Redbooks

ibm.com/redbooks

Implementing the IBM SAN Volume Controller with IBM Spectrum Virtualize Version 8.4.2

Abdulqadir S Ali

Carsten Larsen

Christian Schroeder

Corne Lottering

David Green

Guillaume Legmar

Hartmut Lonzer

John Nycz

Jon Herd

Reginald D'Souza

Shu Mookerjee

Tiago Bastos

Vasfi Gucer



Storage

IBM

Redbooks



IBM Redbooks

Implementing the IBM SAN Volume Controller with IBM Spectrum Virtualize Version 8.4.2

December 2021

SG24-8507-00

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

First Edition (December 2021)

This edition applies to IBM Spectrum Virtualize Version 8.4.2.

© Copyright International Business Machines Corporation 2021. 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

Notices xv
Trademarks xvi

Preface xvii
Authors xvii
Now you can become a published author, too! xxi
Comments welcome xxi
Stay connected to IBM Redbooks xxi

Chapter 1. Introduction and system overview 1
1.1 Storage virtualization terminology 2
1.2 Latest changes and enhancements 5
1.2.1 IBM Spectrum Virtualize V8.4.2 5
1.2.2 IBM Spectrum Virtualize V8.4.0 14
1.3 IBM SAN Volume Controller architecture 15
1.3.1 IBM SAN Volume Controller architectural overview 15
1.3.2 IBM Spectrum Virtualize 18
1.3.3 IBM SAN Volume Controller topology 19
1.4 IBM SAN Volume Controller models 21
1.4.1 IBM SAN Volume Controller SV2 and SA2 21
1.4.2 IBM SAN Volume Controller SV1 23
1.4.3 IBM SAN Volume Controller model comparisons 23
1.5 IBM SAN Volume Controller components 25
1.5.1 Nodes 26
1.5.2 I/O groups 26
1.5.3 System 27
1.5.4 Expansion enclosures 27
1.5.5 Flash drives 29
1.5.6 MDisks 32
1.5.7 Cache 34
1.5.8 Quorum disk 36
1.5.9 Storage pool 38
1.5.10 Volumes 39
1.5.11 Easy Tier 41
1.5.12 Hosts 42
1.5.13 Array 43
1.5.14 Encryption 43
1.5.15 iSCSI and iSCSI Extensions over RDMA 44
1.5.16 IBM Real-time Compression 45
1.5.17 Data Reduction Pools 45
1.5.18 IP replication 46
1.5.19 IBM Spectrum Virtualize copy services 46
1.5.20 Synchronous or asynchronous Remote Copy 47
1.5.21 FlashCopy and Transparent Cloud Tiering 48
1.6 Business continuity 49
1.6.1 Business continuity with Stretched Clusters 50
1.6.2 Business continuity with Enhanced Stretched Cluster 50
1.6.3 Business continuity with HyperSwap 51
1.6.4 Business continuity with three-site replication 51

© Copyright IBM Corp. 2021.

1.6.5 Automatic hot spare nodes . . . 52
1.7 Management and support tools . . . 54
1.7.1 IBM Assist On-site and Remote Support Assistance . . . 54
1.7.2 Event notifications . . . 55
1.8 Useful IBM SAN Volume Controller web links . . . 57

Chapter 2. Planning . . . 59
2.1 General planning rules . . . 60
2.2 Planning for availability . . . 61
2.3 Physical installation planning . . . 62
2.4 Planning for system management . . . 63
2.4.1 User password creation options . . . 63
2.5 Connectivity planning . . . 64
2.6 Fibre Channel SAN configuration planning . . . 65
2.6.1 Physical topology . . . 65
2.6.2 Zoning . . . 66
2.6.3 N_Port ID Virtualization . . . 67
2.6.4 Inter-node zone . . . 67
2.6.5 Back-end storage zones . . . 67
2.6.6 Host zones . . . 68
2.6.7 Zoning considerations for Metro Mirror and Global Mirror . . . 69
2.6.8 Port designation recommendations . . . 69
2.7 IP SAN configuration planning . . . 71
2.7.1 Portsets . . . 72
2.7.2 iSCSI and iSER protocols . . . 73
2.7.3 Priority flow control . . . 73
2.7.4 RDMA clustering . . . 74
2.7.5 iSCSI back-end storage attachment . . . 75
2.7.6 IP network host attachment . . . 75
2.7.7 Native IP replication . . . 76
2.7.8 Firewall planning . . . 77
2.8 Planning topology . . . 77
2.8.1 High availability . . . 77
2.8.2 Three-Site replication . . . 78
2.9 Back-end storage configuration . . . 79
2.10 Internal storage configuration . . . 80
2.11 Storage pool configuration . . . 80
2.11.1 Child pools . . . 83
2.11.2 Storage pool and cache relationship . . . 84
2.11.3 Provisioning policies . . . 85
2.12 Volume configuration . . . 86
2.12.1 Planning for image mode volumes . . . 86
2.12.2 Planning for standard-provisioned volumes . . . 86
2.12.3 Planning for thin-provisioned volumes . . . 87
2.12.4 Planning for compressed volumes . . . 87
2.12.5 Planning for deduplicated volumes . . . 87
2.13 Host attachment planning . . . 89
2.13.1 Queue depth . . . 89
2.13.2 SAN boot support . . . 89
2.13.3 Planning for large deployments . . . 89
2.13.4 Planning for SCSI Unmap . . . 89
2.14 Planning copy services . . . 90
2.14.1 FlashCopy guidelines . . . 90

2.14.2 Planning for Metro Mirror and Global Mirror 91
2.15 Throttles 93
2.16 Data migration 93
2.17 Ansible Automation for IBM SAN Volume Controller 95
2.18 Container Storage Integration 96
2.19 Safeguarded Copy function 97
2.19.1 Prerequisites 97
2.20 Performance monitoring with IBM Storage Insights 98
2.21 Configuration backup procedure 101

Chapter 3. Initial configuration 103
3.1 Prerequisites 104
3.2 System initialization 105
3.2.1 System initialization process 105
3.3 System setup 109
3.3.1 System setup wizard 109
3.4 Base configuration 124
3.4.1 Configuring Remote Direct Memory Access clustering 124
3.4.2 Adding a node or hot spare node 128
3.4.3 Changing the system topology 132
3.4.4 Configuring quorum disks or applications 135
3.4.5 Configuring the local Fibre Channel port masking 139
3.5 Configuring management access 141
3.5.1 Configuring secure communications 141
3.5.2 Configuring password policies 144
3.5.3 Configuring user authentication 147

