license can be performed in one of the following ways:

- Automatic activation: This method is used when you have the authorization code and the workstation that is being used to activate the license can access an external network. In this case, you must enter only the authorization code. The license key is automatically obtained from the internet and activated in the IBM Storage Virtualize system.
- Manual activation: If you cannot activate the license automatically because any of the requirements are not met, you can follow the instructions that are provided in the GUI to obtain the license key from the web and activate in the IBM Storage Virtualize system.

Both methods are available during the initial system setup and when the system is in use.

Note: Three USB keys are required per FlashSystem when Encryption is enabled.

Obtaining an encryption license

You must purchase an encryption license before you activate encryption. If you did not purchase a license, contact an IBM marketing representative or IBM Business Partner to purchase an encryption license.

When you purchase a license, you receive a function authorization document with an authorization code that is printed on it. This code allows you to proceed by using the automatic activation process.

If the automatic activation process fails or if you prefer the use of the manual activation process, see Entitled Systems Support (ESS) to retrieve your license keys.

Ensure that the following information is available:

- Machine type (MT)
- Serial number (S/N)
- Machine signature
- Authorization code

For more information about how to retrieve the machine signature of a node, see “Manual license activation” on page 1111.

Starting the activation process during initial system setup

One of the steps in the initial setup enables encryption license activation. The system asks, “Was the encryption feature purchased for this system?”

To activate encryption at this stage, complete the following steps:

1. Select **Yes** (see Figure 12-68).

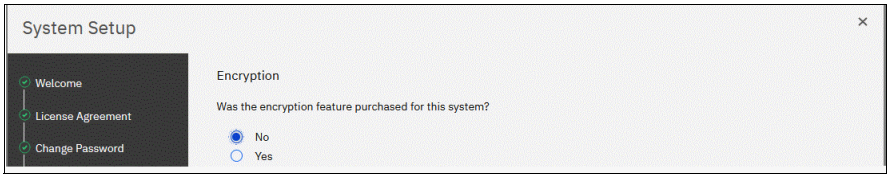


Figure 12-68 Encryption activation during initial system setup

The Encryption window displays information about your storage system, as shown in Figure 12-69.

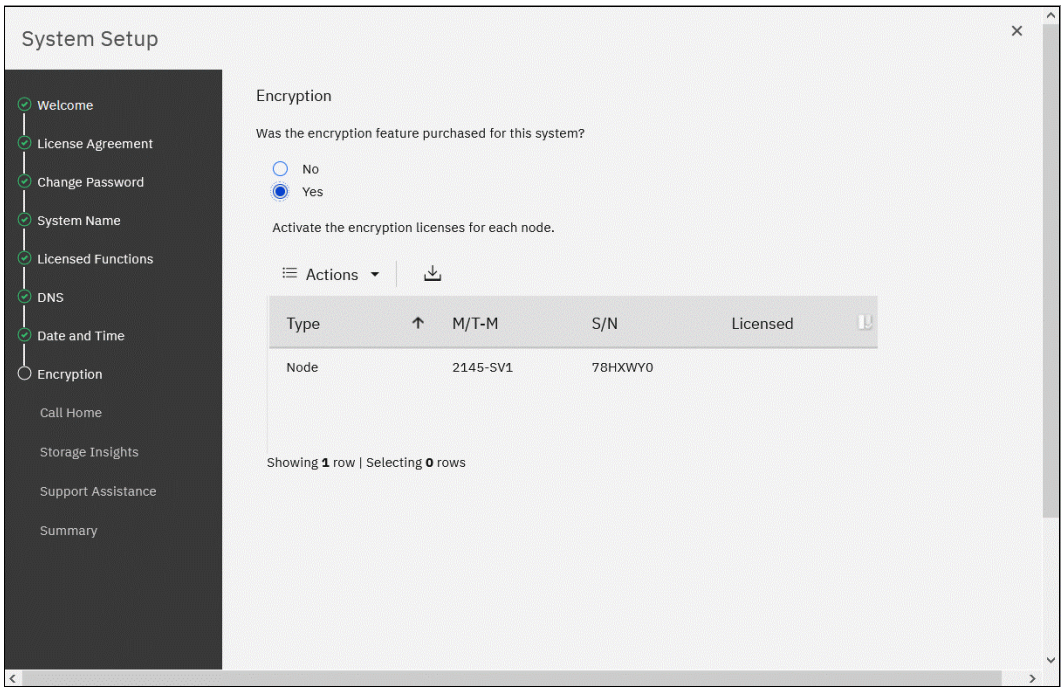


Figure 12-69 Information storage system during initial system setup

Note: If you configured a cluster on one of the nodes and did not add the node to the cluster, the initial setup configuration shows only one node. Therefore, the license for the newly added node must be added later.

2. Right-click the node to open a menu that features two license activation options: Activate License Automatically and Activate License Manually (see Figure 12-70 on page 1106). Use either option to activate encryption.

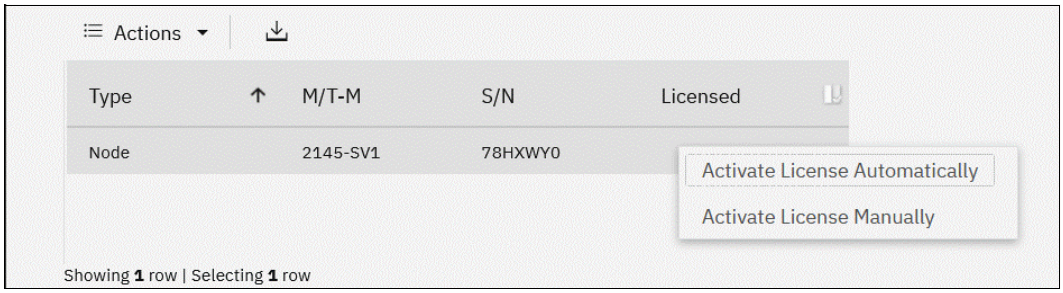


Figure 12-70 Selecting license activation method

For more information about how to complete the automatic activation process, see “Activating the license automatically” on page 1108.

For more information about how to complete a manual activation process, see “Manual license activation” on page 1111.

3. After the activation process is complete, a green checkmark is shown in the column that is labeled “Licensed” that is next to a node for which the license was enabled. Click **Next to** proceed with the initial system setup (see Figure 12-71).

Note: Every enclosure needs an active encryption license before you can enable encryption on the system. Attempting to add a nonlicensed enclosure to an encryption-enabled system fails.

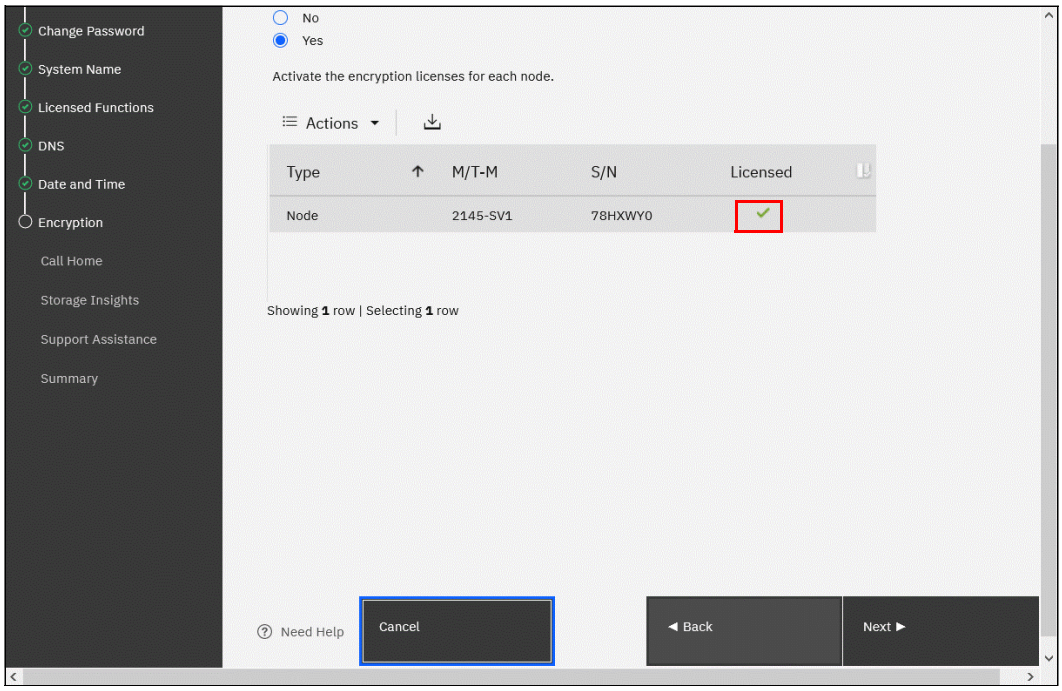


Figure 12-71 Successful encryption license activation during initial system setup

Starting the activation process on a running system

To activate encryption on a running system, complete the following steps:

1. Click **Settings** → **System** → **Licensed Functions**.
2. Click **Encryption Licenses**, as shown in Figure 12-72.

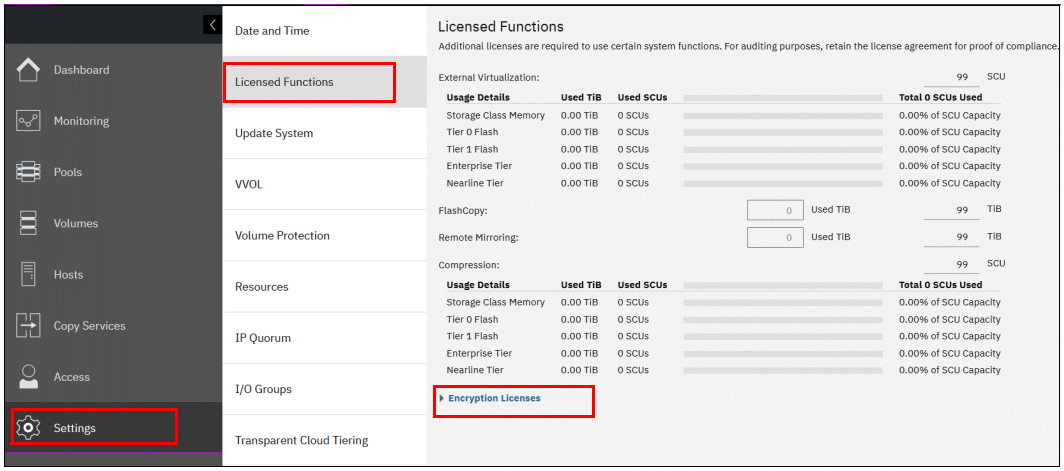


Figure 12-72 Expanding Encryption Licenses section on the Licensed Functions window

3. The Encryption Licenses window displays information about your nodes. Right-click the enclosure on which you want to install an encryption license. A menu opens that includes two license activation options: Activate License Automatically and Activate License Manually (see Figure 12-73). Use either option to activate encryption.

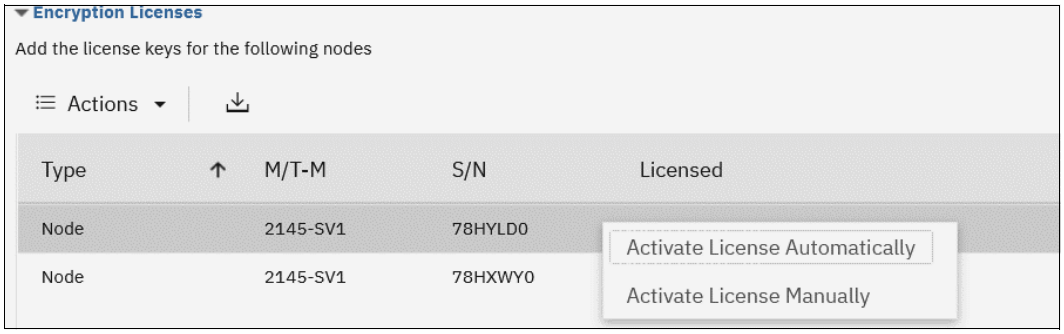


Figure 12-73 Select the node on which you want to enable the encryption

For more information about how to complete an automatic activation process, see “Activating the license automatically” on page 1108. For more information about how to complete a manual activation process, see “Manual license activation” on page 1111.

After the activation process is complete, a green check checkmark is shown in the column that is labeled “Licensed” for the node, as shown in Figure 12-74.

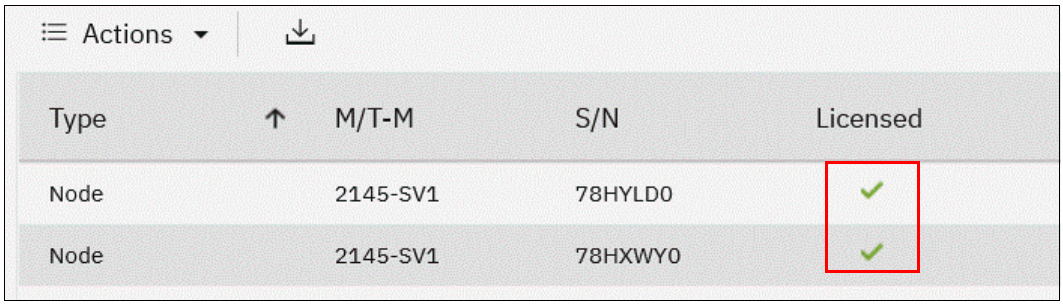


Figure 12-74 Successful encryption license activation on a running system

Activating the license automatically

The automatic license activation is the faster method to activate the encryption license for IBM Storage Virtualize. You need the authorization code and the workstation that is used to access the GUI that can access the external network.

Note: The PC that was used to connect to the GUI and activate the license must connect to the internet.

To activate the encryption license for a node automatically, complete the following steps:

1. Select **Activate License Automatically** to open the Activate License Automatically window, as shown in Figure 12-75.

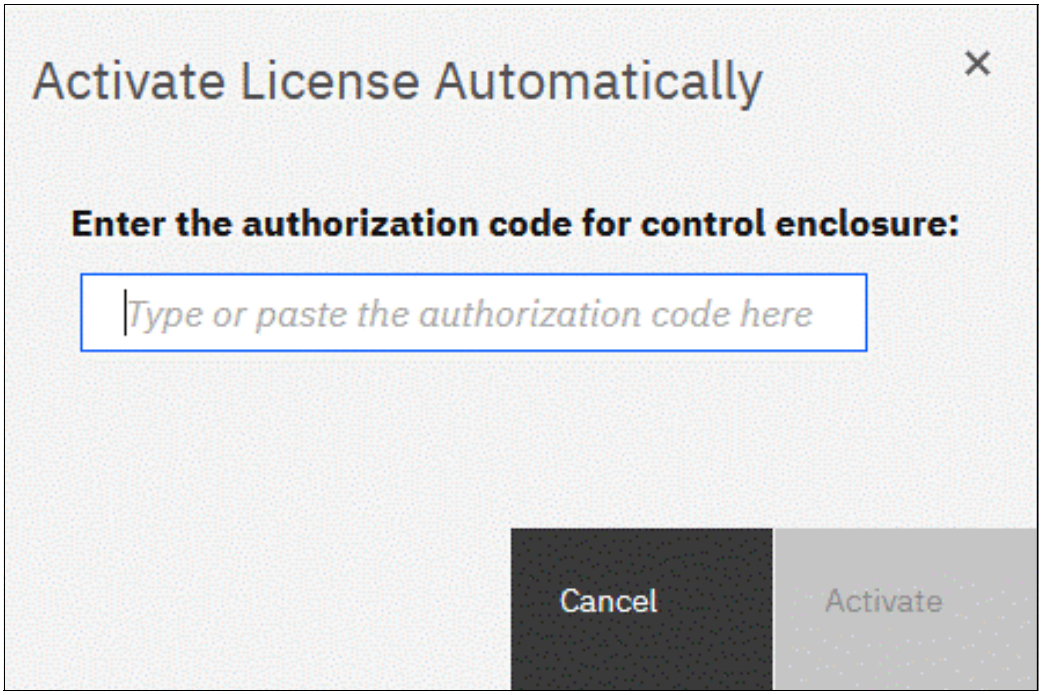


Figure 12-75 Encryption license Activate License Automatically window

2. Enter the authorization code that is specific to the node that you selected, as shown in Figure 12-76. Click **Activate**.

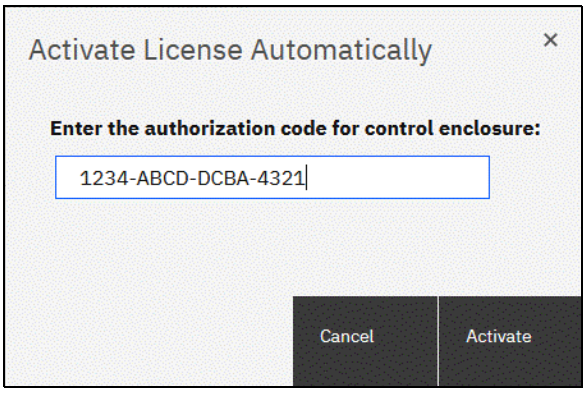


Figure 12-76 Entering an authorization code

The system connects to IBM to verify the authorization code and retrieves the license key. Figure 12-77 shows a window that is displayed during this connection process. If everything works correctly, the procedure takes less than a minute.

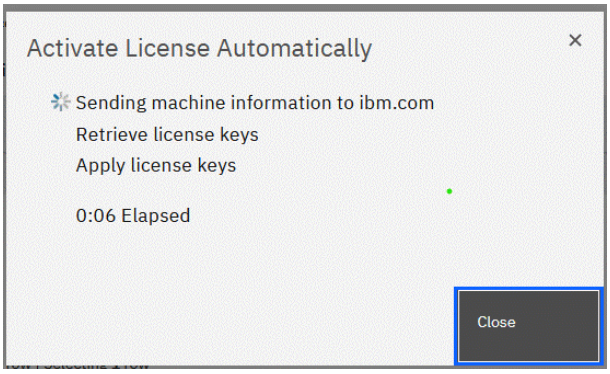


Figure 12-77 Activating encryption

After the license key is retrieved, it is automatically applied, as shown in Figure 12-78.

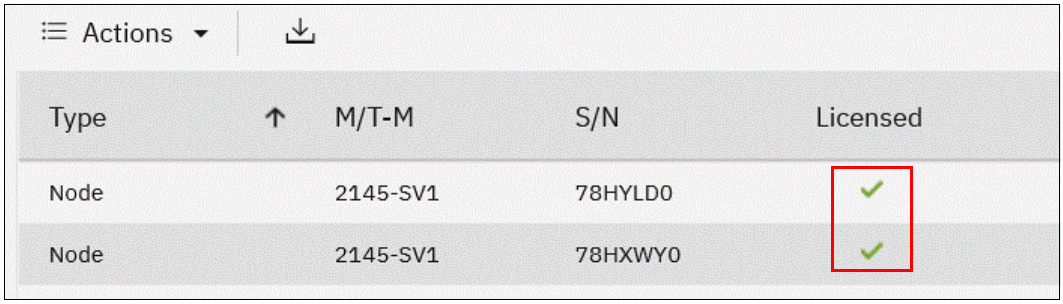


Figure 12-78 Successful encryption license activation

Problems with automatic license activation

If connection problems occur with the automatic license activation procedure, the system times out after 3 minutes with an error (see Figure 12-79).

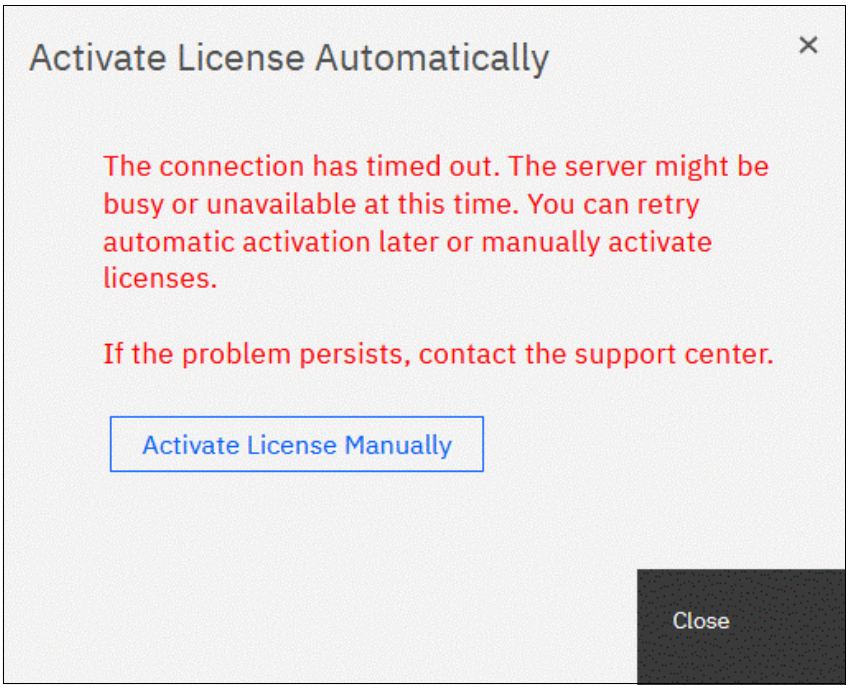


Figure 12-79 Authorization code failure

To remedy the problem, check whether the PC that is used to connect to the IBM SAN Volume Controller GUI and activate the license can access the internet. If you cannot complete the automatic activation procedure, use the manual activation procedure that is described in “Manual license activation” on page 1111.

Although authorization codes and encryption license keys use the same format (four groups of four hexadecimal digits), you can use only one of them in the suitable activation process. If you use a license key when the system expects an authorization code, the system displays an error message.

Manual license activation

To manually activate the encryption license for a node, complete the following steps:

1. Select **Activate License Manually** to open the Manual Activation window, as shown in Figure 12-80.

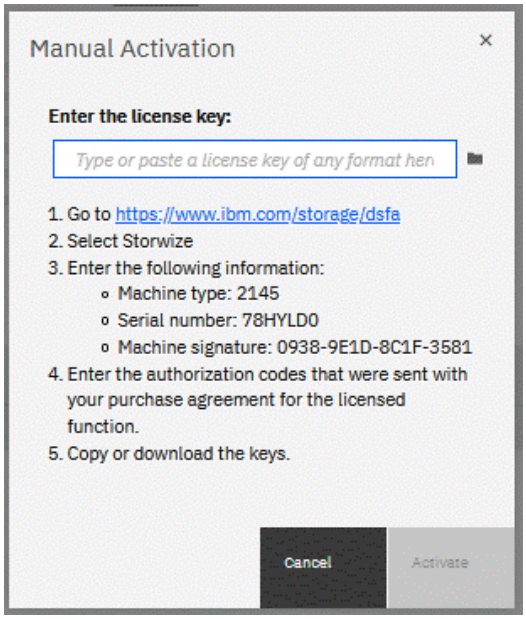


Figure 12-80 Manual encryption license activation window

2. If you have not done so already, obtain the encryption license for the node. The information that is required to obtain the encryption license is displayed in the Manual Activation window. Use this data and follow the instructions that are provided in “Obtaining an encryption license” on page 1104.
3. You can enter the license key by entering it, pasting it, or clicking the folder icon and uploading the license key file to the storage system that was downloaded from DSFA. Click **Activate**.

   After the task completes successfully, the GUI shows that encryption is licensed for the specified node, as shown in Figure 12-81.

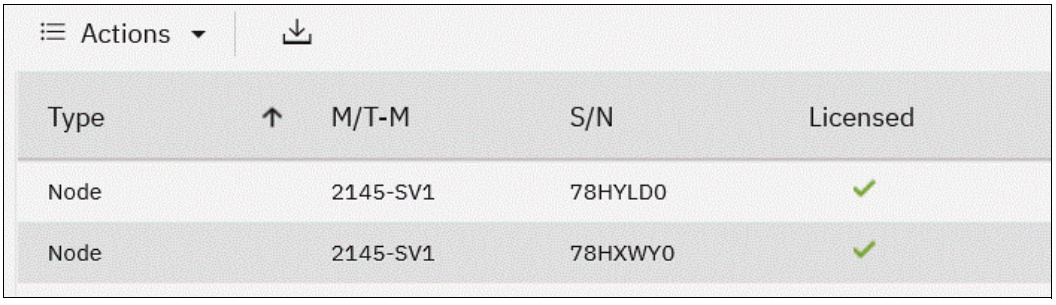


Figure 12-81 Successful encryption license activation

Problems with manual license activation

Although authorization codes and encryption license keys use the same format (four groups of four hexadecimal digits), you can use only one of them in the suitable activation process. If you use an authorization code when the system expects a license key, the system displays an error message, as shown in Figure 12-82.

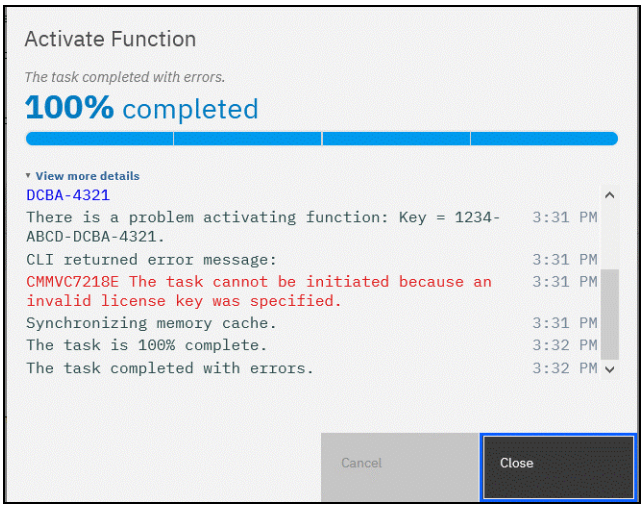


Figure 12-82 License key failure

12.8.2 Enabling encryption

This section describes the process to create and store system master access key copies, which are also referred to as *encryption keys*. These keys can be stored on any or both of two key providers: USB flash drives or a key server.

The following types of key servers are supported by IBM Storage Virtualize:

- IBM Security Guardium Key Lifecycle Manager (SGKLM), which was introduced in IBM Storage Virtualize V7.8.
- Gemalto SafeNet KeySecure, which was introduced in IBM Storage Virtualize V8.2.
- Thales CipherTrust Manager, which is a re-branded version of Gemalto SafeNet KeySecure. Support was added in IBM Storage Virtualize V8.4.1

For more information about supported key servers, see IBM Storage Virtualize Supported Key Servers.

IBM Storage Virtualize code V8.1 introduced the ability to define a maximum of four encryption key servers, which is a preferred configuration because it increases key provider availability. In this version, support for simultaneous use of USB flash drives and a key server was added.

Organizations that use encryption key management servers might consider parallel use of USB flash drives as a backup solution. During normal operation, such drives can be disconnected and stored in a secure location. However, during a catastrophic loss of encryption servers, the USB flash drives can still be used to unlock the encrypted storage.

The key server and USB flash drive characteristics that are described next might help you to choose the type of encryption key provider that you want to use.

Key servers can have the following characteristics:

- Physical access to the system is not required to perform a rekey operation.
- Support for businesses that have security requirements that preclude use of USB ports.
- Hardware security modules (HSMs) can be used to generate an encryption key.
- Keys can be replicated between servers and automatic backups can be performed.
- Implementations follow an open standard (Key Management Interoperability Protocol [KMIP]) that aids in interoperability.
- Operations that are related to key management can be audited.
- Encryption keys and physical access to storage systems can be managed separately.

USB flash drives have the following characteristics:

- Physical access to the system is required to process a rekey operation.
- No moving parts with almost no read or write operations to the USB flash drives.
- Inexpensive to maintain and use.
- Convenient and easy to have multiple identical USB flash drives available as backups.

Important: Maintaining confidentiality of the encrypted data hinges on the security of the encryption keys. Pay special attention to ensure secure creation, management, and storage of the encryption keys.

Starting the Enable Encryption wizard

After the license activation step is successfully completed on IBM SAN Volume Controller nodes, you can now enable encryption. You can enable encryption after completion of the initial system setup by using the GUI or CLI.

The GUI can be used in two ways to start the Enable Encryption wizard. The first method is by clicking **Run Task** that is next to Enable Encryption on the Suggested Tasks window, as shown in Figure 12-83.

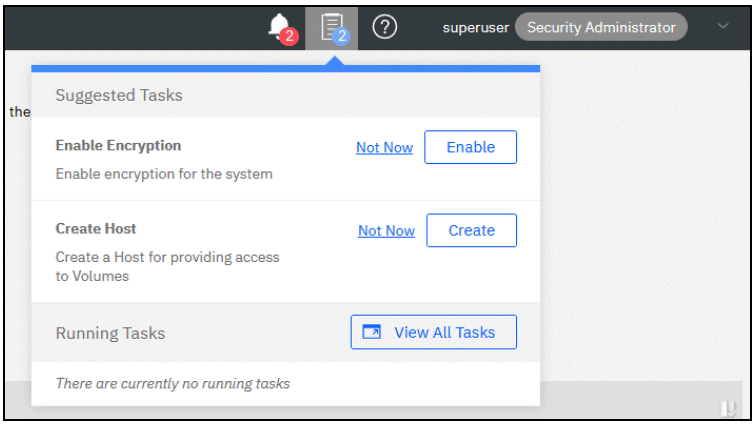


Figure 12-83 Enable Encryption from the Suggested Tasks *window*

You can also click **Settings** → **Security** → **Encryption** and then click **Enable Encryption**, as shown in Figure 12-84.

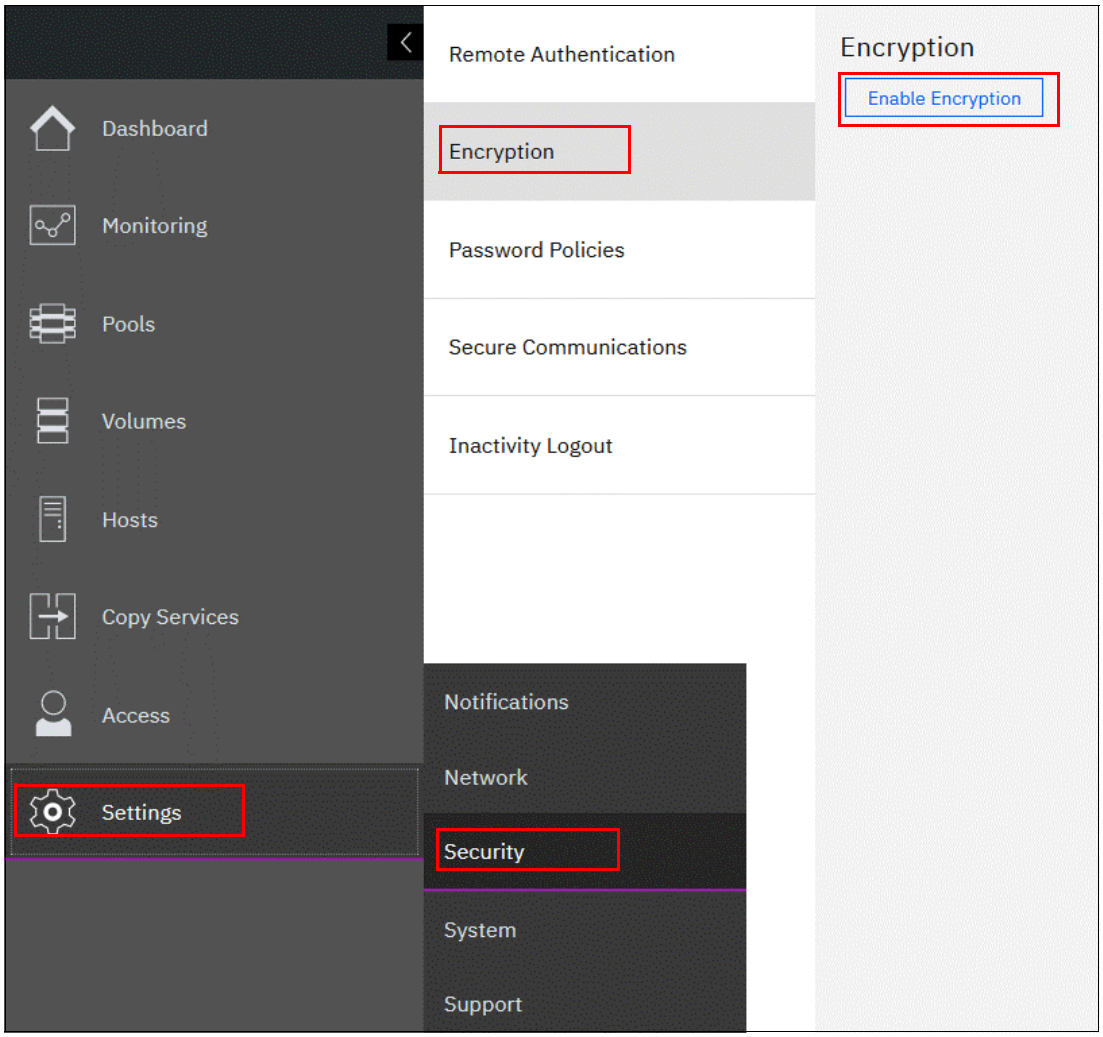


Figure 12-84 Enable Encryption from the Security window

The Enable Encryption wizard starts by prompting you to select the encryption key provider to use for storing the encryption keys, as shown in Figure 12-85. You can enable one or both providers.

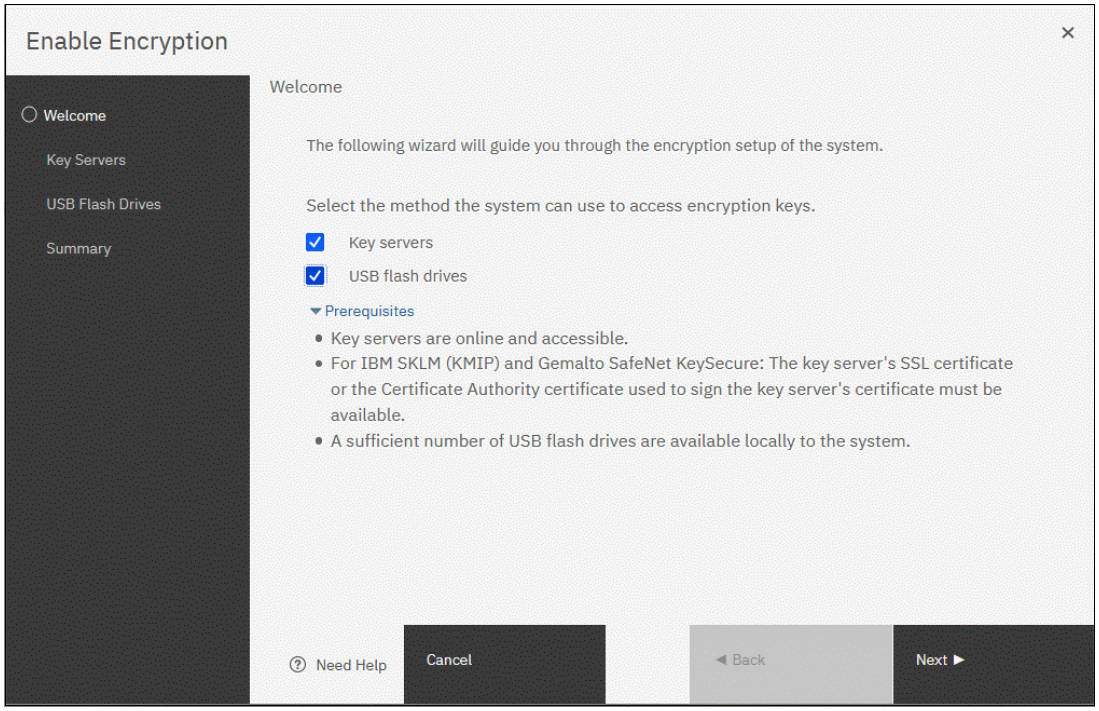


Figure 12-85 Enable Encryption wizard Welcome window

The next section presents a scenario in which both encryption key providers are enabled at the same time. For more information about how to enable encryption by using only USB flash drives, see “Enabling encryption by using USB flash drives” on page 1115.

For more information about how to enable encryption by using key servers as the sole encryption key provider, see 12.8.3, “Enabling encryption by using key servers” on page 1119.

Enabling encryption by using USB flash drives

Note: The system needs at least three USB flash drives to be present before you can enable encryption by using this encryption key provider. IBM USB flash drives are preferred and can be obtained from IBM with the feature name *Encryption USB Flash Drives (Four Pack)*. Other flash drives might also work. You can use any USB ports in any node of the cluster.

The use of USB flash drives as the encryption key provider requires a minimum of three USB flash drives to store the generated encryption keys. Because the system attempts to write the encryption keys to any USB flash drive that is inserted into a node port, it is critical to maintain physical security of the system during this procedure.

While the system enables encryption, you are prompted to insert USB flash drives into the system. The system generates and copies the encryption keys to all available USB flash drives.

Ensure that each copy of the encryption key is valid before you write any user data to the system. The system validates any key material on a USB flash drive when it is inserted into the canister. If the key material is invalid, the system logs an error.

If the USB flash drive is unusable or fails, the system does not display it as output. Figure 12-87 on page 1117 shows an example where the system detected and validated three USB flash drives.

If your system is in a secure location with controlled access, one USB flash drive for each canister can remain inserted in the system. If a risk of unauthorized access exists, all USB flash drives with the master access keys must be removed from the system and stored in a secure place.

Securely store all copies of the encryption key. For example, any USB flash drives that are holding an encryption key copy that is not left plugged into the system can be locked in a safe. Similar precautions must be taken to protect any other copies of the encryption key that are stored on other media.

Notes: Generally, create at least one extra copy on another USB flash drive for storage in a secure location. You can also copy the encryption key from the USB drive and store the data on other media, which can provide extra resilience and mitigate risk that the USB drives that were used to store the encryption key come from a faulty batch.

Every encryption key copy must be stored securely to maintain the confidentiality of the encrypted data.

A minimum of one USB flash drive with the correct master access key is required to unlock access to encrypted data after a system restart, such as a system-wide restart or power loss. No USB flash drive is required during a warm restart, such as a node that is exiting service mode or a single node restart. The data center power-on procedure must ensure that USB flash drives that contain encryption keys are plugged into the storage system before it is powered on.

During power-on, insert USB flash drives into the USB ports on two supported canisters to safeguard against failure of a node, node's USB port, or USB flash drive during the power-on procedure.

To enable encryption by using USB flash drives as the only encryption key provider, complete the following steps:

1. In the Enable Encryption wizard Welcome tab, select **USB flash drives** and then click **Next**, as shown in Figure 12-86 on page 1117.

   If fewer than three USB flash drives are inserted into the system, you are prompted to insert more drives. The system reports how many more drives must be inserted.

   Note: The Next option remains disabled until at least three USB flash drives are detected.

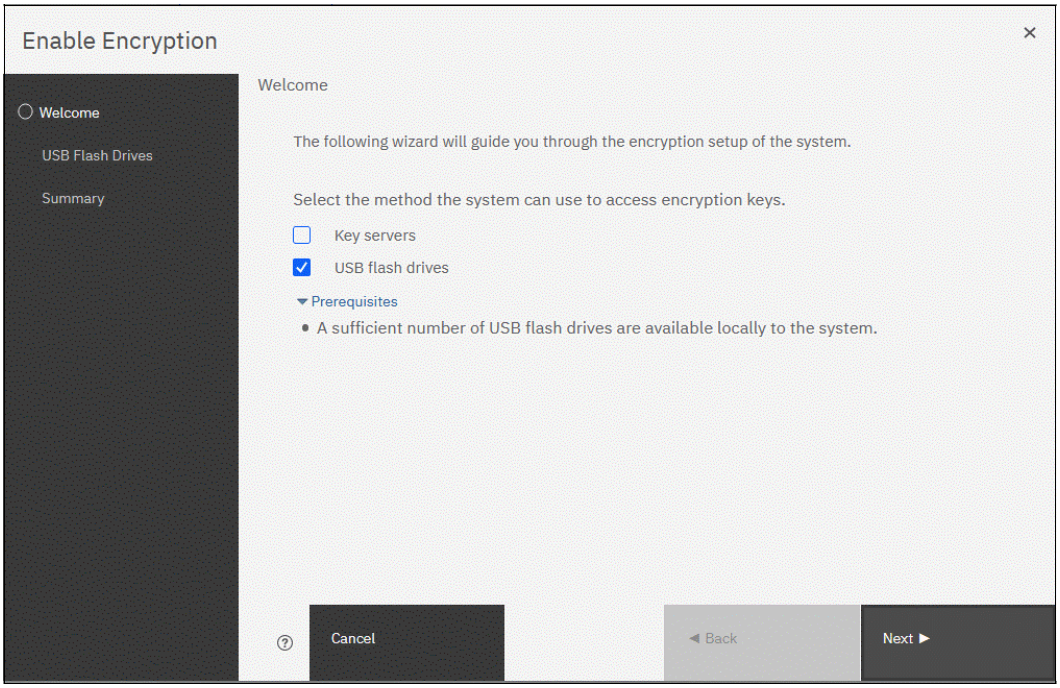


Figure 12-86 Selecting USB flash drives in the Enable Encryption wizard

2. Insert the USB flash drives into the USB ports as requested.

   After the minimum required number of drives is detected, the encryption keys are automatically copied on the USB flash drives, as shown in Figure 12-87.

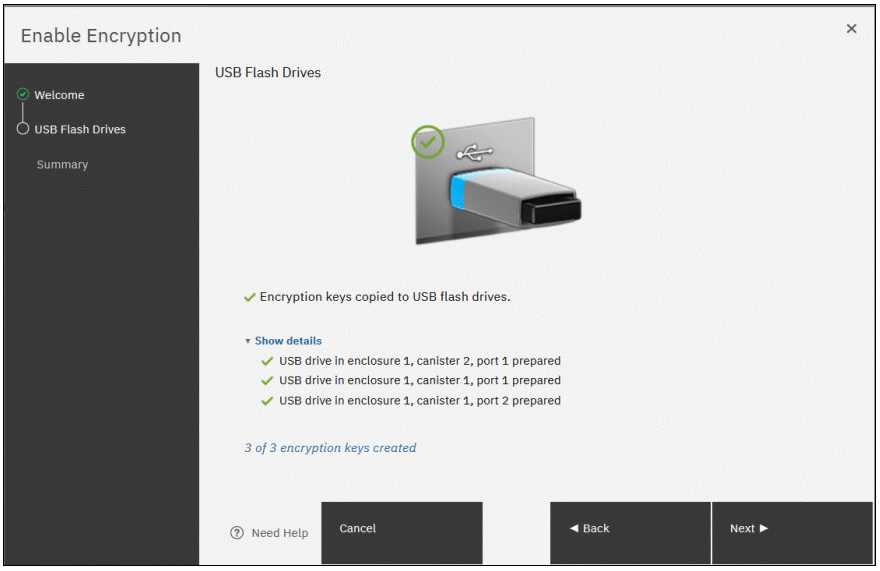


Figure 12-87 Writing the master access key to USB flash drives

You can continue to add USB flash drives or replace the drives that are plugged in to create copies. When done, click **Next**.