Chapter 4. IBM Spectrum Virtualize GUI 155
4.1 Normal operations by using the GUI 156
4.1.1 Accessing the GUI 156
4.2 GUI introduction 160
4.2.1 Task menu 161
4.2.2 Suggested tasks 162
4.2.3 Notification icons and help 162
4.3 System Hardware - Overview window 167
4.3.1 Content-based organization 168
4.4 Monitoring menu 171
4.4.1 System Hardware overview 172
4.4.2 IBM Easy Tier Reports 174
4.4.3 Events 175
4.4.4 Performance 177
4.4.5 Background Tasks 178
4.5 Pools 178
4.6 Volumes 179
4.7 Hosts 180
4.8 Copy Services 180
4.9 Access 181
4.9.1 Ownership groups 181
4.9.2 Users by groups 189
4.9.3 Audit log 197
4.10 Settings 198
4.10.1 Notifications menu 199
4.10.2 Network 203

4.10.3 Using the management GUI 208
4.10.4 Security 218
4.10.5 System menus 222
4.10.6 Support menu 242
4.10.7 GUI Preferences menu 250
4.11 Other frequent tasks in the GUI 257
4.11.1 Renaming components 257
4.11.2 Changing the system topology 260
4.11.3 Restarting the GUI service 267

Chapter 5. Using storage pools 271
5.1 Working with storage pools 272
5.1.1 Creating storage pools 274
5.1.2 Managed disks in a storage pool 277
5.1.3 Actions on storage pools 278
5.1.4 Child pools 285
5.1.5 Encrypted storage pools 292
5.2 Working with external controllers and MDisks 292
5.2.1 External storage controllers 292
5.2.2 Actions on external storage controllers 295
5.2.3 Working with external MDisks 296
5.2.4 Actions for external MDisks 299
5.3 Working with internal drives and arrays 306
5.3.1 Working with drives 307
5.3.2 RAID and distributed RAID 314
5.3.3 Creating arrays 318
5.3.4 Actions on arrays 323

Chapter 6. Creating and provisioning volumes 331
6.1 Introduction to volumes 332
6.2 Volume characteristics 332
6.2.1 Volume type 333
6.2.2 Managed mode and image mode 334
6.2.3 VSize 336
6.2.4 Performance 338
6.2.5 Volume copies 338
6.2.6 I/O operations data flow 341
6.2.7 Storage efficiency 343
6.2.8 Encryption 349
6.2.9 Cache mode 349
6.2.10 I/O throttling 349
6.2.11 Volume protection 350
6.2.12 Secure data deletion 351
6.3 Virtual volumes 351
6.4 Volumes in multi-site topologies 351
6.4.1 HyperSwap topology 352
6.4.2 Stretched topology 353
6.5 Operations on volumes 354
6.5.1 Creating volumes 354
6.5.2 Creating custom volumes 363
6.5.3 Creating volumes in multi-site topologies 367
6.5.4 I/O throttling 373
6.5.5 Volume protection 380

6.5.6 Modifying a volume 380
6.5.7 Deleting a volume 393
6.5.8 Mapping a volume to a host 395
6.5.9 Modify I/O Group or Non-disruptive Volume Move (NDVM) 400
6.5.10 Migrating a volume to another storage pool 405
6.6 Volume operations by using the CLI 414
6.6.1 Displaying volume information 414
6.6.2 Creating a volume 414
6.6.3 Creating a thin-provisioned volume 417
6.6.4 Creating a volume in image mode 417
6.6.5 Adding a volume copy 418
6.6.6 Splitting a mirrored volume 425
6.6.7 Modifying a volume 427
6.6.8 Deleting a volume 427
6.6.9 Volume protection 428
6.6.10 Expanding a volume 428
6.6.11 HyperSwap volume modification with CLI 430
6.6.12 Mapping a volume to a host 431
6.6.13 Listing volumes that are mapped to the host 432
6.6.14 Listing hosts that are mapped to the volume 433
6.6.15 Deleting a volume to host mapping 433
6.6.16 Migrating a volume 434
6.6.17 Migrating a fully managed volume to an image mode volume 435
6.6.18 Shrinking a volume 435
6.6.19 Listing volumes that use MDisks 436
6.6.20 Listing MDisks that are used by the volume 436
6.6.21 Listing volumes that are defined in the storage pool 437
6.6.22 Listing storage pools in which a volume has its extents 437
6.6.23 Tracing a volume from a host back to its physical disks 439

Chapter 7. Hosts 443
7.1 Host attachment overview 444
7.2 Host objects overview 445
7.3 NVMe over Fibre Channel 446
7.4 N_Port ID Virtualization support 447
7.4.1 NPIV prerequisites 450
7.4.2 Verifying the NPIV modes state for a new system installation 450
7.4.3 Enabling NPIV on a system 452
7.5 IP multi-tenancy 457
7.5.1 Limitations and restrictions 457
7.5.2 Prerequisites 459
7.5.3 Configuring the portset 461
7.5.4 Assigning an IP address to a portset 463
7.5.5 Assigning the portset to a host object 464
7.6 Hosts operations by using the GUI 465
7.6.1 Creating hosts 465
7.6.2 Host clusters 478
7.6.3 Actions on hosts 482
7.6.4 Actions on host clusters 492
7.6.5 Host management views 499
7.7 Performing hosts operations by using CLI 502
7.7.1 Creating a host by using the CLI 502
7.7.2 Host administration by using the CLI 504

7.7.3 Adding and deleting a host port by using the CLI 507
7.7.4 Host cluster operations 509
7.7.5 Adding a host or host cluster to an ownership group 511
7.8 Host attachment practical examples 512
7.8.1 Prerequisites 512
7.8.2 Fibre Channel host connectivity and capacity allocation. 512
7.8.3 iSCSI host connectivity and capacity allocation 516
7.8.4 NVMe over Fabric host connectivity example 520
7.9 Container Storage Interface drivers 526

Chapter 8. Storage migration 527
8.1 Storage migration overview 529
8.1.1 Interoperability and compatibility 530
8.1.2 Prerequisites 530
8.2 Storage migration wizard 531
8.3 Migrating data between systems non-disruptively 548
8.3.1 Non-disruptive volume migration. 549

Chapter 9. Advanced features for storage efficiency 559
9.1 IBM Easy Tier 560
9.1.1 Easy Tier concepts 560
9.1.2 Implementing and tuning Easy Tier. 566
9.1.3 Monitoring Easy Tier activity 571
9.2 Thin-provisioned volumes 579
9.2.1 Concepts 579
9.2.2 Implementation 580
9.3 Unmap 581
9.3.1 SCSI unmap command 581
9.3.2 Back-end SCSI unmap 581
9.3.3 Host SCSI unmap 582
9.3.4 Offload IO throttle 583
9.4 Data Reduction Pools 584
9.4.1 Introduction to DRP 584
9.4.2 DRP benefits 585
9.4.3 Planning for DRPs 586
9.4.4 Implementing DRP with compression and deduplication 588
9.5 Saving estimation for compression and deduplication 596
9.5.1 Evaluate compression savings by using IBM Comprestimator 596
9.5.2 Evaluating compression and deduplication. 598
9.6 Overprovisioning and data reduction on external storage. 599