3. The number of keys that were created is shown in the Summary tab, as shown in Figure 12-88. Click **Finish** to finalize the encryption enablement.



Figure 12-88 Commit the encryption enablement

You receive a message confirming that the encryption is now enabled on the system, as shown in Figure 12-89.

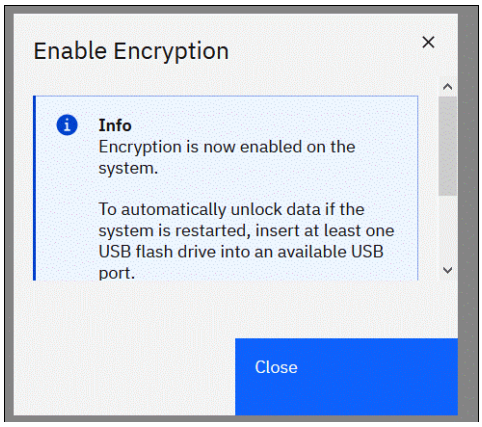


Figure 12-89 Encryption enabled message that uses USB flash drives

4. Confirm that encryption is enabled and verify which key providers are in use by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-90.

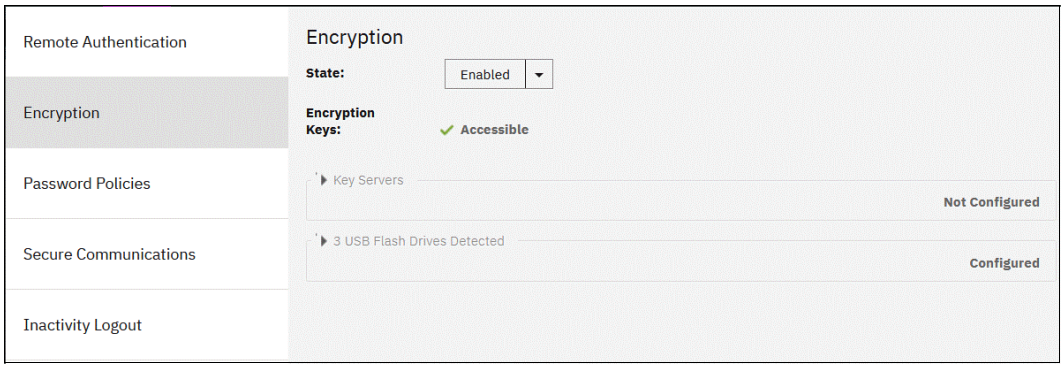


Figure 12-90 Encryption view showing by using USB flash drives as the enabled provider

12.8.3 Enabling encryption by using key servers

A key server is a centralized system that receives and then distributes encryption keys to its clients, which can include IBM Storage Virtualize systems.

IBM Storage Virtualize supports the use of the following key servers as encryption key providers:

- Security Guardium Key Lifecycle Manager (SGKLM)
- Gemalto SafeNet KeySecure
- Thales CipherTrust Manager

Note: Support for SGKLM was introduced in IBM Storage Virtualize V7.8. Support for Gemalto SafeNet KeySecure was introduced in IBM Storage Virtualize V8.2.1.

All these products support Key Management Interoperability Protocol (KMIP), which is a standard for management of cryptographic keys.

Note: Make sure that the key management server function is fully independent from encrypted storage that has encryption that is managed by this key server environment. Failure to observe this requirement might create an encryption deadlock. An *encryption deadlock* is a situation in which none of the key servers in the environment can become operational because some critical part of the data in each server is stored on a storage system that depends on one of the key servers to unlock access to the data.

IBM Storage Virtualize code V8.1 and later supports up to four key server objects that are defined in parallel. However, only one key server type can be enabled at one time.

Another characteristic when working with key servers is that it is not possible to migrate from one key server type directly to another. If you want to migrate from one type to another, you first must migrate from your current key server to USB encryption, and then migrate from USB to the other type of key server.

12.8.4 Enabling encryption by using SGKLM

Before you create a key server object in the storage system, the key server must be configured. Ensure that you complete the following tasks on the SGKLM server before you enable encryption on the storage system:

- Configure the SGKLM server to use Transport Layer Security version 1.2. The default setting is `TLSv1`, but IBM Storage Virtualize supports only version 1.2. Therefore, set the value of security protocols to `SSL_TLSv2` (which is a set of protocols that includes TLSv1.2) in the SGKLM server configuration properties.
- Ensure that the database service is started automatically on start.
- Ensure that at least one Secure Sockets Layer (SSL) certificate is available for browser access.
- Create a `Storage_VIRT` device group for IBM Storage Virtualize systems.

For more information about completing these tasks, see IBM Security Guardium Key Lifecycle Manager documentation.

Access to the key server that is storing the correct master access key is required to enable access to encrypted data in the system after a system restart. System restart can be a system-wide restart or power loss. Access to the key server is not required during a warm restart, such as a node that is exiting service mode or a single node restart.

The data center power-on procedure must ensure key server availability before the storage system that is using encryption is started. If a system with encrypted data is restarted and cannot access the encryption keys, the encrypted storage pools are offline until the encryption keys are detected.

To enable encryption by using an SGKLM key server, complete the following steps:

1. Ensure that service IPs are configured on all your nodes.
2. In the Enable Encryption wizard Welcome tab, select **Key servers** and then click **Next**, as shown in Figure 12-91 on page 1121.

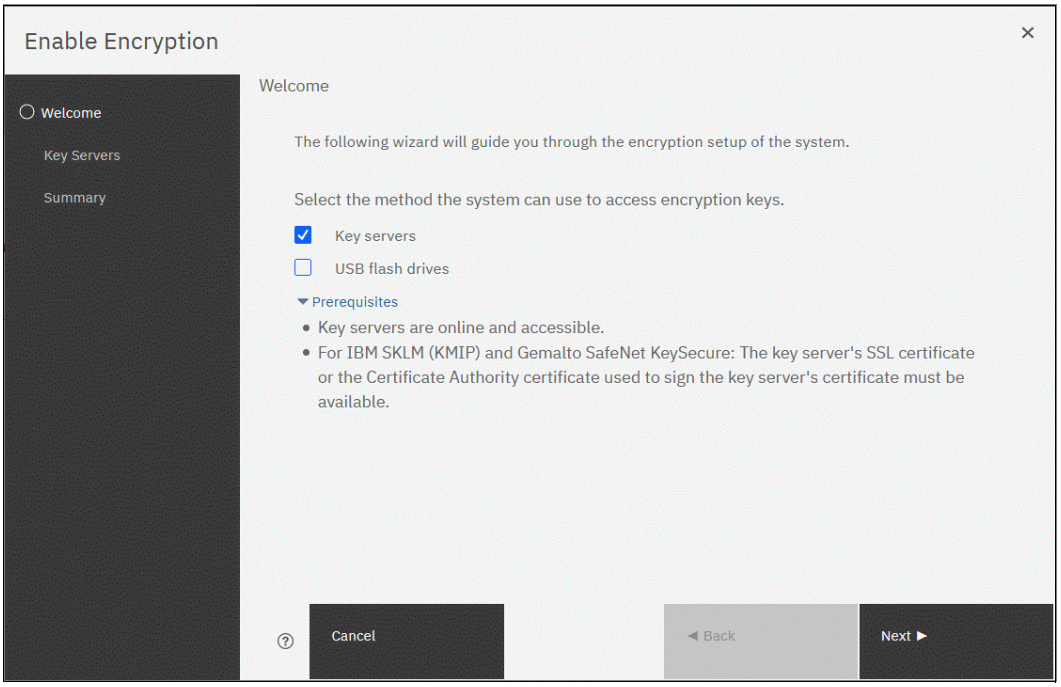


Figure 12-91 Selecting Key server as the only provider in the Enable Encryption wizard

3. Select **IBM SKLM (with KMIP)** as the key server type, as shown in Figure 12-92.

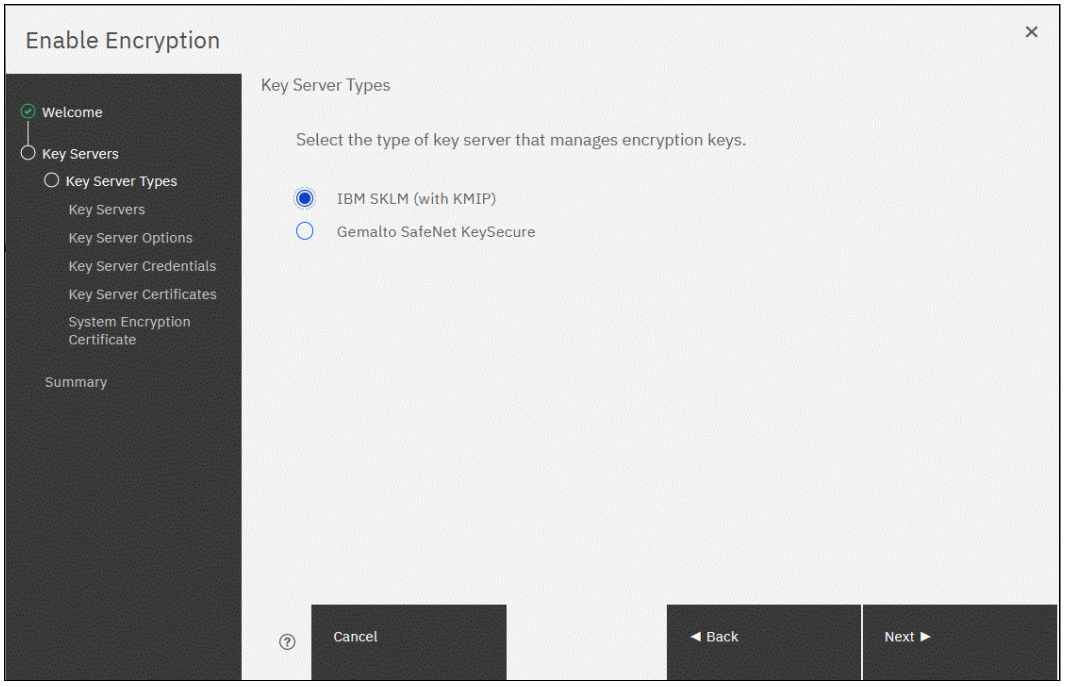


Figure 12-92 Selecting SGKLM as the key server type

4. The wizard moves to the Key Servers tab, as shown in Figure 12-93. Enter the name and IP address of the key servers. The first key server that is specified must be the primary SGKLM key server.

Note: The supported versions of SGKLM (up to Version 4.1.1, which was the latest code version available at the time of this writing) differentiate between the primary and secondary key server role. The Primary SGKLM server as defined on the Key Servers window of the Enable Encryption wizard must be the server that is defined as the primary by SGKLM administrators.

The key server name serves only as a label. Only the provided IP address is used to contact the server. If the key server's TCP port number differs from the default value for the KMIP protocol (that is, 5696), enter the port number. An example of a complete primary SGKLM configuration is shown in Figure 12-93.

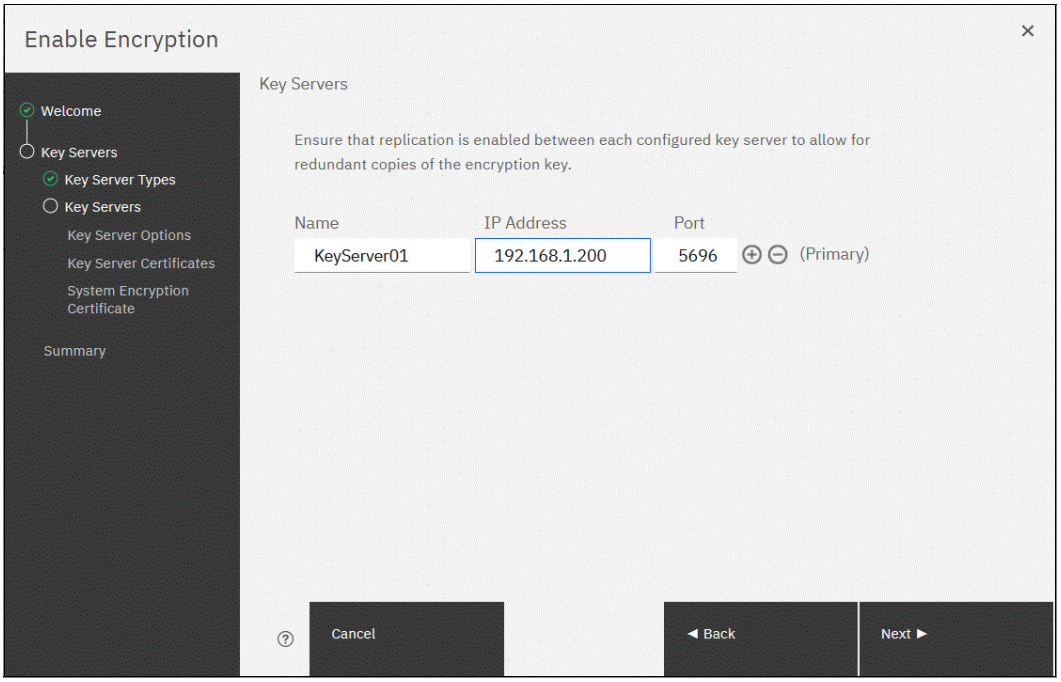


Figure 12-93 Configuration of the primary SGKLM server

5. If you want to add secondary SGKLM servers, click the “+” symbol and enter the data for secondary SGKLM servers, as shown in Figure 12-94. You can define up to four SGKLM servers. Click **Next** when you are done.

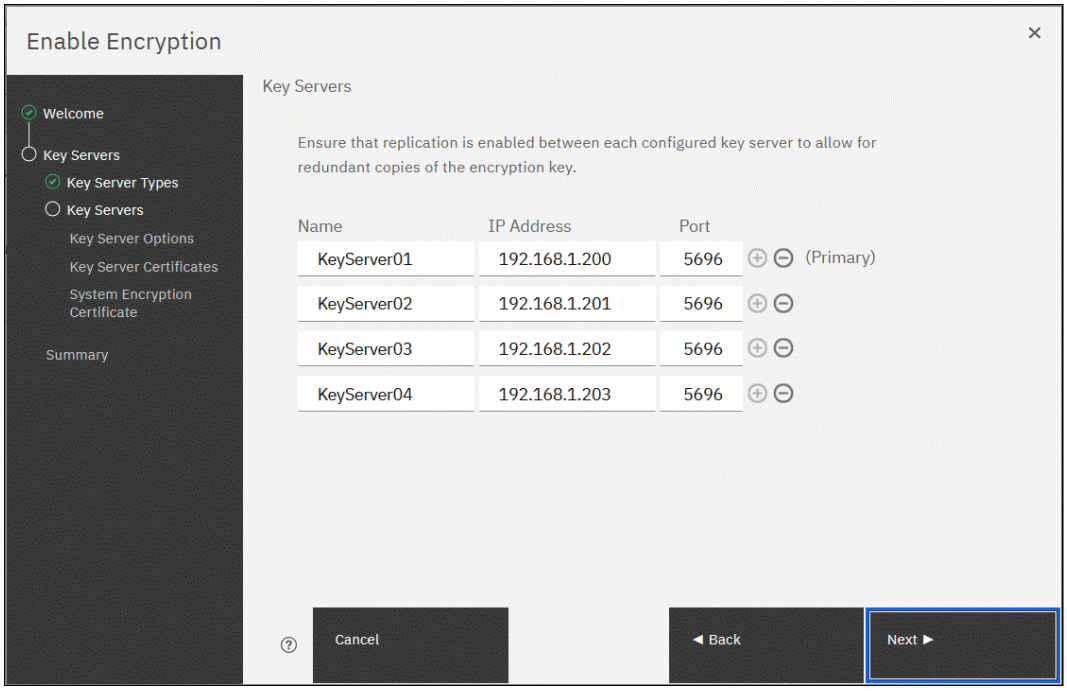


Figure 12-94 Configuring multiple SGKLM servers

6. The next window in the wizard is a reminder that IBM Storage_VIRT device group that is dedicated for IBM Storage Virtualize systems must exist on the SGKLM key servers. Make sure that this device group exists and click **Next** to continue, as shown in Figure 12-95.

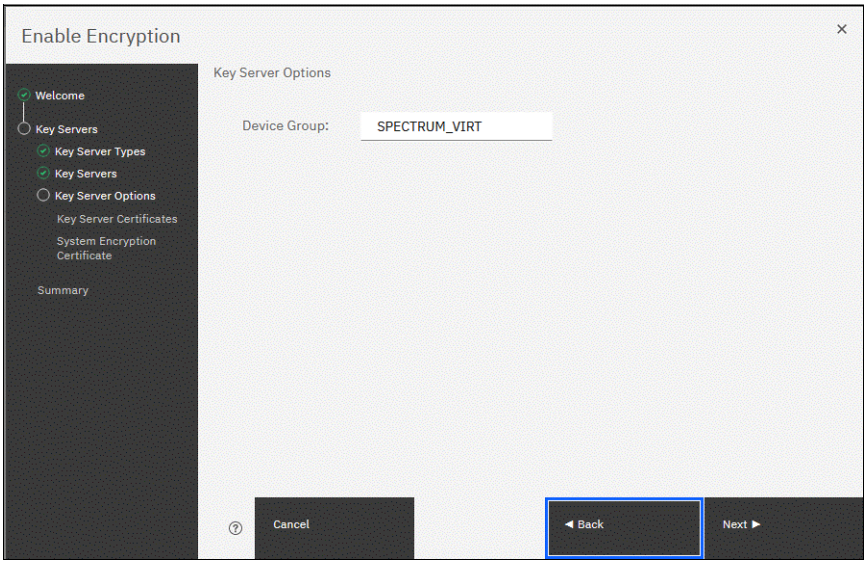


Figure 12-95 Checking key server device group

7. Enable secure communication between the IBM Storage Virtualize system and the SGKLM key servers by uploading the key server certificate from a trusted third-party certificate authority (CA) or by using a self-signed certificate. The self-signed certificate can be obtained directly from each of the key servers.

   Note: Key server TLS certificates to be imported into an IBM Storage Virtualize system need to be in PEM format (base64-encoded).

   After uploading any of the certificates in the window that is shown in Figure 12-96, click **Next**.

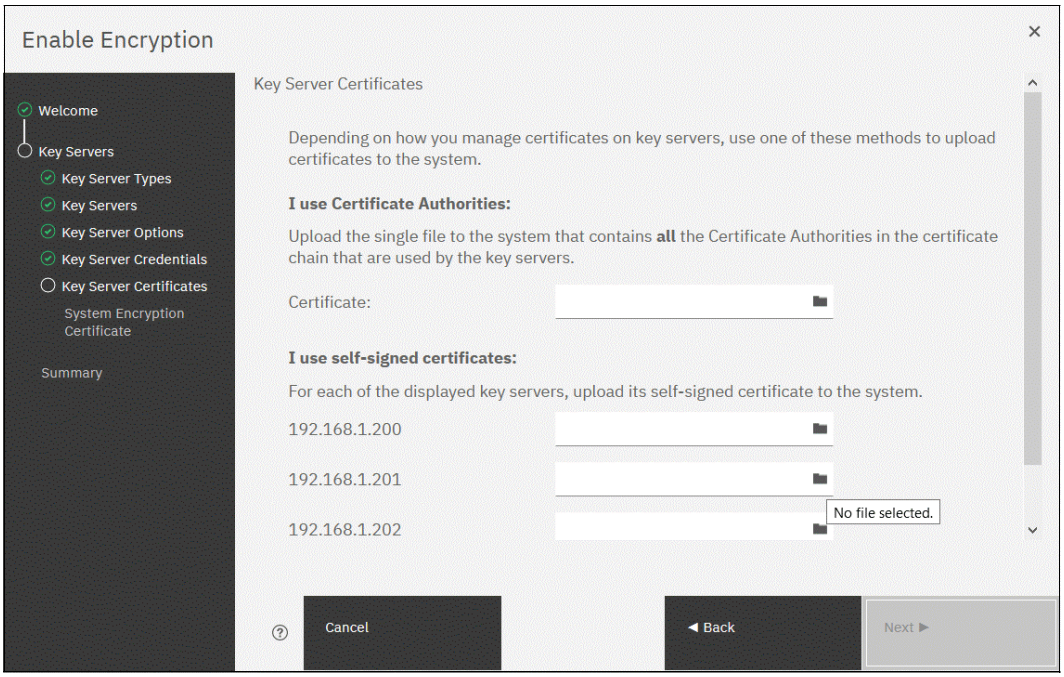


Figure 12-96 Uploading key servers or certificate authority SSL certificate

8. Configure the SGKLM key server to trust the public key certificate of the IBM Storage Virtualize system. You can download the system's public TLS certificate by clicking **Export Public Key**, as shown in Figure 12-97 on page 1125. Install this certificate in the SGKLM key server in the IBM Storage_VIRT device group.

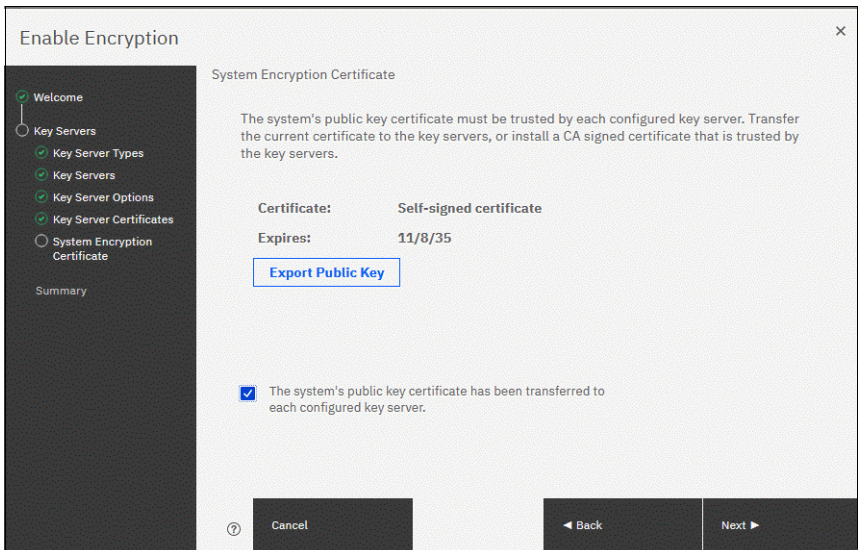


Figure 12-97 Downloading the IBM Storage Virtualize SSL certificate

9. When the IBM Storage Virtualize system public key certificate is installed on the SGKLM key servers, acknowledge this installation by clicking the checkbox that is below the Export Public Key button and then click **Next**.
10. The key server configuration is shown in the Summary tab, as shown in Figure 12-98. Click **Finish** to create the key server object and finalize the encryption enablement.



Figure 12-98 Finish enabling encryption by using SGKLM key servers

11. If no errors occur while the key server object is created, you receive a message that confirms that the encryption is now enabled on the system. Click **Close**.

12. Confirm that encryption is enabled in **Settings** → **Security** → **Encryption**, as shown in Figure 12-99. The `Online` state indicates which SGKLM servers are detected as available by the system.

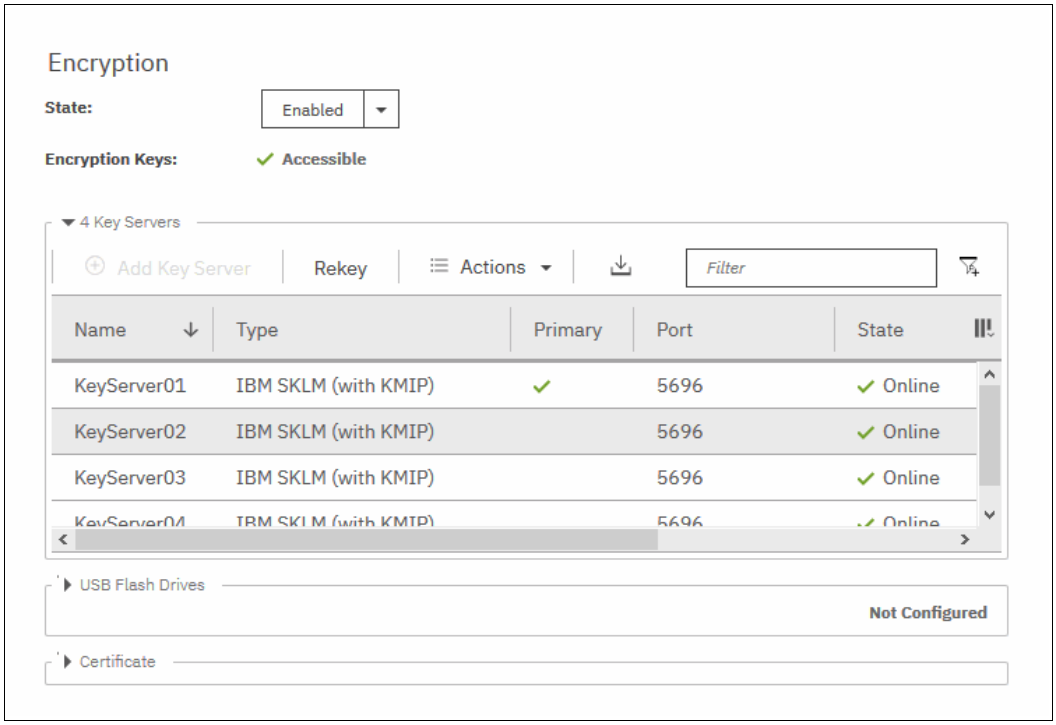


Figure 12-99 Encryption enabled with only SGKLM servers as encryption key providers

12.8.5 Enabling encryption by using SafeNet KeySecure or Thales CipherTrustManager

IBM Storage Virtualize V8.2.1 introduced support for Gemalto SafeNet KeySecure, which is a third-party key management server. It can be used as an alternative to SGKLM.

Starting with version 8.4.1, IBM Storage Virtualize also supports Thales CipherTrust Manager, which is a rebranded version of Gemalto SafeNet KeySecure. The instructions for configuring either version are the same.

IBM Storage Virtualize supports Gemalto SafeNet KeySecure version 8.3.0 and later, and by using only KMIP protocol. It is possible to configure up to four SafeNet KeySecure servers in IBM Storage Virtualize for redundancy (they can coexist with USB flash drive encryption).

It is not possible to have both SafeNet KeySecure and SGKLM key servers that are configured at the same time in IBM Storage Virtualize. It is also not possible to migrate directly from one type of key server to another (from SGKLM to SafeNet KeySecure or vice versa). If you want to migrate from one type to another, first migrate to USB flash drives encryption, and then migrate to the other type of key servers.

KeySecure uses an active-active clustered model. All changes to one key server are instantly propagated to all other servers in the cluster.

Although KeySecure uses KMIP protocol as IBM SGKLM does, an option is available to configure the username and password for IBM Storage Virtualize and KeySecure server authentication, which is not possible when the configuration is performed with SGKLM.

The certificate for client authentication in SafeNet KeySecure can be self-signed or signed by a CA.

To enable encryption in IBM Storage Virtualize by using a Gemalto SafeNet KeySecure key server, complete the following steps:

1. Ensure that the service IPs are configured on all of your nodes.
2. In the Enable Encryption wizard Welcome tab, select **Key servers** and then click **Next**, as shown in Figure 12-100.

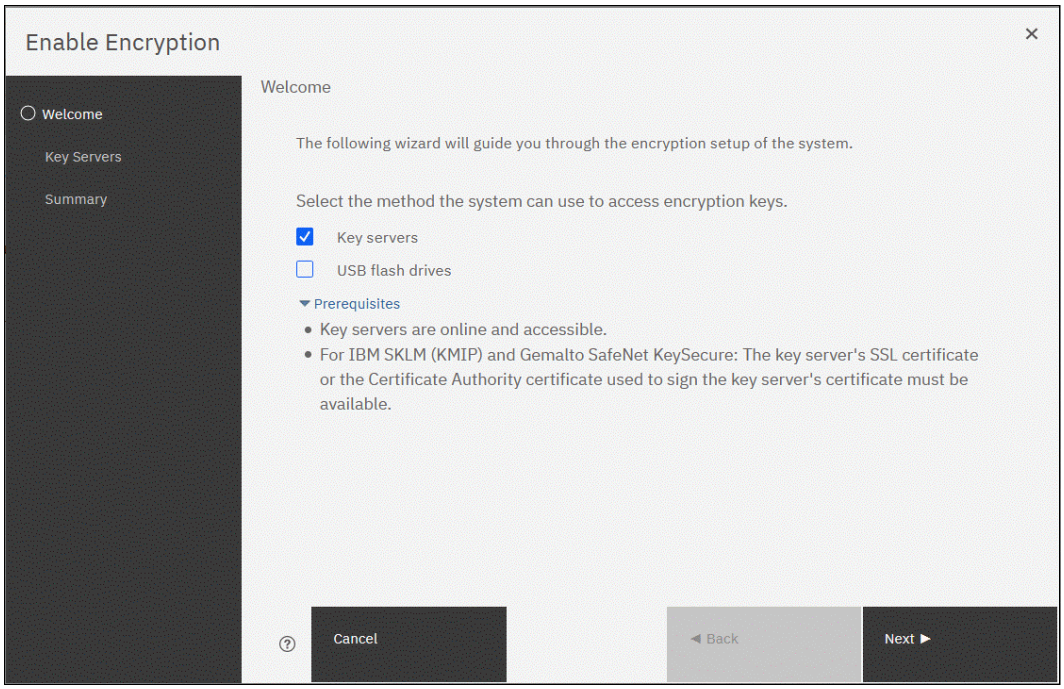


Figure 12-100 Selecting Key servers as the only provider in the Enable Encryption wizard

3. In the next window, you can choose between IBM SGKLM or Gemalto SafeNet KeySecure server types, as shown in Figure 12-101 on page 1128. Select **Gemalto SafeNet KeySecure** and then click **Next**.

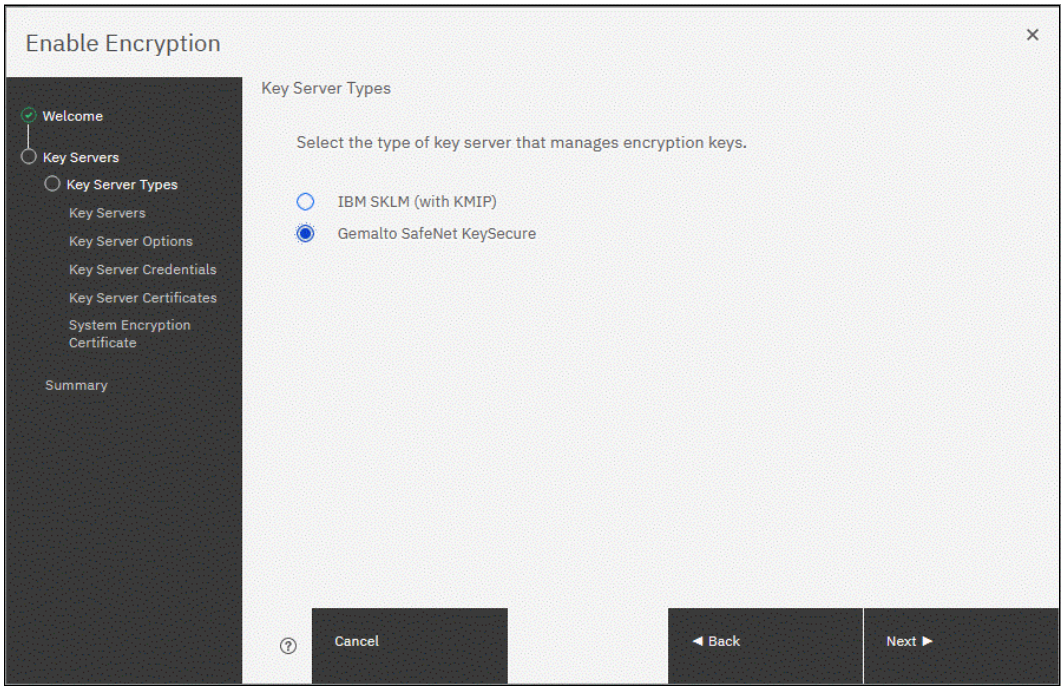


Figure 12-101 Selecting Gemalto SafeNet KeySecure as key server type

4. Add up to four SafeNet KeySecure servers in the next wizard window, as shown in Figure 12-102. For each key server, enter the name, IP address, and TCP port for KMIP protocol. The default value is 5696. Because the server name is only a label, it does not need to be the real hostname of the server.

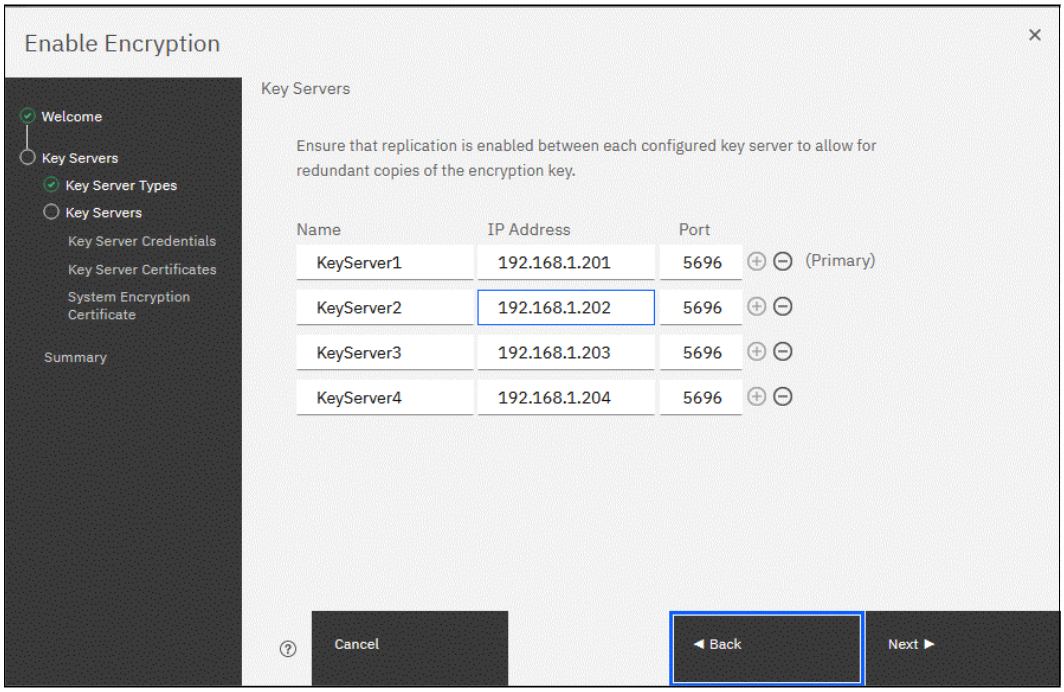


Figure 12-102 Configuring multiple SafeNet KeySecure servers

Although Gemalto SafeNet KeySecure uses an active-active clustered model, IBM Storage Virtualize prompts you for a primary key server. The primary key server

represents only the KeySecure server that is used for key create and rekey operations. Therefore, any of the clustered key servers can be selected as the primary.

Selecting a primary key server is beneficial for load balancing. Any four key servers can be used to retrieve the master key.

The next window in the wizard prompts you for the key server credentials (username and password), as shown in Figure 12-103. This setting is optional because it depends on how SafeNet KeySecure servers are configured.

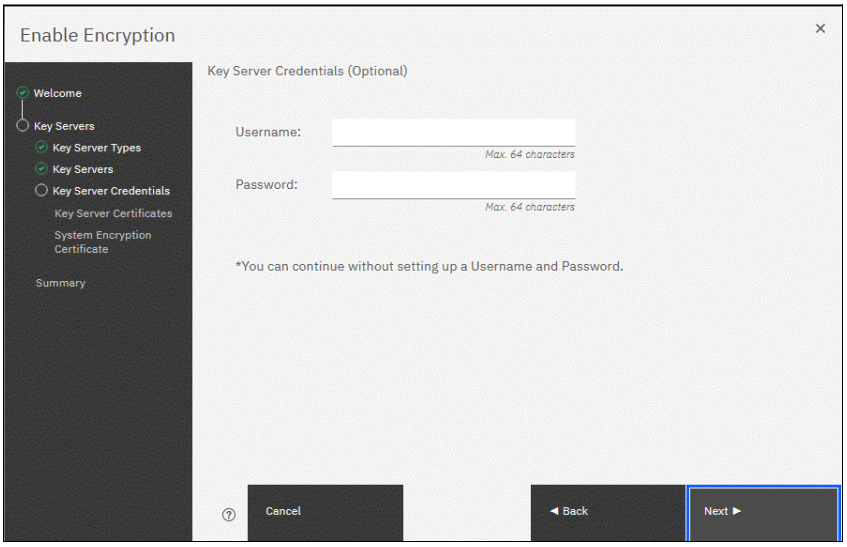


Figure 12-103 Key server credentials input (optional)

5. Enable secure communication between the IBM Storage Virtualize system and the SafeNet KeySecure key servers by uploading the key server certificate from a trusted third-party CA or by using a self-signed certificate. The self-signed certificate can be obtained from each of the key servers directly. After uploading any of the certificates in the window that is shown in Figure 12-104, click **Next**.

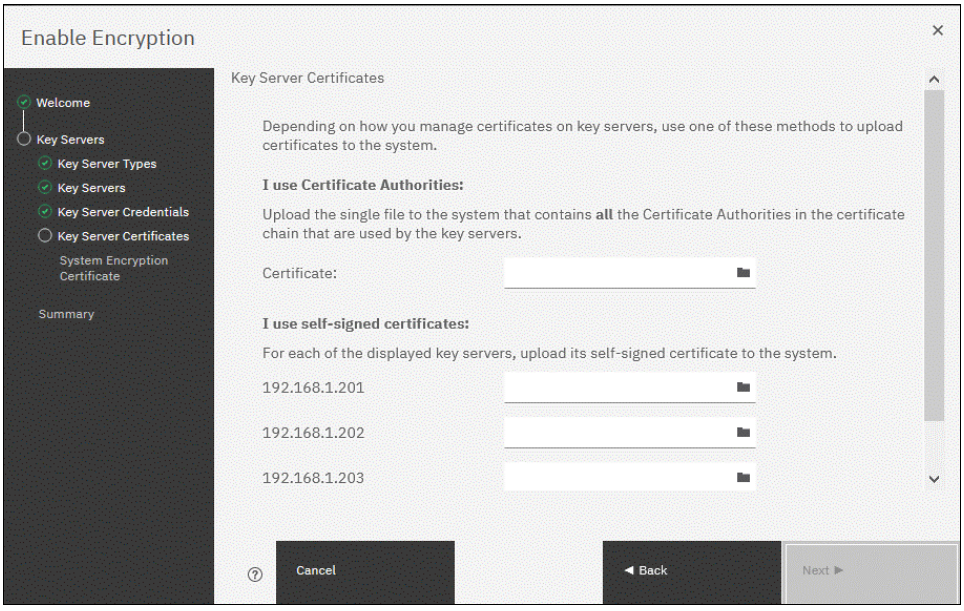


Figure 12-104 Uploading SafeNet KeySecure key servers certificate

6. Configure the SafeNet KeySecure key servers to trust the public key certificate of the IBM Storage Virtualize system. You can download the IBM Storage Virtualize system public SSL certificate by clicking **Export Public Key**, as shown in Figure 12-104 on page 1129. After adding the public key certificate to the key servers, select the checkbox and then click **Next**.

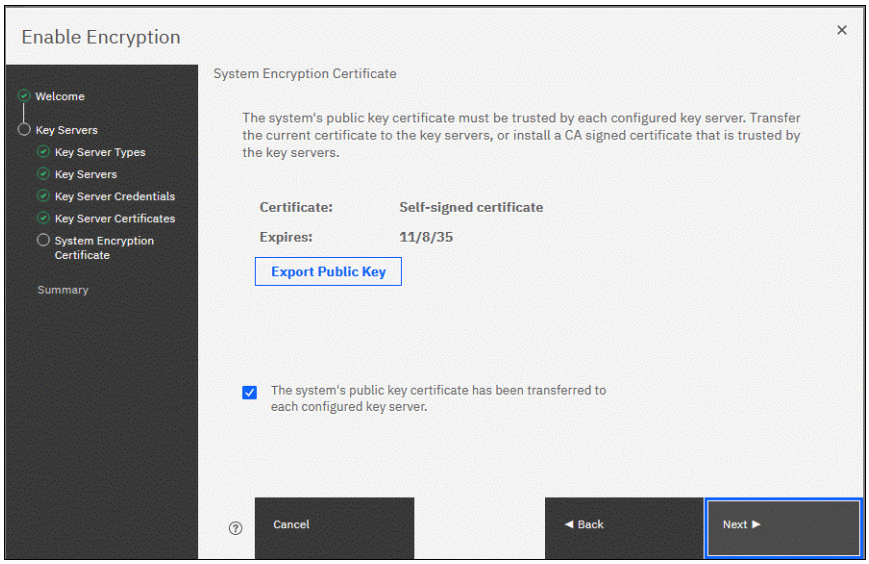


Figure 12-105 Downloading the IBM Storage Virtualize SSL certificate

7. The key server configuration is shown in the Summary tab, as shown in Figure 12-106. Click **Finish** to create the key server object and finalize the encryption enablement.



Figure 12-106 Finish enabling encryption by using SafeNet KeySecure key servers

8. If no errors occurred while creating the key server object, you receive a message that confirms that the encryption is now enabled on the system. Click **Close**.
9. Confirm that encryption is enabled in **Settings** → **Security** → **Encryption**, as shown in Figure 12-107 on page 1131. Check whether the four servers are shown as online, which indicates that all four SafeNet KeySecure servers are detected as available by the system.