Chapter 10. Advanced Copy Services 605
10.1 Maximum configuration updates 606
10.2 IBM Safeguarded Copy 607
10.3 IBM FlashCopy 608
10.3.1 Business requirements for FlashCopy 608
10.3.2 FlashCopy principles and terminology 610
10.3.3 FlashCopy mapping 610
10.3.4 Consistency groups 611
10.3.5 Crash consistent copy and hosts considerations 612
10.3.6 Grains and bitmap: I/O indirection. 613
10.3.7 Interaction with cache 620
10.3.8 Background Copy Rate 620
10.3.9 Incremental FlashCopy 622

10.3.10 Starting FlashCopy mappings and consistency groups 623
10.3.11 Multiple target FlashCopy 625
10.3.12 Reverse FlashCopy 630
10.3.13 FlashCopy and image mode volumes 632
10.3.14 FlashCopy mapping events 633
10.3.15 Thin-provisioned FlashCopy 635
10.3.16 Serialization of I/O by FlashCopy 636
10.3.17 Event handling 636
10.3.18 Asynchronous notifications 637
10.3.19 Interoperation with Metro Mirror and Global Mirror 637
10.3.20 FlashCopy attributes and limitations 638
10.3.21 Expanding Volumes In a FlashCopy Mapping 639
10.4 Managing FlashCopy by using the GUI 640
10.4.1 FlashCopy presets 640
10.4.2 FlashCopy window 642
10.4.3 Creating a FlashCopy mapping 645
10.4.4 Single-click snapshot 655
10.4.5 Single-click clone 657
10.4.6 Single-click backup 659
10.4.7 Creating a FlashCopy consistency group 660
10.4.8 Creating FlashCopy mappings in a consistency group 661
10.4.9 Showing related Volumes 664
10.4.10 Moving FlashCopy mappings across consistency groups 665
10.4.11 Removing FlashCopy mappings from consistency groups 666
10.4.12 Modifying a FlashCopy mapping 667
10.4.13 Renaming FlashCopy mappings 668
10.4.14 Deleting FlashCopy mappings 671
10.4.15 Deleting a FlashCopy consistency group 672
10.4.16 Starting FlashCopy mappings 674
10.4.17 Stopping FlashCopy mappings 675
10.4.18 Memory allocation for FlashCopy 676
10.5 Transparent Cloud Tiering 678
10.5.1 Considerations for using Transparent Cloud Tiering 679
10.5.2 Transparent Cloud Tiering as backup solution and data migration 679
10.5.3 Restoring data by using Transparent Cloud Tiering 680
10.5.4 Transparent Cloud Tiering restrictions 680
10.6 Implementing Transparent Cloud Tiering 681
10.6.1 Domain Name System configuration 681
10.6.2 Enabling Transparent Cloud Tiering 681
10.6.3 Creating cloud snapshots 684
10.6.4 Managing cloud snapshots 687
10.6.5 Restoring cloud snapshots 688
10.7 Volume mirroring and migration options 691
10.8 Remote Copy 693
10.8.1 IBM SAN Volume Controller and IBM FlashSystems system layers 694
10.8.2 Multiple IBM Spectrum Virtualize systems replication 695
10.8.3 Importance of write ordering 698
10.8.4 Remote Copy intercluster communication 700
10.8.5 Metro Mirror overview 701
10.8.6 Synchronous Remote Copy 702
10.8.7 Metro Mirror features 702
10.8.8 Metro Mirror attributes 703
10.8.9 Practical use of Metro Mirror 703

10.8.10 Global Mirror overview 704
10.8.11 Asynchronous Remote Copy 705
10.8.12 Global Mirror features 706
10.8.13 Using Global Mirror with change volumes. 709
10.8.14 Distribution of work among nodes. 710
10.8.15 Background copy performance 711
10.8.16 Thin-provisioned background copy 711
10.8.17 Methods of synchronization 712
10.8.18 Practical use of Global Mirror 712
10.8.19 IBM Spectrum Virtualize HyperSwap topology 712
10.8.20 Consistency Protection for Global Mirror and Metro Mirror. 713
10.8.21 Valid combinations of FlashCopy, Metro Mirror, and Global Mirror 714
10.8.22 Remote Copy configuration limits 714
10.8.23 Remote Copy states and events 715
10.9 Remote Copy commands 722
10.9.1 Remote Copy process 722
10.9.2 Listing available system partners 723
10.9.3 Changing the system parameters 723
10.9.4 System partnership 724
10.9.5 Creating a Metro Mirror/Global Mirror consistency group 725
10.9.6 Creating a Metro Mirror/Global Mirror relationship 726
10.9.7 Changing Metro Mirror/Global Mirror relationship 726
10.9.8 Changing Metro Mirror/Global Mirror consistency group 726
10.9.9 Starting Metro Mirror/Global Mirror relationship 726
10.9.10 Stopping Metro Mirror/Global Mirror relationship 727
10.9.11 Starting Metro Mirror/Global Mirror consistency group 727
10.9.12 Stopping Metro Mirror/Global Mirror consistency group 727
10.9.13 Deleting Metro Mirror/Global Mirror relationship 728
10.9.14 Deleting Metro Mirror/Global Mirror consistency group. 728
10.9.15 Reversing Metro Mirror/Global Mirror relationship 728
10.9.16 Reversing Metro Mirror/Global Mirror consistency group 729
10.10 Native IP replication 729
10.10.1 Enhancements 729
10.10.2 Native IP replication technology 731
10.10.3 IP partnership limitations. 732
10.10.4 IP Partnership and data compression 734
10.10.5 VLAN support 735
10.10.6 IP partnership and terminology 736
10.10.7 States of IP partnership 736
10.10.8 Remote Copy groups 737
10.10.9 Supported configurations 739
10.11 Managing Remote Copy by using the GUI 751
10.11.1 Creating Fibre Channel partnership 753
10.11.2 Creating Remote Copy relationships 755
10.11.3 Creating a consistency group 762
10.11.4 Renaming Remote Copy relationships 763
10.11.5 Renaming a Remote Copy consistency group 764
10.11.6 Moving standalone Remote Copy relationships to consistency group 765
10.11.7 Removing Remote Copy relationships from consistency group 766
10.11.8 Starting Remote Copy relationships 767
10.11.9 Starting a Remote Copy consistency group 768
10.11.10 Switching a relationship copy direction 768
10.11.11 Switching a consistency group direction 770

10.11.12 Stopping Remote Copy relationships . . . 771
10.11.13 Stopping a consistency group . . . 772
10.11.14 Deleting Remote Copy relationships . . . 773
10.11.15 Deleting a consistency group . . . 774
10.12 Remote Copy memory allocation . . . 775
10.13 Troubleshooting Remote Copy . . . 776
10.13.1 1920 error . . . 776
10.13.2 1720 error . . . 778

Chapter 11. Ownership groups . . . 781
11.1 Ownership groups principles of operations . . . 782
11.2 Implementing ownership groups on a new system . . . 783
11.2.1 Creating an ownership group . . . 783
11.2.2 Assigning users to an ownership group . . . 784
11.2.3 Creating ownership group resources . . . 785
11.2.4 Listing ownership group resources . . . 787
11.2.5 Actions on ownership groups . . . 788
11.3 Migrating objects to ownership groups . . . 788

Chapter 12. Encryption . . . 793
12.1 General types of encryption across IBM Spectrum Virtualize . . . 794
12.1.1 Externally virtualized storage . . . 794
12.1.2 Serial-attached SCSI internal storage . . . 794
12.1.3 Non-Volatile Memory Express internal storage . . . 794
12.2 Planning for encryption . . . 795
12.3 Defining encryption of data at-rest . . . 796
12.3.1 Encryption methods . . . 796
12.3.2 Encrypted data . . . 796
12.3.3 Encryption keys . . . 799
12.3.4 Encryption licenses . . . 800
12.4 Activating encryption . . . 800
12.4.1 Obtaining an encryption license . . . 801
12.4.2 Starting the activation process during initial system setup . . . 801
12.4.3 Starting the activation process on a running system . . . 804
12.4.4 Activating the license automatically . . . 805
12.4.5 Manual license activation . . . 808
12.5 Enabling encryption . . . 809
12.5.1 Starting the Enable Encryption wizard . . . 810
12.5.2 Enabling encryption by using USB flash drives . . . 812
12.5.3 Enabling encryption by using key servers . . . 816
12.5.4 Enabling encryption by using both providers . . . 829
12.6 Configuring more providers . . . 832
12.6.1 Adding key servers as a second provider . . . 832
12.6.2 Adding USB flash drives as a second provider . . . 834
12.7 Migrating between providers . . . 835
12.7.1 Changing from USB flash drive provider to encryption key server . . . 835
12.7.2 Changing from encryption key server to USB flash drive provider . . . 836
12.7.3 Migrating between different key server types . . . 836
12.8 Recovering from a provider loss . . . 838
12.9 Using encryption . . . 839
12.9.1 Encrypted pools . . . 839
12.9.2 Encrypted child pools . . . 841
12.9.3 Encrypted arrays . . . 842

12.9.4 Encrypted MDisks . . . 843
12.9.5 Encrypted volumes . . . 847
12.9.6 Restrictions . . . 848
12.10 Rekeying an encryption-enabled system . . . 848
12.10.1 Rekeying by using a key server . . . 849
12.10.2 Rekeying by using USB flash drives . . . 849
12.11 Disabling encryption . . . 851

Chapter 13. Reliability, availability, serviceability, monitoring and troubleshooting 853
13.1 Reliability, availability, and serviceability . . . 854
13.1.1 IBM SAN Volume Controller nodes . . . 855
13.1.2 Power . . . 859
13.1.3 Shutting down an IBM SAN Volume Controller cluster . . . 860
13.2 Removing or adding a node to or from the system . . . 862
13.3 Configuration backup . . . 865
13.3.1 Force configuration backup by using the CLI . . . 866
13.3.2 Saving the backup by using the GUI . . . 868
13.4 IBM Spectrum Virtualize Code Upgrade . . . 869
13.4.1 Spectrum Virtualize Version Numbering . . . 870
13.4.2 Updating the Spectrum Virtualize system software . . . 870
13.4.3 Precautions before the update . . . 870
13.4.4 IBM Spectrum Virtualize upgrade test utility . . . 871
13.4.5 Performing the upgrade . . . 872
13.4.6 Updating IBM Spectrum Virtualize with a hot spare node . . . 877
13.4.7 Updating the IBM SAN Volume Controller system manually . . . 878
13.5 Health checker feature . . . 880
13.6 Troubleshooting and fix procedures . . . 881
13.6.1 Managing event log . . . 882
13.6.2 Running a fix procedure . . . 883
13.6.3 Resolving alerts in a timely manner . . . 886
13.6.4 Event log details . . . 886
13.7 Monitoring and Event Notification . . . 889
13.7.1 Call Home function and email notification . . . 890
13.7.2 Disabling and enabling notifications . . . 897
13.7.3 Remote Support Assistance . . . 897
13.7.4 SNMP configuration . . . 901
13.7.5 Syslog notifications . . . 903
13.8 Audit log . . . 905
13.9 Collecting support information by using the GUI, CLI, and USB . . . 908
13.9.1 Collecting information by using the GUI . . . 908
13.9.2 Collecting logs by using the CLI . . . 913
13.9.3 Collecting logs by using USB . . . 914
13.9.4 Uploading files to the Support Center . . . 915
13.10 Service Assistant Tool . . . 916
13.11 IBM Storage Insights Monitoring . . . 920
13.11.1 Capacity monitoring . . . 921
13.11.2 Performance monitoring . . . 922
13.11.3 Logging support tickets by using IBM Storage Insights . . . 924
13.11.4 Managing support tickets by using IBM Storage Insights and uploading logs 931

Appendix A. Performance data and statistics gathering . . . 933
IBM SAN Volume Controller performance overview . . . 934
Performance considerations . . . 934

IBM Spectrum Virtualize performance perspectives 935
Performance monitoring 936
Collecting performance statistics 936
Real-time performance monitoring 938
Performance data collection and IBM Spectrum Control 949

Appendix B. Command-line interface setup 951
CLI setup 952
Basic setup on a Windows host 952
Basic setup on a Mac, UNIX or Linux host 961

Appendix C. Terminology 965
Commonly used terms 966

Related publications 997
IBM Redbooks 997
Help from IBM 997

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. 2021.

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 http://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®
Db2®
DB2®
DS8000®
Easy Tier®
FlashCopy®
Guardium®
HyperSwap®
IBM®
IBM Cloud®
IBM FlashCore®
IBM FlashSystem®
IBM Garage™
IBM Research®
IBM Security™
IBM Spectrum®
Insight®
Orchestrate®
Passport Advantage®
PowerHA®
Redbooks®
Redbooks (logo) ®
Resilient®
Storwize®
Tivoli®
XIV®

The following terms are trademarks of other companies:

Performance View, are trademarks or registered trademarks of Kenexa, an IBM Company.

SoftLayer, are trademarks or registered trademarks of SoftLayer, Inc., an IBM Company.

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

Built with IBM® Spectrum® Virtualize software, IBM SAN Volume Controller is an enterprise-class storage system that helps organizations achieve better data economics by supporting the new large scale workloads that are critical to success. IBM SAN Volume Controller systems can handle the massive volumes of data from mobile and social applications, enable flexible hybrid cloud deployments, and deliver the performance and scalability that is needed to gain insights from the latest analytics technologies.