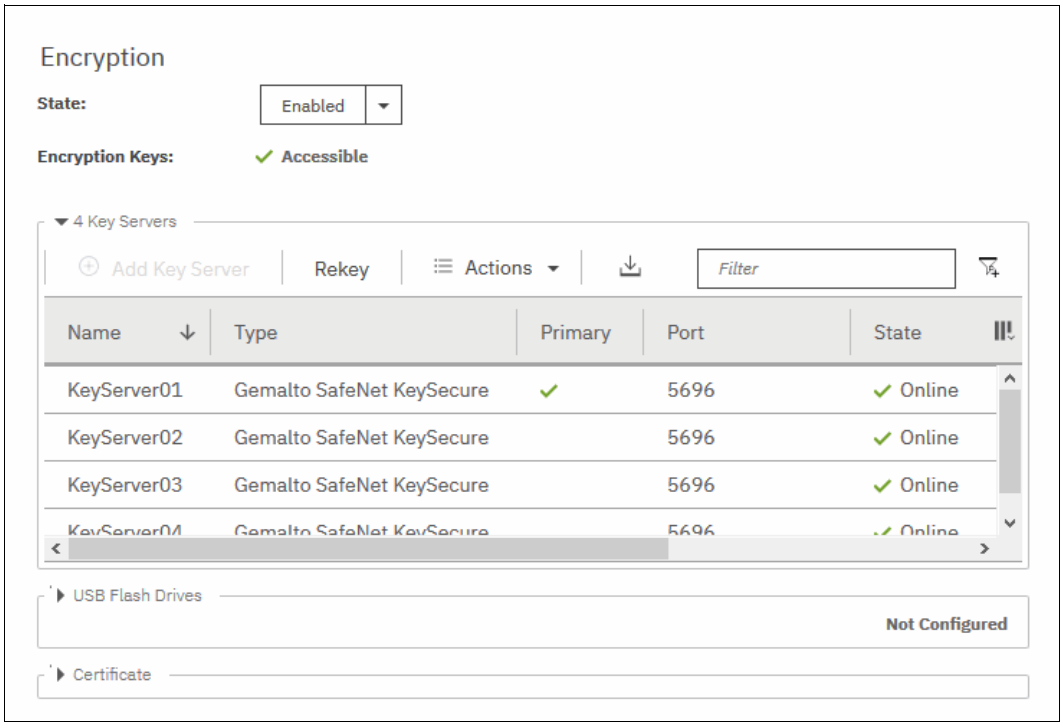


Figure 12-107 Encryption enabled with 4 SafeNet KeySecure key servers

12.8.6 Enabling encryption by using both providers

IBM Storage Virtualize allows parallel use of USB flash drives and one type of key server as encryption key providers. It is possible to configure both providers at the same time by using the enable encryption wizard. To perform this configuration process, the system must meet the requirements of the key server and USB flash drive encryption key providers.

Note: Make sure that the key management server function is fully independent from an encrypted storage that includes encryption that is managed by this key server environment. Failure to observe this requirement might create an encryption deadlock. An *encryption deadlock* is a situation in which none of the key servers in the environment can become operational because some critical part of the data in each server is stored on an encrypted storage system that depends on one of the key servers to unlock access to the data.

IBM Storage Virtualize code V8.1 and later supports up to four key server objects that are defined in parallel.

Before you start to enable encryption by using USB flash drives and a key server, confirm the requirements that are described in “Enabling encryption by using USB flash drives” on page 1115, and 12.8.3, “Enabling encryption by using key servers” on page 1119.

To enable encryption by using a key server and USB flash drive, complete the following steps:

1. Ensure that service IPs are configured on all your nodes.
2. In the Enable Encryption wizard Welcome tab, select **Key servers** and **USB flash drives** and then click **Next** (see Figure 12-108 on page 1132).

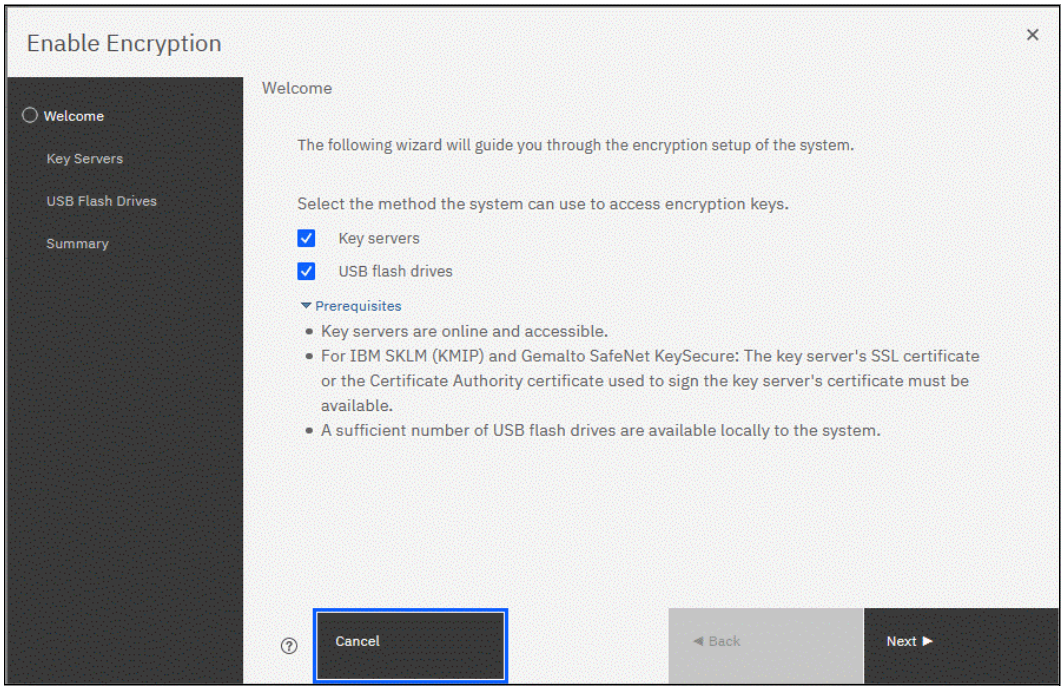


Figure 12-108 Selecting Key servers and USB flash drives in the Enable Encryption wizard

3. The Key Server Types window opens, as shown in Figure 12-109. Select the key server type that manages the encryption keys.

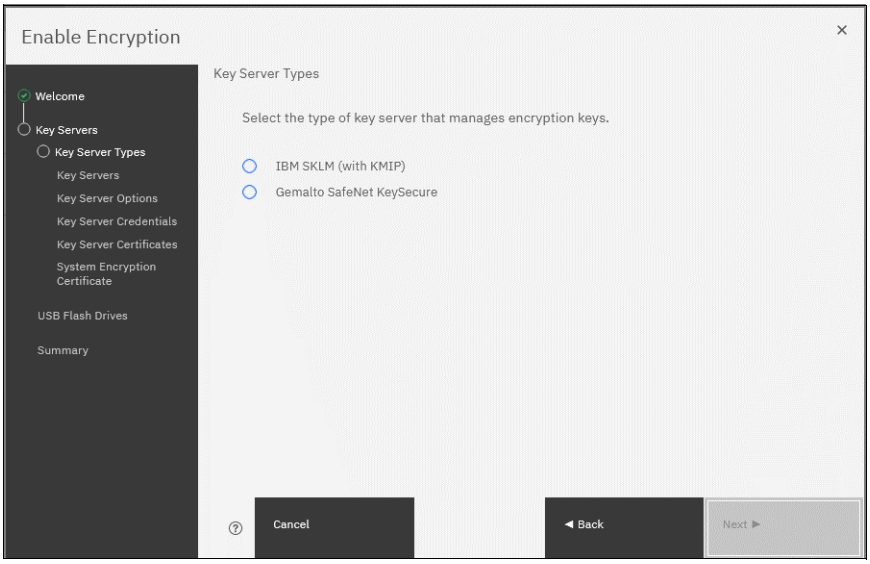


Figure 12-109 Selecting the key server type

The next windows that are displayed are the same as described in 12.8.3, “Enabling encryption by using key servers” on page 1119, depending on the type of key server selected.

When the key servers details are entered, the USB flash drive encryption configuration is displayed. In this step, master encryption key copies are stored in the USB flash drives. If fewer than three drives are detected, the system prompts you to insert more USB flash drives (see Figure 12-110). You cannot proceed until the required minimum number of USB flash drives is detected by the system.

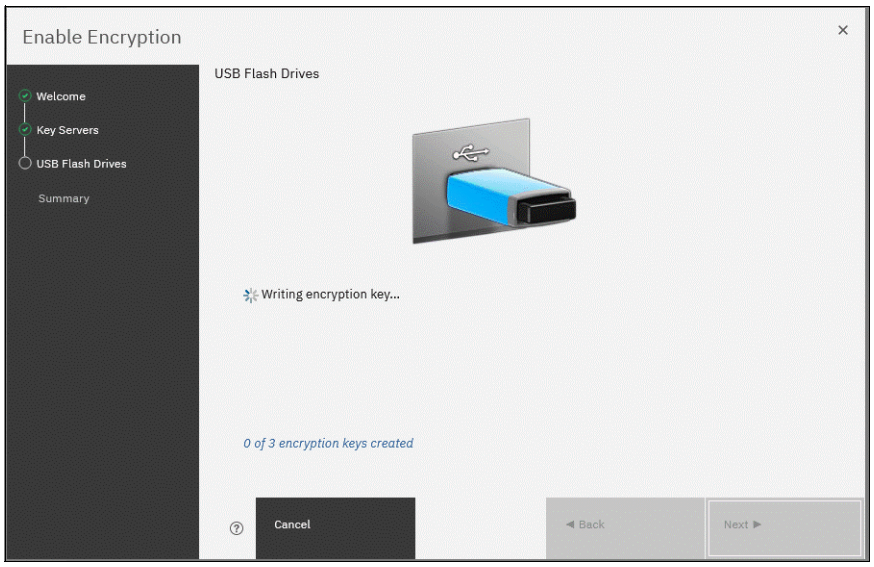


Figure 12-110 Prompt to insert USB flash drives

After at least three USB flash drives are detected, the system writes the master access key to each of the drives (see Figure 12-110). Notice that the system attempts to write the encryption key to any flash drive it detects. Therefore, it is crucial to maintain the physical security of the system during this procedure.

4. After copying encryption keys to USB flash drives, a window is shown that includes the summary of the configuration that is implemented on the system. Click **Finish** to create the key server object and finalize the encryption enablement.
5. If no errors occur while creating the key server object, the system displays a window in which it is confirmed that the encryption is now enabled on the system and that both encryption key providers are enabled.
6. You can confirm that encryption is enabled and verify which key providers are in use by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-111 on page 1134. Notice the `Online` state of the key servers and `Validated` state of the USB ports where USB flash drives are inserted to ensure that they are correctly configured.

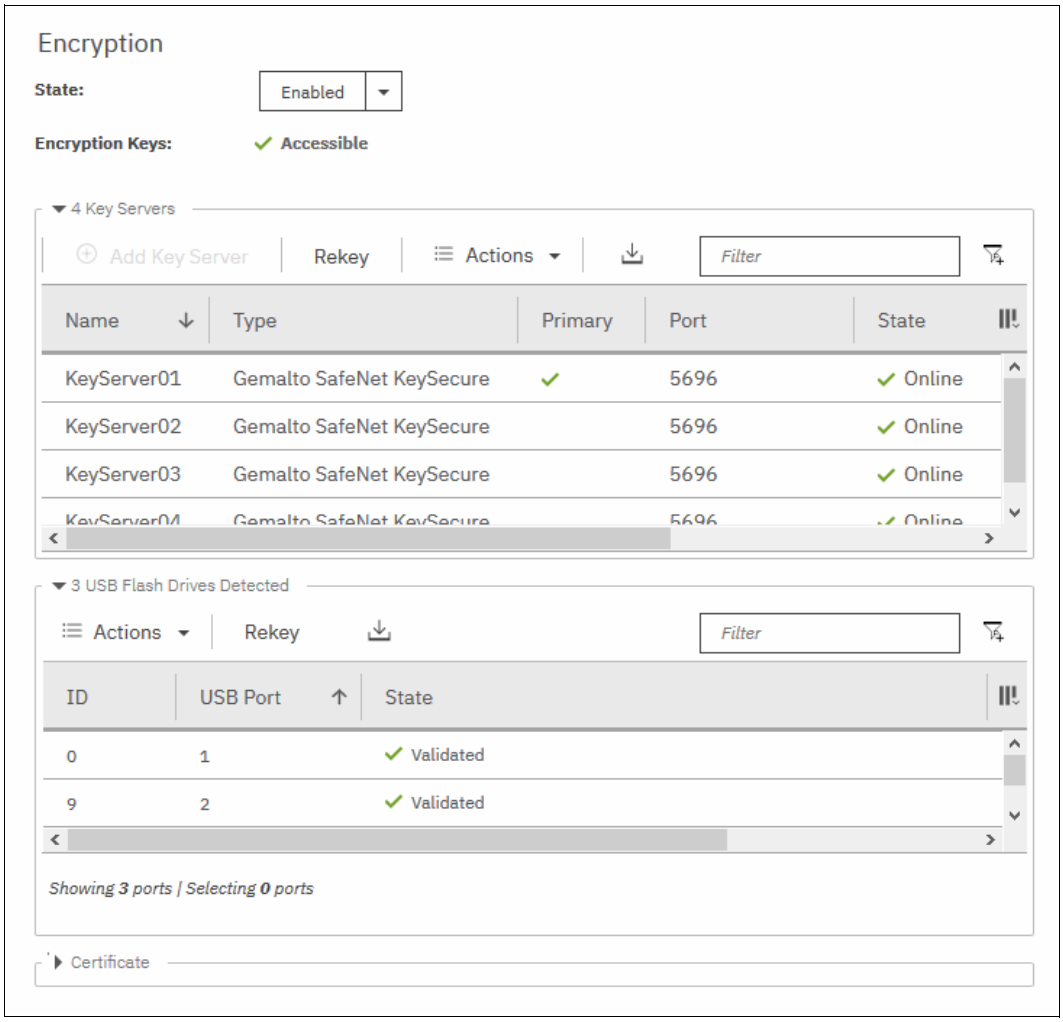


Figure 12-111 Encryption enabled with both USB flash drives and key servers

12.8.7 Configuring more providers

After the system is configured with a single encryption key provider, a second provider can be added.

Note: If you set up encryption of your storage system when it was running IBM Storage Virtualize code version earlier than V7.8.0, you must rekey the master encryption key before you can enable second encryption provider when you upgrade to code version V8.1 or later.

Adding key servers as a second provider

If the storage system is configured with the USB flash drive provider, it is possible to configure SGKLM or SafeNet KeySecure servers as a second provider.

To enable key servers as a second provider, complete the following steps:

1. Select **Settings** → **Security** → **Encryption**. Expand the Key Servers section and click **Configure**, as shown in Figure 12-112. To enable the key server as a second provider, the system must detect at least one USB flash drive with a current copy of the master access key.

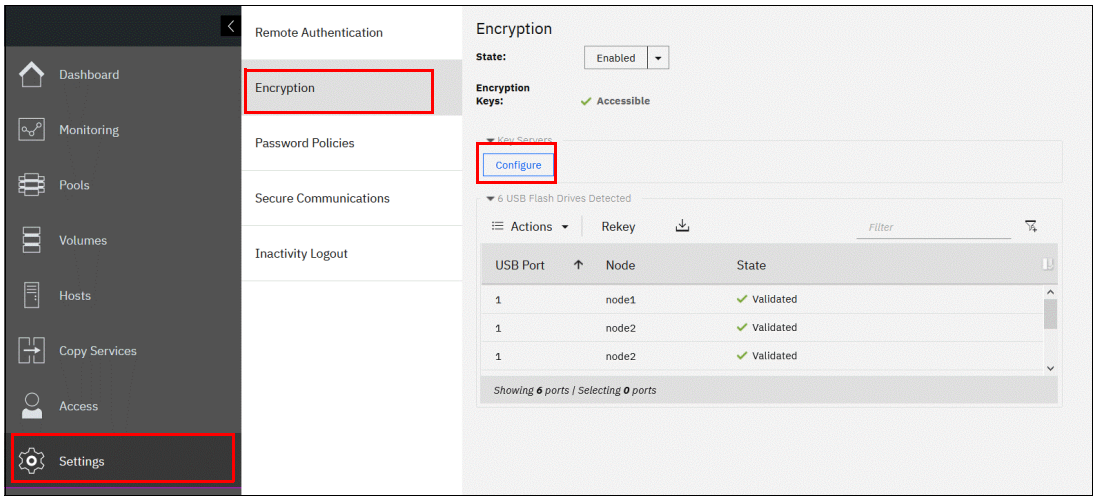


Figure 12-112 Enable key servers as a second provider

2. Complete the steps that are required to configure the key server provider, as described in 12.8.3, “Enabling encryption by using key servers” on page 1119. The difference in the process that is described in that section is that the wizard gives you an option to disable USB flash drive encryption, which aims to migrate from the USB flash drive to key server provider.

   Select **No** to enable both encryption key providers, as shown in Figure 12-113.

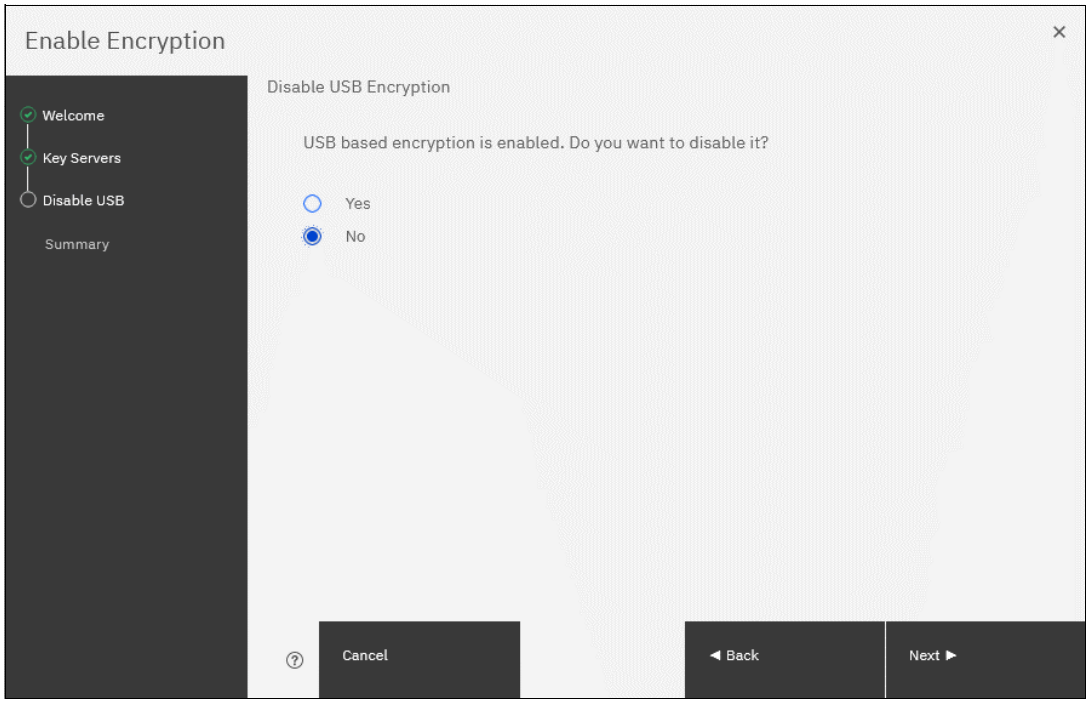


Figure 12-113 Do not disable USB flash drive encryption key provider

This choice is confirmed on the summary window before the configuration is committed, as shown in Figure 12-114.



Figure 12-114 Configuration summary before committing

3. After you click **Finish**, the system configures key servers as a second encryption key provider. Successful completion of the task is confirmed by a message. Click **Close**.
4. You can confirm that encryption is enabled and verify which key providers are in use by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-115. Notice the `Online` state of the key servers and the `Validated` state of the USB ports where USB flash drives are inserted to make sure that they are correctly configured.

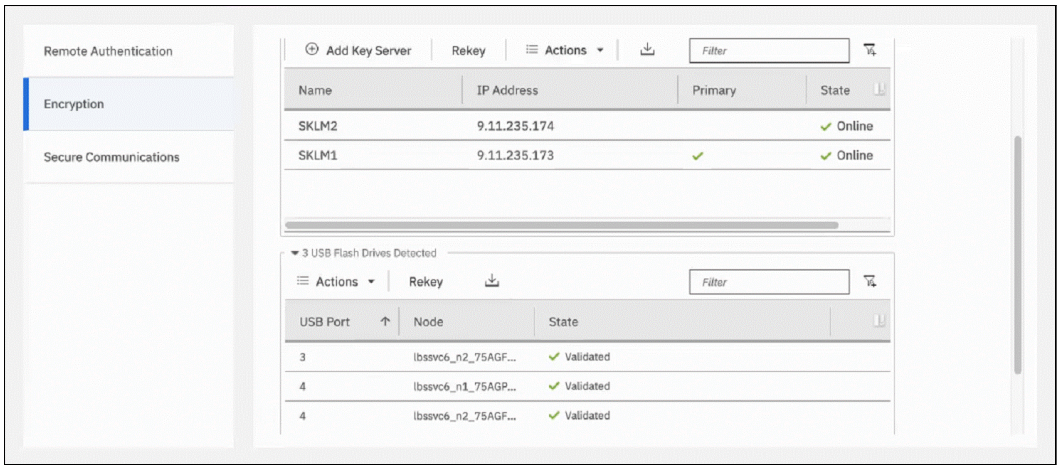


Figure 12-115 Encryption enabled with two key providers available

12.8.8 Adding USB flash drives as a second provider

If the storage system is configured with an SGKLM or SafeNet KeySecure encryption key provider, it is possible to configure USB flash drives as a second provider.

To enable USB flash drives as a second provider, complete the following steps:

1. Select **Settings** → **Security** → **Encryption**. Expand the **USB Flash Drives** section and click **Configure**. To enable USB flash drives as a second provider, the system must access key servers with the current master access key.
2. After you click **Configure**, you are presented with a wizard that is similar to the one that is described in “Enabling encryption by using USB flash drives” on page 1115. You cannot disable key server providers during this process.
3. After successful completion of the process, a message is displayed that confirms that both encryption key providers are enabled.
4. You can confirm that encryption is enabled and verify which key providers are in use by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-116. Notice the `Online` state of the key servers and the `Validated` state of the USB ports where USB flash drives are inserted to make sure that they are correctly configured.

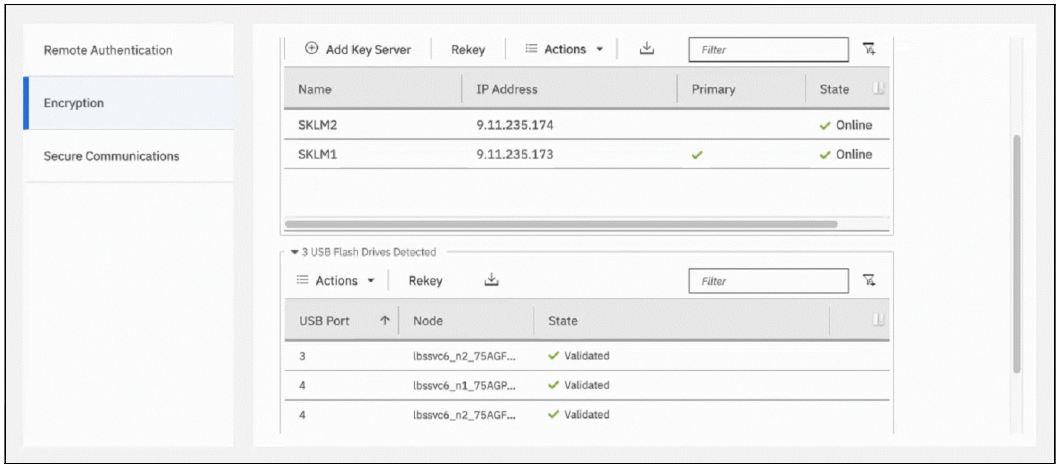


Figure 12-116 Example of encryption enabled with two key providers available

12.8.9 Migrating between providers

IBM Storage Virtualize V8.1 introduced support for simultaneous use of USB flash drives and a key server as encryption key providers. The system also allows migration from configuration by using only USB flash drive provider to key servers provider, and vice versa.

If you want to migrate from one key server type to another (for example, migrating from SGKLM to SafeNet KeySecure or vice versa), direct migration is not possible. In this case, it is required first to migrate from the current key server type to a USB flash drive, and then migrate to the other type of key server.

Changing from USB flash drive provider to encryption key server

The system is designed to facilitate changing from USB flash drives encryption key provider to encryption key server provider.

If you follow the steps that are described in “Adding key servers as a second provider” on page 1134, but when completing step 2 on page 1135, select **Yes** instead of **No** (see

Figure 12-117). This action deactivates the USB flash drives provider, and the procedure completes with only key servers that are configured as key provider.

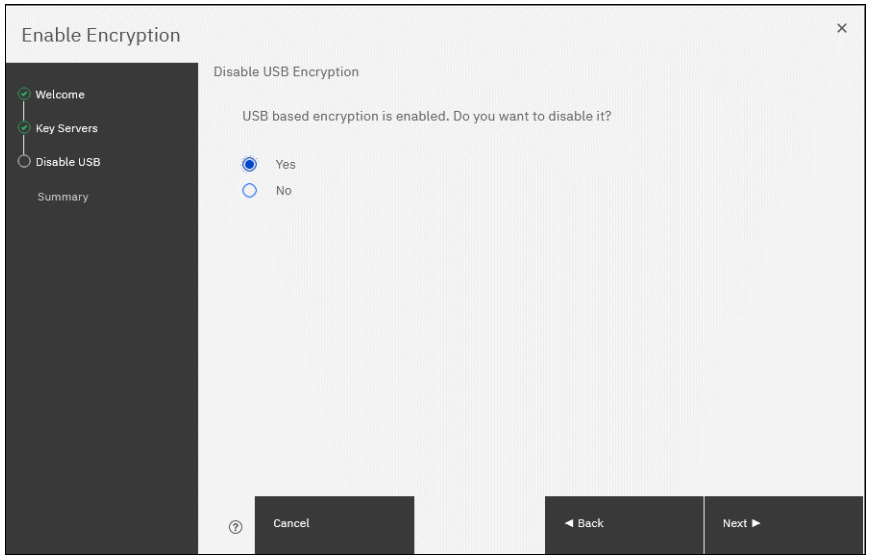


Figure 12-117 Disable USB flash drive provider while changing to SGKLM provider

12.8.10 Changing from encryption key server to USB flash drive provider

Change in the other direction (that is, from the use of encryption key servers provider to USB flash drives provider) is not possible by using only the GUI.

To change the direction, add USB flash drives as a second provider by completing the steps described in 12.8.8, “Adding USB flash drives as a second provider” on page 1137.

Then run the following command in the CLI:

```
chencryption -usb validate
```

To make sure that USB drives contain the correct master access key, disable the encryption key server provider by running the following command:

```
chencryption -keyserver disable
```

This command disables the encryption key server provider, which effectively migrates your system from encryption key server to USB flash drive provider.

12.8.11 Migrating between different key server types

The migration between different key server types cannot be performed directly from one type of key server to another. USB flash drives encryption must be used to facilitate this process.

If you want to migrate from one type of key server to another, you first must migrate from your current key servers to USB encryption, and then migrate from USB to the other type of key servers.

The procedure to migrate from one key server type to another is shown here. In this example, we migrate an IBM Storage Virtualize system that is servers that are configured with IBM SGKLM key servers (Figure 12-118) to SafeNet KeySecure servers.

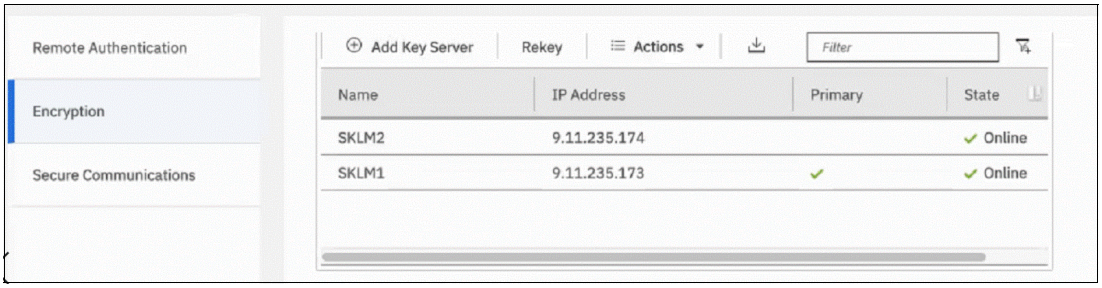


Figure 12-118 IBM Storage Virtualize encryption configured with IBM SGKLM servers

Complete the following steps to migrate to Gemalto SafeNet KeySecure:

1. Migrate from key server encryption to USB flash drives encryption, as described in “Changing from encryption key server to USB flash drive provider” on page 1138. After this step, only USB flash drives encryption is configured, as shown in Figure 12-119.

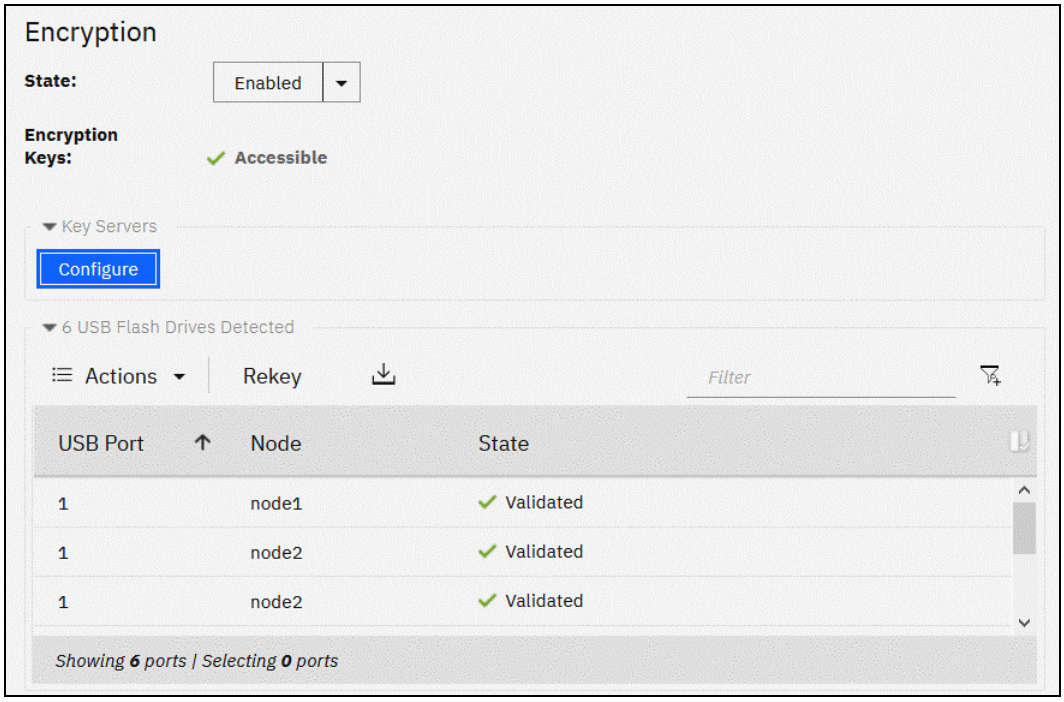


Figure 12-119 IBM SAN Volume Controller encryption configured with USB flash drives

2. Migrate from USB flash drives encryption to the other key server type encryption (in this example, Gemalto SafeNet KeySecure), following the steps that are described in “Changing from USB flash drive provider to encryption key server” on page 1137. After completing this step, the other key server type is configured as an encryption provider in IBM Storage Virtualize, as shown in Figure 12-120.

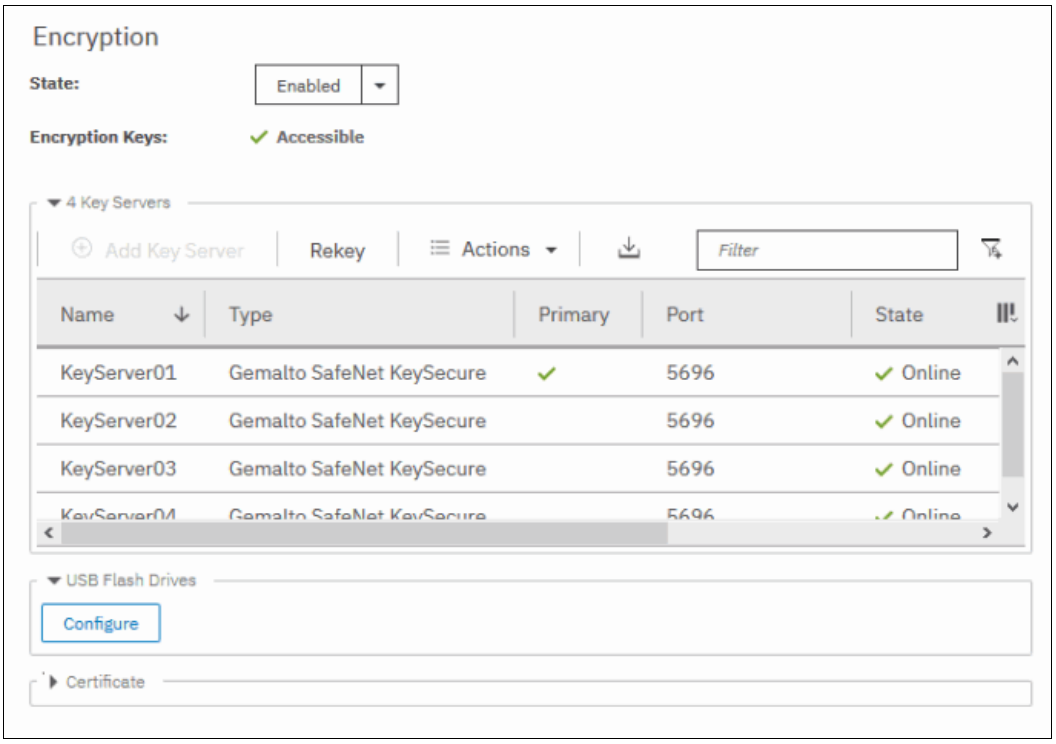


Figure 12-120 IBM SAN Volume Controller encryption configured with SafeNet KeySecure

12.8.12 Recovering from a provider loss

If both encryption key providers are enabled, and you lose one of them (by losing all copies of the encryption key that is kept on the USB flash drives or by losing all SGKLM servers), you can recover from this situation by disabling the provider to which you lost the access. To disable the unavailable provider, you must have access to a valid master access key on the remaining provider.

If you lost access to the encryption key server provider, run the following command:

```
chencryption -keyserver disable
```

If you lost access to the USB flash drives provider, run the following command:

```
chencryption -usb disable
```

If you want to restore the configuration with both encryption key providers, follow the instructions that are described in 12.8.7, “Configuring more providers” on page 1134.

Note: If you lose access to all encryption key providers that are defined in the system, no method is available to recover access to the data that was protected by the master access key.

12.9 Using encryption

The design for encryption is based on the concept that a system is fully encrypted or not encrypted. Encryption implementation is intended to encourage solutions that contain only encrypted volumes or only unencrypted volumes. For example, after encryption is enabled on the system, all new objects (for example, pools) are created as encrypted by default.

Some unsupported configurations are actively policed in the code. For example, no support exists for creating unencrypted child pools from encrypted parent pools. However, consider the following exceptions:

- During the migration of volumes from unencrypted to encrypted volumes, a system might report encrypted and unencrypted volumes.
- It is possible to create unencrypted arrays from CLI by manually overriding the default encryption setting.

Notes: Encryption support for Distributed Redundant Array of Independent Disks (DRAID) is available in IBM Storage Virtualize code V7.7 and later.

You must decide whether to encrypt or not encrypt an object when it is created. You cannot change this setting later. To change the encryption state of stored data, you must migrate from an encrypted object, for example, a pool, to an unencrypted one, or vice versa. Volume migration is the only way to encrypt any volumes that were created before enabling encryption on the system.

12.9.1 Encrypted pools

For more information about how to open the Create Pool window, see Chapter 5, “Using storage pools” on page 365.

After encryption is enabled, any new pool is created by default as encrypted, as shown in Figure 12-121.

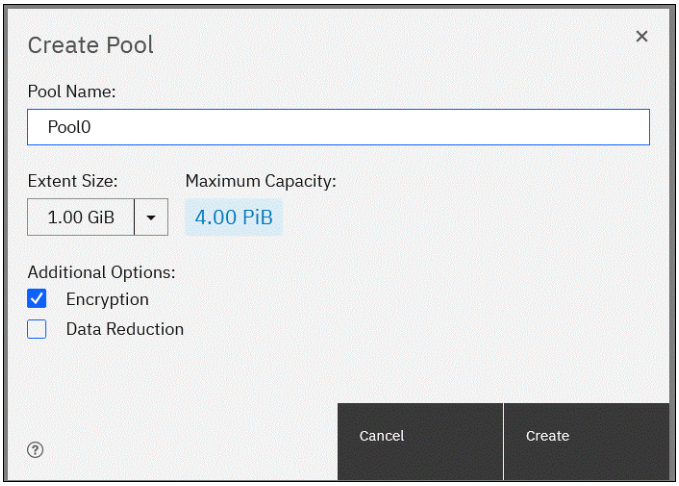


Figure 12-121 Create Pool window basic

You can click **Create** to create an encrypted pool. All storage that is added to this pool is encrypted.

You can customize the Pools view in the management GUI to show pool encryption status. Select **Pools** → **Pools**, and then select **Actions** → **Customize Columns** → **Encryption**, as shown in Figure 12-122.

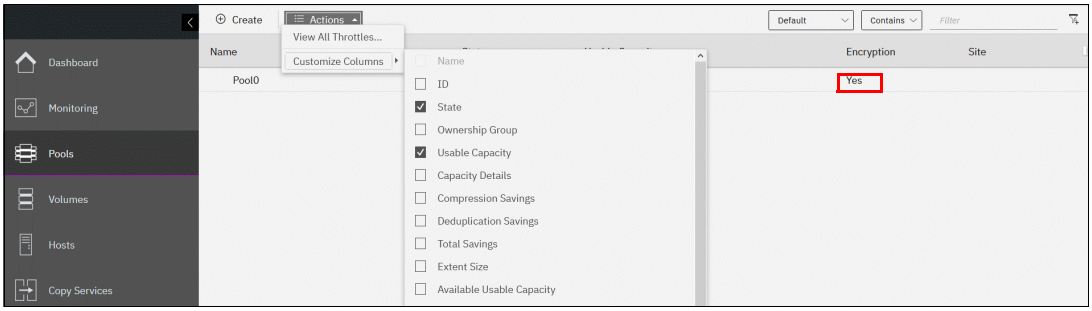


Figure 12-122 Pool encryption state

If you create an unencrypted pool but you add only encrypted arrays or self-encrypting MDisks to the pool, the pool is reported as `encrypted` because all extents in the pool are encrypted. The pool reverts to the `unencrypted` state if you add an unencrypted array or MDisk.

By default, if encryption is enabled on the storage, newly added internal MDisks (arrays) are created as encrypted and the pool is reported as encrypted, unless any unencrypted MDisks exist in the pool.

Important: Unencrypted pools allow encrypted and unencrypted MDisks to be added in one pool. If you remove all unencrypted MDisks from an unencrypted pool so that the pool contains only encrypted MDisks, the pool is reported as encrypted.

Data is encrypted based on the MDisks encryption status. A pool with encrypted MDisks allows unencrypted MDisks to be added. If unencrypted MDisks are added, the pool reverts to its unencrypted state. In this case, data partially ends up on an unencrypted MDisk.

In general, mixing MDisks with different encryption status should be avoided. However, it can be allowed for temporary migration scenarios.

Adding encrypted storage to encrypted pools is described in 12.9.2, “Encrypted child pools” on page 1143. You can mix and match storage encryption types in a pool. Figure 12-123 on page 1143 shows an example of an encrypted pool that contains storage by using different encryption methods.

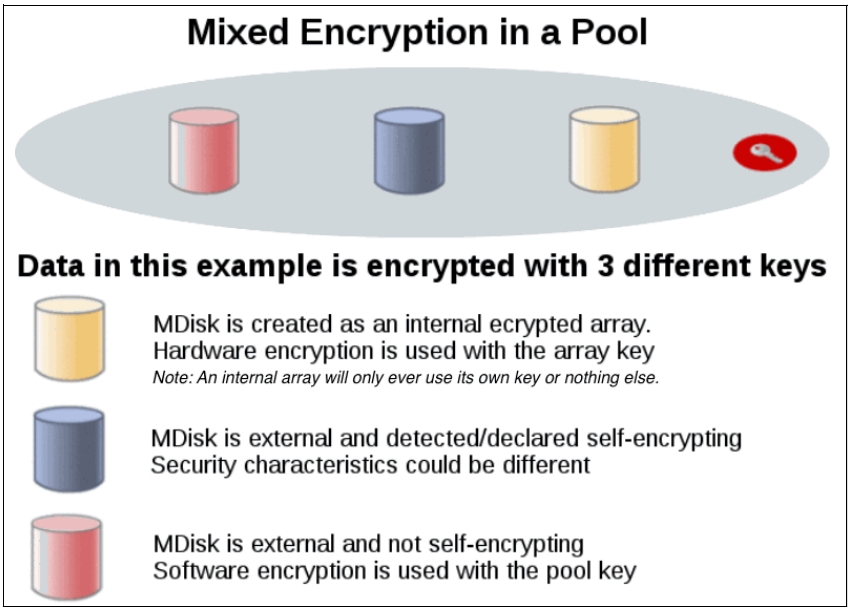


Figure 12-123 Mix and match encryption in a pool

12.9.2 Encrypted child pools

For more information about how to open the Create Child Pool window, see Chapter 5, “Using storage pools” on page 365.

If the parent pool is encrypted, every child pool must also be encrypted. The GUI enforces this requirement by automatically selecting **Encryption Enabled** in the Create Child Pool window and preventing changes to this setting, as shown in Figure 12-124.