This IBM Redbooks® publication introduces the features and functions of an IBM San Volume Controller system that is powered by IBM Spectrum Virtualize V8.4.2 by using several examples. This book is aimed at pre-sales and post-sales technical support and marketing and storage administrators. It helps you understand and implement the architecture, and how to take advantage of its industry-leading functions and features.

Authors

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



Abdulqadir S Ali is a Technical Lead for Spectrum Virtualize SME - North America's group and Project Manager for IBM Storage Technical Client Council. Abdulqadir joined IBM in 2012 as a Level I (SE) Support for IBM Storage and Host Support. Abdulqadir joined the Level II (SME) team and then became the Technical Lead in 2018. Abdulqadir also is a Platinum Advocate for a some of the biggest medical, financial, and retail IBM customers. Abdulqadir is passionate about support and providing the best possible IBM Support experience to our Worldwide Customers.



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.

© Copyright IBM Corp. 2021. All rights reserved.



Christian Schroeder is providing support with passion to clients worldwide, covering IBM storage products of the Spectrum Virtualize and FlashSystem products family. He has been with IBM for more than 20 years, working in product support for various platforms as IBM System x servers, SAN switches, IBM SAN Volume Controller, IBM Storwize®, and IBM FlashSystem. Christian co-authored several IBM Redbooks publications that cover IBM System x servers and SAN Volume Controller V6.3.0.



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 FlashSystem products.



David Green works with the IBM SAN Central team troubleshooting performance and other problems on storage networks. He has authored, or contributed to, several IBM Redbooks publications. He is a regular speaker at IBM Technical University. You can find his blog at Inside IBM Storage Networking where he writes about all things related to Storage Networking and IBM Storage Insights.



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



Hartmut Lonzer is the IBM FlashSystem Territory Account Manager for Germany (D), Austria (A), and Switzerland (CH) (DACH). Before this position, he was OEM Alliance Manager for Lenovo in IBM Germany. He works at the IBM Germany headquarters 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 for 43 years.



John Nycz is an Advanced Subject Matter Expert for IBM Spectrum 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 member of IBM's Spectrum Virtualization Support Team.



Jon Herd is an IBM Executive Technical Advisor working for the European Storage Competency Center (ESCC), Germany. He covers the UK, Ireland, and Europe, advising customers on a portfolio of IBM storage products, including IBM FlashSystem products. He also leads special projects for senior and executive management and is the SME lead for new product introduction in the ESCC. Jon has been with IBM for more than 47 years, and has held various technical roles, including Europe, Middle East, and Africa (EMEA) level support on mainframe servers and technical education development. He has written many IBM Redbooks publications about IBM FlashSystem products and is an IBM Redbooks Platinum level author. He holds IBM certifications in Product Services at a Thought Leader L3 level, and 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).



Reginald D'Souza is the IBM Storage Technical Sales Specialist with IBM Australia. He has 20 years of experience as an IT infrastructure specialist managing and delivering projects spanning different industry verticals. He is focused on software-defined storage and modern data protection solutions assisting clients in their hybrid cloud journey. Reginald also is an IBM L2 Certified IT Specialist.



Shu Mookerjee is a Storage Technical Specialist in the US, focusing on the Financial Services Sector in Texas, Oklahoma, Louisiana, and Arkansas. He has been with IBM for over 20 years and has worked in various roles, including sales, management, and technology. Shu has experience with multiple storage technologies, including the IBM DS8000, IBM Storwize, XIV, FlashSystems, IBM SAN Volume Controller, IBM Cloud® Object Storage, IBM Tape Systems, and IBM SAN switches. He holds L2 (Expert) Professional Certifications as a Certified Technical Specialist and as a Certified Storage Systems Specialist.



Tiago Bastos is a storage area network (SAN) 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 a service (SaaS) 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.



Vasfi Gucer is an IBM Technical Content Services Project Leader with IBM Garage™ for Systems. He has more than 25 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, cloud computing and cloud storage technologies for the last 8 years. Vasfi also is an IBM Certified Senior IT Specialist, Project Management Professional (PMP), IT Infrastructure Library (ITIL) V2 Manager, and ITIL V3 Expert.

Thanks to the following for their contributions that made this book possible:

Evelyn Perez, Suri Polisetti
IBM Hursley, UK

Roger Kasten
IBM USA

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

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

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

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://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 family.

This chapter includes the following topics:

- 1.1, “Storage virtualization terminology” on page 2
- 1.2, “Latest changes and enhancements” on page 5
- 1.3, “IBM SAN Volume Controller architecture” on page 15
- 1.4, “IBM SAN Volume Controller models” on page 21
- 1.5, “IBM SAN Volume Controller components” on page 25
- 1.6, “Business continuity” on page 49
- 1.7, “Management and support tools” on page 54
- 1.8, “Useful IBM SAN Volume Controller web links” on page 57

© Copyright IBM Corp. 2021. All rights reserved.

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.

We describe storage virtualization in the following way:

- 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 away 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 Spectrum 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 Spectrum Virtualize uses the Small Computer System Interface (SCSI) protocol to communicate with its clients and 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 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 (OS) 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 Spectrum 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 Spectrum Virtualize by external storage usually 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.

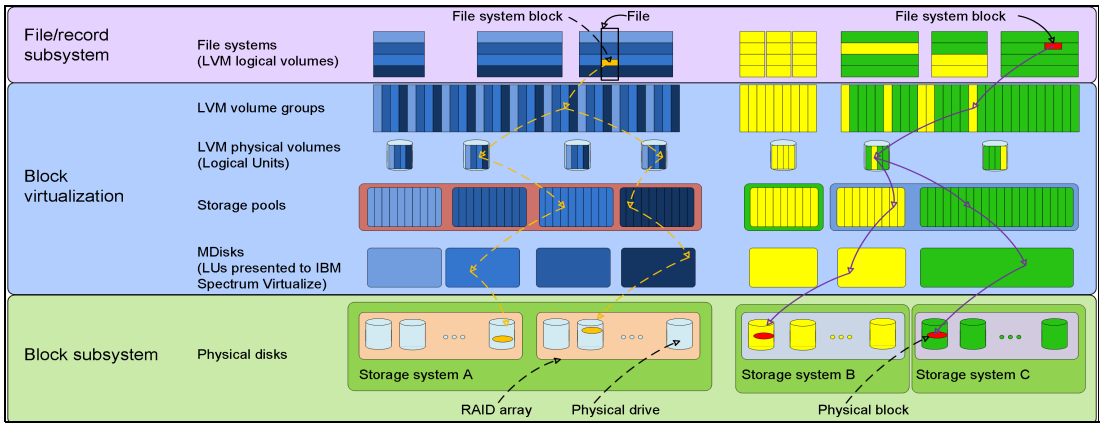


Figure 1-1 Block-level virtualization overview

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.