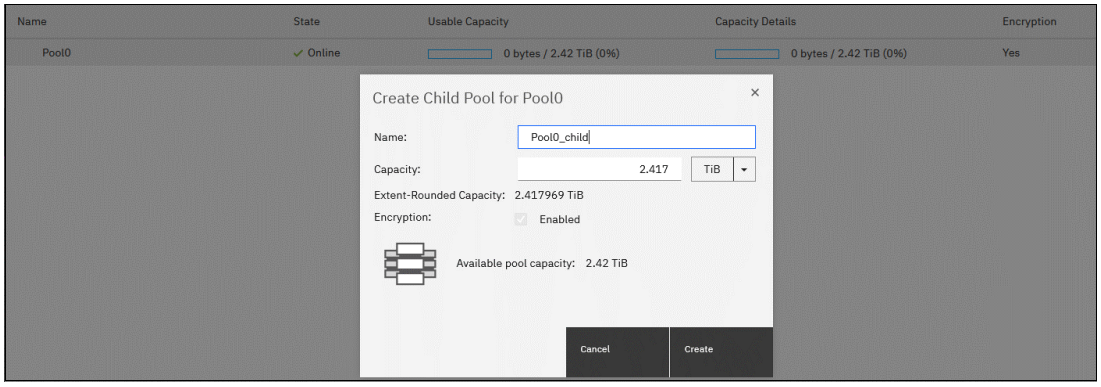


Figure 12-124 Create a child pool of an encrypted parent pool

However, if you want to create encrypted child pools from an unencrypted storage pool that contains a mix of internal arrays and external MDisks, the following restrictions apply:

- The parent pool cannot contain any unencrypted internal arrays. If any unencrypted internal array is in the unencrypted pool, it is created as unencrypted when you try to create a child pool and select the option to set as encrypted.
- All IBM SAN Volume Controller nodes in the system must support software encryption and have the encryption license activated.

Note: An encrypted child pool that is created from an unencrypted parent storage pool reports as unencrypted if the parent pool contains any unencrypted internal arrays. Remove these arrays to ensure that the child pool is fully encrypted.

If you modify the Pools view, you see the encryption status of child pools, as shown in Figure 12-125. The example shows an encrypted child pool with nonencrypted parent pool.

.

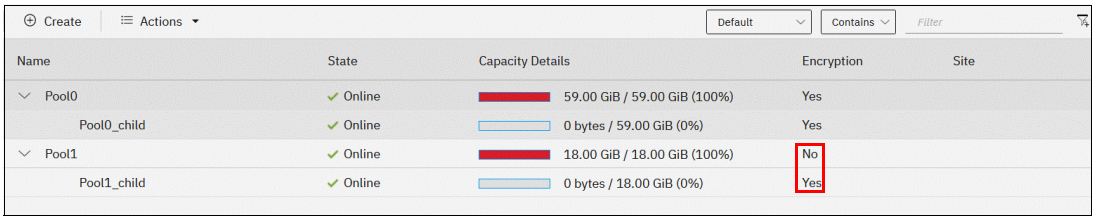


Figure 12-125 Child pool encryption state

12.9.3 Encrypted arrays

For more information about how to add internal storage to a pool, see Chapter 5, “Using storage pools” on page 365.

After encryption is enabled, all newly built arrays are hardware-encrypted by default. In this example, the GUI does not allow you to create an unencrypted array. To create an unencrypted array, the CLI must be used.

Example 12-9 shows how to create an unencrypted array by using the CLI.

Example 12-9 Creating an unencrypted array by using CLI with IBM SAN Volume Controller

```
IBM_SAN:ITSO-SVC:superuser>svctask mkarray -drive 6:4 -level raid1 -sparegoal 0
-strip 256 -encrypt no Pool2
MDisk, id [2], successfully created
IBM_SAN:ITSO-SVC:superuser>
```

Note: It is not possible to add unencrypted arrays to an encrypted pool.

You can customize the MDisks by Pools view to show array encryption status. Select **Pools** → **MDisk by Pools**, and then click **Actions** → **Customize Columns** → **Encryption**. You can also right-click the table header to customize columns and select **Encryption**, as shown in Figure 12-126.

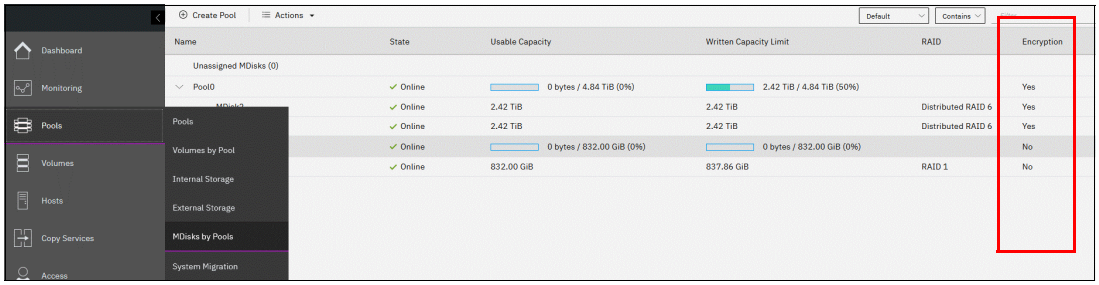


Figure 12-126 Array encryption state

You can also check the encryption state of an array by reviewing its drives in **Pools** → **Internal Storage** view. The internal drives that are associated with an encrypted array are assigned an encrypted property that can be seen, as shown in Figure 12-127.

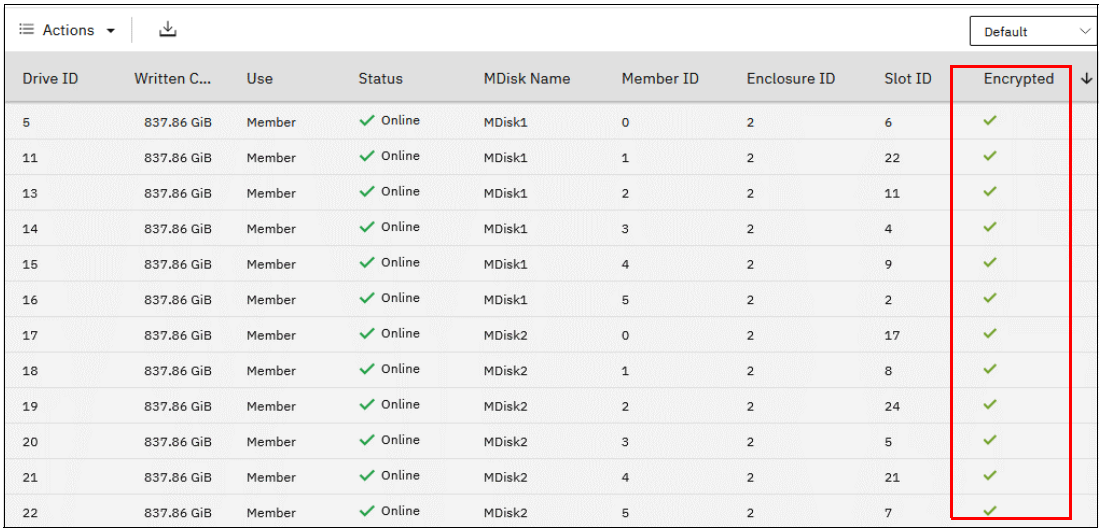


Figure 12-127 Drive encryption state

12.9.4 Encrypted MDisks

For more information about how to add external storage to a pool, see Chapter 5, “Using storage pools” on page 365.

Each MDisk that belongs to external storage that is added to an encrypted pool or child pool is automatically encrypted by using the pool or child pool key, unless the MDisk is detected or declared as self-encrypting.

The user interface gives no method to see which extents contain encrypted data and which do not. However, if a volume is created in a correctly configured encrypted pool, all data that is written to this volume is encrypted.

You can use the MDisk by Pools view to show the object encryption state by selecting **Pools** → **MDisk by Pools**. Figure 12-128 shows an example in which self-encrypting MDisk is in an unencrypted pool.

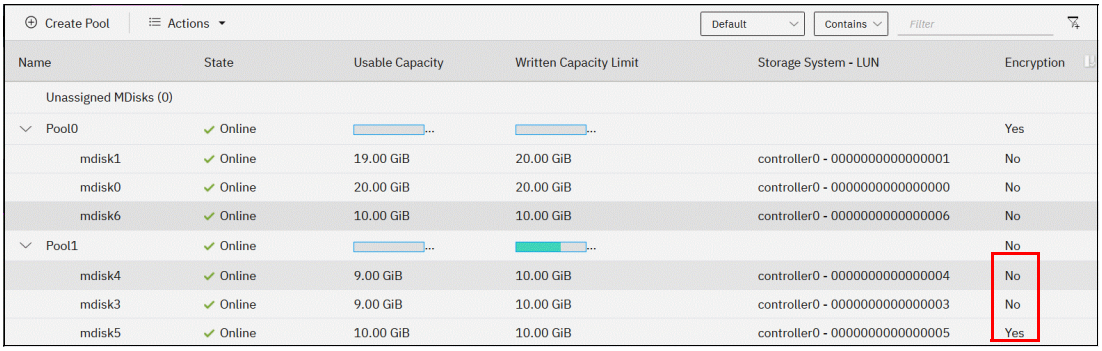


Figure 12-128 MDisk encryption state

Note: When working with MDisks encryption, take extra care when configuring MDisks and pools.

If the MDisk was used earlier for storage of unencrypted data, the extents can contain stale unencrypted data. This issue occurs because file deletion only checkmarks disk space as free. The data is *not* removed from the storage. Therefore, if the MDisk is not self-encrypting and was a part of an unencrypted pool and later was moved to an encrypted pool, it contains stale data from its previous life.

Another mistake that can occur is to mis-configure an external MDisk as self-encrypting, while in reality it is not self-encrypting. In that case, the data that is written to this MDisk is not encrypted by IBM SAN Volume Controller because IBM SAN Volume Controller expects that the storage system that is hosting the MDisk encrypts the data. At the same time, the MDisk does not encrypt the data because it is not self-encrypting; therefore, the system has unencrypted data on an extent in an encrypted storage pool.

However, all data that is written to any MDisk that is a part of a correctly configured encrypted storage pool is going to be encrypted.

Self-encrypting MDisks

When adding external storage to a pool, be exceptionally diligent when declaring the MDisk as self-encrypting. Correctly declaring an MDisk as self-encrypting avoids wasting resources, such as CPU time. However, when used incorrectly, it might lead to unencrypted data at-rest.

To declare an MDisk as self-encrypting, select **Externally encrypted** when adding external storage in the Add Storage view, as shown in Figure 12-129.

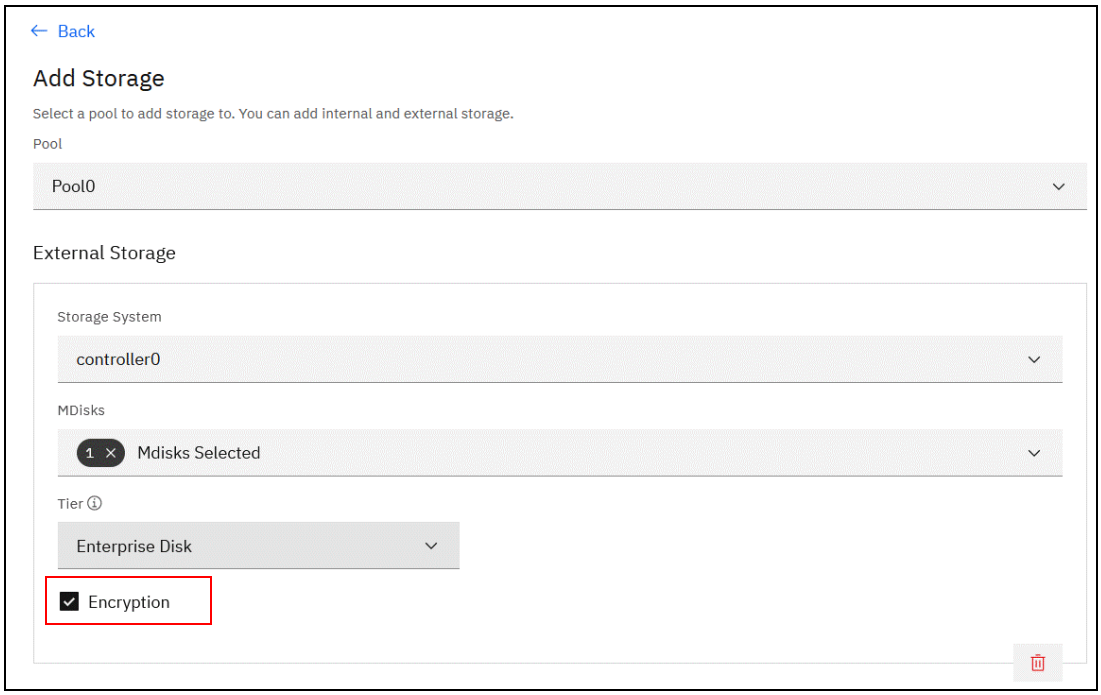


Figure 12-129 Declaring MDisk as externally encrypted

IBM Storage Virtualize products can detect that an MDisk is self-encrypting by using the SCSI Inquiry page C2. MDisks that are provided by other IBM Storage Virtualize products report this page correctly. The Encryption option is selected for those MDisks.

Note: You can override the external encryption setting of a detected MDisk as self-encrypting and configure it as unencrypted by running the **`chmdisk -encrypt no`** CLI command. However, do so only if you plan to decrypt the data on the backend or if the backend uses inadequate data encryption.

It is also possible to override the external encryption setting of a detected MDisk by right-clicking the MDisk within **Pools** → **External Storage** and selecting it (see Figure 12-130).

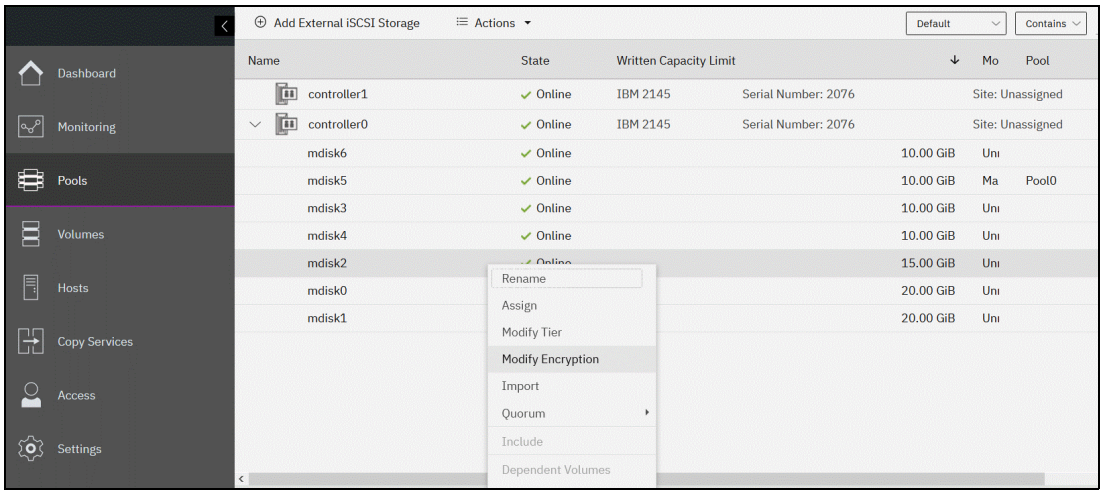


Figure 12-130 Overriding external encryption setting for external MDisk

If the external encryption setting is overridden for self-encrypting MDisk, the Encryption checkbox is not automatically set in the Add Storage dialog when this MDisk is added to the pool (see Figure 12-129 on page 1146).

To check whether an MDisk was declared as self-encrypting, select **Pools** → **MDisk by Pools** and verify the information that is in the Encryption column, as shown in Figure 12-131.

Name	State	Usable Capacity	Written Capacity Limit	Storage System - LUN	Encryption
Unassigned MDisks (3)					
Pool0	Online				Yes
mdisk5	Online	10.00 GiB	10.00 GiB	controller0 - 0000000000000005	Yes
mdisk1	Online	19.00 GiB	20.00 GiB	controller0 - 0000000000000001	No
mdisk0	Online	19.00 GiB	20.00 GiB	controller0 - 0000000000000000	No
mdisk3	Online	9.00 GiB	10.00 GiB	controller0 - 0000000000000003	No
Pool1	No Storage				No

Figure 12-131 MDisk self-encryption state

The value that is shown in the Encryption column shows the property of objects in respective rows. That means that in the configuration that is shown in Figure 12-131, `Pool0` is encrypted. Therefore, every volume that is created from this pool is encrypted.

However, that pool is formed by four MDisks, out of which one is self-encrypting and others are not. Therefore, a value of `No` next to `mdisk1`, `mdisk0`, and `mdisk3` does not imply that

encryption of `Pool0` is in any way compromised. Instead, it indicates only that encryption of the data that is placed on `mdisk1`, `mdisk0`, and `mdisk3` is done by using software encryption.

Data that is placed on `mdisk5` is encrypted by the back-end storage that is providing these MDisks.

Note: You can change the self-encrypting attribute of an MDisk that is unmanaged or member of an unencrypted pool. However, you cannot change the self-encrypting attribute of an MDisk after it is added to an encrypted pool.

12.9.5 Encrypted volumes

For more information about how to create and manage volumes, see Chapter 6, “Volumes” on page 419.

The encryption status of a volume depends on the pool encryption status. Volumes that are created in an encrypted pool are automatically encrypted.

You can modify Volumes view to show if the volume is encrypted. Select **Volumes** → **Volumes**. Then click **Actions** → **Customize Columns** → **Encryption** to customize the view to show volumes encryption status, as shown in Figure 12-132.

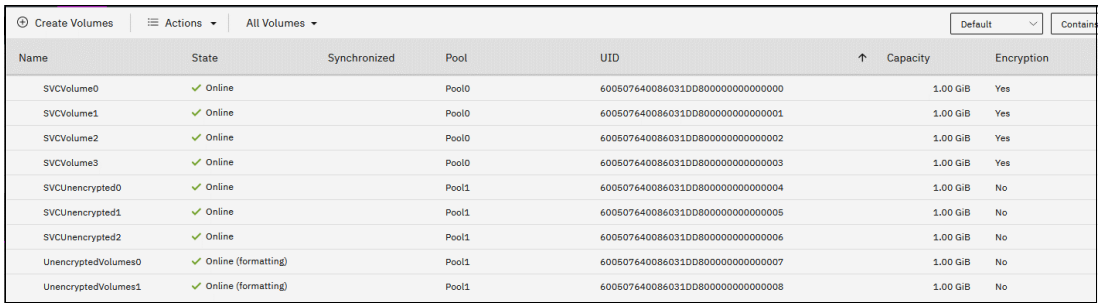


Figure 12-132 Volume view customization

A volume is reported as encrypted only if all the volume copies are encrypted, as shown in Figure 12-133.

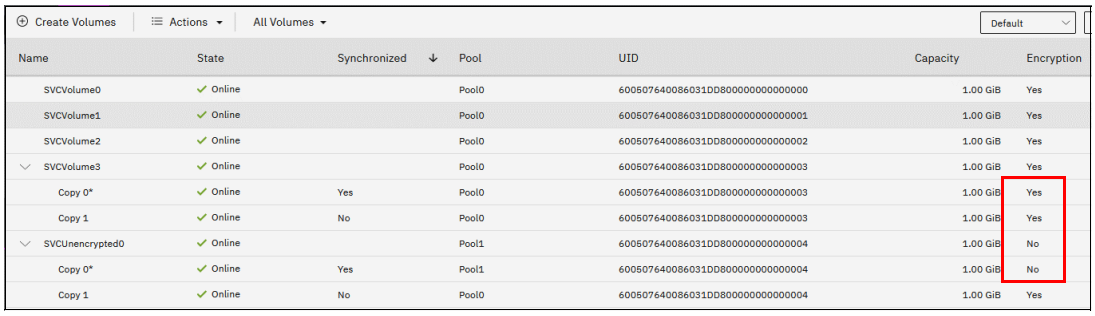


Figure 12-133 Volume encryption status depending on volume copies encryption

When creating volumes, make sure to select encrypted pools to create encrypted volumes, as shown in Figure 12-134.

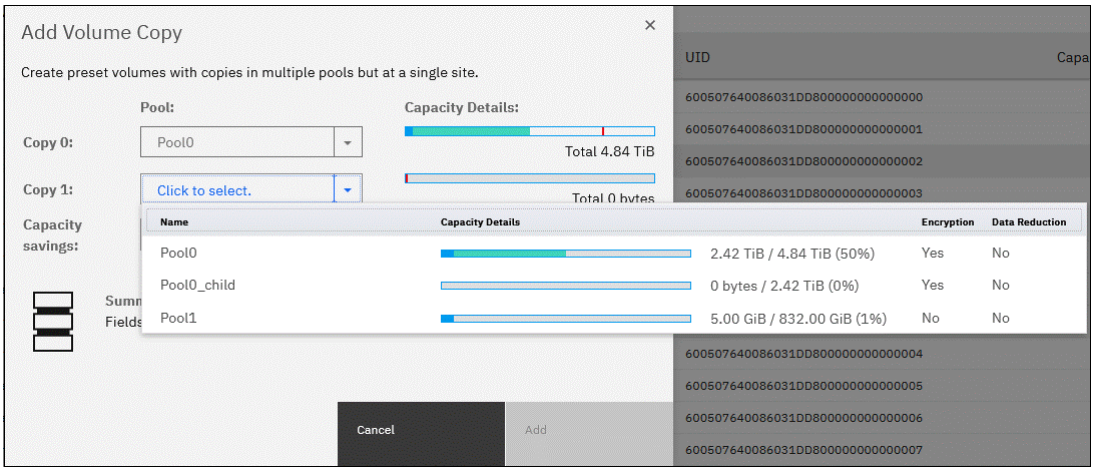


Figure 12-134 Create an encrypted volume by selecting an encrypted pool

You cannot change an unencrypted volume to an encrypted version of itself dynamically. However, this conversion is possible by using one of the following migration options:

- Migrate a volume to an encrypted pool or child pool.
- Mirror a volume to an encrypted pool or child pool and delete the unencrypted copy.

For more information about these methods, see 12.7, “Encryption” on page 1098.

12.9.6 Restrictions

The following restrictions apply to encryption:

- Image mode volumes cannot be in encrypted pools.
- You cannot add external non self-encrypting MDisks to encrypted pools unless all nodes in the system support encryption.

12.9.7 Rekeying an encryption-enabled system

Changing the master access key is a security requirement. *Rekeying* is the process of replacing a current master access key with a newly generated one. The rekey operation works whether encrypted objects exist.

The rekeying operation requires access to a valid copy of the original master access key on an encryption key provider that you plan to rekey. Use the rekey operation according to the schedule that is defined in your organization’s security policy and whenever you suspect that the key might be compromised.

If you have USB and key server enabled, rekeying is done separately for each of the providers.

Important: Before you create a master access key, ensure that all nodes are online and that the current master access key is accessible.

No method is available to directly change data encryption keys. If you must change the data encryption key that is used to encrypt data, the only available method is to migrate that data to a new encrypted object (for example, an encrypted child pool). Because the data encryption keys are defined per encrypted object, such migration forces a change of the key that is used to encrypt that data.

Rekeying by using a key server

Ensure that all the configured key servers can be reached by the system and that service IPs are configured on all your nodes.

To rekey the master access key that is kept on the key server provider, complete the following steps:

1. Select **Settings** → **Security** → **Encryption**. Ensure that Encryption Keys show that all configured SGKLM servers are reported as `Accessible`. Click **Key Servers** to expand the section.
2. Click **Rekey**, as shown in Figure 12-135.

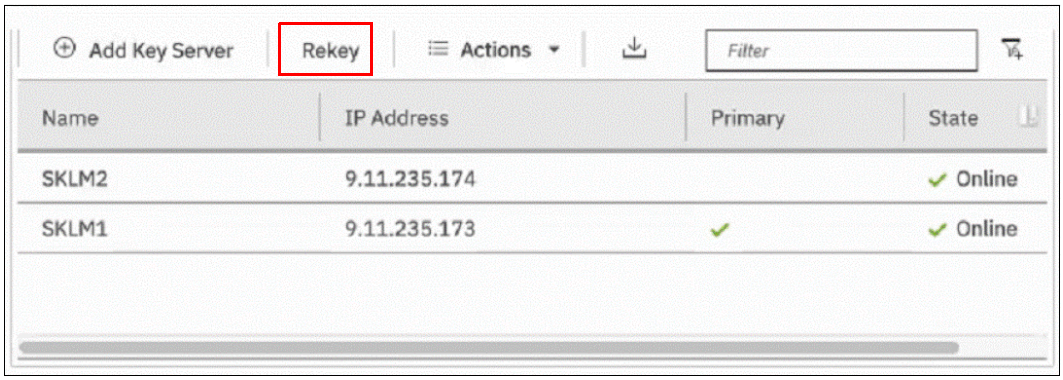


Figure 12-135 Start rekey on SGKLM key server

3. In the next window, confirm the rekey operation.

> **Note:** The rekey operation is performed on only the primary key server that is configured in the system. If more key servers are configured apart from the primary key, they do not hold the updated encryption key until they obtain it from the primary key server. To restore encryption key provider redundancy after a rekey operation, replicate the encryption key from the primary key server to the secondary key servers.

You receive a message confirming that the rekey operation was successful.

Rekeying by using USB flash drives

During the rekey process, new keys are generated and copied to the USB flash drives. These keys are then used instead of the current keys. The rekey operation fails if at least one of the USB flash drives does not contain the current key. To rekey the system, you need at least three USB flash drives to store the master access key copies.

After the rekey operation is complete, update all other copies of the encryption key, including copies stored on other media. Take the same precautions to securely store all copies of the new encryption key as when you were enabling encryption for the first time.

To rekey the master access key on USB flash drives, complete the following steps:

1. Select **Settings** → **Security** → **Encryption**. Click **USB Flash Drives** to expand the section.
2. Verify that all USB drives are plugged into the system, detected, and show as `Validated`, as shown in Figure 12-136. Click **Rekey**. You need at least three USB flash drives, with at least one reported as `Validated` to process with rekey.

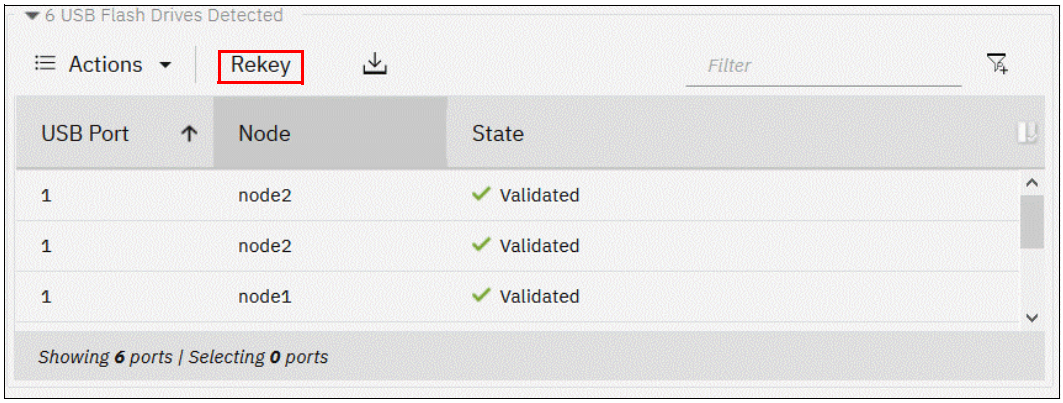


Figure 12-136 Start rekey on USB flash drives provider

3. If the system detects a validated USB flash drive and at least three available USB flash drives, new encryption keys are automatically copied on the USB flash drives, as shown in Figure 12-137. Click **Commit** to finalize the rekey operation.

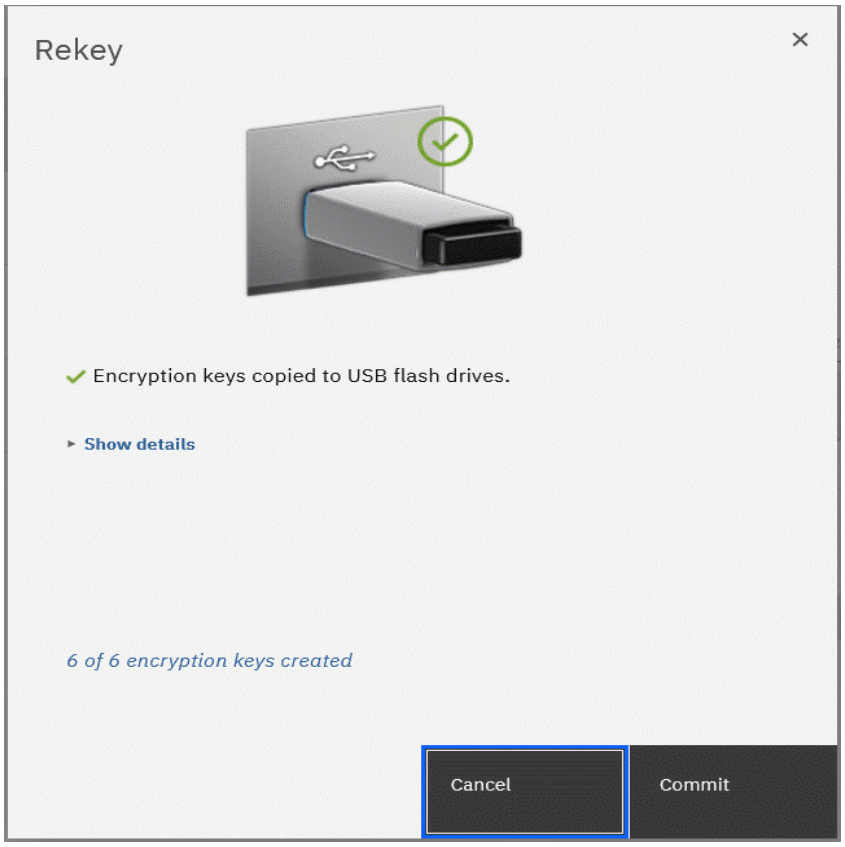


Figure 12-137 Writing new keys to USB flash drives

4. You receive a message confirming the rekey operation was successful. Click **Close**.

12.9.8 Disabling encryption

You are prevented from disabling encryption if any encrypted objects are defined apart from self-encrypting MDisks. You can disable encryption in the same way whether you use USB flash drives, key server, or both providers.

To disable encryption, complete the following steps:

1. Select **Settings** → **Security** → **Encryption** and click **Enabled**. If no encrypted objects exist, a menu is displayed. Click **Disabled** to disable encryption on the system. Figure 12-138 shows an example for a system with both encryption key providers configured.

   You receive a message confirming that encryption was disabled. Figure 12-139 on page 1153 shows the message when a key server is used.

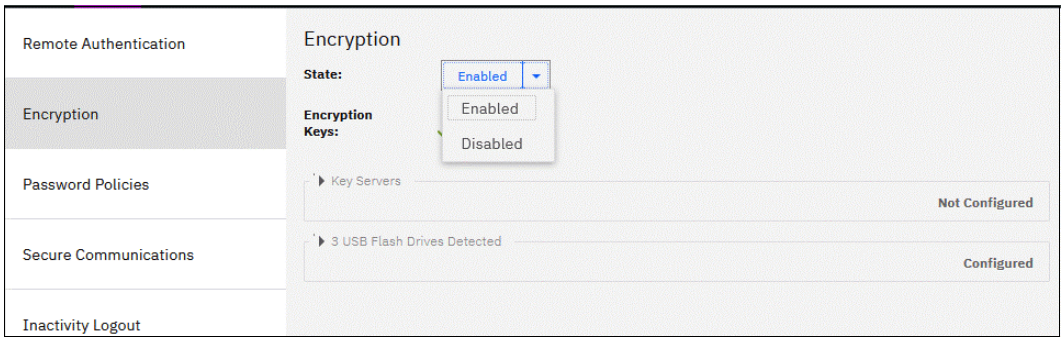


Figure 12-138 Disabling encryption on a system with both providers

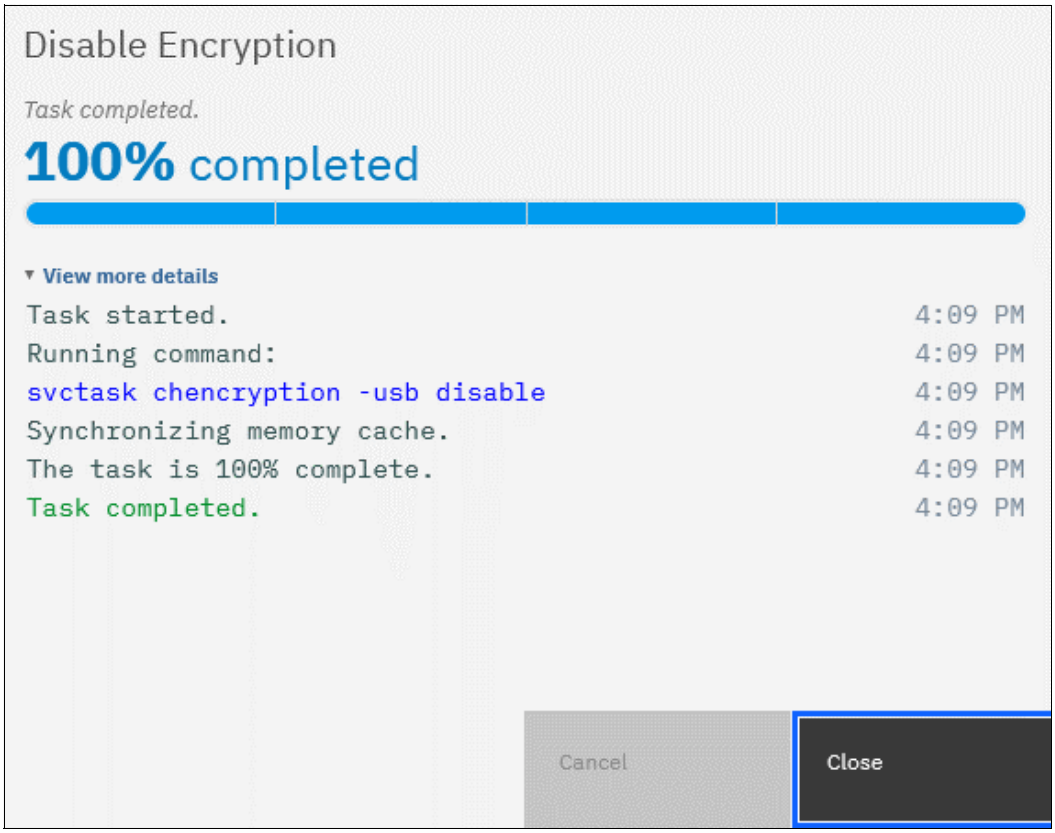


Figure 12-139 Encryption disabled

13

Automation and scripting

This chapter provides information about scripting and automation tasks that can occur in an IBM Storage Virtualize environment that uses Ansible.

This chapter includes the following topics:

- 13.1, “REST API on IBM Storage Virtualize” on page 1156
- 13.2, “Scripting” on page 1164
- 13.3, “Automation with Red Hat Ansible” on page 1180

© Copyright IBM Corp. 2024.

13.1 REST API on IBM Storage Virtualize

The IBM Storage Virtualize Representational State Transfer (REST) application programming interface (API) consists of command targets that are used to retrieve system information and to create, modify, and delete system resources. These command targets allow command parameters to pass through unedited to the IBM Storage Virtualize command line interface (CLI), which handles parsing parameter specifications for validity and error reporting. Hypertext Transfer Protocol Secure (HTTPS) is used to communicate with the REST API server.

The easiest way to interact with the storage system by using the REST API is the curl utility to make an HTTPS command request with a valid configuration node URL destination. For more information, see curl: command line tool and library. Open TCP port 7443 and include the API version in combination with the keyword `rest` followed by the IBM Storage Virtualize target command that you want to run.

Each curl command uses the following format:

```
curl –k –X POST –H <header_1> –H <header_2> ... -d <JSON input>
https://StorageVirtualize_ip_address:7443/rest/<api_version>/target
```

For the command, the following definitions apply:

- POST is the only HTTPS method that the Storage Virtualize REST API supports.
- Headers <*header_1*> and <*header_2*> are individually specified HTTP headers, such as, `Content-Type` and `X-Auth-Username`.
- Use of the parameter **`-d`** followed by the input in Java script Object Notation (JSON). For example, '{"raid_level": "raid5"}' to provide required more configuration information.
- <*StorageVirtualize_ip_address*> is the IP address of the IBM Storage Virtualize storage to which you are sending requests.
- <*target*> is the target object of commands, which includes any object IDs, names, and parameters.
- <*api_version*> specifies which version of the API should get used. The latest API version is v1.

Note: Compatibility with an earlier version is implemented by auto redirection of nonversioned requests to v0. It is recommended to use versioned endpoints for guaranteed behavior.

Consider the following examples:

- `https://StorageVirtualize_ip_address:7443/rest/target` uses API v0
- `https://StorageVirtualize_ip_address:7443/`**`v0`**`/rest/target` uses API v0
- `https://StorageVirtualize_ip_address:7443/`**`v1`**`/rest/target` uses API v1

Authentication

Aside from data encryption, the HTTPS server requires authentication of a valid username and password for each API session. It is required to specify two authentication header fields for your credentials: `X-Auth-Username` and `X-Auth-Password`.

Initial authentication requires that you POST the authentication target (**`/auth`**) with the username and password. The REST API server returns a hexadecimal token. A single session lasts a maximum of two active hours or 30 inactive minutes, whichever occurs first.

Note: The **chsecurity** command configures the amount of time (in minutes) before a token expires. The allowed range is 10 - 120 minutes. The default value is 60 minutes.

When your session ends because of inactivity (or if you reach the maximum time that is allotted), error code 403 indicates the loss of authorization. Use the **/auth** command target to reauthenticate by using the username and password.

The following example shows the correct procedure for authenticating. You authenticate by first producing an authentication token and then use that token in all future commands until the session ends.

For example, the following command passes the authentication command to IBM SAN Volume Controller node IP address 192.168.10.20 at port 7443 by using API version v1:

```
curl –k –X POST –H 'Content-Type: application/json' –H 'X-Auth-Username: myuser'
–H 'X-Auth-Password: mypassw0rd' https://192.168.10.20:7443/rest/v1/auth
```

Note: Make sure that you format the request correctly by using spaces after each colon in each header; otherwise, the command fails.

This request yields an authentication token, which can be used for all subsequent commands, as shown in the following example:

```
{"token": "38823f60c758dca26f3eaac0ffee42aadc4664964905a6f058ae2ec92e0f0b63"}
```

The X-Auth-Token header in combination with the authentication token replaces the username and password for all further actions. The token is valid for one session, but the session times out after two hours of activity or 30 minutes of inactivity. Repeat the authentication process for creating another token.

Example commands

In this section, we discuss some of example commands.

Using the authentication token

Most actions can be taken only after authentication. The following example of displaying the system information demonstrates the use of the previously generated token in place of the authentication headers that are used in the authentication process:

```
curl –k –X POST –H 'Content-Type: application/json' –H 'X-Auth-Token:
38823f60c758dca26f3eaac0ffee42aadc4664964905a6f058ae2ec92e0f0b63'
https://192.168.10.20:7443/rest/v1/lssystem
```

Note: If you use curl, you do not receive the HTTPS error code that is displayed if you do not specify the **-f** option.

Specifying more parameters

Although querying any information does not require that any other parameters are required within the REST call, this mandatory requirement exists for any action that intends to modify an object or create an object; for example, host, hostcluster, array, or VDisk.

The following example demonstrates the use of the **-d** parameter to specify the parameters and associated values that are required for creating a mirrored volume:

```
curl –k –X POST –H 'Content-Type: application/json' –H 'X-Auth-Token:
38823f60c758dca26f3eaac0ffee42aadc4664964905a6f058ae2ec92e0f0b63'
–d '{"name":"myVDisk1", "copies":"2", "mdiskgrp":"mdiskgrp0:mdiskgrp1",
"size":"200", "vtype":"striped", "unit":"gb", "rsize":"5%" }'
https://192.168.10.20:7443/rest/v1/mkvdisk
```

This REST call is equivalent to running the following **mkvdisk** command (it produces the same output):

```
mkvdisk -name myVDisk1 -copies 2 -mdiskgrp mdiskgrp0:mdiskgrp1 -size 200 -vtype
striped -unit gb -rsize 5%
```

The parameters and values that can or must be specified when a REST target is used, such as **mkvdisk** or **mkhost**, are identical to those within the CLI.

Note: Parameters (keys) and values must be specified in JavaScript Object Notation (JSON) notation.

JSON is a lightweight data-interchange text format that is language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, and Python.

JSON data is written as key/value pairs. A key/value pair consists of a field key, followed by a colon, followed by a value, whereby key and value must be placed in double quotation marks.

One or more key/value pairs build an object, which begins with a left brace ({) and ends with a right brace (}), as shown in Example 13-1.

Example 13-1 JSON notation for creating a thin provisioned mirrored VDisk

```
{
  "name": "myVDisk1",
  "copies": "2",
  "mdiskgrp": "mdiskgrp0:mdiskgrp1",
  "size": "200",
  "vtype": "striped",
  "unit": "gb",
  "rsize": "5%"
}
```

For more information about JSON, see Introducing JSON.

Rate limiting

Rate limiting helps with security and the prevention of an attack, such as a denial of service in which unlimited work is sent to the system. The rate limiting is implemented at millisecond granularity and creates a return code (`429 - too many requests`) when a violation occurs. REST API rate limits are listed in Table 13-1 on page 1159.

Table 13-1 REST API rate limits

Limit	Type	Value
Maximum active connections per cluster	REST API	4
Maximum requests per second to the `/auth` endpoint	REST API	3 per second
Maximum requests per second to the `/non-auth` endpoint	REST API	10 per second
Number of simultaneous CLIs in progress	System	1

REST API Explorer

To view REST API documentation, see Spectrum Virtualize RESTful API. Support can also be found directly on the system within the REST API Explorer.

The REST API Explorer is based on the Swagger UI and runs within a browser. It offers an easy way to become familiar with the API and to test the commands that it contains.

To access the REST API Explorer, enter the following URL in a browser:

`https://<StorageVirtualize_ip_address | FQDN>:7443/rest/explorer`

Figure 13-1 shows the grouping of available actions within the RET API Explorer.

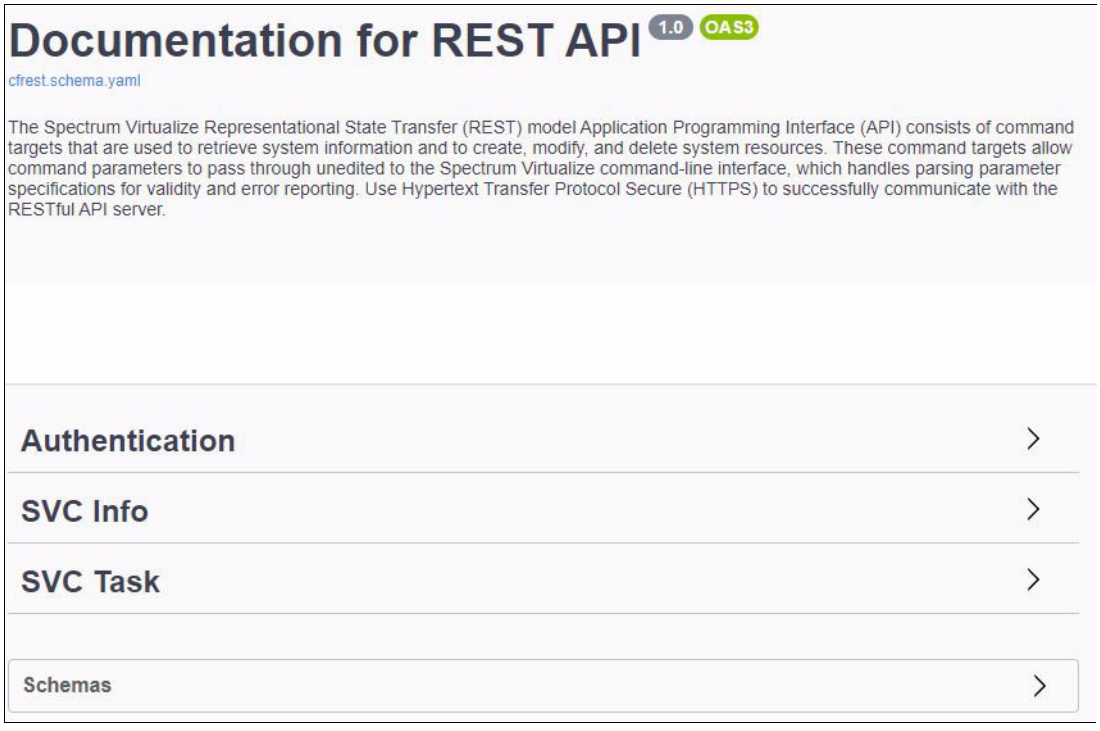


Figure 13-1 REST API Explorer actions

The use of the REST API Explorer also requires generating a token, which can be used for all further actions. Figure 13-2 shows how to create an authentication token within the IBM SAN Volume Controller information and IBM SAN Volume Controller Task actions.

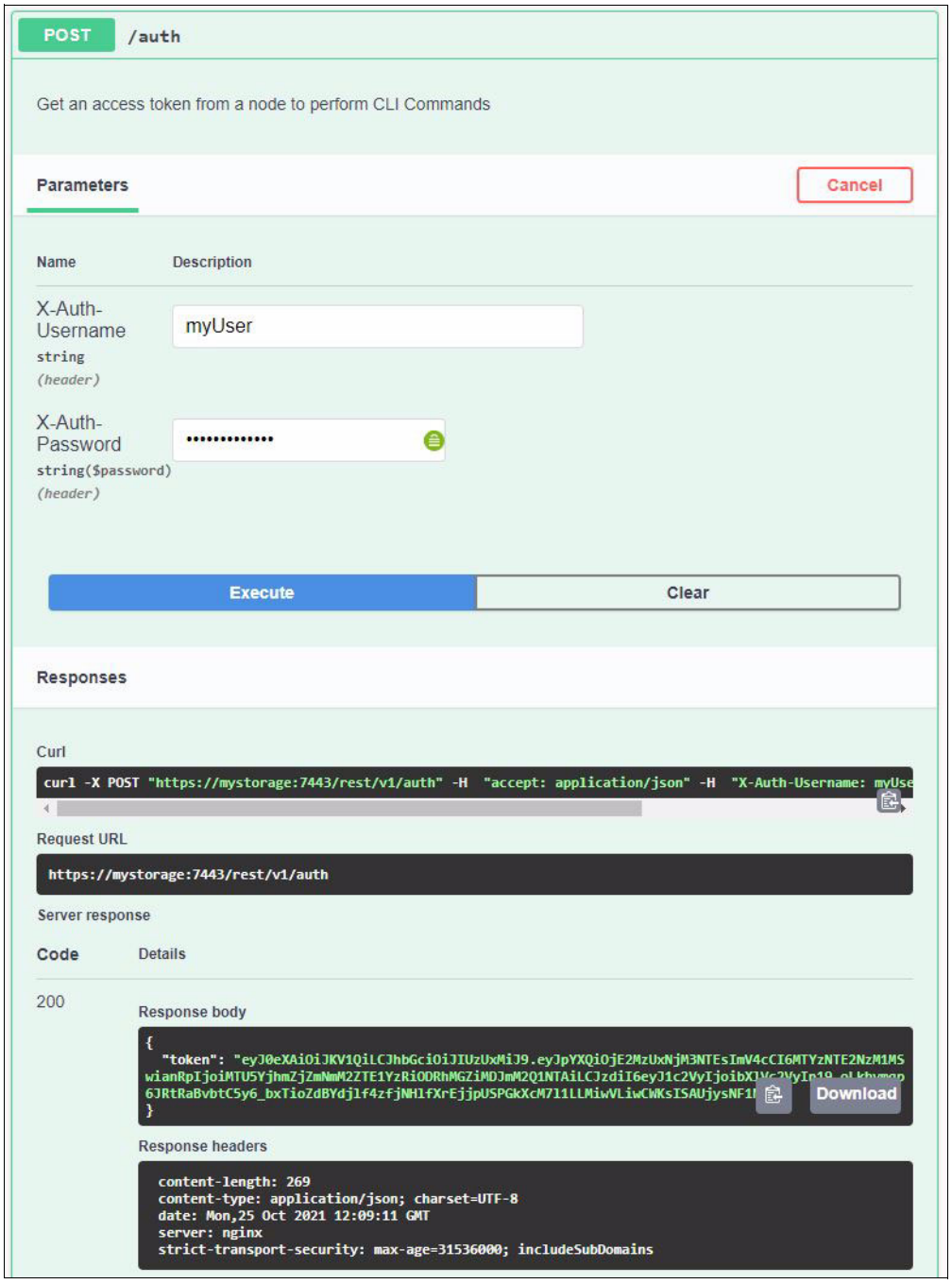


Figure 13-2 REST API Explorer authentication

Figure 13-2 also shows the following outputs after successful authentication:

- The curl command that is required to carry out the action
- Request URL, which was addressed during the execution of the action
- Server response in from of the Response body and the response header

The token is displayed in JSON notation in the response body.

By using the generated authentication token, more actions can be completed in the REST API Explorer.

Figure 13-3 shows the **mkvdisk** task in the REST API Explorer. All accepted parameters are listed in the request body. These parameters can then be adapted or deleted according to the requirements for creating the VDisk.

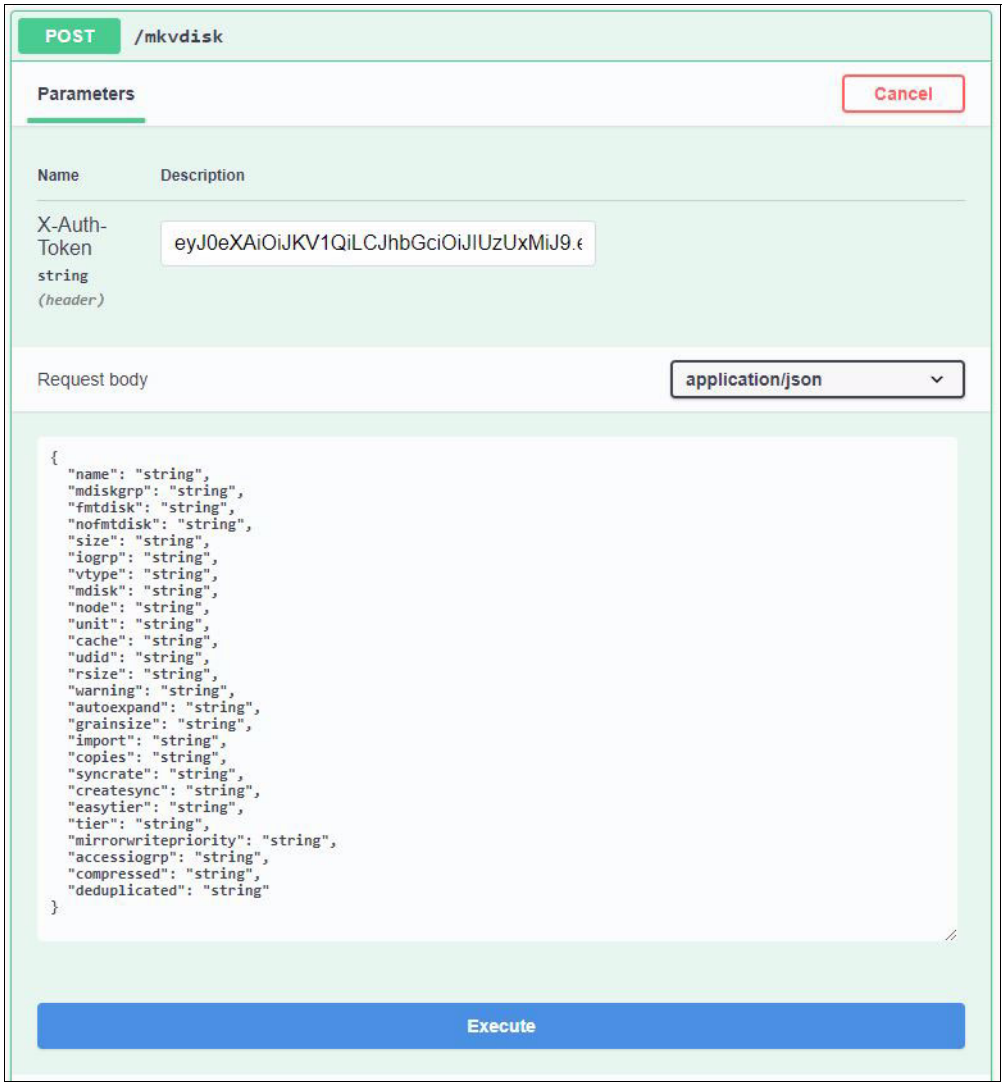


Figure 13-3 REST API Explorer /mkvdisk

Figure 13-4 shows an example for the Request body to create a mirrored VDisk and the output within the Response body.

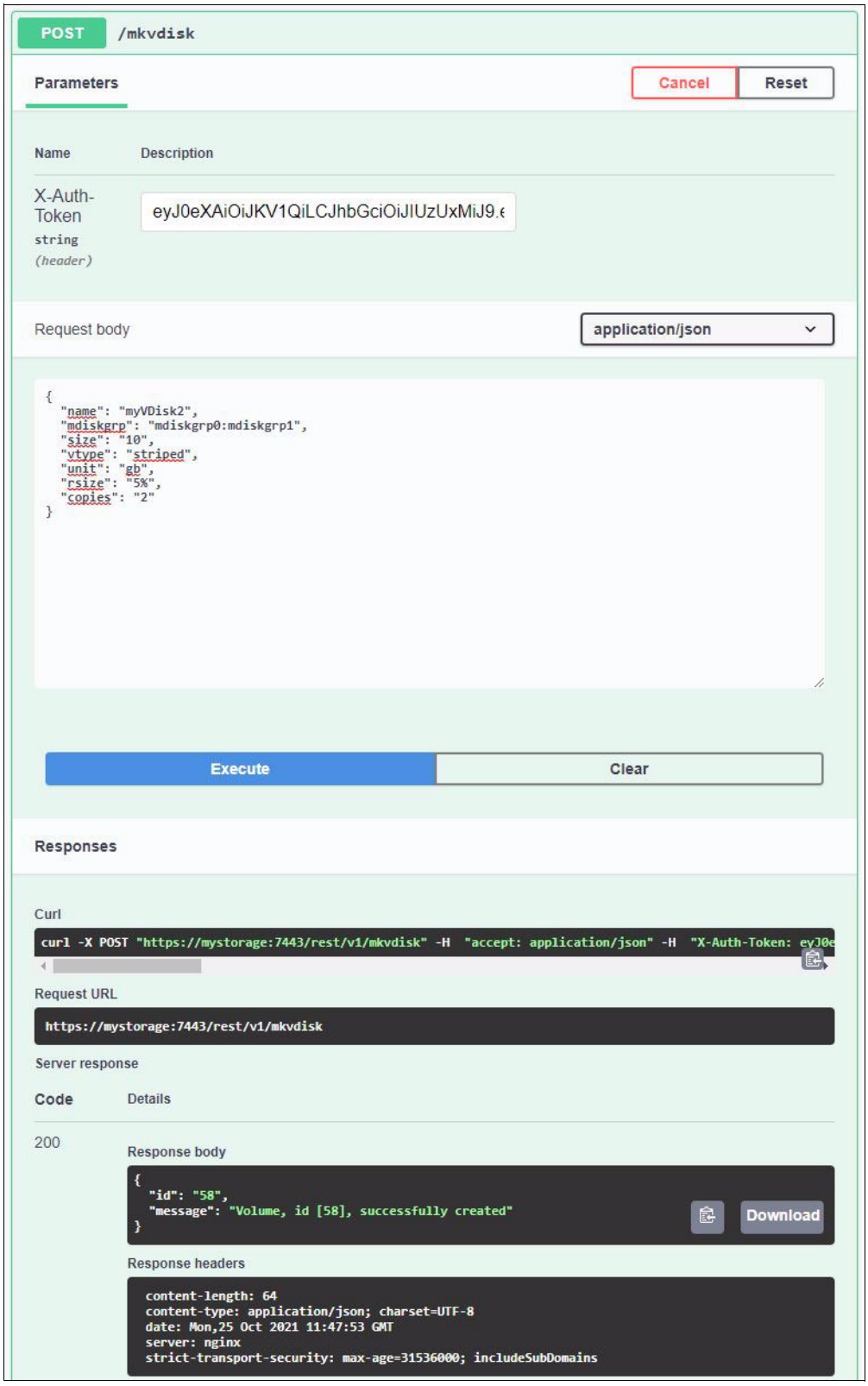


Figure 13-4 REST API Explorer /mkvdisk output

REST API HTTP messages

When an issue exists with the information that you provided, an error message appears.

Different types of error messages can appear, depending on the issue. The only error messages that are described in this document are HTTP errors. For more information about other error messages, see this IBM Documentation web page.

The following HTTP error codes are returned to the user by the REST API in response to a problem with the request:

- `400: bad request`

  The command did not specify a required parameter or gave a parameter value that did not pass the REST API checks.

- `401: unauthorized`

  The command requires a successful authentication.

- `403: forbidden`

  The user did not send a valid authentication token to interact with the specified URL.

- `404: not found`

  The command attempted to issue a request to a URL that does not exist.

- `405: method not allowed`

  The command attempted to use an HTTP method that is invalid for the specified URL.

- `409: conflict`

  The sent request conflicts with the current state of the system.

- `429: too many requests`

  Too many requests violate the rate limiting.

- `500: something went wrong on the server`

  A Storage Virtualize command error was forwarded from the REST API.

- `502: bad gateway`

  The API received an invalid response from the upstream system.

For more information about the use of the REST API, see Spectrum Virtualize RESTful API.

For more information about other examples, see the following web pages:

- IBM Storage Virtualize Interfacing Using the RESTful API
- Tips and tricks using the Storage Virtualize REST API

Audit logging

Commands that are started by the REST API are auditable, such as actions that are started by the CLI or GUI. The Origin field within the output of the **`cataudit log`** CLI command shows the source interface of the command that was run.

13.2 Scripting

This section describes some methods that can be used to access the IBM Storage Virtualize Controller family by using scripts. These methods can be used for configuration, reporting, and administration tasks.

IBM Storage Virtualize Controller family supports the following methods or protocols for running configuration tasks and monitoring, in addition to the traditional web-based graphical user interface (GUI):

- Secure Shell (SSH)
- SMI-S
- HTTPS and REST API on IBM Storage Virtualize
- HTTPS and REST API on IBM Storage Control

13.2.1 Scripting principles

The following practices are recommended for scripting:

- Always use secure protocols, such as SSH and HTTPS.
- Use SSH-keys for authentication if possible and protect the SSH-keys.
- Use dedicated users for monitoring and configuring and administering purposes.
- Assign only the required permissions according to the purpose of the configured user.
- Implement error handling in the scripts.

13.2.2 Secure Shell

SSH is a network protocol that enables secure communication and operation over an insecure network.

All members of the IBM Storage Virtualize storage products feature a CLI, which is accessible by using the SSH protocol.

Note: The SSH protocol authenticates users by using passwords or SSH keys (by using an asymmetric cryptography method). For security reasons, it is recommended to use and protect the configured SSH keys.

The system supports up to 32 interactive SSH sessions on the management IP address simultaneously.

Note: After an SSH interactive session times out, the session is automatically closed. The session timeout limit is set to 15 minutes by default. The limit value can be changed by using the **`chsecurity`** command. For more information, see chsecurity.

To connect to the system, the SSH client requires a user login name and an SSH password (or the key pair if you require command line access without entering a password). Authenticate to the system by using a management username and password.

When you use an SSH client to access a system, you must use your username and password. The system uses the password (and if not a password, the SSH key pair) to authorize the user who is accessing the system.

General tips

Consider the following general tips when SSH is used:

- Use of **svcinfo** and **svctask**

  Some small differences exist between the CLIs of the different products; for example, **lsnodecanister** is used on IBM Storage FlashSystem Controllers, and **lsnode** is used on IBM Storage Virtualize controllers.

- Use of **-delim parameter** on **ls-commands**

  Parsing the output of a **ls-command** becomes much easier because it inserts a single, selectable character between each field instead of several spaces. The colon (:) is a good choice for a delimiter for all output that does not contain any IPv6 addresses.

- Use of **-nohdr** on **ls-commands**

  The use of the **-nohdr** parameter suppresses the output of the header so that the required code for skipping the first line of the output is bypassed.

Using SSH in bash/ksh

The use of an SSH client within a shell is a common way of running a specified command, but not a login shell on a remote system. Instead of opening an interactive session, SSH runs a command on the remote system, forwards the output to the local computer, and then exits the session.

Running commands remotely by way of SSH provides a way to write and use advanced scripts to collect data from an IBM Storage Virtualize system. It also continues processing that data on a local computer in combination with other available tools and utilities.

Running commands remotely by way of SSH allows SSH to be used in a shell and piping the output to any external program for parsing, filtering, and data processing.

Example 13-2 shows how to use SSH to run a command (**svcinfo lssystem**) on the IBM Storage Virtualize system (`mystorage`) with the privileges of `myuser` and piping the output to filter only for lines containing `unmap`.

Example 13-2 Using ssh and grabbing selected information

```
ssh myuser@mystorage 'svcinfo lssystem | grep unmap'
```

Using SSH in Windows command line

For the use of the SSH in combination with the Windows operating system, the PuTTY Plink utility or the optional available OpenSSH client feature (which integrates into the standard command line) must be installed.

PuTTY Plink (see Example 13-3) enables authentication by using SSH keys by using or configuring the PuTTY authentication agent (Pageant).

Example 13-3 Using PuTTY PLink

```
cd C:\Program Files\PuTTY\
plink.exe myuser@mystorage 'svcinfo lssystem | findstr unmap'
```

To enable the Windows OpenSSH client for authentication by using SSH keys, the keys must be placed within the following directory structure `C:\User\<username>\.ssh\`.

Using SSH with Python

Python requires the use of another external module to connect to IBM Storage Virtualize by using the SSH protocol.

The `paramiko` open source module is popular and can be installed by using `pip`.

Example 13-4 shows the simple use of `paramiko` within a Python script by connecting to `myStorage` by using the `myUser` user and running the commands to display the configured host and controller objects.

Example 13-4 Using paramiko within Python

```
#!/usr/bin/python

import paramiko
mystorage = 'myStorage'
myuser = 'myUser'

ssh = paramiko.SSHClient()
ssh.set_missing_host_key_policy(paramiko.AutoAddPolicy())
ssh.connect(hostname=mystorage, username=myuser)

command1 = 'lshost -delim :'
command2 = 'lscontroller -delim :'

stdin, stdout, stderr = ssh.exec_command(command1)
data = stdout.read()

errors = stderr.read()
if data:
        print(data)
if errors:
        print(errors)

print("--------------------------------------------------------------------------------\n");

stdin, stdout, stderr = ssh.exec_command(command2)
data = stdout.read()

errors = stderr.read()
if data:
        print(data)
if errors:
        print(errors)

ssh.close()
```

The `set_missing_host_key_policy(paramiko.AutoAddPolicy())` method defines how to proceed if the remote system SSH fingerprint is not known locally.

The `connect(hostname=target, username=user)` method connects with the storage system. If keys-based authentication is configured correctly, keys are checked automatically and a session with SSH server is established.

Several options are available with `client.connect()`. For example, certificates can be specified by using `pkey` or `key_filename` arguments, and set the user password with password argument if better authentication methods cannot be used.

The following example shows how to specify user and password when a connection is created. This type is the most insecure because the password is saved in plain text in the script. Therefore, this approach is not recommended:

```
client.connect(hostname='myStorage', username='myUser', password='myPassword')
```

Note: If you do not want to manage session handling and `paramiko` methods, you can use the IBM Storage Virtualize Python Client (`pysvc`), which is available for download at this GitHub web page.

Using SSH with Perl

Perl requires the usage of an extra external module to connect to IBM Storage Virtualize by using the SSH protocol.

`Net::OpenSSH` is a Secure Shell client package that is implemented on the OpenSSH binary client, which is installed by using CPAN.

Example 13-5 shows the use of `Net::OpenSSH` within a Perl script by connecting to `myStorage` by using the user `myUser` and running the commands to display the configured host and controller objects.

Example 13-5 Using Net::OpenSSH within Perl

```
#!/usr/bin/perl

use strict;
use Net::OpenSSH;

my $host = "myStorage";
my $user = "myUser";

my $command1 = "lshost -delim :";
my $command2 = "lscontroller -delim :";

my $ssh = Net::OpenSSH->new("$user\@$host", forward_agent => 1);
$ssh->error and die "SSH connection failed: " . $ssh->error;
print "Connected to $host\n";

my @vDisk = $ssh->capture($command1) or die "Unable to run command";
my @controller = $ssh->capture($command2) or die "Unable to run command";

print @vDisk;
print
"--------------------------------------------------------------------------------\n";
print @controller;

$ssh->disconnect();
```

The `forwared_agent=>1` option defines the use of the `ssh-agent` authentication agent for the SSH key-based authentication.

13.2.3 SMI-S

The *Storage Management Initiative Specification* (SMI-S) is a common standard that was developed and maintained by the Storage Network Industry Association (SNIA). SMI-S also was ratified as an ISO standard.

The main objective of SMI-S is the management of heterogeneous storage systems across different vendors.

Because SMI-S was available before the REST API was introduced, several products, such as IBM Storage Protect Snapshot, still use this interface.

SMI-S consists of the following three main components:

- Common Information Model (CIM)

  The CIM is an open standard that defines how managed elements are represented as a set of objects and their relationships in an IT environment.

- Web-based Enterprise Management standards (WBEM)

  WBEM is a set of standards that enable computers and other network devices to be managed by using a standard web browser.

- Service Location Protocol (SLP)

  The SLP is a service discovery protocol that allows computers and other devices to find services in a LAN without prior configuration.

Python requires the use of an extra external module `pywbem` to connect to IBM Storage Virtualize by using the SMI-S interface.

The script that is shown in Example 13-6 shows the basic use of `pywbem`.

Example 13-6 Basic use of pywbem

```
#!/usr/bin/python

import pywbem
import getpass

mystorage = 'myStorage'
url = 'https://' + mystorage

username = 'myUser'
password = getpass.getpass()

wbemc = pywbem.WBEMConnection(url, (username, password), 'root/ibm', no_verification=True)
cluster = wbemc.EnumerateInstances('IBMTSSVC_Cluster')
print(cluster[0].items())
```

In this example, `WBEMConnection()` establishes HTTPS connection with WBEM services of IBM Storage Virtualize controller. Here, the target storage system URL is specified by the URL argument. The username and password and the CIM namespace (`root/ibm`) to query are also provided in the next lines.

Note: The `getpass` module is not necessary to work with SMI-S because its purpose is to securely read passwords from standard input with the terminal echo function switched off to hide what is entered.

The `no_verification=True` argument disables SSL certificate verification. That is, it forces the script to trust any certificate that is provided by the WBEM server.

After the connection is successfully established, instances of a specific CIM class can be enumerated by using the `EnumerateInstances()` method, which returns a complex data structure (a list of `CIMInstance()` classes). As shown in Example 13-6 on page 1168, it is done over the `IBMTSSVC_Cluster` class, which represents system-level information that is comparable with the results of running the **`lssystem`** command.

Different CIM classes are available for comprehensive management of the IBM SAN Volume Controller system, including the following examples:

- `IBMTSSVC_Cluster`: System level information
- `IBMTSSVC_Node`: Information about nodes
- `IBMTSSVC_ConcreteStoragePool`: MDisk groups
- `IBMTSSVC_BackendVolume`: MDisks
- `IBMTSSVC_StorageVolume`: VDisk information

This section gives a brief overview of these CIM classes to illustrate SMI-S capabilities, but it does not provide a full list of these classes or their descriptions. For more information about IBM SAN Volume Controller WBEM/CIM classes, their purposes, and relationship diagrams, see *IBM Spectrum Virtualize: Interfacing Using the RESTful API*.

The last line of the script parses and prints the data. However, it is not the only way to complete the job. Python is a flexible language and it performs work in different ways. Several approaches of processing the data that is acquired by `EnumerateInstances()` for several CIM classes are listed in Example 13-7.

Example 13-7 Parsing EnumerateInstances() output for classes cluster, nodes, and storage pools

```
print('Cluster information')
cluster = wbemc.EnumerateInstances('IBMTSSVC_Cluster')
print(cluster[0]['ElementName'])

for c_prop in cluster[0]:
    print('\t{prop}: "{val}"'.format(prop=c_prop, val=cluster[0].properties[c_prop].value))
print('Nodes information')

nodes = wbemc.EnumerateInstances('IBMTSSVC_Node')
for node in nodes:
    print(node['ElementName'])
    for n_prop in node:
        print('\t{prop}: "{val}"'.format(prop=n_prop, val=node[n_prop]))
print('Pools information')

pools = wbemc.EnumerateInstances('IBMTSSVC_ConcreteStoragePool')
print('PoolID', 'NumberOfBackendVolumes', 'ExtentSize', 'UsedCapacity',
      'RealCapacity', 'VirtualCapacity', 'TotalManagedSpace', sep=',')

for pool in pools:
    print(
        pool['ElementName'], pool['NumberOfBackendVolumes'], pool['ExtentSize'],
        pool['UsedCapacity'], pool['RealCapacity'], pool['VirtualCapacity'],
        pool['TotalManagedSpace'], sep=','
    )
```

By using similar, yet different approaches, the `Cluster information` and `Nodes information` sections of the example parse data in key/value pairs to show all acquired data. However, the `Pools information` part filters data to print selected fields only. It wastefully ignores all other fields.

For some classes, such as `IBMTSSVC_StorageVolume`, full enumeration of all the instances can be slow and can generate several megabytes of unnecessary data. This data must be prepared by the storage system, passed over the network, and then parsed by the script. However, it is possible to significantly reduce such data flows by requesting a limited amount of necessary information only.

As shown in Example 13-8, by using the `ExecQuery()` method, the WBEM server can be requested in a convenient query language, which is similar to SQL.

Example 13-8 Querying only required data by using the ExecQuery() method

```
print('Vdisks:')
vdisks = wbemc.ExecQuery(
    'DMTF:CQL',
    "SELECT VolumeId, VolumeName, NumberOfBlocks FROM IBMTSSVC_StorageVolume"
    " WHERE VolumeName LIKE 'vdisk.'"
)
for vd in vdisks:
    print(vd['VolumeId'], vd['VolumeName'], vd['NumberOfBlocks'], sep=',')
```

Two dialects (CIM Query Language [DMTF:CQL] and WBEM Query Language [WQL]) are recognized by PyWBEM and both can be used with IBM Storage Virtualize. However, we use the DMTF:CQL syntax in the examples in this chapter. The DMTF specification (DSP0202) for CQL can be found in *CIM Query Language Specification*.

One of the advantages of SMI-S on IBM SAN Volume Controller is its capability to collect performance data of various storage system components by using “Statistic” family CIM classes, as shown in the following examples:

- `IBMTSSVC_BackendVolumeStatistics`
- `IBMTSSVC_FCPortStatistics`
- `IBMTSSVC_NodeStatistics`
- `IBMTSSVC_StorageVolumeStatistics`

A detailed example of performance data collecting, and processing script that includes commentaries is shown in Example 13-9. It works with `IBMTSSVC_StorageVolumeStatistics` to retrieve VDisks statistics, as shown in Example 13-9.

Example 13-9 Accessing performance metrics by using the PyWBEM module

```
import pywbem
import getpass
import time

mystorage = 'myStorage'
myuser = 'myUser'
mypassword = getpass.getpass()
url = 'https://' + mystorage

ofs = ',' # Output field separator
header = ['InstanceID', 'ReadIOs', 'WriteIOs', 'TotalIOs',
   'KBytesRead', 'KBytesWritten', 'KBytesTransferred']
frequency = 5 # Performance collection interval in minutes
def vdisks_perf(wbem_connection, hdr):
```

```
    """Get performance statistics for vdisks"""
    # Form "select" request string
    request = "SELECT " + ','.join(hdr) + " FROM IBMTSSVC_StorageVolumeStatistics"
    result = []
    # Request WBEM
    vd_stats = wbem_connection.ExecQuery('DMTF:CQL', request)
    # parse reply and form a table
    for vds in vd_stats:
        # Handle 'InstanceID' in a specific way
        vde = [int(vds.properties[hdr[0]].value.split()[1])]
        # Collect the rest of numeric performance fields
        for fld in header[1:]:
            vde.append(int(vds.properties[fld].value))
        result.append(vde)
    return result

def count_perf(new, old, interval):
"""Calculate performance delta divided by interval to get per second values"""
    result = []
    for r in range(0, len(new)):
        row = [new[r][0]]                       # InstanceID
        for c in range(1, len(new[0])):
            row.append(round(float(new[r][c] - old[r][c]) / interval, 2))
result.append(row)
    return result
def print_perf(stats, hdr):
    """Printout performance data matrix"""
    # Print header
    print(ofs.join(str(fld) for fld in hdr))
    # Print performance table
    for ln in stats:
        print('{}{}{}'.format(ln[0], ofs, ofs.join(str(fld) for fld in ln[1:])))

# Connect with WBEM/CIM services
wbemc = pywbem.WBEMConnection(url, (myuser, mypassword), 'root/ibm', no_verification=True)

# Infinite performance processing loop
new_perf = vdisks_perf(wbemc, header)
while True:
    old_perf = new_perf
    new_perf = vdisks_perf(wbemc, header)
    delta_perf = count_perf(new_perf, old_perf, frequency * 60)
    print_perf(delta_perf, header)
    time.sleep(frequency * 60)
```

13.2.4 HTTPS and REST API on IBM Storage Virtualize

In this section, we discuss various ways in which the REST API can be used by using Curl, Python, and Perl.

We do not provide a recommendation for which programming language is to be used regarding the REST API.

Although Curl offers to test individual REST API calls quickly, Python and Perl are suitable for more complex tasks in which several REST API calls are to be run depending on each other.

Curl

Table 13-2 lists the `curl` command options.

Table 13-2 Options of the curl command

Command option	Description	Notes
`curl`	This executable sends the request to the server.	
`-k`	By default, every SSL connection curl makes is verified to be secure. This option allows curl to proceed and operate, even for server connections otherwise considered insecure.	You do not need this option if you use a signed SSL certificate.
`-H 'Key:Value'`	Send the information in the quotation marks as a header. Key is the name of the header. It describes what specific header is being sent. Value is the value for the key.	
`X-Auth-Username`	The username that you use to log in.	Only used for initial authentication.
`X-Auth-Password`	The password that you use to log in.	Only used for initial authentication.
`X-Auth-Token`	The authentication token that is used to authenticate the REST calls after authentication is complete.	Only used for running commands, not for the authentication.
`Content-Type:application/json`	Tells the server to send the result back in JSON format.	
`https://{{cluster IP or DNS name}}:7443/rest/v1/auth`	The URI to which you send an authentication request.	
`https://{{cluster IP or DNS name}}:7443/v1/{{cli command}}`	The URI to which you send a CLI command.	
`-d '{{DATA}}'`	The `-d` flag is used to send the CLI options, encoded in JSON.	

Creating an authentication token

Example 13-10 shows how to authenticate at the REST API endpoint. The successful authentication creates an authentication token for further use with the REST API.

Example 13-10 Creating a JSON Web Token (JWT)

```
curl -k -X POST -H 'Content-Type:application/json' -H 'X-Auth-Username: MyUser' -H
'X-Auth-Password: MyPassword' https://myStorage:7443/rest/v1/auth

{"token": "4d8916c21058db218d623df51c33f5f01cefeafc988ed7af78c1c51b4a104212"}
```

Query for all configured MDisks

Example 13-11 shows how to use the REST API to get a list of all MDisks by using the formerly generated authentication token.

Example 13-11 Get all M0Disks

```
curl -k -X POST -H 'Content-Type:application/json' -H 'X-Auth-Token:
4d8916c21058db218d623df51c33f5f01cefeafc988ed7af78c1c51b4a104212'
https://myStorage:7443/rest/v1/lsmdisk

[{ "id": "0", "name": "mdisk0", "status": "online", "mode": "array", "mdisk_grp_id": "0",
"mdisk_grp_name": "Pool0", "capacity": "21.7TB", "ctrl_LUN_#": "", "controller_name"
: "", "UID": "", "tier": "tier1_flash", "encrypt": "no", "site_id": "", "site_name": "",
"distributed": "yes", "dedupe": "no", "over_provisioned": "no", "supports_unmap": "ye
s" }]
```

Because this output is difficult to read, add "`| python -m json.tool`" to get a better readable output (see Example 13-12).

Example 13-12 Piping the output to python for getting better readable JSON output

```
curl -k -X POST -H 'Content-Type:application/json' -H 'X-Auth-Token:
4d8916c21058db218d623df51c33f5f01cefeafc988ed7af78c1c51b4a104212'
https://10.1.1.10:7443/rest/v1/lsmdisk | python -m json.tool

[
    {
        "UID": "",
        "capacity": "21.7TB",
        "controller_name": "",
        "ctrl_LUN_#": "",
        "dedupe": "no",
        "distributed": "yes",
        "encrypt": "no",
        "id": "0",
        "mdisk_grp_id": "0",
        "mdisk_grp_name": "Pool0",
        "mode": "array",
        "name": "mdisk0",
        "over_provisioned": "no",
        "site_id": "",
        "site_name": "",
        "status": "online",
        "supports_unmap": "yes",
        "tier": "tier1_flash"
    }
]
```

Python

The script that is shown in Example 13-13 on page 1174 shows an example of the authentication and creation of an access token for further use in the context of querying all available MDisks.

The output of the script provides the following information about each MDisk:

- Name
- Name of the providing controller
- Name of the MDisk group
- Capacity
- Status

The script uses the `getpass` module to prompt for the password and prevent the storage of the credentials in clear text within the script.

Example 13-13 Using the REST API by using Python

```
#!/usr/bin/python

import json
import requests
import getpass

myStorage = 'myStorage'
myUser = 'myUser'
myPassword = getpass.getpass()

### disable SSL verification
ssl_verify = False

### ignore warning for SSL not being used
from requests.packages.urllib3.exceptions import InsecureRequestWarning
requests.packages.urllib3.disable_warnings(InsecureRequestWarning)

### get session token
tokenRequest = requests.post('https://' + myStorage + ':7443/rest/v1/auth',
        headers={
                'Content-type':         'application/json',
                'X-Auth-Username':      myUser,
                'X-Auth-Password':      myPassword
                },
        params="", data="", verify=ssl_verify)

### convert to JSON
_token = json.loads(tokenRequest.text)
token = _token['token']

### get mdisks
mdiskRequest = requests.post('https://' + myStorage + ':7443/rest/v1/lsmdisk',
        headers={
                'Content-type':         'application/json',
                'X-Auth-token':         token
                },
        params="", data="", verify=ssl_verify)

_mdisks = json.loads(mdiskRequest.text)

print( '{:32.32s} {:20.20s} {:32.32s} {:8.8s} {:10.10s}' \
        .format("name","controller_name","mdisk_grp_name","capacity","status") )

for mdisk in _mdisks:
        print( '{:32.32s} {:20.20s} {:32.32s} {:8.8s} {:10.10s}' \
.format(mdisk['name'],mdisk['controller_name'],mdisk['mdisk_grp_name'],mdisk['capacity'],md
isk['status']) )
```

The use of the `verify=False` option allows insecure SSL connections. By default, every SSL connection is verified to be secure. This option allows the request to get proceed; otherwise, the connection is considered insecure. If you use a signed SSL certificate, you do *not* need this option.

Perl

The script that is shown in Example 13-14 shows an example of the authentication and creation of an access token for further use in the context of querying all available MDisks.

The output of the script provides the following information about each MDisk:

- Name
- Name of the providing controller
- Name of the MDisk group
- Capacity
- Status

The script uses the `IO::Prompter` module to prompt for the password and prevent the storage of the credentials in clear text within the script.

Example 13-14 Using the REST API by using Perl

```
#!/usr/bin/perl

use strict;
use JSON;
use REST::Client;
use IO::Prompter;

my $myStorage = 'myStorage';
my $myUser = 'myUser';

my $myPassword = prompt 'Please enter your password:', -echo=>"*";

my $restURL = 'https://' . $myStorage . ':7443/rest/v1/';

### get the session token
my $tokenRequest = REST::Client->new();
$tokenRequest->addHeader('Content-type', 'application/json');
$tokenRequest->addHeader('X-Auth-Username' , $myUser);
$tokenRequest->addHeader('X-Auth-Password', $myPassword);
$tokenRequest->getUseragent()->ssl_opts('verify_hostname' => 0);
$tokenRequest->POST($restURL . '/auth');
my $token = decode_json($tokenRequest->responseContent())->{'token'};

### get the list of mdisks
my $mdiskRequest = REST::Client->new();
$mdiskRequest->addHeader('Content-type', 'application/json');
$mdiskRequest->addHeader('X-Auth-Token', $token);
$mdiskRequest->getUseragent()->ssl_opts('verify_hostname' => 0);
$mdiskRequest->POST($restURL . '/lsmdisk');

my $mdiskList = $mdiskRequest->responseContent();
my @mdiskListJSON = @{decode_json($mdiskList)};

for my $key (@mdiskListJSON) {
        printf "%32s %20s %32s %8s %10s\n",
                $key->{'name'},
                $key->{'controller_name'},
                $key->{'mdisk_grp_name'},
```

```
                $key->{'capacity'},
                $key->{'status'};
}
```

The use of the `getUseragent()->ssl_opts('verify_hostname' => 0)` method allows insecure SSL connections. By default, every SSL connection is verified to be secure. This option allows the request to proceed; otherwise, the connection is considered insecure. If you use a signed SSL certificate, you do *not* need this option.

PowerShell

A PowerShell script will inherit the security settings for TLS/SSL from the operating system if they are not explicitly set in the script. This could cause communication failures if the TLS/SSL setting does not match the configuration of **`lssecurity`** (sslprotocol) on the Storage Virtualize system.

By default, a Storage Virtualize system uses self-signed SSL certificates.

During deployment a customer may choose to use the same certificates or replace them with CA signed certificates. Additional consideration is required in the PowerShell script when working with self-signed certificates.

The examples in this chapter are very basic because the intention is mainly to demonstrate the command syntax for different use cases with PowerShell.

For an actual deployment of a script, use recommended practices.

For example, use PowerShell credentials, secure strings or hash files on disk for credentials, instead of having them in plain text in the script.

It is possible to access the REST API using two PowerShell cmdlets.

- `Invoke-WebRequest`
- `Invoke-RestMethod`

The examples within this chapter will focus on the `Invoke-RestMethod`.

Table 13-3 shows a syntax comparison of Curl, PowerShell, and PowerShell Core

Table 13-3 Syntax comparison

Description	curl	Windows PowerShell Invoke-RestMethod	PowerShell Core Invoke-RestMethod
SSL/TLS version	--tlsX.Y	Separate command	-SslProtocol
SSL verification	-k	Custom scripting needed	-SkipCertificateCheck
HTTP method	X	-Method	-Method
Headers	-H	-Headers	-Headers
Content Type Header	-H	-ContentType	-ContentType
Data	-d	-Body	-Body
URL/URI	argument	-Uri	-Uri

There are some differences in the PowerShell language between Windows PowerShell and PowerShell (Core). The differences are most notable in the availability and behavior of

PowerShell cmdlets between Windows and non-Windows platforms and the changes resulting from the differences between the .NET Framework and the .NET Core.

PowerShell on Linux and macOS uses .NET Core, which is a subset of the full .NET Framework on Microsoft Windows. This is important because PowerShell provides direct access to the underlying framework types and methods.

This difference becomes visible when using the `Invoke-RestMethod cmdlet` to determine a specific SSL/TLS version or SSL verification behavior.

While the `Invoke-RestMethod cmdlet` from PowerShell Core provides dedicated options for this, using the cmdlet in conjunction with Windows PowerShell requires a separate command or additional scripting.

Examples

The script that is shown in Example 13-15 shows an example of the authentication and creation of an access token for further use in the context of querying all available MDisks using the Windows PowerShell `cmdlet Invoke-RestMethod`.