Importantly, the mapping can be dynamic: With IBM Easy Tier, IBM Spectrum Virtualize can automatically change 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 Spectrum 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 Spectrum Virtualize.

Storage virtualization is implemented on many layers. Figure 1-1 on page 3 shows an example where a file system block is mirrored by the host's OS (left side of the figure) by using features of the logical volume manager (LVM) or the IBM Spectrum Virtualize system at the storage pool level (as shown on the right side of Figure 1-1 on page 3). Although the result is similar (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 Spectrum 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 Spectrum 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 Spectrum 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 functionality to take immutable copies, and aids in the recovery from ransomware or customer internal actions that might result in loss of data.
- 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 Spectrum 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 thin provisioning, deduplication, and compression for space efficiency, storing more data on the storage that is owned. IBM Spectrum Virtualize can enable significant 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

Note: IBM Real-time Compression (RtC) is only available for earlier generation engines. The newer IBM SAN Volume Controller engines (SV2 and SA2) 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 Spectrum Virtualize running on IBM SAN Volume Controller and IBM FlashSystem family is a mature, 10th-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 Spectrum Virtualize can improve the usage of your storage resources, simplify storage management, and improve the availability of business applications.

1.2 Latest changes and enhancements

In this section, the latest changes and enhancements in IBM Spectrum Virtualize V8.4.2 are discussed.

1.2.1 IBM Spectrum Virtualize V8.4.2

IBM Spectrum Virtualize V8.4.2 provides more features and updates to the IBM Spectrum Virtualize family of products of which IBM FlashSystem is part. Version 8.4.2 features the following major software changes:

- Code release schedule
- DRP and performance improvements
- Safeguarded Copy (SGC)
- Volume mobility
- IP improvements
- RESTful API improvements
- Cloud driver updates (Ansible)
- GUI updates
- Miscellaneous items:
  - FlashCopy Volume Expand
  - Volume scalability
  - Remote support interface improvements
  - Child pool throttling
  - Upgrade Checker for Drive Firmware updates
  - PCI-NVMe Inter Canister Messaging
- Storage as a Service (STaaS)

These changes are described next.

Code release schedule

The following new code release schedule is featured:

- Two types of releases are available: Continuous Development Releases (CDR) and Long Term Releases (LTR).
- There are four slated updates each year:
  - One LTR:
    - New features include a formal announcement.
    - All of the enhancements of the previous CDRs are included.
    - Adopts the major numbering scheme; for example, levels such as 8.4, 8.5, and 8.6.
    - Includes all Program Temporary Fixes (PTFs) of problems found with existing features going forward.
  - Three CDRs:
    - No formal announcement.
    - New features are announced, but do not have PTFs.
    - PTFs generally are added to the next CDR, unless critical.
    - The fixes are incorporated into the next CDR.
    - Adopts a minor numbering scheme that is used; for example, levels such as 8.4.1 and 8.4.2.

DRP and performance improvements

The following DRP and performance improvements are included:

- Overall write performance:
  - The FS 5200 introduced write performance improvements in 1Q21.
  - With Release 8.4.2, the same applies to FS9200 and FS7200:.

    Streamlining the inter-canister system messaging with PCI-NVMe-based messaging (replacing the established SCSI-based messaging).
  - These DRP and performance improvements translate to performance benefits for all customers and use cases.
- DRP performance enhancements:
  - FS9200:
    - Random I/O performance improvements for read and write.
    - 256 KB blocks sequential performance improvements for read and write.
    - Significant write latency reduction because of intra-canister link messaging,
    - SHA1 off-load, and many other DRP optimizations.
  - FS7200:
    - OLP performance improvements.
    - Significant write benefits because of intra-canister link messaging,
    - SHA1 off-load, and many other DRP optimizations.

– FS5200:
  - Some performance improvements with Release 8.4.2, with many DRP optimizations.
  - The FS5200 incorporated the benefits from streamlining intra-canister link messaging at its initial release earlier in 2021.
  - Modest additional OLTP performance improvements with Release 8.4.2.

Safeguarded Copy

SGC is a virtual air gap mechanism that uses FlashCopy functions to take immutable copies, and aids in the recovery from ransomware or internal “bad actors” who want to destroy data.

Note: The Safeguarded Copy function is available with IBM Spectrum Virtualize software 8.4.2, and is *not* supported for the FlashSystem 5000, Storwize V5030E, Storwize V7000 Gen2, and Storwize V7000 Gen2+ models.

It is a feature or solution with which you can create point-in time copies (“Granularity”) of active production data that cannot be altered or deleted (so that is “Immutability” or protected copies). It requires a user with the right privilege access to modify the SGC expiration settings (Separation of Duties). Lastly, SGC uses Copy Management software for testing and ease of recovery of copies.

Safeguarded Copy use cases

The Safeguarded Copy use cases are available:

- 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.

- 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 recovery option.

- Offline Backup

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

Safeguarded Copy is the first step toward an IBM Cyber Resilient® solution.

Focusing mainly on the feature set from a customer’s view, we must address the following three pillars, as shown in Figure 1-2 on page 8.