The output of the script provides the following information about each MDisk:

- Name
- Name of the providing controller
- Name of the MDisk group
- Capacity
- Status

Example 13-15 Using the REST API by using Windows PowerShell

```
# Storage and according credentials
$myStorage = 'myStorage'
$myUser = 'myUser'
$myPassword = Read-Host "Enter Password" -AsSecureString

# Force TLS V1.2
[System.Net.ServicePointManager]::SecurityProtocol = [Net.SecurityProtocolType]::Tls12

#Code to handle self signed certificates
if (-not
([System.Management.Automation.PSTypeName]'ServerCertificateValidationCallback').Type)
{
$certCallback = @"
    using System;
    using System.Net;
    using System.Net.Security;
    using System.Security.Cryptography.X509Certificates;
    public class ServerCertificateValidationCallback
    {
        public static void Ignore()
        {
            if(ServicePointManager.ServerCertificateValidationCallback ==null)
            {
                ServicePointManager.ServerCertificateValidationCallback +=
                    delegate
                    (
                        Object obj,
                        X509Certificate certificate,
                        X509Chain chain,
                        SslPolicyErrors errors
                    )
```

```
                    {
                        return true;
                    };
                }
            }
        }

"@
    Add-Type $certCallback
}
[ServerCertificateValidationCallback]::Ignore()
# Get auth token
$token = Invoke-RestMethod -Method 'Post' -Headers @{'X-Auth-Username' = $($myUser);
'X-Auth-Password' = $($myPassword)} -ContentType 'application/json' -Uri
https://$($myStorage):7443/rest/auth

# Get mdisks
Invoke-RestMethod -Method 'Post' -Headers @{'X-Auth-Token' = $Token.token} -ContentType
'application/json' -Uri https://$($myStorage):7443/rest/lsmdisk | Format-Table name,
controller_name, mdisk_grp_name, capacity, status
```

The script that is shown in Example 13-16 shows an example of the authentication and creation of an access token for further use in the context of querying all available MDisks using the PowerShell Core `cmdlet Invoke-RestMethod`.

Because the cmdlet of PowerShell Core does provide the parameters `SslProtocol` and `SkipCeritificateCheck`, there is no need for additional code within the script to workaround that circumstances.

The output of the script provides the following information about each MDisk:

- Name
- Name of the providing controller
- Name of the MDisk group
- Capacity
- Status

Example 13-16 Using the REST API by using PowerShell Core

```
# Storage and according credentials
$myStorage = 'myStorage'
$myUser = 'myUser'
$myPassword = Read-Host "Enter Password" -AsSecureString

# Get auth token
$token = Invoke-RestMethod -SslProtocol Tls12 -SkipCertificateCheck-Method 'Post' -Headers
@{'X-Auth-Username' = $($myUser); 'X-Auth-Password' = $($myPassword)} -ContentType
'application/json' -Uri https://$($myStorage):7443/rest/auth

# Get mdisks
Invoke-RestMethod -SslProtocol Tls12 -SkipCertificateCheck -Method 'Post' -Headers
@{'X-Auth-Token' = $Token.token} -ContentType 'application/json' -Uri
https://$($myStorage):7443/rest/lsmdisk | Format-Table name, controller_name,
mdisk_grp_name, capacity, status
```

The script that is shown in Example 13-17 shows an example of the authentication and creation of an access token for further use in the context of creating a vDisk using the Windows PowerShell `cmdlet Invoke-RestMethod`.

Example 13-17 Using the REST API by using Windows PowerShell - providing data as body

```
# Storage and according credentials
$myStorage = 'myStorage'
$myUser = 'myUser'
$myPassword = Read-Host "Enter Password" -AsSecureString

# Force TLS V1.2
[System.Net.ServicePointManager]::SecurityProtocol = [Net.SecurityProtocolType]::Tls12

#Code to handle self signed certificates
[Code to ignore self signed certificate errors goes here, omitted for brevity]

# Get auth token
$token = Invoke-RestMethod -Method 'Post' -Headers @{'X-Auth-Username' = $($myUser);
'X-Auth-Password' = $($myPassword)} -ContentType 'application/json' -Uri
https://$($myStorage):7443/rest/auth

#Define a hash table with the parameters that you would like to pass to the REST API call
in "-body"
$body = @{
"mdiskgrp" = 1
"unit" = "gb"
"size" = 100
"name" = "my_vDisk_01"
}
#It is required to convert the content defined within the body to JSON format
$body_json = $body | ConvertTo-Json

# Get mdisks
Invoke-RestMethod -Method 'Post' -Headers @{'X-Auth-Token' = $Token.token} -ContentType
'application/json' -Body $body_json -Uri https://$($myStorage):7443/rest/mkmdisk
```

The script that is shown in Example 13-17 shows an example of the authentication and creation of an access token for further use in the context of list pool properties. This example shall demonstrate how to pass a Boolean value to `-Body` from a dictionary, for example a parameter like, `-bytes` does not have an argument so passing of a boolean value is required.

Example 13-18 Using the REST API by using Windows PowerShell - passing boolean values

```
# Storage and according credentials
$myStorage = 'myStorage'
$myUser = 'myUser'
$myPassword = Read-Host "Enter Password" -AsSecureString

# Force TLS V1.2
[System.Net.ServicePointManager]::SecurityProtocol = [Net.SecurityProtocolType]::Tls12

#Code to handle self signed certificates
[Code to ignore self signed certificate errors goes here, omitted for brevity]

# Get auth token
$token = Invoke-RestMethod -Method 'Post' -Headers @{'X-Auth-Username' = $($myUser);
'X-Auth-Password' = $($myPassword)} -ContentType 'application/json' -Uri
https://$($myStorage):7443/rest/auth
```

```
#Define a hash table with the parameters that you would like to pass to the REST API call
in "-body"
$body = @{
   "bytes" = [bool]::Parse('true')
}

#It is required to convert the content defined within the body to JSON format
$body_json = $body | ConvertTo-Json

# Get mdisks
Invoke-RestMethod -Method 'Post' -Headers @{'X-Auth-Token' = $Token.token} -ContentType
'application/json' -Body $body_json -Uri https://$($myStorage):7443/rest/mkmdiskgrp
```

13.2.5 HTTPS and REST API on IBM Storage Control

You can use the REST API for IBM Storage Control to access information about resources and to generate custom capacity, configuration, and performance reports.

The main advantage of this method is that you can access all information about the entire SAN and storage infrastructure that is managed by the IBM Storage Control server (see Example 13-19).

Example 13-19 Using the IBM Storage Control REST API by using Python

```
#!/usr/bin/python

import requests
import getpass
username = 'myUser'
password = getpass.getpass()

url = 'https://storagecontrol-server:9569/srm/'

sesssion = requests.Session()
sesssion.verify = False

response = session.post(url + 'j_security_check',
                     data={'j_username': username, 'j_password': password})
response.raise_for_status()

response = session.get(url + 'REST/api/v1/' + 'StorageSystems')
response.raise_for_status()

print(response.json())
```

13.3 Automation with Red Hat Ansible

Automation is a priority for maintaining today's busy storage environments. Automation software allows for the creation of repeatable sets of instructions. It also reduces the need for human interaction with computer systems.

Red Hat Ansible and other third-party automation tools are becoming increasingly used across the enterprise IT environments. It is not unexpected that their use in storage environments is to become more popular.

13.3.1 Red Hat Ansible

The IBM SAN Storage Virtualize Controller family includes integration with Red Hat Ansible Automation Platform. This integration allows IT to create an Ansible playbook that automates repetitive tasks across an organization in a consistent way, which helps improve outcomes and reduces errors.

Ansible is an agentless automation management tool that uses the SSH protocol. As of this writing, Ansible can be run from any machine with Python 2 (version 2.7) or Python 3 (versions 3.5 and higher) installed. Supported platforms for Ansible include Red Hat, SUSE, Debian, CentOS, macOS, and any of the Berkeley Software Distribution (BSD) versions.

Note: Windows is *not* supported for the Ansible control node.

13.3.2 Red Hat Ansible editions

The following Red Hat Ansible editions are available:

- Ansible Core

  Ansible Core is the command line tool that is installed from community repositories or the official Red Hat repositories for Ansible.

- Ansible Tower

  Ansible Tower is the GUI tool that is used to run Ansible tasks. Tower requires a license that is based on the number of systems Ansible Tower is to manage. Ansible Tower is available as Standard or Premium Edition (24x7 support is included in the Premium Edition).

13.3.3 Requirements

Ansible server (Control Node) features the following requirements:

- Python 2 (version 2.7) or Python 3 (versions 3.5 and higher)

  Note: Some plug-ins that run on the control node include other requirements. These requirements are listed in the plug-in documentation.

- Host requirements:
  - Although you do not need a daemon on your managed nodes, you need a way for Ansible to communicate with them.
  - For most managed nodes, Ansible makes a connection over SSH and transfers modules by using SFTP. If SSH works but SFTP is not available on some of your managed nodes, you can switch to SCP in `ansible.cfg`.
  - For any machine or device that can run Python, you also need Python 2 (version 2.6 or later) or Python 3 (version 3.5 or later).

Note: Some modules feature more requirements that must be met on the ‘target’ machine (the managed node). These requirements are listed in the module documentation.

13.3.4 Essential terminology in an Ansible environment

Ansible environment features the following essential terminology:

- Ansible Galaxy: A hub for finding and sharing Ansible content.
- Ansible server: The machine with Ansible installed, which runs all tasks and playbooks.
- Playbook: A framework where Ansible automation tasks are defined (written in YAML).
- Task: A section that contains a single procedure that you want to be run.
- Tag: A name that you can assign to a task.
- Play: The execution of a playbook.
- Hosts: The devices that you manage with Ansible.
- Modules: A command or set of commands that are made for execution on the client side.
- Handler: A task that is called only if a notifier is present.
- Notifier: A section that is assigned to a task that calls a handler if the output is changed.
- Inventory: A file that contains Ansible client/server data.
- Fact: Information that is retrieved from the client from global variables by using the `gather-facts` operation.
- Roles: A structured way of grouping tasks, handlers, variables, and other properties.
- Container: Ansible Container uses Ansible roles to build images, initialize projects, and add services to projects.

13.3.5 Automating IBM Storage with Ansible

IBM data storage provides simple storage solutions that address modern data requirements and provides a solution to your hybrid multicloud strategy.

With the speed, scale, and complexity of hybrid multicloud and even traditional on-premises environments, automation became a priority.

IBM Storage FlashSystem family for hybrid multicloud includes integration with Red Hat Ansible Automation Platform. It allows IT to create an Ansible playbook that automates the tasks that are repeated across an organization in a consistent way, which helps improve outcomes and reduces risk.

It also standardizes how IT and application owners interact together and features the following benefits:

- With Red Hat Ansible Automation Platform and IBM Storage, customers can easily automate tasks, such as configuration management, provisioning, workflow orchestration, application deployment, and lifecycle management.
- By using Red Hat Ansible Automation Platform and IBM Storage, customers can reduce system inconsistencies with the automation modules.
- Red Hat Ansible Automation Platform can also be used to configure end-to-end infrastructure in an orchestrated fashion.
- Ansible provides a single pane of glass visibility to multicluster, multicloud environments, which allows lines of business to use playbooks to accomplish their goals without needing to understand the details of how the work is being done.

IBM is a Red Hat-certified support module vendor that provides simple management for the following commands that are used in the IBM Storage Virtualize Ansible Collection:

- Collect facts: Collect basic information, including hosts, host groups, snapshots, consistency groups, and volumes
- Manage hosts: Create, delete, or modify hosts
- Manage volumes: Create, delete, or extend the capacity of volumes
- Manage MDisk: Create or delete a managed disk
- Manage pool: Create or delete a pool (managed disk group)
- Manage volume map: Create or delete a volume map
- Manage consistency group snapshot: Create or delete consistency group snapshots
- Manage snapshot: Create or delete snapshots
- Manage volume clones: Create or delete volume clones

This collection provides a series of Ansible modules and plug-ins for interacting with the IBM Storage Virtualize family storage products. The modules in the IBM Storage Virtualize Ansible collection use the REST API to connect to the IBM Storage Virtualize storage system. These products include:

- IBM SAN Volume Controller
- IBM Storage FlashSystem family members that are built with IBM Storage Virtualize
- IBM Storwize family
- IBM Storage Virtualize for Public Cloud

For more information, see *Automate and Orchestrate Your IBM FlashSystem Hybrid Cloud with Red Hat Ansible*, REDP-5598.

For IBM Storage Virtualize modules, Ansible version 2.9 or higher is required. For more information about IBM Storage Virtualize modules, see the Ansible web page ibm.spectrum_virtualize.

13.3.6 Getting started

The Ansible Collection (`ibm.storage_virtualize`) provides a series of Ansible modules and plug-ins for interacting with the IBM Storage Virtualize family storage products,.

As of this writing, the Ansible collection for IBM Storage Virtualize is available in version 2.0.

All information in this section is based on this version.

Prerequisites for using the modules

`Paramiko` must be installed to use i**`bm_svctask_command`** and **`ibm_svcinfo_command`** modules.

`Paramiko` is a Python (2.7, 3.4+) implementation of the SSHv2 protocol, and provides client and server functions.

Although `Paramiko` is a Python C extension for low-level cryptography, it is a pure Python interface around SSH networking concepts.

Current limitations

The modules in the IBM Storage Virtualize Ansible collection use the REST API to connect to the IBM Storage Virtualize storage system.

This collection includes the following limitations:

- The use of the REST API to list more than 2000 objects might create a loss of service from the API side because it automatically restarts because of memory constraints.
- The Ansible collection can run on all IBM Storage Virtualize storage versions that are 8.1.3, except versions 8.3.1.3, 8.3.1.4 and 8.3.1.5.
- It is not possible to access the REST API by using a cluster IPv6 address.
- When the Storage Virtualize Ansible v1.8.0 collection was released, it is possible only to automate the license agreements acceptance, including EULA, for Licensed Machine Code (LMC) systems. For non-LMC systems, the license agreement acceptance is to be presented with a GUI setup wizard upon user-interface login, whether the Ansible modules were used for initial configuration.

Prerequisites

Ensure that the following prerequisites are met:

- Ansible is installed and configured on a controller node.
- Ansible Galaxy Collection `ibm.storage_virtualize` is installed on the same controller node.
- Network access is available from the controller node to IBM Storage Virtualize Management IP.
- A user with sufficient authorization to create or delete objects on IBM Storage Virtualize.
- IBM Storage Virtualize operates at version 8.1.3 or higher.

Installing or upgrading Ansible Galaxy Collection ibm.storage_virtualize

To install the IBM Storage Virtualize collection that is hosted in Galaxy, use the following command:

```
ansible-galaxy collection install ibm.storage_virtualize
```

To upgrade to the latest version of the IBM Storage Virtualize collection, use the following command:

```
ansible-galaxy collection install ibm.storage_virtualize --force
```

Functions provided by IBM Storage Virtualize Ansible modules

The `ibm.storage_virtualize` collection provides the following modules:

- `ibm_svc_auth`: Generates an authentication token for a user on the IBM Storage Virtualize family storage system.
- `ibm_svc_complete_initial_setup`: Completes the initial setup configuration for Licensed Machine Code (LMC) systems. For non-LMC systems, logging in to the user interface is required to complete the automation of Day 0 configuration.
- `ibm_svc_host`: Manages hosts that are on the IBM Storage Virtualize system.
- `ibm_svc_hostcluster`: Manages the host cluster that is on the IBM Storage Virtualize system.
- `ibm_svc_info`: Collects information about the IBM Storage Virtualize system.
- `ibm_svc_initial_setup`: Manages initial setup configuration on the IBM Storage Virtualize system.
- `ibm_svc_manage_callhome`: Manages configuration of the Call Home feature on the Storage Virtualize system.

- `ibm_svc_manage_consistgrp_flashcopy`: Manages the FlashCopy consistency groups that are on the IBM Storage Virtualize system.
- `ibm_svc_manage_cv`: Manages the change volume in remote copy replication that is on the IBM Storage Virtualize system.
- `ibm_svc_manage_flashcopy`: Manages the FlashCopy mappings that are on the IBM Storage Virtualize system.
- `ibm_svc_manage_ip`: Manages IP provisioning on the IBM Storage Virtualize system.
- `ibm_svc_manage_migration`: Manages volume migration between clusters on IBM Storage Virtualize systems.
- `ibm_svc_manage_mirrored_volume`: Manages the mirrored volumes that are on the IBM Storage Virtualize system.
- `ibm_svc_manage_ownershipgroup`: Manages ownership groups on the IBM Storage Virtualize system.
- `ibm_svc_manage_portset`: Manages IP portset on IBM Storage Virtualize system.
- `ibm_svc_manage_replication`: Manages the remote copy replication that is on the IBM Storage Virtualize system.
- `ibm_svc_manage_replicationgroup`: Manages the remote copy consistency group on the IBM Storage Virtualize system.
- `ibm_svc_manage_safeguarded_policy`: Manages safeguarded policy configuration on the IBM Storage Virtualize system.
- `ibm_svc_manage_sra`: Manages the remote support assistance configuration on the IBM Storage Virtualize system.
- `ibm_svc_manage_user`: Manages users on the IBM Storage Virtualize system.
- `ibm_svc_manage_usergroup`: Manages user groups on the IBM Storage Virtualize system.
- `ibm_svc_manage_volume`: Manages the standard volumes on the IBM Storage Virtualize system.
- `ibm_svc_manage_volumegroup`: Manages the volume groups that are on the IBM Storage Virtualize system.
- `ibm_svc_mdisk`: Manages the MDisks for IBM Storage Virtualize system.
- `ibm_svc_mdiskgrp`: Manages pools for IBM Storage Virtualize system.
- `ibm_svc_start_stop_flashcopy`: Starts or stops the FlashCopy mapping and consistency groups that are on the IBM Storage Virtualize system.
- `ibm_svc_start_stop_replication`: Starts or stops the remote copy relationship or group on the IBM Storage Virtualize system.
- `ibm_svc_vol_map`: Manages the volume mapping for IBM Storage Virtualize system.
- `ibm_svcinfo_command`: Runs the **`svcinfo`** CLI command on the IBM Storage Virtualize system over an SSH session.
- `ibm_svctask_command`: Runs the **`svctask`** CLI commands on the IBM Storage Virtualize system over and SSH session.
- `ibm_sv_manage_awss3_cloudaccount`: Manages Amazon S3 cloud account configuration on IBM Storage Virtualize system.
- `ibm_sv_manage_cloud_backup`: Manages cloud backup on the IBM Storage Virtualize system.
- `ibm_sv_manage_fc_partnership`: Manages Fibre Channel (FC) partnership on the IBM Storage Virtualize system.

- `ibm_sv_manage_fcportsetmember`: Manages addition or removal of ports from the Fibre Channel (FC) portsets on the IBM Storage Virtualize system.
- `ibm_sv_manage_ip_partnership`: Manages IP partnership configuration on the IBM Storage Virtualize system.
- `ibm_sv_manage_provision_policy`: Manages provisioning policy configuration on IBM Storage Virtualize system.
- `ibm_sv_manage_replication_policy`: Manages policy-based replication configuration on the IBM Storage Virtualize system.
- `ibm_sv_manage_snapshot`: Manages snapshots (mutual consistent images of a volume) on the IBM Storage Virtualize system.
- `ibm_sv_manage_snapshotpolicy`: Manages snapshot policy configuration on the IBM Storage Virtualize system.
- `ibm_sv_manage_ssl_certificate`: Exports an existing system certificate on to IBM Storage Virtualize system.
- `ibm_sv_manage_truststore_for_replication`: Manages certificate trust stores for replication on the IBM Storage Virtualize system.
- `ibm_sv_restore_cloud_backup`: Restores cloud backups on IBM Storage Virtualize system.
- `ibm_sv_switch_replication_direction`: Switches the replication direction on the IBM Storage Virtualize system.

Note: Beginning with version 1.6.0, the `ibm_svc_vdisk` module is considered a deprecated feature. A new module (`ibm_svc_manage_volume`) was introduced to manage standard volumes.

Getting help for IBM Storage Virtualize Ansible modules

To get the online documentation for a specific module that is displayed, use the following command:

```
ansible-doc <collection-name>.<module-name>
```

The output of the help includes all permissible options and some examples of how to use the module (see Example 13-20).

Example 13-20 Example displaying online help

```
ansible-doc ibm.storage_virtualize.ibm_svc_manage_volume

> IBM_SVC_MANAGE_VOLUME
Ansible interface to manage 'mkvolume', 'rmvolume', and 'chvdisk' volume commands.

  * This module is maintained by The Ansible Community
OPTIONS (= is mandatory):

- buffersize
        Specifies the pool capacity that the volume will reserve as a buffer for
thin-provissioned and compressed volumes.
        Parameter 'thin' or 'compressed' must be specified to use this parameter.
        The default buffer size is 2%.
        `thin' or `compressed' is required when using `buffersize'.
        Valid when `state=present' to create a volume.
        [Default: (null)]
```

```
        type: str

= clustername
        The hostname or management IP of the Storage Virtualize storage system.

        type: str
:
:
:
```

13.3.7 Securing credentials in Ansible

While working with Ansible, you can create several playbooks, inventory files, variable files, and so on. Some of the files might contain sensitive data, such as access credentials. To protect this kind of data, Ansible provides the Ansible Vault, which helps to prevent this data from being exposed. Sensitive data and passwords are kept in an encrypted file rather than in plain text files.

13.3.8 Creating an Ansible playbook

The playbook consistently automates the tasks that are repeated across an organization, which improves outcomes and reduces risk. It also standardizes how IT and application owners interact.

In this section, we discuss creating an Ansible playbook. The creation of the playbook is based on the use case that is used here.

For a new VMware ESX cluster that consists of two new servers, two VDisks are to be created and mapped to the host cluster object.

Note: The `idempotency` property might be included in a mathematics or computer science operation. It roughly means that an operation can be carried out multiple times without changing the result.

The IBM Storage Virtualize Ansible modules provide `idempotency` in Ansible playbooks.

The IBM Storage Virtualize Ansible modules check whether the object to be created exists in the defined state and does not attempt to create it again.

Table 13-4 lists the variable parameters and their values for the example playbook.

Table 13-4 Variable parameters and their values for the example playbook

Attribute	Value
Name of new host cluster	ESX-Cluster-1
Name of new host 1	ESX-Host-1
WWPNs of new host 1	100000109C400798, 1000001AB0440446
Name of new host 2	ESX-Host-2
WWPNs of new host 2	100000109B600424, 1000001BC0660146
Name of VDisk 1	Datastore1

Attribute	Value
Name of VDisk 2	Datastore2

The steps that are used to create an Ansible playbook are described next.

Step 1: Authentication

Example 13-21 shows the required YAML notation for the part of the playbook to authenticate at the IBM Storage Virtualize REST API to obtain a token for further use. To avoid storing the password in clear text within the playbook, the password was encrypted in a vault.

Example 13-21 YAML notation for obtaining an authentication token

```
vars:
    clustername: <Cluster management ip | hostname>
    domain: <FQDN>
    username: myuser
    password: !vault |
              $ANSIBLE_VAULT;1.1;AES256
62653531313434393266646438306537396264306433653638343439643136333238383139616561
65303734306362653166396262343763363066303433333640a326332626564656233323336333239
39633132656631353030386430663736363631656438343364346235653534316333333233333531
31663432636265383 60a6336646162643261336433393363333636383232323739623938393566 37
              6138
tasks:
    - name: Obtain an authentication token
      register: result
      ibm_svc_auth:
        clustername: "{{ clustername }}"
        domain: "{{ domain }}"
        username: "{{ username }}"
        password: "{{ password }}"
```

For more information about how to work with Ansible vaults, see Protecting sensitive data with Ansible vault.

Step 2: Creating the host cluster object

Example 13-22 shows the required YAML notation for the part of the playbook to create an empty host cluster object.

Example 13-22 YAML notation for creating an empty host cluster

```
- name: Define a new host cluster
      ibm_svc_hostcluster:
        clustername: "{{ clustername }}"
        domain: "{{ domain }}"
         token: "{{ result.token }}"
         log_path: "{{ log_path }}"
        name: <hostcluster_name>
        state: present
```

Step 3: Creating an FC host

Example 13-23 on page 1189 shows the required YAML notation for the part of the playbook to create an FC host object.

Example 13-23 YAML notation for creating a new FC host object

```
- name: Define a new FC host
      ibm_svc_host:
        clustername: "{{ clustername }}"
        domain: "{{ domain }}"
        token: "{{ result.token }}"
        log_path: "{{ log_path }}"
        name: "{{ hostname }}"
        state: present
        fcwwpn: "{{ fcwwpn(s) }}"
        iogrp: 0:1:2:3
        protocol: scsi
        type: generic
         hostcluster: "{{ hostcluster_name }}"
```

Step 4: Creating a thin-provisioned volume

Example 13-24 shows the required YAML notation for the part of the playbook to create a thin-provisioned volume.

Example 13-24 YAML notation to create a thin-provisioned volume

```
- name: Create a thin-provisioned volume
      ibm_svc_manage_volume:
        clustername: "{{ clustername }}"
        domain: "{{ domain }}"
        token: "{{ result.token }}"
        log_path: "{{ log_path }}"
        name: "volume_name"
        state: "present"
        pool: "<pool_name>"
        size: "<size>"
        unit: "<size_unit>"
        thin: true
        buffersize: 10%
```

Step 5: Mapping the new volume to the host cluster object

Example 13-25 shows the required YAML notation for the part of the playbook to map the new volume to the hostcluster.

Example 13-25 YAML notation to map a volume to the hostcluster

```
    - name: Map a volume to a hostcluster
      ibm_svc_vol_map:
        clustername: "{{ clustername }}"
        domain: "{{ domain }}"
        token: "{{ result.token }}"
        log_path: "{{ log_path }}"
        volname: <volume_name>
        hostcluster: <hostcluste_-name>
        state: present
```

If a SCSI-Id must be specified, use the scsi: `<scsi-id>` parameter.

Putting it all together

Example 13-26 shows the combined required tasks for the use of the IBM Storage Virtualize collection to create hostcluster volumes (this use case is featured in this chapter) to be used as a playbook by Ansible. All customized lines of the playbook are highlighted in bold in the example.

Example 13-26 Complete playbook for specified use-case

```
- name: Using Storage Virtualize collection to create hostcluster - hosts -
volumes
  hosts: localhost
  collections:
    - ibm.storage_virtualize
  gather_facts: no
  connection: local

# definition of global variables
  vars:
    clustername: mySVC
    domain: mydomain.com
    username: myuser
    password: !vault |
              $ANSIBLE_VAULT;1.1;AES256
62653531313434393266646438306537396264306433653638343439643136333238383139616561
65303734306362653166396262343763363066303433333640a326332626564656233323336333239
39633132656631353030386430663736363631656438343364346235653534316333333233333531
31663432636265383 60a6336646162643261336433393363333636383232323739623938393566376138
    log_path: /tmp/redbook-example.log

# define variables for running the playbook
# hostcluster
    hostcluster_name: ESX-Cluster-1

# host 1
    host1_name: ESX-Host-1
    host1_fcwwpn: 100000109C400798{{":"}}1000001AB0440446

# host 2
    host2_name: ESX-Host-2
    host2_fcwwpn: 100000109B600424{{":"}}1000001BC0660146

# pools to use for volume mirror
    pool1_name: pool1
    pool2_name: pool2

#  volume 1
    volume1_name: Datastore1
    volume1_size: '10'
    volume1_size_unit: tb

#  volume 2
    volume2_name: Datastore2
    volume2_size: '10'
    volume2_size_unit: tb
```

```
tasks:
# creating an authentication token for further usage within the playbook
    - name: Obtain an authentication token
      register: result
      ibm_svc_auth:
        clustername: "{{  clustername  }}"
        domain: "{{  domain }}"
        username: "{{  username }}"
        password: "{{  password }}"
        log_path: "{{  log_path  }}"

# create the hostcluster object
    - name: Create the hostcluster
      ibm_svc_hostcluster:
        clustername: "{{  clustername  }}"
        domain: "{{  domain  }}"
        token: "{{  result.token  }}"
        log_path: "{{  log_path  }}"
        name: "{{  hostcluster_name  }}"
        state: present

# create first host object
    - name: Define first FC host
      ibm_svc_host:
        clustername: "{{ clustername }}"
        domain: "{{ domain }}"
        token: "{{ result.token }}"
        log_path: "{{ log_path }}"
        name: "{{ host1_name }}"
        state: present
        fcwwpn: "{{ host1_fcwwpn }}"
        iogrp: 0:1:2:3
        protocol: scsi
        type: generic
        hostcluster: "{{ hostcluster_name }}"

# create second host object
    - name: Define second FC host
      ibm_svc_host:
        clustername: "{{ clustername }}"
        domain: "{{ domain }}"
        token: "{{ result.token}}"
        log_path: "{{ log_path }}"
        name: "{{ host2_name }}"
        state: present
        fcwwpn: "{{ host2_fcwwpn }}"
        iogrp: 0:1:2:3
        protocol: scsi
        type: generic
        hostcluster: "{{ hostcluster_name }}"

# create first mirrored thin-provisioned volume
    - name: Create first thin-provisioned volume
      ibm_svc_manage_volume:
        clustername: "{{ clustername }}"
```

```
        domain: "{{ domain }}"
        token: "{{ result.token }}"
        log_path: "{{ log_path }}"
        name: "{{ volume1_name }}"
        state: "present"
        pool: "{{  pool1_name }}:{{ pool2_name }}"
        size: "{{ volume1_size }}"
        unit: "{{ volume1_size_unit }}"
        thin: true
        buffersize: 10%

# create second mirrored thin-provisioned volume
    - name: Create second thin-provisioned volume
      ibm_svc_manage_volume:
        clustername: "{{ clustername }}"
        domain: "{{ domain }}"
        token: "{{ result.token }}"
        log_path: "{{ log_path }}"
        name: "{{ volume2_name }}"
        state: "present"
        pool: "{{ pool1_name }}:{{ pool2_name }}"
        size: "{{ volume2_size }}"
        unit: "{{ volume2_size_unit }}"
        thin: true
        buffersize: 10%

# mapping of first volume to the hostcluster
    - name: Map first volume to the hostcluster
      ibm_svc_vol_map:
        clustername: "{{ clustername }}"
        domain: "{{ domain }}"
        token: "{{ result.token }}"
        log_path: "{{ log_path }}"
        volname: "{{ volume1_name }}"
        hostcluster: "{{ hostcluster_name }}"
        state: present

# mapping of second volume to the hostcluster
    - name: Map second volume to the hostcluster
      ibm_svc_vol_map:
        clustername: "{{ clustername }}"
        domain: "{{ domain }}"
        token: "{{ result.token }}"
        log_path: "{{ log_path }}"
        volname: "{{ volume2_name }}"
        hostcluster: "{{ hostcluster_name }}"
        state: present
```

13.3.9 More automation

The use case that is described in this chapter can be extended by completing the following steps:

1. Create the required FC zoning.
2. Scan the HBA for the newly created volumes.
3. Create a VMFS data store on the discovered volumes.
4. Create one or more virtual machines (VMs).

For more information about the Brocade FOS FC collection on Ansible Galaxy, see the Ansible web page brocade.fos.

For more information about the `community.vmware` Ansible Collection on Ansible Galaxy, see this Ansible web page community.vmware.

A

Command line interface setup

This appendix describes access configuration to the command line interface (CLI) by using the local Secure Shell (SSH) authentication method.

This appendix includes the following topics:

- "Overview" on page 1196
- "Basic setup on a Windows host" on page 1197
- "Basic setup on a Mac, UNIX, or Linux host" on page 1205

© Copyright IBM Corp. 2024. All rights reserved.

Overview

The IBM Storage Virtualize system features a powerful CLI that offers more options and flexibility as compared to the GUI. This appendix describes how to configure a management system by using the SSH protocol to connect to the IBM Storage Virtualize system for issuing commands by using the CLI.

For more information about the CLI, see IBM Documentation.

Note: If a task completes in the GUI, the associated CLI command is always displayed in the details, as shown throughout this book.

In the IBM Storage Virtualize GUI, authentication is performed by supplying a username and password. The CLI uses SSH to connect from a host to the IBM Storage Virtualize system. A private and a public key pair or username and password is necessary.

Using SSH keys with a passphrase is more secure than a login with a username and password because authenticating to a system requires the private key and the passphrase. By using the other method, only the password is required to obtain access to the system.

When SSH keys are used without a passphrase, it becomes easier to log in to a system because you must provide only the private key when performing the login and you are not prompted for password. This option is less secure than using SSH keys with a passphrase.

To enable CLI access with SSH keys, complete the following steps:

1. Generate a public key and a private key as a pair.
2. Upload a public key to the IBM Storage Virtualize system by using the GUI.
3. Configure a client SSH tool to authenticate with the private key.
4. Establish a secure connection between the client and the system.

SSH is the communication vehicle between the management workstation and the IBM Storage Virtualize system. The SSH client provides a secure environment from which to connect to a remote machine. It uses the principles of public and private keys for authentication.

SSH keys are generated by the SSH client software. The SSH keys include a public key, which is uploaded and maintained by the storage system, and a private key, which is kept private on the workstation that is running the SSH client. These keys authorize specific users to access the administration and service functions on the system.

Each key pair is associated with a user-defined ID string that can consist of up to 256 characters. Up to 100 keys can be stored on the system. New IDs and keys can be added, and unwanted IDs and keys can be deleted. To use the CLI, an SSH client must be installed on that system. To use the CLI with SSH keys, the SSH client is required. An SSH key pair also must be generated on the client system, and the client's SSH public key must be stored on the IBM Storage Virtualize system.

Basic setup on a Windows host

The SSH client on a Windows host that is used in this book is PuTTY. A PuTTY key generator can also be used to generate the private and public key pair. The PuTTY client can be downloaded at no cost from Download PuTTY.

Download the following tools:

- PuTTY SSH client: `putty.exe`
- PuTTY key generator: `puttygen.exe`

Generating a public and private key pair

To generate a public and private key pair, complete the following steps:

1. Start the PuTTY key generator to generate the public and private key pair, as shown in Figure A-1.

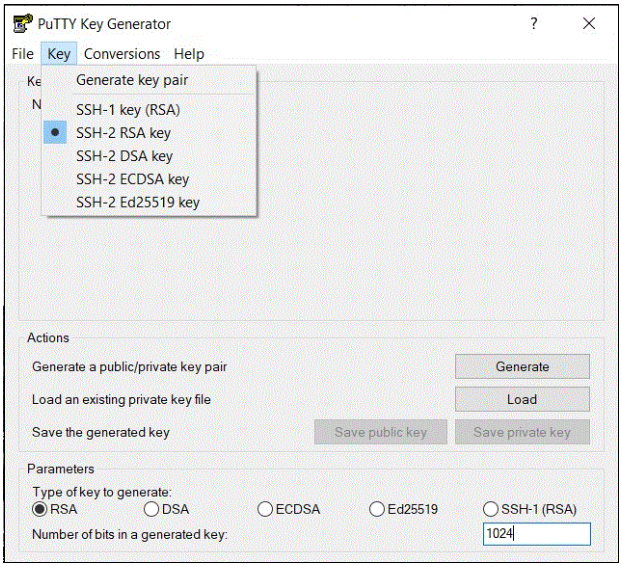


Figure A-1 PuTTY key generator

Select the following options:

- **SSH-2 RSA**
- Number of bits in a generated key: **1024**

> **Note:** Larger SSH keys, such as 2048 bits, are also supported.

2. Click **Generate** and move the cursor over the blank area to generate keys (see Figure A-2).

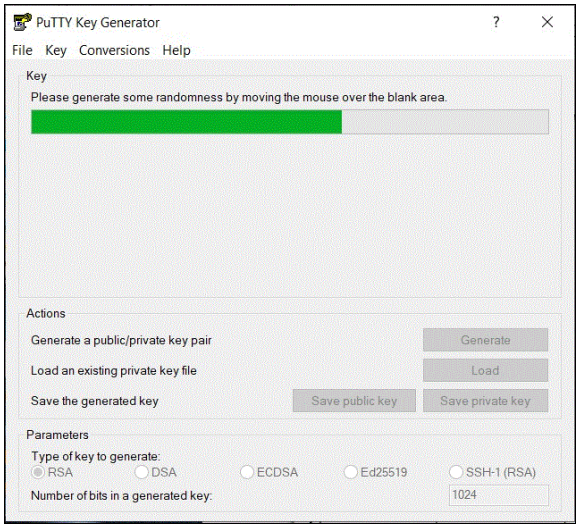


Figure A-2 Generating keys

> **To generate keys**: The blank area that is indicated by the message is the large blank rectangle in the GUI inside the Key field. Continue to move the mouse pointer over the blank area until the progress bar reaches the far right. This action generates random characters based on the cursor location to create a unique key pair.

3. After the keys are generated, save them for later use. Click **Save public key**.
4. You are prompted to enter a name (for example, `sshkey.pub`) and a location for the public key (for example, `C:\Keys\`). Enter this information and click **Save**.

   Ensure that you record the SSH public key name and location because this information must be specified later.

> **Public key extension:** By default, the PuTTY key generator saves the public key with no extension. Use the string `pub` for naming the public key. For example, add the extension `.pub` to the name of the file to easily differentiate the SSH public key from the SSH private key.

5. Click **Save private key**. A warning message is displayed (see Figure A-3). Click **Yes** to save the private key without a passphrase.

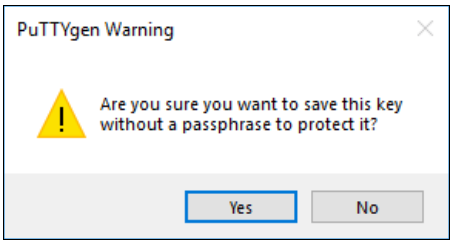


Figure A-3 Confirming the security warning

Note: It is possible to use a passphrase for an SSH key. Although this action increases security, it generates an extra step to log in with the SSH key because it requires the passphrase input.

6. When prompted, enter a name (for example, `sshkey.ppk`), select a secure place as the location, and click **Save**.

Private Key Extension: The PuTTY key generator saves the PuTTY private key (PPK) with the `.ppk` extension. This is a proprietary format for PuTTY and the keys are not interchangeable with OpenSSH clients. There is a utility to convert keys between PuTTY and OpenSSH if you want to use the same keys between the two environments.

7. Close the PuTTY key generator.

Uploading the SSH public key to the IBM Storage System

After you create your SSH key pair, upload your SSH public key onto the IBM Storage System. Complete the following steps:

1. Open the user section in the GUI, as shown in Figure A-4.

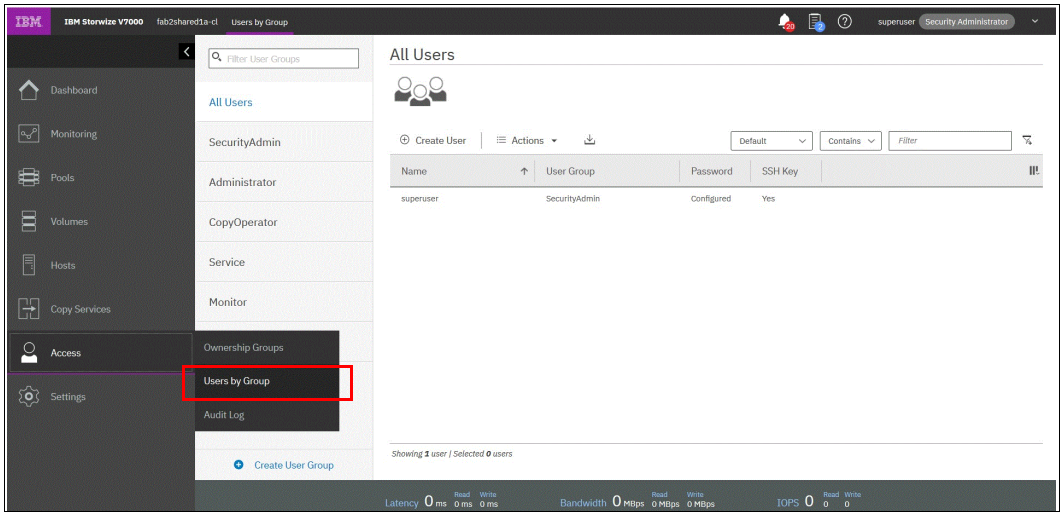


Figure A-4 Opening the user section

2. Right-click the username for which you want to upload the key and click **Properties** (see Figure A-5).

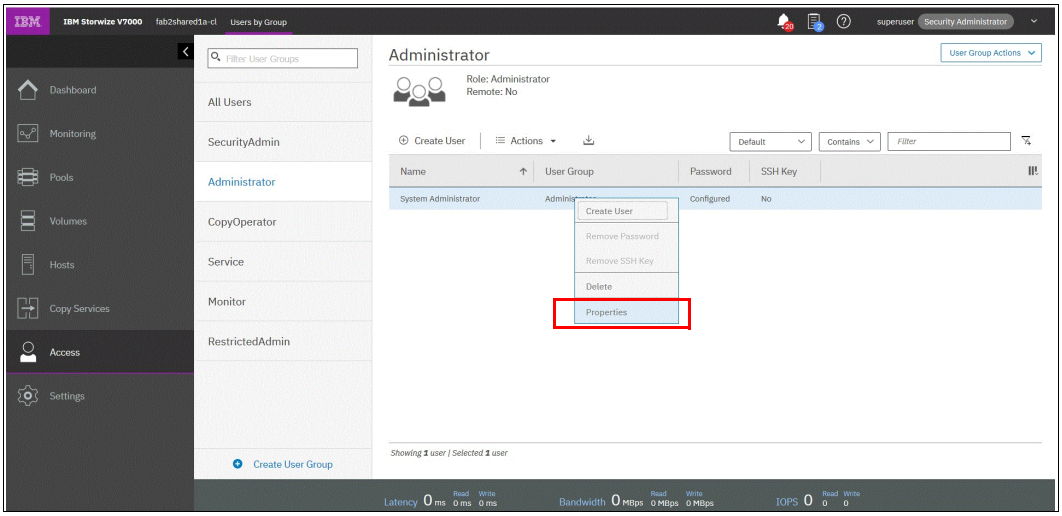


Figure A-5 User properties

3. To upload the public key, click **Browse**, open the folder where you stored the public SSH key, and select the key.
4. Click **OK** and the key is uploaded, as shown in Figure A-6.

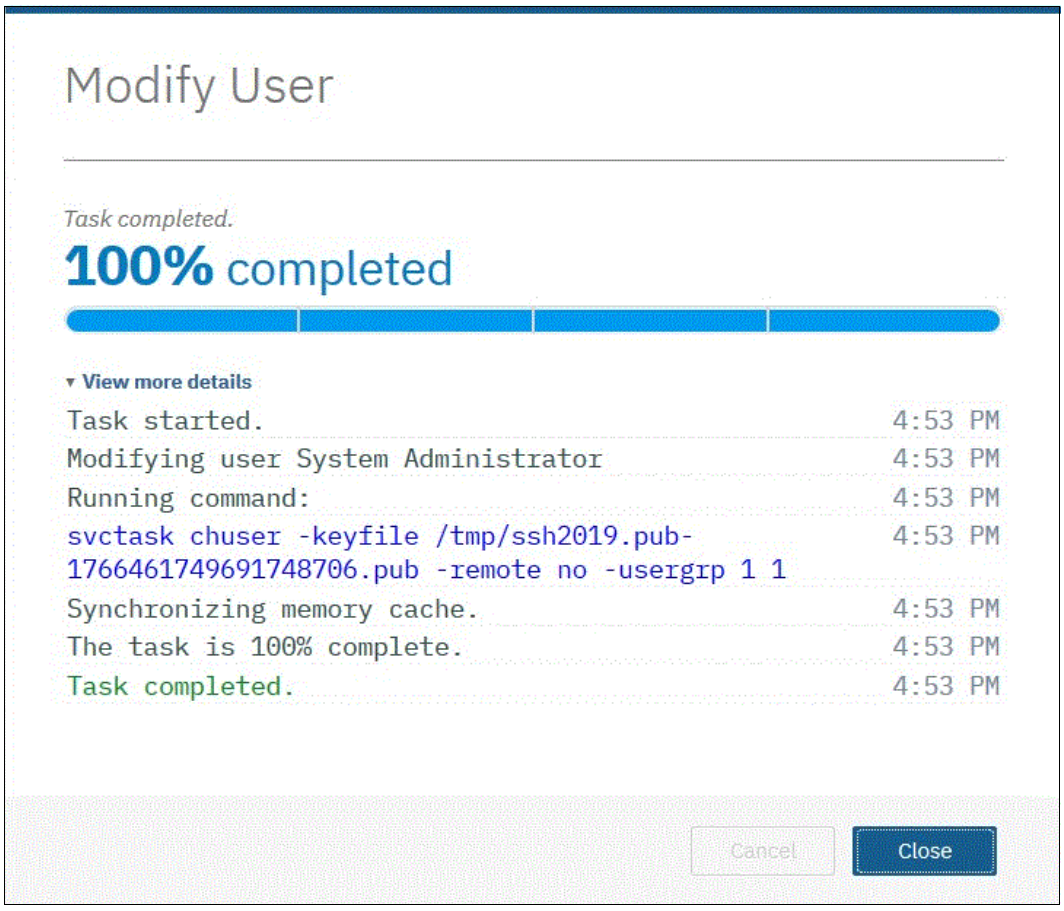


Figure A-6 Confirming the SSH key upload

5. Check in the GUI to ensure that the SSH key is imported successfully (see Figure A-7).

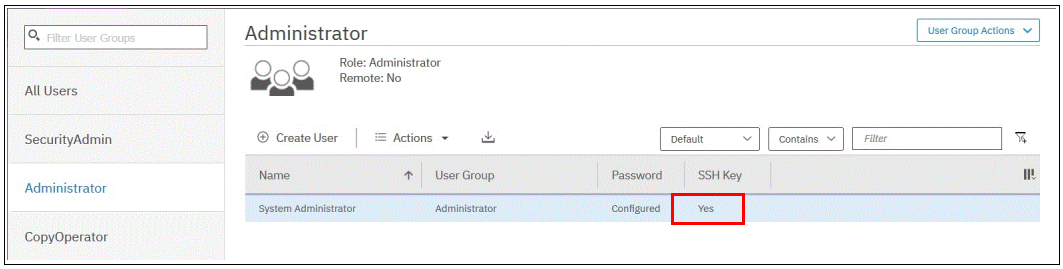


Figure A-7 Key successfully imported

Configuring the SSH client

Before the CLI can be used, the SSH client must be configured. Complete the following steps:

1. Start PuTTY. The PuTTY Configuration window opens (see Figure A-8).

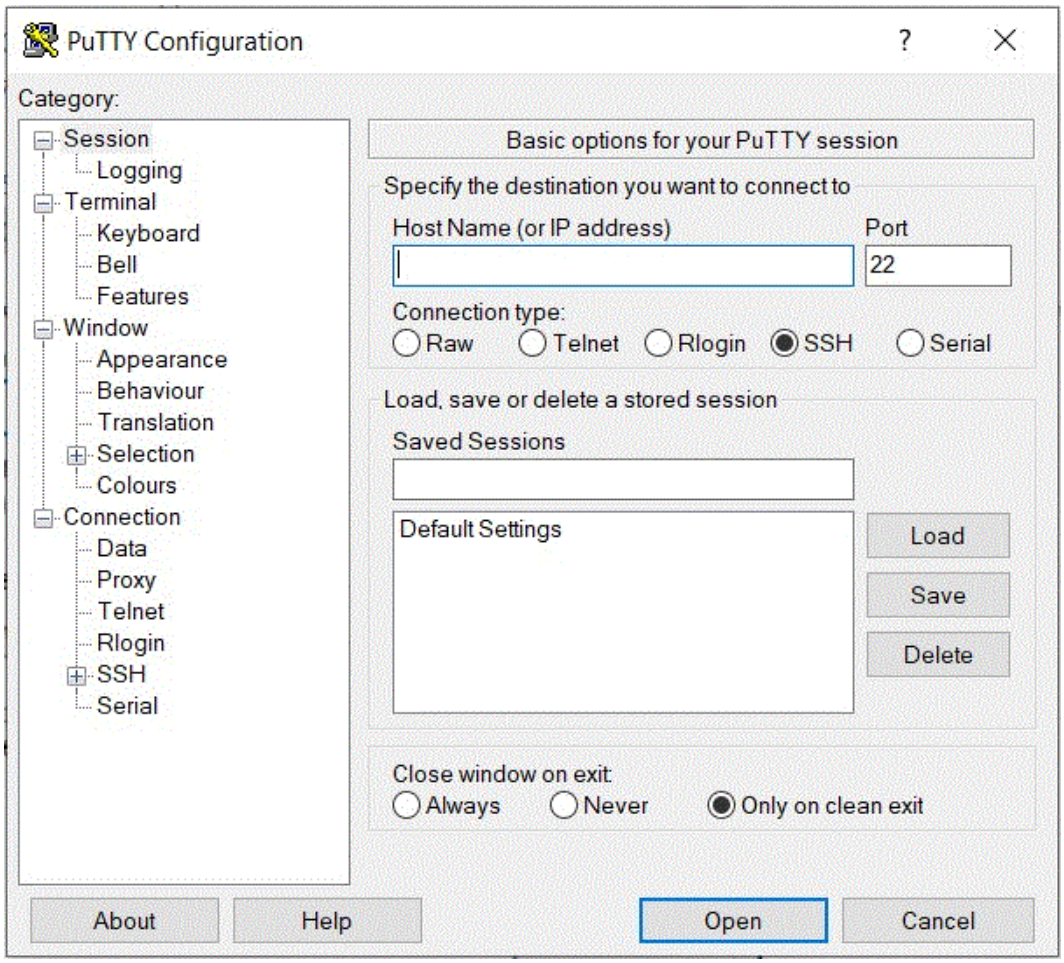


Figure A-8 PuTTY Configuration

2. In the upper right, select **SSH** as the connection type. In the “Close window on exit” section, select **Only on clean exit** (see Figure A-9 on page 1202), which ensures that if any connection errors occur that they are displayed on the user’s window.

3. In the Category window, on the left side of the PuTTY Configuration window, select **Connection** → **Data**, as shown on Figure A-9. In the “Auto-login username” field, enter the IBM Storage Virtualize user ID that was used when uploading the public key. The `admin` account was used in the example that is shown in Figure A-5 on page 1200.

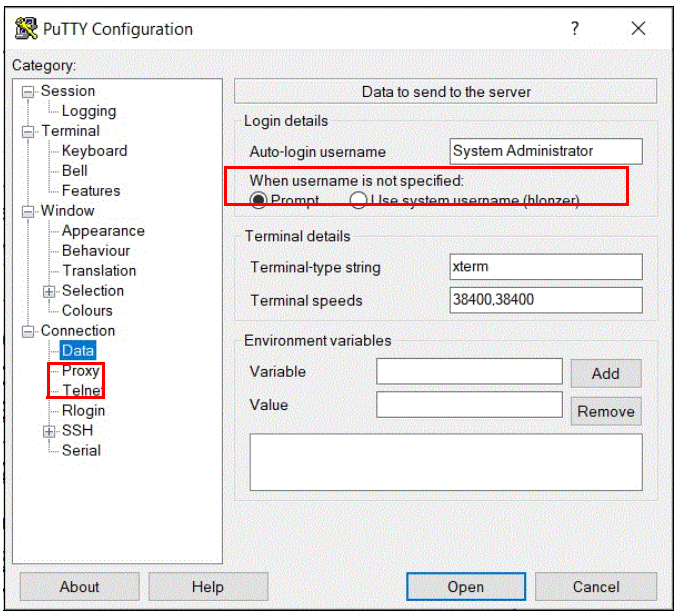


Figure A-9 PuTTY Auto-login username

4. In the Category window, on the left side of the PuTTY Configuration window (see Figure A-10), select **Connection** → **SSH** to open the PuTTY SSH Configuration window. In the SSH protocol version section, select **2**.

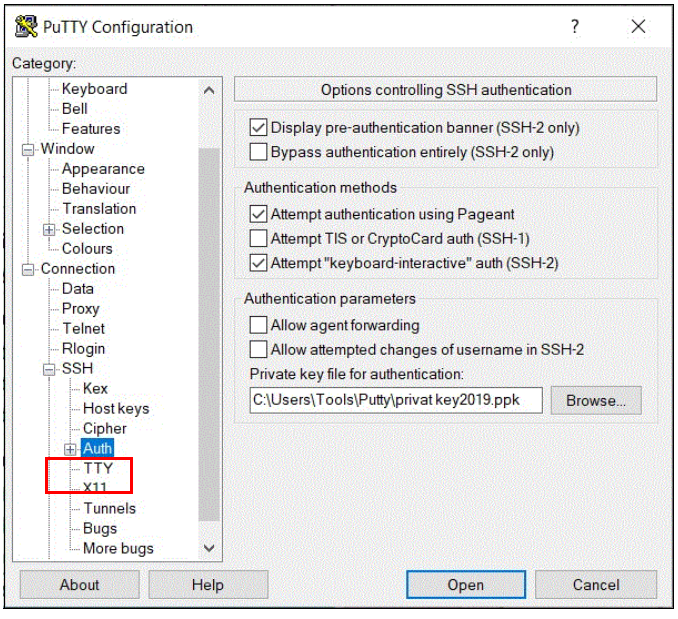


Figure A-10 SSH protocol Version 2

5. In the Category window on the left, select **Connection** → **SSH** → **Auth**. More options are displayed for controlling SSH authentication.

6. In the “Private key file for authentication” field in Figure A-11, browse to or enter the fully qualified directory path and file name of the SSH client private key file that was created (in this example, `C:\Users\Tools\Putty\privatekey2019.ppk` is used).

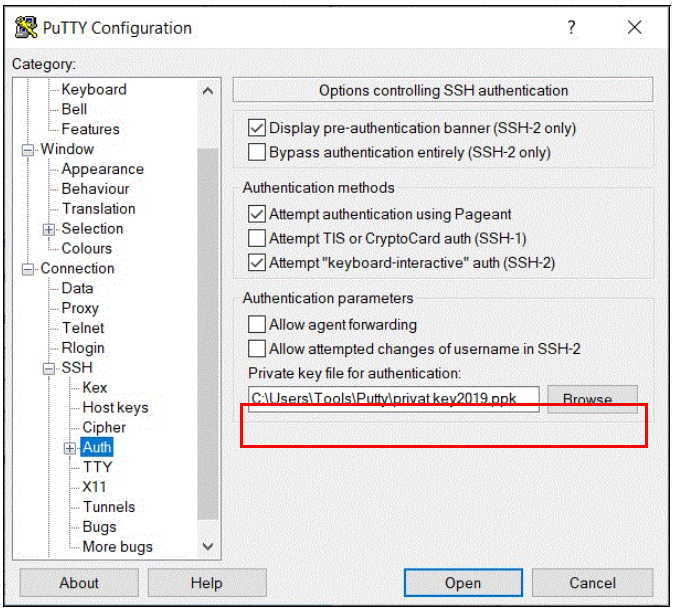


Figure A-11 SSH authentication

7. In the Category window, click **Session** to return to the “Basic options for your PuTTY session” view.
8. Enter the following information in the fields in the right pane (see Figure A-12):
   - Host Name (or IP address): Specify the hostname or system IP address of the IBM Storage Virtualize system.
   - Saved Sessions: Enter a session name.
9. Click **Save** to save the new session (see Figure A-12).

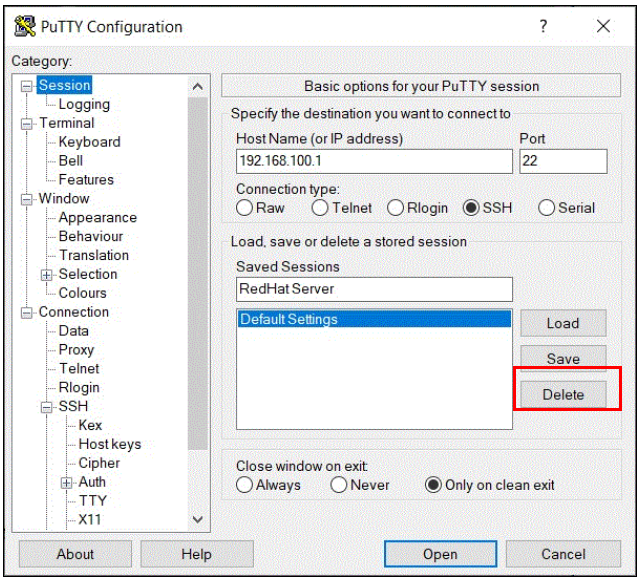


Figure A-12 Session information

10. Select the new session and click **Open** to connect to the IBM Storage Virtualize system, as shown in Figure A-13.

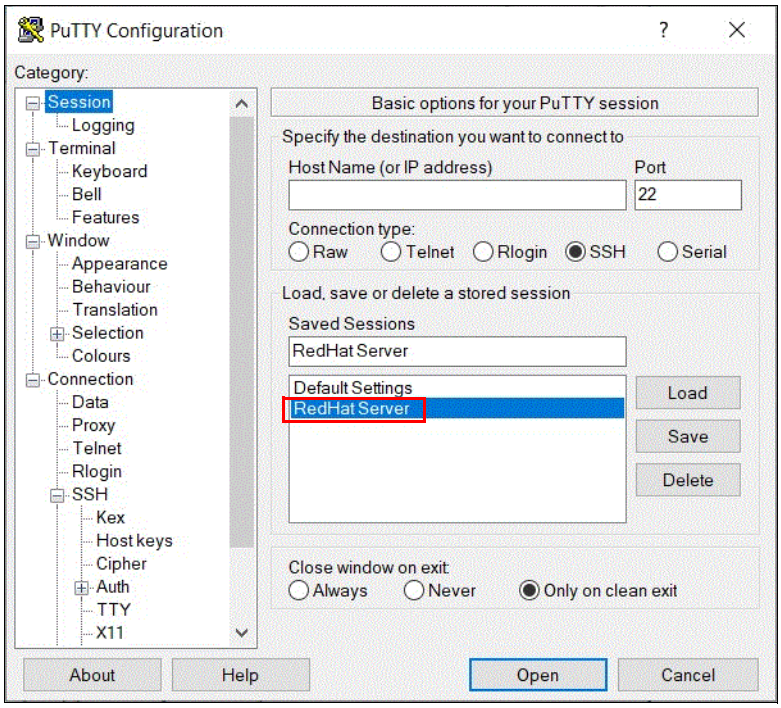


Figure A-13 Connecting to a system

11. If a PuTTY Security Alert opens as shown in Figure A-14, confirm it by clicking **Yes**.

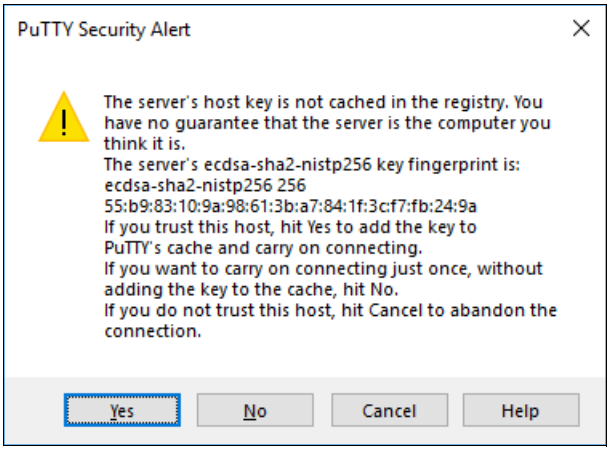


Figure A-14 Confirming the security alert

12. As shown in Figure A-15, PuTTY now connects to the system automatically by using the user ID that was specified earlier, without prompting for password.

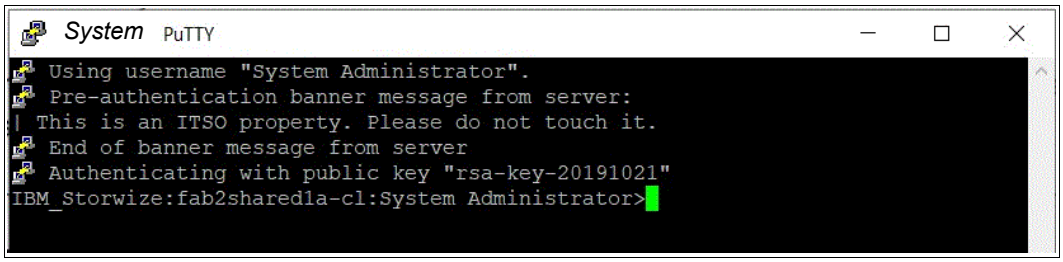


Figure A-15 PuTTY login

The CLI is now configured for IBM Storage Virtualize system administration.

13.4 Basic setup on a Mac, UNIX, or Linux host

The OpenSSH client is the most common tool that is used on Mac, UNIX, or Linux operating systems (OSs). It is installed by default on most of these types of OSs. If OpenSSH is not installed on your system, download it from OpenSSH: Portable Release.

The OpenSSH suite consists of various tools. The following tools are used to generate the SSH keys, transfer the SSH keys to a remote system, and establish a connection to an IBM Storage Virtualize device by using SSH:

- `ssh`: OpenSSH SSH client
- `ssh-keygen`: Tool to generate SSH keys
- `scp`: Tool to transfer files between hosts

Generating a public and private key pair

To generate a public and private key pair to connect to an IBM Storage Virtualize system without entering the user password, run the **`ssh-keygen`** tool, as shown in Example A-1.

Example: A-1 SSH keys generation with ssh-keygen

```
# ssh-keygen -t rsa -b 1024
Generating public/private rsa key pair.
Enter file in which to save the key (//.ssh/id_rsa): /.ssh/sshkey
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /.ssh/sshkey.
Your public key has been saved in /.ssh/sshkey.pub.
The key fingerprint is:
55:5e:5e:09:db:a4:11:01:b9:57:96:74:0c:85:ed:5b root@hostname.ibm.com
The key's randomart image is:
+--[ RSA 1024]----+
|           .+=BO*|
|           + oB*+|
|          . oo+o |
|         . . .  E|
|        S   .   o|
|                 |
|                 |
|                 |
|                 |
+-----------------+
#
```

In `ssh-keygen`, the parameter **`-t`** refers to the type of SSH key (RSA in Example A-1 on page 1205) and **`-b`** is the size of SSH key in bits (in Example A-1 on page 1205, 1024 bits was used).

You must also specify the path and name for the SSH keys. The name that you provide is the name of the private key. The public key has the same name, but with extension `.pub`. In Example A-1 on page 1205, the path is `/.ssh/`, the name of the private key is `sshkey`, and the name of the public key is `sshkey.pub`.

Note: Using a passphrase for the SSH key is optional. If a passphrase is used, security is increased, but more steps are required to log in with the SSH key because the user must enter the passphrase.

Uploading the SSH public key to the IBM Storage System

To upload the new SSH public key to IBM Storage Virtualize by using the GUI, see “Uploading the SSH public key to the IBM Storage System” on page 1199.

To upload the public key by using the CLI, complete the following steps:

1. On the SSH client (for example, AIX or Linux host), run **`scp`** to copy the public key to the IBM Storage System. The basic syntax for the command is:

   ```
   scp <file> <user>@<hostname_or_IP_address>:<path>
   ```

   The directory `/tmp` in the IBM Storage Virtualize active configuration node can be used to store the public key temporarily. Example A-2 shows the command to copy the newly generated public key to the IBM Storage Virtualize system.

   Example: A-2 SSH public key copy to an IBM Storage System

   ```
   # scp /.ssh/sshkey.pub admin@192.168.100.1:/tmp/
   Password:*********
   sshkey.pub
   100%  241     0.2KB/s   00:00
   #
   ```

2. Log in to the storage system by using SSH and run the **`chuser`** command (as shown in Example A-3) to associate the public SSH key with a user.

   Example: A-3 Importing the SSH public key to a user

   ```
   IBM_Storage System:ITSO:admin>chuser -keyfile /tmp/sshkey.pub admin
   IBM_Storage System:ITSO:admin>lsuser admin
   id 4
   name admin
   password yes
   ssh_key yes
   remote no
   usergrp_id 1
   usergrp_name Administrator
   IBM_Storage System:ITSO:admin>
   ```

   When running the **`lsuser`** command as shown in Example A-3, it is indicated that a user has a configured SSH key in the field `ssh_key`.

Connecting to an IBM Storage Virtualize system

Now that the SSH key is uploaded to the IBM Storage Virtualize system and assigned to a user account, you can connect to the device by running the **ssh** command with the following options:

```
ssh -i <SSH_private_key> <user>@<IP_address_or_hostname>
```

Example A-4 shows the SSH command that is running from an AIX server and connecting to the storage system with an SSH private key and no password prompt.

Example: A-4 Connecting to IBM Storage System with an SSH private key

```
# ssh -i /.ssh/sshkey admin@192.168.100.1
IBM_Storage System:ITSO:admin>
```

B

Terminology

This appendix lists the IBM SAN Volume Controller, IBM FlashSystem, and the IBM Storage Virtualize terms that are commonly used in this book.

For more information about the complete set of terms that are related to the IBM SAN Volume Controller, see the glossary in the IBM SAN Volume Controller Documentation.

For more information about the terms that related to the IBM FlashSystem, see the glossary in the IBM FlashSystem Documentation.

© Copyright IBM Corp. 2024. All rights reserved.

Commonly encountered terms

This book uses the common IBM Storage Virtualize, IBM SAN Volume Controller, and IBM FlashSystem terminology that is listed here.

Access mode
One of the modes in which a logical unit (LU) in a disk controller system can operate. The three access modes are image mode, managed space mode, and unconfigured mode. See also “Image mode” on page 1225, “Managed mode” on page 1228, and “Unconfigured mode” on page 1239.

Activation key
See “License key” on page 1227.

Address Resolution Protocol
A protocol that dynamically maps an IP address to a network adapter address in a local area network.

Advisory lock
A type of lock that a process holds on a region of a file that signals any other process to not use or lock the region or an overlapping region. Other processes are not forced to comply.

Agent code
An open-systems policy that interprets Common Information Model (CIM) requests and responses as they are transferred between the client application and the device.

Allocatable extent limit
A maximum total capacity for the system. The allocatable extent limit is calculated from pool extent sizes.

Array
An ordered collection or group of physical devices (disk drive modules, SSDs, or FCMs) that are used to define logical volumes or devices. An array is a group of drives that is designated to be managed with a redundant array of independent disks (RAID).

Asymmetric virtualization
A virtualization technique in which the virtualization engine is outside the data path and performs a metadata-style service. The metadata server contains all the mapping and locking tables, and the storage devices contain only data. See also “Symmetric virtualization” on page 1238.

Asynchronous replication
A type of replication in which control is given back to the application when the write operation is made to the source volume. Later, the write operation is made to the target volume. See also “Synchronous replication” on page 1238.

Audit Log
An unalterable record of all commands or user interactions that are issued to the system.

Authenticated user
A user who has logged in to the system with a valid account (user ID and password).

Authentication
The mechanism by which a system determines what permissions that a particular authenticated user has, so the user can ccess specific resources or actions. See also “Authorization”.

Authorization
The mechanism by which a system determines what permissions that a particular authenticated user has to access specific resources or actions. See also “Authentication”.

Authorization code
An alphanumeric code generated for administrative functions, such as password resets or two-factor authentication bypass.

Automatic data placement mode
An Easy Tier operating mode in which the host activity on all the volume extents in a pool are “measured,” a migration plan is created, and then automatic extent migration is performed.

Auxiliary volume
The auxiliary volume that contains a mirror of the data on the master volume. See also “Master volume” on page 1228, and “Reliability, availability, and serviceability” on page 1234.

Available (usable) capacity
See “Capacity” on page 1212.

Back end and front end
The view of an IBM SAN Volume Controller or IBM FlashSystem where the drives or the virtualized external storage systems are visible to the controllers. See also “Front end and back end” on page 1223.

Boot drive
A drive that includes the required software to start a system.

Cache
Storage or memory that is used to improve access times to instructions, data, or both. For example, data that is in cache memory is normally a copy of data that is elsewhere in slower, less expensive storage, such as on a disk or on another network node.

Caching input/output group
The caching input/output (I/O group) is the I/O group in the system that performs the cache function for a volume.

Call Home
A communication link that is established between a product and a service provider. The product can use this link to call IBM or another service provider when the product requires service. With access to the machine, service personnel can perform service tasks, such as viewing error and problem logs or initiating trace and memory dump retrievals.

Canister

A single processing unit within a storage system.

Capacity

IBM applies the following definitions to capacity:

- Available capacity

  The amount of usable capacity that is not yet used in a system, pool, array, or managed disk (MDisk).

- Data reduction

  A set of techniques that can be used to reduce the amount of usable capacity that is required to store data. Examples of data reduction include data deduplication and compression.

- Data reduction savings

  The total amount of usable capacity that is saved in a system, pool, or volume through the application of an algorithm, such as compression or deduplication on the written data. This saved capacity is the difference between the written capacity and the used capacity.

- Effective capacity

  The amount of provisioned capacity that can be created in a system or pool without running out of usable capacity given the current data reduction savings being achieved. This capacity equals the usable capacity times the data reduction rat.

- Overhead capacity

  An amount of usable capacity that is occupied by metadata in a system or pool and other data that is used for system operations.

- Overprovisioned ratio

  The ratio of provisioned capacity to usable capacity in the pool or system.

- Overprovisioning

  The result of creating more provisioned capacity in a storage system or pool than there is usable capacity. Overprovisioning occurs when thin provisioning or data reduction techniques ensure that the used capacity of the provisioned volumes is less than their provisioned capacity.

- Physical capacity

  Physical capacity indicates the total capacity in all storage on the system. Physical capacity includes all the storage that the system can virtualize and assign to pools.

- Provisioned capacity

  Total capacity of all volumes and volume copies in a pool or system.

- Provisioning limit - maximum provisioned capacity - overprovisioning limit

  In some storage systems, restrictions in the storage hardware or configured by the user define the limit of the maximum provisioned capacity in a pool or system.

- Raw capacity

  The reported capacity of the drives in the system before formatting or RAID is applied.

- Standard provisioning

  The ability to completely use a volume's capacity for that specific volume.

- Standard provisioned volume

  A volume that uses all the storage at creation.

- Thin-provisioning savings

  The total amount of usable capacity that is saved in a pool, system, or volume by using usable capacity when needed as a result of write operations. The capacity that is saved is the difference between the provisioned capacity minus the written capacity.

- Total capacity savings

  The total amount of usable capacity that is saved in a pool, system, or volume through thin-provisioning and data reduction techniques. The capacity that is saved is the difference between the used usable capacity and the provisioned capacity.

- Usable capacity

  The amount of capacity that is provided for storing data on a system, pool, array, or MDisk after formatting and RAID techniques are applied. Usable capacity is the total of used and available capacity. For example, 50 TiB used, 50 TiB available is a usable capacity of 100 TiB.

- Used capacity

  The amount of usable capacity that is taken up by data or capacity in a system, pool, array, or MDisk after data reduction techniques are applied.

- Written capacity

  The amount of usable capacity that might be used to store written data in a pool or system before data reduction is applied.

- Written capacity limit

  The largest amount of capacity that can be written to a drive, array, or MDisk. The limit can be reached even when usable capacity is still available.

Capacity licensing

A licensing model that licenses features with a price-per-terabyte model. Licensed features are IBM FlashCopy, Metro Mirror (MM), Global Mirror (GM), and virtualization. See also “FlashCopy” on page 1222, “Metro Mirror” on page 1229, and “Virtualized storage” on page 1240. With 8.5, only license External Virtualization is discussed.

Capacity recycling

The amount of provisioned capacity that can be recovered without causing stress or performance degradation. This capacity identifies the amount of resources that can be reclaimed and provisioned to other objects in an environment.

Capacity threshold

The amount of physical configured storage space that is available for stored capacity or reserved capacity must exceed before a notification is sent. See also “Total usable physical capacity” on page 1239.

Certificate

A digital document that binds a public key to the identity of the certificate owner, which enables the certificate owner to be authenticated. A certificate is issued by a certificate authority (CA) and is digitally signed by that authority.

Certificate chain
If the system certificate is signed by an intermediate certificate authority (CA), then the full chain of certificates must be installed. To install the signed certificate and CA certificates, create a single file that contains the full chain of certificates. The file should include the signed certificate and the intermediate CA certificates. The root CA certificate can be included, but is optional.

Chain
A set of enclosures that is attached to provide redundant access to the drives inside the enclosures. Each control enclosure can have one or more chains.

Challenge Handshake Authentication Protocol
An authentication protocol that protects against eavesdropping by encrypting the username and password.

Change volume
A volume that is used in GM that holds earlier consistent revisions of data when changes are made and are crash consistent.

Channel extender
A device that is used for long-distance communication that connects other storage area network (SAN) fabric components. Generally, channel extenders can involve protocol conversion to asynchronous transfer mode (ATM), IP, or another long-distance communication protocol.

Child pool
Used to control capacity allocation for volumes that are used for specific purposes. Rather than being created directly from MDisks, child pools are created from existing capacity that is allocated to a parent pool. As with parent pools, volumes can be created that specifically use the capacity that is allocated to the child pool. Child pools are similar to parent pools with similar properties. Child pools can be used for volume copy operation. See also “Parent pool” on page 1231.

Clone
A copy of a volume on a server at a particular point in time (PiT). The contents of the copy can be customized while the contents of the original volume are preserved.

Cloud account
An agreement with a cloud service provider (CSP) to use storage or other services at that service provider. Access to the cloud account is granted by presenting valid credentials.

Cloud container
A virtual object that includes all of the elements, components, or data that is common to a specific application or data.

Cloud service provider
The company or organization that provides off- and on-premises cloud services, such as storage, server, and network. IBM Storage Virtualize includes built-in software capabilities to interact with CSPs such as IBM Cloud, Amazon S3, and deployments of OpenStack Swift.

Cloud tenant
A group or an instance that provides common access with the specific privileges to an object, software, or data source.

Clustered system
A group of up to eight IBM Storage Systems canisters (two in each system) that presents a single configuration, management, and service interface to the user.

Cold extent
An extent that is so infrequently accessed that it does not need the performance benefit if it is moved to a faster media.

Command line interface (CLI)
A computer interface in which the input and output are text-based.

Compression
A function that removes repetitive characters, spaces, strings of characters, or binary data from the data that is being processed and replaces characters with control characters. Compression reduces the amount of storage space that is required for data.

Compression accelerator
Hardware onto which the work of compression is offloaded from the microprocessor.

Configuration node
While the cluster is operational, a single node in the cluster is appointed to provide configuration and service functions over the network interface. This node is termed the configuration node. This configuration node manages the data that describes the clustered-system configuration and provides a focal point for configuration commands. If the configuration node fails, another node in the cluster transparently assumes that role.

Consistency group
A group of copy relationships between virtual volumes or data sets that are maintained with the same time reference so that all copies are consistent in time. A consistency group can be managed as a single entity.

Container
A software object that holds or organizes other software objects or entities.

Contingency capacity
For thin-provisioned volumes that are configured to automatically expand, the unused real capacity that is maintained. For thin-provisioned volumes that are not configured to automatically expand, the difference between the used capacity and the new real capacity.

Control enclosure
A hardware unit that includes the enclosure chassis, node canisters, drives, and system function.

Copied state

Copied is a FlashCopy state that indicates that a copy was triggered after the copy relationship was created. The Copied state indicates that the copy process is complete, and the target disk has no further dependency on the source disk. The time of the last trigger event is normally displayed with this status.

Copy-on-write

A snapshot method that uses read and write operations to maintain an image of the data. See also “Redirect-on-write” on page 1233.

Copyback

A process that moves data back to its expected or preferred location to maintain an array in a more efficient configuration after a failed drive is replaced.

Counterpart SAN

A non-redundant portion of a redundant storage area network (SAN). A counterpart SAN provides all the connectivity of the redundant SAN but without the redundancy. Each counterpart SAN provides an alternative path for each SAN-attached device. See also “Redundant Storage Area Network” on page 1234.

Cross-volume consistency

A consistency group property that ensures consistency between volumes when an application issue dependent write operations that span multiple volumes.

Customer-replaceable unit (CRU)

An assembly or part that can be replaced in its entirety by a user when any one of its components fails.

Data consistency

A characteristic of the data at the target site where the dependent write order is maintained to ensure the recoverability of applications.

Data deduplication

A method of reducing storage needs by eliminating redundant data. Only one instance of the data is retained on storage media. Other instances of the same data are replaced with a pointer to the retained instance.

Data encryption key

Used to encrypt data. It is created automatically when an encrypted object, such as an array, a pool, or a child pool, is created. It is stored in secure memory and it cannot be viewed or changed. The data encryption key is encrypted by using the master access key.

Data migration

The movement of data from one physical location to another physical location without the disruption of application I/O operations.

Data reduction

A set of techniques that can be used to reduce the amount of physical storage that is required to store data. An example of data reduction includes data deduplication and compression. See also “Data reduction pool” on page 1217 and “Capacity” on page 1212.

Data reduction savings

The total amount of usable capacity that is saved in a system, pool, or volume through the application of an algorithm, such as compression or deduplication on the written data. This saved capacity is the difference between the written capacity and the used capacity. See also “Capacity” on page 1212 and “Data reduction” on page 1217

Data reduction pool

Data reduction pools are an allocation of data blocks that can free unused (unmapped or overwritten) capacity at a fine grain.

Dense wavelength division multiplexing

A technology that places many optical signals onto one single-mode fiber by using slightly different optical frequencies. DWDM enables many data streams to be transferred in parallel.

Deduplication

See “Data deduplication” on page 1216.

Dependent write operation

Must be applied in the correct order to maintain cross-volume consistency.

Destage

To move data from cache to a nonvolatile storage medium.

Directed maintenance procedure

The fix procedures, which are also known as directed maintenance procedures (DMPs), ensure that you fix any outstanding errors in the error log. To fix errors, from the Monitoring window, click **Events**. The Next Recommended Action is displayed at the top of the Events window. Select **Run This Fix Procedure** and follow the instructions.

Discovery

The automatic detection of a network topology change, for example, new and deleted nodes or links.

Disk tier

MDisks (logical unit numbers (LUNs)) that are presented to the IBM Storage cluster likely have different performance attributes because of the type of disk or RAID array on which they are installed. The MDisks can be on Storage Class Memory (SCM), Flash Core Modules (FCM) SSDs, Small Computer System Interface (SCSI) (SAS) disk, nearline (NL) SAS. Therefore, a storage tier attribute is assigned to each MDisk, and the default is `generic_hdd`.

Distributed redundant array of independent disks

An alternative RAID scheme where the number of drives that are used to store the array can be greater than the equivalent, typical RAID scheme. The same data stripes are distributed across a greater number of drives, which increases the opportunity for parallel I/O and improves the overall performance of the array. This technology removes the need of separate spare disks. The spare capacity is spread over all disks in the DRAID area. See also “Rebuild area” on page 1233.

Domain name server

Domain name server (DNS) is a server program that supplies name-to-address conversion by mapping domain names to IP addresses.

Domain name system

Domain name system is the distributed database system that maps domain names to IP addresses.

Drive class

A combination of drive technology and speed, which uniquely defines a class of drives that have approximately the same performance characteristics in a pool:

- Storage Class Memory

  Storage Class Memory tier exists when the pool contains drives that use persistent memory technologies that improve the endurance and speed of current flash storage device technologies.

- Flash Class

  Flash Class exists when the pool contains high-performance flash drives.

- Enterprise Class

  Enterprise Class exists when the pool contains enterprise-class Disks, which are disk drives that are optimized for performance.

- Nearline Class

  Nearline Class exists when the pool contains nearline-class Disks, which are disk drives that are optimized for capacity.

Drive technology

A category of a drive that pertains to the method and reliability of the data storage techniques being used on the drive. Possible values include enterprise (ENT) drive, NL drive, or solid-state drive (SSD).

Dual Inline Memory Module

A small circuit board with memory-integrated circuits containing signal and power pins on both sides of the board. The following terms are associated with DIMMs:

- Channel: The memory modules are installed into matching banks, which are usually color-coded on the system board. These separate channels enable the memory controller to access each memory module. For the Intel Cascade Lake architecture, there are six DIMM Memory channels per CPU, and each memory channel has two DIMMs. The memory bandwidth is tied to each of these channels, and the speed of access for the memory controller is shared across the pair of DIMMs in that channel.
- Slot: Generally, the physical slot that a DIMM can fit into, but in this context, a slot is DIMM0 or DIMM1, which refers to the first or second slot within a channel on the system board. There are two slots per memory channel on the IBM SAN Volume Controller SV2 hardware. On the system board, DIMM0 is the blue slot and DIMM1 is the black slot within each channel.
- Rank: A single-rank DIMM has one set of memory chips that is accessed while writing to or reading from the memory. A dual-rank DIMM is like having two single-rank DIMMs on the same module, with only one rank accessible at a time. A quad-rank DIMM is, effectively, two dual-rank DIMMs on the same module. The 32G DIMMS are dual rank.

Dynamic random access memory

A memory in which the cells require repetitive application of control signals to retain stored data.

Easy Tier

A volume performance function within the IBM Storage family that provides automatic data placement of a volume's extents in a multitiered storage pool. The pool normally contains a mix of SCM, flash drives and HDDs. Easy Tier measures host I/O activity on the volume's extents and migrates hot extents onto the flash drives to ensure the maximum performance.

Effective capacity

The amount of provisioned capacity that can be created in a system or pool without running out of usable capacity given the current data reduction savings being achieved. This capacity equals the usable capacity that is divided by the data reduction savings percentage. See also "Capacity" on page 1212.

Encryption deadlock

A specific case of not being able to access keys because the key server is trying to boot from storage for which it has the key. See also "Encryption recovery key"

Encryption key

Also known as *master access key*, it is created and stored on USB flash drives or on a key server when encryption is enabled. The master access key is used to decrypt the data encryption key.

Encryption key label

The list of encryption key labels used by the storage system to identify keys that are to be used on the key server.

Encryption key manager/server

An internal or external system that receives and then serves encryption keys or certificates to a storage system.

Encryption recovery key

Enables a method to recover where the normal encryption key servers are not available, or from an encryption deadlock situation.

Encryption of data-at-rest

The inactive encryption data that is stored physically on the storage system.

Enhanced stretched system

A stretched system is an extended high availability (HA) method that is supported by the SAN Volume Controller to enable I/O operations to continue after the loss of half of the system. Enhanced Stretched Systems provide several primary benefits:

- In addition to the automatic failover that occurs when a site fails in a standard stretched system configuration, an Enhanced Stretched System provides a manual override that can be used to select which of the two sites continues operation.
- Enhanced Stretched Systems intelligently route I/O traffic between nodes and controllers to reduce the amount of I/O traffic between sites, and to minimize the effect on host application I/O latency.
- Enhanced Stretched Systems include an implementation of more policing rules to ensure that the correct configuration of a standard stretched system is used.

Evaluation mode

An Easy Tier operating mode in which the host activity on all the volume extents in a pool are “measured” only. No automatic extent migration is performed.

Event (error)

An occurrence of significance to a task or system. Events can include the completion or failure of an operation, user action, or a change in the state of a process.

Event code

A value that is used to identify an event condition to a user. This value might map to one or more event IDs or to values that are presented on the service window. This value is used to report error conditions to IBM and to provide an entry point into the service guide.

Event ID

A value that is used to identify a unique error condition that was detected by the IBM Storage System. An event ID is used internally in the cluster to identify the error.

Excluded condition

The excluded condition is a status condition. It describes an MDisk that the IBM Storage System decided is no longer sufficiently reliable to be managed by the cluster. The user must issue a command to include the MDisk in the cluster-managed storage.

Exabyte (EB)/Exbibyte (EiB)

A exabyte (EB) is for processor storage, real and virtual storage, and channel volume. For disk storage capacity and communications volume, it is 10 to the power of 18 = 1,000,000,000,000,000,000 bytes. An Exbibyte (EiB) is two to the power of 60 or 1,152,921,504,606,846,976 bytes.

Extent
A fixed-size unit of data that is used to manage the mapping of data between MDisks and volumes. The size of the extent can range from 16 MB–8 GB.

External storage
Refers to MDisks that are SCSI LUs that are presented by storage systems that are attached to and managed by the clustered system.

Failback
The restoration of an appliance to its initial configuration in the primary data center after the detection and repair of a failed network or component.

Failover
An automatic operation that switches to a redundant or standby system or node in a software, hardware, or network interruption. See also “Failback”.

Feature activation code
An alphanumeric code that activates a licensed function on a product. See also “License key” on page 1227.

Fibre Channel
A technology for transmitting data between computer devices. It is especially suited for attaching computer servers to shared storage devices and for interconnecting storage controllers and drives. See also “Zoning” on page 1242.

Fibre Channel Arbitrated Loop
An implementation of the FC standards that uses a ring topology for the communication fabric, as described in American National Standards Institute (ANSI) INCITS 272-1996 (R2001). In this topology, two or more FC end points are interconnected through a looped interface.