- Separation of Duties
- Protected Copies
- Automation

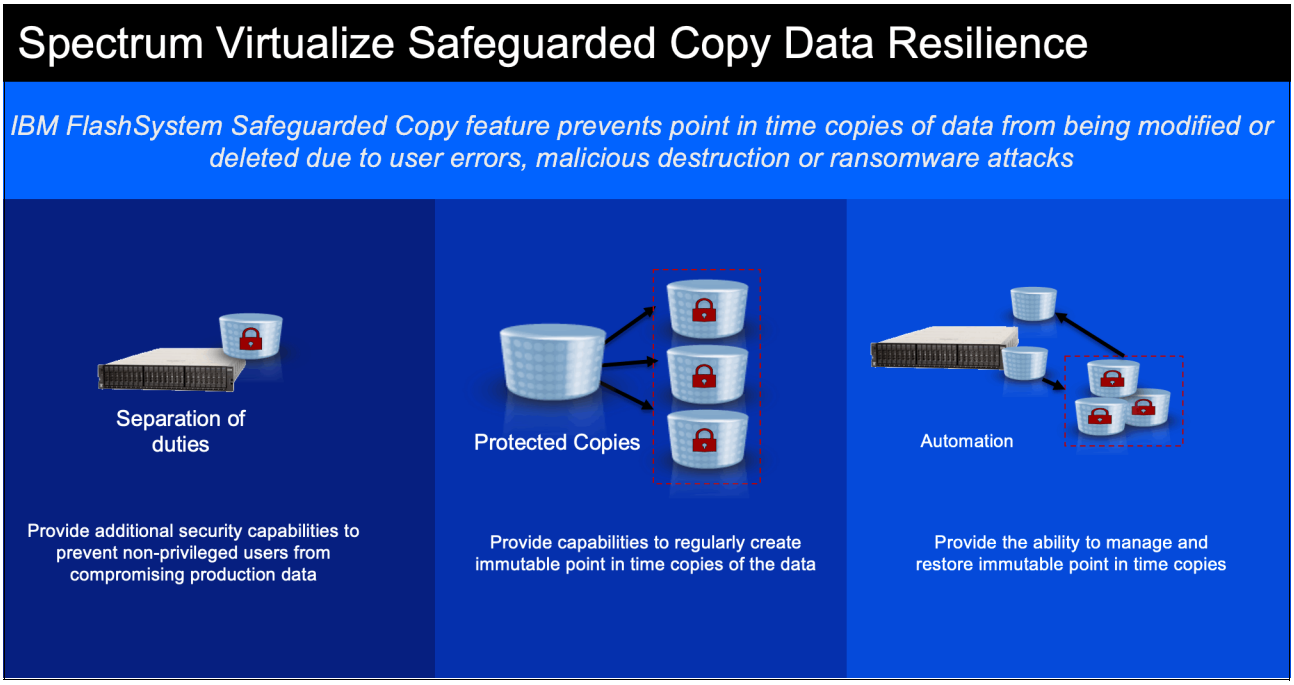


Figure 1-2 Safeguarded Copy Data Resilience

Figure 1-2 shows the Spectrum Virtualize Safeguarded Copy Data Resilience examples.

- Separation of duties
  - Traditional backup and restore capabilities are normally storage administrator-controlled and do not protect against intentional (for example, rogue employee) or non-intentional attacks.
  - Ability to treat primary and backup differently. Protecting current backups, or securing and hardening the existing 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, or 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:

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

Volume mobility

Volume mobility is similar to non-disruptive volume moves between I/O groups. The difference is that with volume mobility, a volume is migrated between systems (for example, IBM SAN Volume Controller to FS9200). It allows a user to non-disruptively move data between systems that do not natively cluster. This feature is a major benefit in upgrading and changing Spectrum Virtualize systems.

Volume migrations also includes the following features:

- Non-disruptive volume migrations between independent systems or clusters
- Non-disruptive migration between non-clustering platforms
- Migration of volumes away from a cluster that is reaching max capacity limits

Volume mobility use enhancements to SCSI (ALUA) path states and the migration is based on Remote Copy (Metro Mirror) functionality.

Volume mobility include some considerations and restrictions, such as the following examples:

- No 3-site support
- Not a DR or HA solution
- No support for consistency groups (no use-case), 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, and SCSI only (No NVMeF)
- No SCSI persistent reservations or Offloaded Data Transfer (ODX)
- IBM i, AIX®, Hyper-V, and Windows are *not* supported
- Remote Mirror license are not required

IP improvements

The following IP improvements are included in this release:

- Multiple IPs/VLANs per port

  Ethernet support on Spectrum Virtualize systems was expanded to allow multiple IP address per port along with VLAN tagging. A new model called *portsets*, replaces the current port groups.

  The following features are supported with multiple IPs/VLANs per port:

  - Enhanced Spectrum Virtualize Ethernet support with more than 1 IPv4 and 1 IPv6 addresses, to be defined per port for use by Ethernet host attach protocols, such as iSCSI, iSER and NVMeF (in the future).
  - VLAN separation for individual IP address or as wanted.
  - New Portset-based configuration model for Ethernet/IP connectivity for iSCSI and iSER Host Attach and IP Replication.

    Note: At the time of this writing, only iSCSI and iSER Host Attach and IP Replication are supported. NVMeF and Fiber Channel are to be included in the future.

  - OBAC-based, per-tenant administration and partitioning model for multi tenant cloud environments.
  - New CLI model for Ethernet network configuration.

- Multiple IP Partnerships

  IBM Spectrum Virtualize V8.4.2 provides the ability to have 3 IP partnerships per system, which allows parity with FC-based systems (it allows up to four systems to be in partnerships with each other). It does not allow some of the any-to-any connectivity of partnerships that FC allows; therefore, the only configurations that are allowed are a star and linear system configuration.

Consider the following points:

- Overall throughput of a Portset on 10Gb stays the same; that is, approximately 500 MBps for 2 – 10 Gb ports.
- In a multiple I/O group system, using a unique remote copy portset on each I/O group enables bandwidth scaling.
- Overall system limit of three partnerships per system.

  Only one can be an IP Partnership with pre-8.4.2.0 code.
- Partnerships on low memory platform nodes share memory resources; therefore, they might result in lesser performance.
- Up to 2 WAN Links can be used per I/O group.
- Portsets replaced port groups.
- Link1 and Link2 (WAN Links) to be equivalent to two port groups.
- Portsets can be shared across IP Partnerships for high bandwidth ports.

  Link1 and Link2 must be separate portsets.
- Portset type *must* be set to replication.

For more information, see IBM Documentation.

RESTful API improvements

The following RESTful API improvements are included in this release:

- JSON Web (JWT) Token feature

  JSON Web Tokens allow for authentication to occur once for a specific period and ensures that sure all commands and data are securely transmitted.

  By default, the user ID and password are used for first access and then, a token is generated and provided to the source system. This token is then used until the token expires.

  Some of the new functions are:

  - Signatures based authentication
  - No memory management
  - Contains encoded metadata:
    - Expiry time
    - Username
  - Unlimited Tokens or user without affecting existing sessions
  - User configurable token expiry:
    - Default: 1h
    - Range: 10 minutes – 2 hours

- Enhanced HTTP error messages:
  - Descriptive error codes
  - Detectable back-end failures
  - No more guessing on error conditions
  - Documented via the OpenAPI explorer interface
- Audit Logging:
  - API-started commands are now auditable
  - New field added to `catauditlog` (Origin)

    Allowed fields are GUI, CLI, and REST
- Rate limiting:
  - Protection against brute-force attacks
  - Highly restricted `/auth` endpoint
  - Non-authenticated endpoints also are rate limited; for example, `/rest/v1/lsvdisk`
  - Millisecond granularity
  - Return codes: 429 – Too Many Requests

Cloud driver updates (Ansible)

One of the more prominent features is the single authentication support for the entire playbook. This features speeds up processing and eliminates contention on the cluster.

Ansible updates are not necessarily tied to a specific Spectrum Virtualize release (unless support exists for features in that release). Another series of updates that were made to the driver were presented in the 8.4.1 Technical Update.