Fibre Channel over IP
A storage network technology that extends the FCP by using IP to connect distributed SANs over long distances.

Fibre Channel port fan-in
FC port fan-in describes the situation in a SAN or FCAL where multiple hosts can log in to and use the same port.

Fibre Channel port logins
FC port logins refer to the number of hosts that can see any one Storage port. The IBM Storage Virtualize has a maximum limit per node port (N_Port) of FC logins that are allowed.

Fibre Channel Protocol
FCP is the serial SCSI command protocol that is used on FC networks.

Field-replaceable unit

Also referred to as *customer-replaceable units* (CRU), they are parts that are replaced in its entirety by a service representative when any one of its components fails. Both CRU and FRU are held as spares by IBM Service.

File Transfer Protocol

In TCP/IP, FTP is an application layer protocol that uses TCP and telnet services to transfer bulk-data files between machines or hosts.

Fix procedure

A maintenance procedure that runs within the product application and provides step-by-step guidance to resolve an error condition.

FlashCopy

Refers to a point-in-time (PiT) copy where a virtual copy of a volume is created. The target volume maintains the contents of the volume at the PiT when the copy was established. Any subsequent write operations to the source volume are not reflected on the target volume.

FlashCopy mapping

The relationship of the source and the target volumes.

FlashCopy relationship

See “FlashCopy mapping” on page 1222.

FlashCopy service

A copy service that duplicates the contents of a source volume on a target volume. In the process, the original contents of the target volume are lost. See also “Point-in-time copy” on page 1231.

FlashCore Module

A family of high-performance flash drives. The FCM design uses the Non-Volatile Memory Express (NVMe) protocol, a PCIe interface, and high-speed NAND memory to provide high throughput and I/O operations per second (IOPS) and low latency. FCM modules are available in different capacities. Hardware-based data compression and self-encryption are supported.

Flash drive

A data storage device, which is typically removable and rewriteable that uses solid-state memory to store persistent data. See also “Flash module”.

Flash module

A modular hardware unit containing flash memory, one or more flash controllers, and associated electronics. See also “Flash drive”.

Flush-through mode

A process in which data is written to a storage device at the same time as the data is cached. See “Write-through mode” on page 1241.

Front end and back end

The IBM FlashSystem and the SAN Volume Controller take MDisks to create pools of capacity from which volumes are created and presented to application servers (hosts). The view of an IBM SAN Volume Controller or FlashSystem from a host perspective, where volumes are visible to hosts. See also “Back end and front end” on page 1211.

Full restore operation

A copy operation where a local volume is created by reading an entire a volume snapshot from a local snap or from cloud storage.

Full snapshot

A type of volume snapshot that contains all the volume data. When a full snapshot is created, an entire copy of the volume data is transmitted to a local snap or to the cloud.

Gigabyte (GB)/Gibibyte (GiB)

A gigabyte (GB) is for processor storage, real and virtual storage, and channel volume. For 1,073,741,824 bytes.

Global Mirror (GM)

A method of asynchronous replication that maintains data consistency across multiple volumes within or across multiple systems. GM is used where distances between the source site and target site cause increased latency beyond what the application can accept. GM is an asynchronous copy service that enables host data on a volume to be mirrored over long distances to a volume in a remote location.

Global Mirror with change volumes

Change volumes are used to record changes to the primary and secondary volumes of a Remote Copy (RC) relationship. A FlashCopy mapping exists between a primary and its change volume, and a secondary and its change volume.

Grain

The unit of data that is represented by a single bit in a FlashCopy bitmap (64 kibibytes (KiB) or 256 KiB) in the IBM Storage System. A grain is also the unit to extend the real size of a thin-provisioned volume (32 KiB, 64 KiB, 128 KiB, or 256 KiB).

Graphical user interface

A computer interface that presents a visual metaphor of a real-world scene, often of a desktop, by combining high-resolution graphics, pointing devices, menu bars and other menus, overlapping windows, icons, and the object-action relationship.

Heartbeat

A signal that one entity sends to another to convey that it is still active.

Hop

One segment of a transmission path between adjacent nodes in a routed network.

Host

A physical or virtual server.

Host bus adapter
An interface card that connects a server to the SAN environment through its internal bus system. Typically, it is referred to as the FC adapter.

Host cluster
A configured set of physical or virtual hosts that share one or more storage volumes to increase the scalability or availability of computer applications.

Host ID
A numeric identifier that is assigned to a group of host FC ports or internet Small Computer Systems Interface (iSCSI) hostnames for LUN mapping. For each host ID, SCSI IDs are mapped to volumes separately. The intent is to have a one-to-one relationship between hosts and host IDs, although this relationship cannot be policed.

Host mapping
Refers to the process of controlling which hosts have access to specific volumes within a cluster. Host mapping is equivalent to LUN masking.

Host object
A logical representation of a host within a storage system that is used to represent the host for configuration tasks.

Host zone
A zone that is defined in the SAN fabric in which the hosts can address the system.

Hot extent
A frequently accessed volume extent that gets a performance benefit if it is moved from a lower storage class onto a higher class, fo example, from NL to Flash or SCM drive.

Hot spare node
An online IBM SAN Volume Controller node that is defined in a cluster but not in any I/O group. During a failure of any online node in any I/O group of clusters, it is automatically swapped with this spare node. After the recovery of an original node finishes, the spare node returns to the standby spare status. This feature is not available for IBM FlashSystems.

IBM FlashCore Module
See “FlashCore Module” on page 1222

IBM HyperSwap
A solution that provides continuous, transparent availability against storage errors and site failures, and is based on synchronous replication.

IBM Remote Support Server and Client

A software toolkit that is in IBM Storage System and opens a secured tunnel to the IBM Remote Support Server. IBM Remote Support Server is in the IBM network and collects key health check and troubleshooting information that is required by IBM support personnel.

IBM SAN Volume Controller

An appliance that is designed for attachment to various host computer systems. The IBM SAN Volume Controller performs block-level virtualization of disk storage. IBM Storage Virtualize is the software engine of IBM SAN Volume Controller and IBM FlashSystem family that performs block-level virtualization of disk and flash storage, while also providing enterprise-level management and data reliability features.

IBM Security Key Lifecycle Manager (SKLM)

Simplifies and automates the encryption key management process to help minimize risk and reduce operational costs of encryption key management. Follow on is the IBM security Guardium Key Lifecycle Manager

IBM Security Guardium Key Lifecycle Manager (GKLM)

Follow on of the IBM Security Key Lifecycle Manager (SKLM).

Image mode

An access mode that establishes a one-to-one mapping of extents in the storage pool (existing LUN or (image mode) MDisk) with the extents in the volume. See also “Managed mode” on page 1228 and “Unconfigured mode” on page 1239.

Image volume

A volume in which a direct block-for-block conversion exists from the MDisk to the volume.

I/O group

Each pair of IBM SAN Volume Controller cluster nodes is known as an input/output (I/O) group. An I/O group has a set of volumes that are associated with it that are presented to host systems. Each IBM SAN Volume Controller node or FlashSystem node is associated with exactly one I/O group. The nodes in an I/O group provide a failover and failback function for each other.

Incremental restore operation

A copy operation where a local volume is modified to match a volume snapshot by reading from storage only the parts of the volume snapshot that differ from the local volume.

Incremental snapshot

A type of volume snapshot where the changes to a local volume relative to the volume's previous snapshot are stored on cloud storage.

Input/output operations per second

A computing benchmark used to compare performance capabilities.

Input/output throttling rate

The maximum rate at which an I/O transaction is accepted for a volume.

Internal storage
An array of MDisks and drives that are held in IBM Storage System enclosures.

Internet Protocol (IP)
A protocol that routes data through a network or interconnected networks. This protocol acts as an intermediary between the higher protocol layers and the physical network.

Internet Small Computer Systems Interface (iSCSI)
A protocol that is used by a host system to manage iSCSI targets and iSCSI discovery. iSCSI initiators use the internet Storage Name Service (iSNS) protocol to locate the appropriate storage resources.

Internet Storage Name Service (iSNS)
The protocol that is used by a host system to manage iSCSI targets and the automated iSCSI discovery, management, and configuration of iSCSI and FC devices. It was defined in Request for Comments (RFC) 4171.

Inter-switch link
The physical connection that carries a protocol for interconnecting multiple routers and switches in a storage area network (SAN)

Inter-switch link hop
A connection between two switches and counted as one ISL hop. The number of hops is always counted on the shortest route between two N-ports (device connections). In an IBM Storage System environment, the number of ISL hops is counted on the shortest route between the pair of canisters that are farthest apart. The IBM Storage System supports a maximum of three ISL hops.

iSCSI
See “Internet Small Computer Systems Interface (iSCSI)” on page 1226.

iSCSI alias
An alternative name for the iSCSI-attached host.

iSCSI initiator
Functions as an iSCSI client. An initiator typically serves the same purpose to a computer as a SCSI bus adapter would, except that, instead of physically cabling SCSI devices (such as HDDs and tape changers), an iSCSI initiator sends SCSI commands over an IP network.

iSCSI name
A name that identifies an iSCSI target adapter or an iSCSI initiator adapter. An iSCSI name can be an iSCSI Qualified Name (IQN) or an extended-unique identifier (EUI). Typically, this identifier has the following format: iqn.datecode.reverse domain.

iSCSI Qualified Name
Special names that identify iSCSI initiators and targets. IQN is one of the three name formats that is provided by iSCSI. The IQN format is `iqn.<yyyy-mm>.<reversed domain name>`. For example, the default for an IBM Storage System canister can be in the following format:

`iqn.1986-03.com.ibm:2076.<clustername>.<nodename>`

iSCSI session
The interaction (conversation) between an iSCSI Initiator and an iSCSI Target.

iSCSI target
A storage resource that us available to the host through a network.

Just a bunch of disks
A group of HDDs that are not configured according to the RAID system to increase fault tolerance and improve data access performance.

Key server
A server that negotiates the values that determine the characteristics of a dynamic virtual private network (VPN) connection that is established between two endpoints. See “Encryption key manager/server” on page 1219.

Latency
The time interval between the initiation of a send operation by a source task and the completion of the matching receive operation by the target task. More generally, latency is the time between a task initiating data transfer and the time that transfer is recognized as complete at the data destination.

Least recently used
Pertains to an algorithm that is used to identify and make available the cache space that contains the data that was least recently used.

Licensed capacity
The amount of capacity on a storage system that a user is entitled to configure.

License key
An alphanumeric code that activates a licensed function on a product.

License key file
A file that contains one or more licensed keys.

Lightweight Directory Access Protocol (LDAP)
An open protocol that uses TCP/IP to provide access to directories that support an X.500 model. It does not incur the resource requirements of the more complex X.500 directory access protocol. For example, LDAP can be used to locate people, organizations, and other resources in an internet or intranet directory.

Local and remote fabric interconnect
The SAN components that are used to connect the local and remote fabrics. Depending on the distance between the two fabrics, they can be single-mode optical fibers that are driven by long wave gigabit interface converters (GBICs) or small form factor pluggable (SFP), or more sophisticated components, such as channel extenders or special SFP modules that are used to extend the distance between SAN components.

Local fabric
Composed of SAN components (switches, cables, and other components) that connect the components (nodes, hosts, and switches) of the local cluster together.

Logical drive
See “Volume” on page 1240.

Logical unit and logical unit number
The LU is defined by the SCSI standards as a LUN. LUN is an abbreviation for an entity that exhibits disk-like behavior, such as a volume or an MDisk.

LUN masking
A process where a host object can detect more LUNs than it is intended to use, and the device-driver software masks the LUNs that are not to be used by this host.

Machine signature
A string of characters that identifies a system. A machine signature might be required to obtain a license key.

Managed disk (MDisk)
A SCSI disk that is presented by a RAID controller and managed by IBM Storage Systems. The MDisk is not visible to host systems on the SAN.

Managed mode
An access mode that enables virtualization functions to be performed. See also “Image mode” on page 1225 and “Virtualized storage” on page 1240.

Management node
A node that is used for configuring, administering, and monitoring a system.

Master volume
In most cases, the volume that contains a production copy of the data and that an application accesses. See also “Auxiliary volume” on page 1211, and “Reliability, availability, and serviceability” on page 1234.

Maximum replication delay
The number of seconds that MM or GM replication can delay a write operation to a volume.

MDisk
See “Managed disk (MDisk)”.

MDisk group (Storage Pool)
See “Storage pool (managed disk group)” on page 1237.

Media Access Control
In networking, the lower of two sublayers of the Open Systems Interconnection model data link layer. The Media Access Control (MAC) sublayer handles access to shared media, such as whether token passing or contention is used.

Megabyte (MB)/Mebibyte (MiB)
For processor storage, real and virtual storage, and channel volume. For disk storage capacity and communications volume, 1 Megabyte is 10 to the power of 6 =1,000,000 bytes; 1 Mebibyte is 2 to the 20th power or 1,048,576 bytes.

Megabytes per second

A unit of data transfer rate equal to 1024 * 1024 bytes.

Metro Mirror

A method of synchronous replication that maintains data consistency across multiple volumes within the system. MM is used when the write latency that is caused by the distance between the source site and target site is acceptable to application performance.

Metro Global Mirror

A cascaded replication solution where MM synchronously copies data to the target site. This MM target is the source volume for GM that asynchronously copies data to a third site. This solution can provide disaster recovery (DR) with no data loss at GM distances when the intermediate site does not participate in the disaster that occurs at the production site.

Mirrored volume

A single virtual volume that has two physical volume copies. The primary physical copy is known within the IBM Storage System as `copy 0` and the secondary copy is known within the IBM Storage System as `copy 1`.

Multifactor authentication

Requires users to provide multiple pieces of information when they log in to the system to prove their identity. Multifactor authentication uses any combination of two or more methods, called factors, to authenticate users to your resources and protect those resources from unauthorized access.

N_Port ID Virtualization

N_Port ID Virtualization (NPIV) is an FC feature whereby multiple FC N_Port IDs can share a single physical N_Port.

Namespace Globally Unique Identifier

The NGUID is defined in the Identify Namespace data structure. The NGUID is composed of an IEEE organizationally unique identifier (OUI), an extension identifier, and a vendor-specific extension identifier and is used in NVMe implementations. The extension identifier and vendor-specific extension identifier are both assigned by the vendor and can be considered as a single field. NGUID is defined in big endian format. The OUI field differs from the OUI Identifier, which is in little endian format.

Nearline SAS drive

A drive that combines the high capacity data storage technology of a SATA drive with the benefits of a SAS interface for improved connectivity.

Node

A single processing unit within a system. For redundancy, multiple nodes are typically deployed to make up a system.

Node canister

A removable hardware unit that includes the node hardware, fabric and service interfaces, and SAS expansion ports. Node canisters are recognized on IBM Storage System products. An IBM SAN Volume Controller appliance performs the same role as a node canister in an IBM FlashSystem, so IBM SAN Volume Controller does not use the phrase node canister.

Node rescue
The process that enables a node with an invalid copy of software to obtain a valid copy from another node on the same FC fabric.

Non-Volatile Memory Express
An open logical-device interface specification for accessing non-volatile storage media that is attached through a PCIe bus.

NVMe Qualified Name
Used to uniquely describe a host or NVMe subsystem for identification and authentication. The NQN for the NVMe subsystem is specified in the Identify Controller data structure. An NQN is permanent for the lifetime of the host or NVMe subsystem.

Object-Based Access Control
See “Ownership Groups”.

Object storage
A general term that refers to the entity in which cloud object storage organizes, manages, and stores units of storage or just *objects*.

Overhead capacity
An amount of usable capacity that is occupied by metadata in a system or pool and other data that is used for system operations.

Overprovisioned
See “Capacity” on page 1212.

Overprovisioned ratio
See “Capacity” on page 1212.

Oversubscription
Refers to the ratio of the sum of the traffic on the initiator N-port connections to the traffic on the most heavily loaded ISLs, where more than one connection is used between these switches. Oversubscription assumes a symmetrical network, and a specific workload that is applied equally from all initiators and sent equally to all targets. A symmetrical network means that all the initiators are connected at the same level, and all the controllers are connected at the same level.

Ownership Groups
Provides a method of implementing a multi-tenant solution on the system. Ownership groups enable the allocation of storage resources to several independent tenants with the assurance that one tenant cannot access resources that are associated with another tenant. Ownership groups restrict access for users in the ownership group to only those objects that are defined within that ownership group.

Parent pool

Receive their capacity from MDisks. All MDisks in a pool are split into extents of the same size. Volumes are created from the extents that are available in the pool. You can add MDisks to a pool at any time either to increase the number of extents that are available for new volume copies or to expand existing volume copies. The system automatically balances volume extents between the MDisks to provide the best performance to the volumes. See also “Child pool” on page 1214.

Partner node

The other node that is in the I/O group to which this node belongs.

Partnership

In Metro Mirror or Global Mirror operations, the relationship between two clustered systems. In a clustered-system partnership, one system is defined as the local system and the other system as the remote system.

Petabyte (PB)/Pebibyte (PiB)

A Petabyte (PB) is for processor storage, real and virtual storage, and channel volume. For disk storage capacity and communications volume, it is 10 to the power of 15 = 1,000,000,000,000,000 bytes. A Pebibyte (PiB) is two to the power of 50 or 1,125,899,906,842,624 bytes.

Performance group

A collection of volumes that are assigned the same performance characteristics. See also “Performance policy”.

Performance policy

A policy that specifies performance characteristics, for example quality of service (QoS). See also “Pool”.

Point-in-time copy

An instantaneous copy that the FlashCopy service makes of the source volume. See also “FlashCopy service” on page 1222.

Pool

See “Storage pool (managed disk group)” on page 1237.

Pool pair

Two storage pools that are required to balance the workload. Each storage pool is controlled by a separate node.

Preferred node

When you create a volume, you can specify a preferred node. Many of the multipathing driver implementations that the system supports use this information to direct I/O to the preferred node. The other node in the I/O group is used only if the preferred node is not accessible. If you do not specify a preferred node for a volume, the system selects the node in the I/O group that has the fewest volumes to be the preferred node. After the preferred node is chosen, it can be changed only when the volume is moved to a different I/O group. The management GUI provides a wizard that moves volumes between I/O groups without disrupting host I/O operations.

Preparing phase
Before you start the FlashCopy process, you must prepare a FlashCopy mapping. The preparing phase flushes a volume's data from cache in preparation for the FlashCopy operation.

Primary volume
In a stand-alone MM or GM relationship, the target of write operations that are issued by the host application. See also "Reliability, availability, and serviceability" on page 1234.

Priority flow control
A link-level flow control mechanism that is based on IEEE standard 802.1Qbb. PFC operates on individual priorities. Instead of pausing all traffic on a link, PFC is used to selectively pause traffic according to its class.

Provisioned capacity
See "Capacity" on page 1212.

Provisioning group
An object that represents a set of MDisks that share physical resources. Provisioning groups are used for capacity reporting and monitoring of over-provisioned storage resources.

Qualifier
A value that provides more information about a class, association, indication, method, method parameter, instance, property, or reference.

A modifier that makes a name unique.

Queue depth
The number of I/O operations that can be run in parallel on a device.

Quorum disk
A disk that contains a reserved area that is used exclusively for system management. The quorum disk is accessed when it is necessary to determine which half of the clustered system continues to read and write data. Quorum disks can either be MDisks or drives.

Quorum index
The quorum index is the pointer that indicates the order that is used to resolve a tie. Nodes attempt to lock the first quorum disk (index 0), followed by the next disk (index 1), and then the last disk (index 2). The tie is broken by the node that locks them first.

Quota
The amount of disk space and number of files and directories that are assigned as upper limits for a specified user, group of users, or file set.

RAID controller
See "Node canister" on page 1229.

Raw capacity
See "Capacity" on page 1212.

Real capacity

Real capacity is the amount of storage that is allocated to a volume copy from a storage pool. See also “Capacity” on page 1212.

Rebuild area

Reserved capacity that is distributed across all drives in a redundant array of drives. If a drive in the array fails, the lost array data is systematically restored into the reserved capacity, which returns redundancy to the array. The duration of the restoration process is minimized because all drive members simultaneously participate in restoring the data. See also “Distributed redundant array of independent disks” on page 1218.

Reclaimable (or reclaimed) capacity

The capacity that is no longer needed. Reclaimable capacity is created when data is overwritten and the new data is stored in a new location, when data is marked as unneeded by a host by using the SCSI **UNMAP** command, or when a volume is deleted.

Recovery key

See “Encryption recovery key” on page 1220.

Redirect-on-write

A snapshot method that uses lighter-weight metadata references and improves, when possible, both the data reduction ratio and system performance. See also “Copy-on-write” on page 1216.

Redundant array of independent disks

Refers to two or more physical disk drives that are combined in an array in a certain way, which incorporates a RAID level for failure protection or better performance. The most common RAID levels are 0, 1, 5, 6, and 10. Some storage administrators refer to the RAID group as traditional RAID (TRAID). For distributed redundant array of independent disks (DRAID), see “Distributed redundant array of independent disks” on page 1218.

RAID 0

A data striping technique, which is commonly called RAID Level 0 or RAID 0 because of its similarity to common, RAID, data-mapping techniques. However, it includes no data protection, so the appellation RAID is a misnomer. RAID 0 is also known as data striping.

RAID 1

A mirroring technique that is used on a storage array in which two or more identical copies of data are maintained on separate mirrored disks.

RAID 10

A collection of two or more physical drives that present to the host an image of one or more drives. Druing a physical device failure, the data can be read or regenerated from the other drives in the RAID due to data redundancy.

RAID 5

An array that has a data stripe, which includes a single logical parity drive. The parity check data is distributed across all the disks of the array.

RAID 6

A RAID level that has two logical parity drives per stripe, which are calculated with different algorithms. Therefore, this level can continue to process read and write requests to all the array's virtual disks (virtual disks (VDisks)) in the presence of two concurrent disk failures.

Redundant Storage Area Network

A SAN configuration in which no single point of failure (SPOF) exists. Therefore, data traffic continues no matter what component fails. Connectivity between the devices within the SAN is maintained (although possibly with degraded performance) when an error occurs.

A redundant SAN design is normally achieved by splitting the SAN into two independent counterpart SANs (two SAN fabrics). In this configuration, if one path of the counterpart SAN is destroyed, the other counterpart SAN path keeps functioning. See also "Counterpart SAN" on page 1216.

Reliability, availability, and serviceability

A combination of design methodologies, system policies, and intrinsic capabilities that, when taken together, balance improved hardware availability with the costs that are required to achieve it.

Reliability is the degree to which the hardware remains free of faults. Availability is the ability of the system to continue operating despite predicted or experienced faults. Serviceability is how efficiently and nondisruptively broken hardware can be fixed.

Remote Copy

See "Global Mirror (GM)" on page 1223 and "Metro Mirror" on page 1229.

Remote fabric

The remote fabric is composed of SAN components (switches, cables, and other components) that connect the components (nodes, hosts, and switches) of the remote cluster together. Significant distances can exist between the components in the local cluster and those components in the remote cluster.

Replication relationship

In MM or GM, a relationship is the association between a master volume and an auxiliary volume. These volumes also have the attributes of a primary or secondary volume. See also "Auxiliary volume" on page 1211, "Master volume" on page 1228, "Primary volume" on page 1232, and "Secondary volume" on page 1235.

Safeguarded Copy

A feature of IBM SAN Volume Controller and IBM FlashSystem that provides the ability to create immutable copies by creating a snapshot of volumes in the system:

- Safeguarded backup

  A Safeguarded backup is a volume in a Safeguarded backup location. As part of the Safeguarded Copy function, you can add volumes to a volume group and assign a Safeguarded policy to that group. The IBM Copy Services Manager applies the policy to all the volumes in the group to create Safeguarded backups. Safeguarded backups are created in the same parent pool as the Safeguarded source volumes. A Safeguarded backup is the target of FlashCopy mapping with Safeguarded source volumes as a source.

- Safeguarded backup location

  A Safeguarded backup location is a child pool in each parent pool where the source volumes are located. The Safeguarded backup location stores Safeguarded backup copies after the Safeguarded policy is assigned to the volume group.

- Safeguarded policy

  A Safeguarded policy is a set of rules that controls the creation, retention, and expiration of Safeguarded backups of source volumes.

- Safeguarded source volume

  A Safeguarded source volume is added to a volume group. After a Safeguarded policy is assigned to the volume group, IBM Copy Services Manager uses this volume as the source copy for the Safeguarded backups. You can set a volume as a Safeguarded source volume while creating a volume in a Safeguarded volume group. You can also assign a volume to a Safeguarded volume group after creating a volume, or create a volume in a volume group and then, assign the Safeguarded policy to the volume group.

- Safeguarded volume group

  A volume group is a set of related volumes that can be managed and configured collectively. A volume group is called a Safeguarded volume group after a volume group is created and assigned with a Safeguarded policy.

SCSI initiator

The system component that starts communications with attached targets.

SCSI target

A device that acts as a subordinate to a SCSI initiator and consists of a set of one or more LUs, each with an assigned LUN. The LUs on the SCSI target are typically I/O devices.

Secondary volume

Pertinent to RC, the volume in a relationship that contains a copy of data that is written by the host application to the primary volume.

Secure Copy Protocol

The secure transfer of computer files between a local and a remote host or between two remote hosts by using the Secure Shell (SSH) protocol.

Secure Sockets Layer certificate

The standard security technology for establishing an encrypted link between a web server and a browser. This link ensures that all data that is passed between the web server and browsers remain private. To create an SSL connection, a web server requires an SSL certificate.

Sequential volume

Every volume that uses extents from a single MDisk.

Serial-attached SCSI

A method that is used in accessing computer peripheral devices that employs a serial (1 bit at a time) means of digital data transfer over thin cables. The method is specified in the ANSI standard that is called SAS. In the business enterprise, SAS is useful for access to mass storage devices, external HDDs.

Service assistant
A user interface that services hardware independent of the storage system.

Service Location Protocol
An internet service discovery protocol that enables computers and other devices to find services in a local area network (LAN) without prior configuration. It was defined in RFC 2608.

Simple Network Management Protocol
A set of protocols for monitoring systems and devices in complex networks. Information about managed devices is defined and stored in a Management Information Base (MIB).

Small Computer System Interface
An ANSI-standard electronic interface with which PCs can communicate with peripheral hardware, such as disk drives, tape drives, CD-ROM drives, printers, and scanners, faster and more flexibly than with previous interfaces.

Snapshot
An image backup type that consists of a PiT view of a volume.

Solid-state drive
A disk that is made from solid-state memory and therefore has no moving parts. Most SSDs use NAND-based flash memory technology. It is defined to the IBM Storage System as a disk tier `generic_ssd`.

Space efficient
See “Thin provisioning” on page 1238.

Space-efficient virtual disk
See “Thin-provisioned volume” on page 1238.

Spare
An extra storage component, such as a drive or tape that is predesignated for use as a replacement for a failed component.

Spare drive
A drive that is reserved in an array for rebuilding a failed drive in a RAID. If a drive fails in a RAID, a spare drive from within that device adapter pair is selected to rebuild it.

Spare goal
The optimal number of spares that are needed to protect the drives in the array from failures. The system logs a warning event when the number of spares that protect the array drops below this number.

Space-efficient volume
See “Thin-provisioned volume” on page 1238.

Stand-alone relationship
In FlashCopy, MM, and GM, relationships that do not belong to a consistency group and that have a null consistency-group attribute.

Standard-provisioned volume
A volume that completely uses storage at creation.

Statesave
Binary data collection that is used for a problem determination by IBM service support.

Storage area network (SAN)
A dedicated storage network that is tailored to a specific environment, which combines servers, systems, storage products, networking products, software, and services.

Storage-class memory
A type of NAND flash that includes a power source to ensure that data is not lost due to a system crash or power failure. SCM treats non-volatile memory as DRAM and includes it in the memory space of the server. Access to data in that space is quicker than access to data in local, PCI-connected SSDs, direct-attached HDDs, or external storage arrays. SCM read/write technology is up to 10 times faster than NAND flash drives and is more durable.

Storage capacity unit
An IBM Storage Virtualize license metric that measures the managed capacity so that the price is differentiated by the technology that is used to store the data.

Storage node
A component of a storage system that provides internal storage or a connection to one or more external storage systems.

Storage pool (managed disk group)
A collection of storage capacity, which is made up of one or more MDisks that provides the pool of storage capacity for a specific set of volumes. A storage pool can contain more than one tier of disk, which is known as a multitier storage pool, which is a prerequisite of Easy Tier automatic data placement.

Stretched system
An extended HA method that is supported by the SAN Volume Controller to enable I/O operations to continue after the loss of half of the system. A stretched system is also sometimes referred to as a *split system*. One half of the system and I/O group is usually in a geographically distant location from the other, often 10 kilometers (6.2 miles) or more. A third site is required to host a quorum application or storage system with a quorum disk.

Stored capacity
The amount of capacity that is used to store data that is written by a host after data reduction See also “Data reduction” on page 1217, and “Total usable physical capacity” on page 1239.

Striped
Pertaining to a volume that is created from multiple MDisks that are in the storage pool. Extents are allocated on the MDisks in the order that is specified.

Support Assistance
A function that is used to provide support personnel remote access to the system to perform troubleshooting and maintenance tasks.

Symmetric virtualization

A virtualization technique in which the physical storage, in the form of a RAID, is split into smaller chunks of storage that are known as extents. These extents are then concatenated by using various policies to make volumes. See also “Asymmetric virtualization” on page 1210.

Synchronous replication

Synchronous replication is a type of replication in which the application write operation is made to both the source volume and target volume before control is given back to the application. See also “Asynchronous replication” on page 1210.

Syslog

A standard for transmitting and storing log messages from many sources to a centralized location to enhance system management.

T10 DIF

T10 DIF is a *Data Integrity Field* (DIF) extension to SCSI to enable end-to-end protection of data from a host application to physical media.

Terabyte (TB)/Tebibyte (TiB)

A TB is for processor storage, real and virtual storage, and channel volume. For disk storage capacity and communications volume, it is 10 to the power of 12 = 1,000,000,000,000 bytes. A Tebibyte (TiB) is two to the power of 40 or 1,099,511,627,776 bytes.

Thin-provisioned volume

A volume that allocates storage when data is written to it.

Thin provisioning

Refers to the ability to define storage, usually a storage pool or volume, with a “logical” capacity size that is larger than the actual physical capacity that is assigned to that pool or volume. Therefore, a thin-provisioned volume is a volume with a virtual capacity that differs from its real capacity.

Thin provisioning savings

See “Capacity” on page 1212.

Throttles

A mechanism to control the amount of resources that are used when the system is processing I/Os on supported objects. The system supports throttles on hosts, host clusters, volumes, copy offload operations, and storage pools. If a throttle limit is defined, the system either processes the I/O for that object or delays the processing of the I/O to free resources for more critical I/O operations.

Throughput

A measure of the amount of information that is transmitted over a network in a period. Throughput is measured in bits per second (bps), kilobits per second (Kbps), or megabits per second (Mbps).

Tie-breaker

When a cluster is split into two groups of nodes, the role of tie-breaker in a quorum device decides which group continues to operate as the system and handle all I/O requests.

Transparent Cloud Tiering
A separately installable feature that provides a native cloud storage tier.

Trial license
A temporary entitlement to use a licensed function.

Total capacity savings
See “Capacity” on page 1212.

Total usable physical capacity
The amount of physical configured storage space that is available for stored capacity or reserved capacity. This capacity can consist of internal storage through arrays and external storage through MDisks. See also reserved capacity, stored capacity. See “Capacity threshold” on page 1213

Unconfigured mode
An access mode in which an external storage MDisk is not configured in the system, so no operations can be performed. See also “Image mode” on page 1225 and “Managed mode” on page 1228.

Unique identifier
An identifier that is assigned to storage system LUs when they are created. It is used to identify the LU regardless of the LUN, the status of the LU, or whether alternative paths exist to the same device. Typically, a UID is used only once.

Unmapped volume capacity
The amount of volume capacity that is not mapped to a host. See also “Capacity” on page 1212.

Usable capacity
The system supports virtualizing capacity that is on external storage systems that are attached to the system.

For external storage systems, administrators must configure out-of-space alerts that notify the system when usable capacity on the external storage system reaches a defined threshold. Without these thresholds defined on the external storage systems, external storage on these systems can become over-provisioned and risk running out of usable capacity that is used for host operations.

Overprovisioned external storage occurs when the sum of the provisioned capacity of all the volumes in a system or pool is greater than the total usable capacity of the system or pool that is allocated from the external storage system.

Used capacity
The amount of usable capacity that is taken up by data or capacity in a system, pool, array, or MDisk after data reduction techniques are applied.

VDisk
See “Virtual disk” or “Volume”

VDisk-to-host mapping
See “Host mapping” on page 1224.

Virtual capacity
The total capacity of all volumes and volume copies in a system or pool.

Virtual disk
See “Volume” on page 1240.

Virtualization
In the storage industry, virtualization is a concept in which a pool of storage is created that contains several storage systems. Storage systems from various vendors can be used. The pool can be split into volumes that are visible to the host systems that use them. See also “Capacity licensing” on page 1213.

Virtualized capacity
The amount of capacity that is contributed to a storage pool by a given provisioning group.

Virtualized storage
Physical storage that has virtualization techniques that are applied to it by a virtualization engine.

Virtual local area network
This tagging separates network traffic at the layer 2 level for Ethernet transport. The system supports VLAN configuration on both IPv4 and IPv6 connections.

Virtual storage area network
A logical fabric entity that is defined within the SAN. It can be defined on a single physical SAN switch or across multiple physical switches or directors. In VMware terminology, the VSAN is defined as a logical layer of storage capacity that is built from physical disk drives that are attached directly into the Elastic Sky X Integrated (ESXi) hosts. This solution is not considered within the scope of this publication.

Vital product data
Information that uniquely defines the system, hardware, software, and microcode elements of a processing system.

Volume
An IBM Storage System logical device that appears to host systems that are attached to the SAN as a SCSI disk. Each volume is associated with exactly one I/O group. A volume has a preferred node within the I/O group.

Volume copy
A physical copy of the data that is stored on a volume. Mirrored volumes have two copies. Non-mirrored volumes have one copy.

Volume protection

To prevent active volumes or host mappings from inadvertent deletion, the system supports a global setting that prevents these objects from being deleted if the system detects that they have recent I/O activity. When you delete a volume, the system checks to verify whether it is part of a host mapping, FlashCopy mapping, or an RC relationship. In these cases, the system fails to delete the volume unless the **-force** parameter is specified. Using the **-force** parameter can lead to unintentional deletions of volumes that are still active. Active means that the system detected recent I/O activity to the volume from any host.

Volume snapshot

A collection of objects that represents the data of a volume at a particular time.

Worldwide ID

A name identifier that is unique worldwide and that is represented by a 64-bit value that includes the IEEE-assigned OUI.

Worldwide name

A 64-bit, unsigned name identifier that is unique.

Worldwide node name

A unique 64-bit identifier for a host containing an FC port. See also “Worldwide port name” on page 1241.

Worldwide port name

A unique 64-bit identifier that is associated with an FC adapter port. The WWPN is assigned in an implementation-independent and protocol-independent manner. See also “Worldwide node name” on page 1241.

Write-through mode

A process in which data is written to a storage device while the data is cached. See also “Flush-through mode” on page 1222.

Written capacity

See “Capacity” on page 1212.

Written capacity limit

The largest amount of capacity that can be written to a drive, array, or MDisk. The limit can be reached even when usable capacity is still available.

Yottabyte (YB)/Yobibyte (YiB)

A Yottabyte (YB) is for processor storage, real and virtual storage, and channel volume. For disk storage capacity and communications volume, it is 10 to the power of 24 = 1,000,000,000,000,000,000,000,000 bytes. A Yobibyte (YiB) is two to the power of 80 or 1,208,925,819,614,629,174,706,176 bytes.

Zettabyte (ZB)/Zebibyte (ZiB)

A Zettabyte (ZB) is for processor storage, real and virtual storage, and channel volume. For disk storage capacity and communications volume, it is 10 to the power of 21 = 1,000,000,000,000,000,000,000 bytes. A Zebibyte (ZiB) is two to the power of 70 or 1,180,591,620,717,411,303,424 bytes.

Zoning

The grouping of multiple ports to form a virtual and private storage network. Ports that are members of a zone can communicate with each other, but are isolated from ports in other zones. See also “Fibre Channel” on page 1221.

C

List of the demonstration videos

Table C-1 lists the demonstration videos referenced in this book. Most of these videos were created by the authors of this IBM Redbooks as part of the project. You can also visit https://www.redbooks.ibm.com/feature/storagevideos.

Table C-1 Demonstration videos

Title of the video	Link	Section in which the video is referenced in the book	QR code
IBM Storage Virtualize V8.6 GUI, including volume group snapshots (with Safeguarded Copy) and policy-based replication	https://ibm.biz/BdMcgN	4.1, “Performing operations by using the GUI” on page 234	
How to enable and activate encryption for IBM FlashSystem with IBM Storage Virtualize V8.6	https://ibm.biz/BdMBMY	12.7, “Encryption” on page 1098	
IBM Storage Virtualize V8.6 NVME over Fabrics/TCP configuration	https://ibm.biz/BdMc89	7.5, “NVMe over TCP” on page 521	

© Copyright IBM Corp. 2024.

Title of the video	Link	Section in which the video is referenced in the book	QR code
IBM Storage Virtualize V8.6 Cisco DUO multi-factor authentication (MFA)	https://ibm.biz/BdMcgm	12.4, “Multifactor authentication” on page 1066	
IBM Storage FlashSystem 5200 overview	https://ibm.biz/Bdy6s2	1.11, “IBM FlashSystem 5200 overview” on page 71	
IBM Storage Virtualize V8.6 Initial Setup: Customer configuration tasks	https://ibm.biz/BdMBL9	3.2, “System initialization” on page 179	
IBM Storage Virtualize V8.6 Initial Setup: SSR configuration tasks	https://ibm.biz/BdMBLQ	3.2, “System initialization” on page 179	
IBM Storage Virtualize V8.6 Initial Setup: Setting up a cluster from the Service IP	https://ibm.biz/BdMBLg	3.2, “System initialization” on page 179	
IBM Storage Virtualize V8.6: Two Person Integrity	https://ibm.biz/BdMcg4	12.6, “Two Person Integrity (TPI)” on page 1093	

Title of the video	Link	Section in which the video is referenced in the book	QR code
Connecting IBM Storage to VMware vSphere	https://ibm.biz/Bdy6sb	“VMware vSphere virtual volumes” on page 90	
Managing datastores provisioned from IBM FlashSystem Storage in the vSphere Client	https://ibm.biz/Bdy6sp	“VMware vSphere virtual volumes” on page 90	
Creating a datastore on IBM FlashSystem Storage, directly from the vSphere Client	https://ibm.biz/Bdy6s8	“VMware vSphere virtual volumes” on page 90	
IBM Storage Virtualize for Public Cloud V8.5 installation on AWS	https://ibm.biz/Bdy6sh	1.12.4, “IBM Storage Virtualize for Public Cloud” on page 84	
IBM Storage Virtualize for Public Cloud V8.5 installation on Azure	https://ibm.biz/Bdy6sJ	1.12.4, “IBM Storage Virtualize for Public Cloud” on page 84	
IBM Storage Virtualize V8.6 Cloud Call Home	https://ibm.biz/BdMcgq	11.8.1, “Email notifications and the Call Home function” on page 1007	

Title of the video	Link	Section in which the video is referenced in the book	QR code
IBM Storage Virtualize V8.6: Handling snapshots using the command line interface	https://ibm.biz/BdMcgb	10.15.1, “Volume group snapshots” on page 896	
IBM Storage Virtualize V8.6: Handling snapshots using the graphical user interface	https://ibm.biz/BdMcgK	10.15.1, “Volume group snapshots” on page 896	
IBM Storage Virtualize V8.6: Policy-based replication	https://ibm.biz/BdMcgp	10.16, “IBM Storage policy-based replication” on page 905	
IBM Storage Virtualize V8.6: Transparent Cloud Tiering	https://ibm.biz/BdmPP9	10.5, “Transparent Cloud Tiering” on page 789	
IBM Storage for Data Resilience simply explained	https://ibm.biz/Bdy6ix	“Safeguarded Copy” on page 18	

Title of the video	Link	Section in which the video is referenced in the book	QR code
Videos for IBM Storage Insights	https://ibm.biz/Bdy6im	11.12, “IBM Storage Insights monitoring” on page 1035	

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *Performance and Best Practices Guide for IBM Storage FlashSystem and IBM SAN Volume Controller: Updated for IBM Storage Virtualize Version 8.6,* SG24-8543
- *IBM Spectrum Virtualize 3-Site Replication*, SG24-8504
- *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430
- *IBM Storage Virtualize and VMware: Integrations, Implementation and Best Practices*, SG24-8549
- *Policy-Based Replication with IBM Storage FlashSystem, IBM SAN Volume Controller and IBM Storage Virtualize*, REDP-5704
- *IBM Storage as a Service (STaaS) Offering Guide*, REDP-5644
- *IBM FlashSystem Safeguarded Copy Implementation Guide*, REDP-5654
- *Automate and Orchestrate Your IBM FlashSystem Hybrid Cloud with Red Hat Ansible,* REDP-5598
- *IBM Spectrum Virtualize and SAN Volume Controller Enhanced Stretched Cluster with VMware*, SG24-8211
- *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317
- *IBM Storage Virtualize, IBM Storage FlashSystem, and IBM SAN Volume Controller Security Feature Checklist*, REDP-5717
- *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- IBM Redbooks Storage videos

  https://www.redbooks.ibm.com/feature/storagevideos

© Copyright IBM Corp. 2024.

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

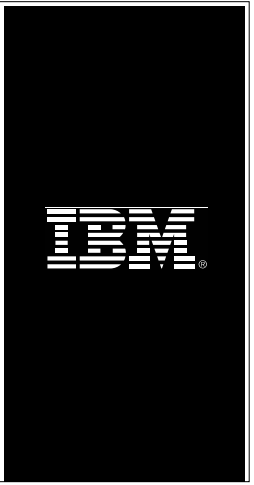
ibm.com/services



Implementation Guide for IBM Storage FlashSystem

SG24-8542-00

ISBN DocISBN



(2.5” spine)
2.5”<->nnn.n”
1315<-> nnnn pages



SG24-8542-00

ISBN 0738426113

Printed in U.S.A.

Get connected