Ansible v1.5 release includes the following new features and updates:

- Manage hostcluster
- Single authentication for the entire playbook
- Check mode support: A preview of what is to occur
- Support for resizing volume for standard volume and HyperSwap volume

GUI updates

The following changes were made to the management GUI and the Service GUI:

- Browser language capability

  The GUI can take advantage of the browser language capability.
- Login message improvements

  Previously, if the login message was long, it was difficult to read and scroll through and was generally unmanageable. In this format, it is easier to read and handle lengthy login messages.
- User ID lock and unlock and password expiry enhancements

  Previously, these enhancements were available from the CLI only but are now incorporated into the GUI.
- Audit log source changes

  Origin and initiator information was added to the audit log (and GUI) to identify the origin of a command and who started it. This change makes it easier to track activity on the system.

- Service Assistant (SA) GUI enhancements:
  - Complete redesign of the SA interface to give a better experience to the user.
  - Includes the following changes:
    - Node Selection interface
    - Node detailed view

Miscellaneous updates

In this section, various miscellaneous updates are described.

FlashCopy Volume Expand

FlashCopy Volume Expand-related enhancements include the following examples:

- Volumes that are associated with user-defined FlashCopy mappings can now be expanded.
- The basic rules for expanding such volumes:
  - Source or target volume can be expanded at any time.
  - For incremental FlashCopy maps, the Target VDisk must be expanded before the source volume can be expanded.
  - Source and target must be same size when mapping is prepared or started.
  - Source can be shrunk, but only to the size of the largest copying-or-stopping target.
  - Target volume cannot be shrunk.
  - When target volume is a Safeguarded Copy, a security administrator must expand it from the command line.
  - Applies to:
    - Volumes in MM, GM, GMCV relationships
    - HyperSwap volumes
  - Does not apply to:
    - Three-site volumes
    - Virtual Volumes (vVols)
    - Transparent Cloud Tiering TCT volumes
  - TCT and FC cannot be used with each other

Volume scalability

Volume scalability-related enhancements include the following examples:

- The volume count on Spectrum Virtualize was increased to approximately 16 K for current and non-entry level platforms.

  Number of VDisks (VDisks + VDisk copies): 10,000 → 15,864 for DH8, SV1, SV2, SA2, FS7x, and FS9x.
- VDisks are all mappable to hosts.
- Number of FlashCopy mappings was increased from 10,000 to 15,864.
- Number of Remote Copy relationships was increased from 5000 to 7,932.

  Note: Both sites must support 10 K or more VDisks.

- HyperSwap capacity increase.

Remote support interface improvements

Remote support interface-related enhancements include the following examples:

- Remote code load

  Today, drive firmware and general fix packages are staged onto the system by using an SFTP connection to IBM Fix Central typically, which is performed by the Remote Code Team. To enable this feature, clients must open firewall holes, which is not satisfactory for most of them.

  This feature covers the migration of the code download function to use the Call Home Using REST API (CHURA) infrastructure, with the addition optional ability to use an HTTP proxy to download packages from `esupport.ibm.com`.

  The main value here is the ability to download selected code bundles (including prerequisites, drive firmware, or ifixes) without the need for the firewall hole to IBM Fix Central, and more importantly, by way of an HTTP proxy.

  The result is a reduced configuration for network administrators and fewer security worries.

- Secure Remote Access (SRA) improvements

  The current method for proxying the SRA connection uses a dedicated SSH proxy. This configuration means that clients must use:

  – Multiple objects for same purpose
  – More port forwarding for clients

  Therefore, SRA can now be configured to use an HTTP proxy, while the existing SRA proxy also is still supported.

  SRA proxy behavior now features the following aspects

  – If the established proxy is configured, SRA uses that proxy regardless of any configured HTTP proxies.
  – If no established proxy is configured and an HTTP proxy is configured, it uses the HTTP proxy.
  – If no established and no HTTP proxies are configured, it attempts to connect directly.

For more information about setting up Call Home and remote support, see this IBM Support web page.

Child pool throttling

Child pool throttling was enhanced and updated to allow control of the sub pools. The parent pool throttle also governs volumes that are in the child pool. On mirrored volumes where the secondary copy is in another pool and that pool includes a throttle, only the throttle on the primary pool applies.

Consider the following points about child pools:

- Placing a throttle on a child pool limits the IOPs and bandwidth to that pool.
- If the parent pool includes a throttle, that throttle is applied as well.
- Parent 5 K IOPs and child pool 7 K IOPS means that the child pool is limited to 5 K IOPs.
- If a volume includes a mirrored copy in another pool, the throttling applies only on to the side of the primary copy.
- Throttles apply on an I/O group basis and the throttle applies to a singular node within that I/O group.

Generally, if multi-pathing is ALUA-based, this issue does not create problems because data is driven to the primary node.

1.2.2 IBM Spectrum Virtualize V8.4.0

IBM Spectrum Virtualize 8.4 provides more features and updates to the IBM Spectrum Virtualize family of products of which IBM SAN Volume Controller is part. The following major software changes are in version 8.4 apply to IBM SAN Volume Controller:

- Data Reduction Pool (DRP) improvements;
  - DRP allows for more flexibility, such as multi-tenancy.
  - FlashCopy with redirect-on-write support use DRP's internal deduplication referencing capabilities to reduce overhead by creating references instead of copying the data. Redirect-on-write (RoW) is an alternative to the copy-on-write (CoW) capabilities.

    Note: At the time of this writing, this capability might be used only for volumes with supported deduplication without mirroring relationships and within the same pool and I/O group. The mode selection (RoW/CoW) is automatic based on these conditions.

  - Compresstimator always on, which allows the systems to sample each volume at regular intervals. It provides the ability to display the compressibility of the data in the GUI and IBM Storage Insights at any time.
  - RAID Reconstruct Read increases reliability and availability by reducing chances of DRP going offline because of fixable array issues. It uses RAID capabilities, and DRP asks for a specific data block reconstruction when detecting a potential corruption.
- Expansion of mirrored vDisks (also known as *volumes*) allows the vDisks capacity to be expanded or reduced online without requiring an offline format and sync. This change improves the availability of the volume for use because the new capacity is available immediately.
- Three-site replication with IBM HyperSwap support provides improved availability for data in three-site implementations. This improvement expands on the Disaster Recovery (DR) capabilities that are inherent in this topology.

  Note: Three-site replication by using Metro Mirror was supported on version 8.3.1 only in limited installations by way of the RPQ process. With 8.4.0, this implementation is generally available.

- Host attachment support with FC-NVMe in HyperSwap systems.
- DNS support for LDAP and NTP with full DNS length (that is, 256 characters)
- Updates to maximum configuration limits, which doubles FlashCopy mapping from 5,000 to 10,000 and increasing HyperSwap volumes limit from 1,250 to 2,000.

1.3 IBM SAN Volume Controller architecture

This section explains the major concepts underlying IBM SAN Volume Controller. It also describes the architectural overview and the terminologies that are used in a virtualized storage environment. Finally, it introduces the software and hardware components and the other functions that are available with Version 8.4.

1.3.1 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 (vDisks 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 to the 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 can have 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 has 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 about the data itself.

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

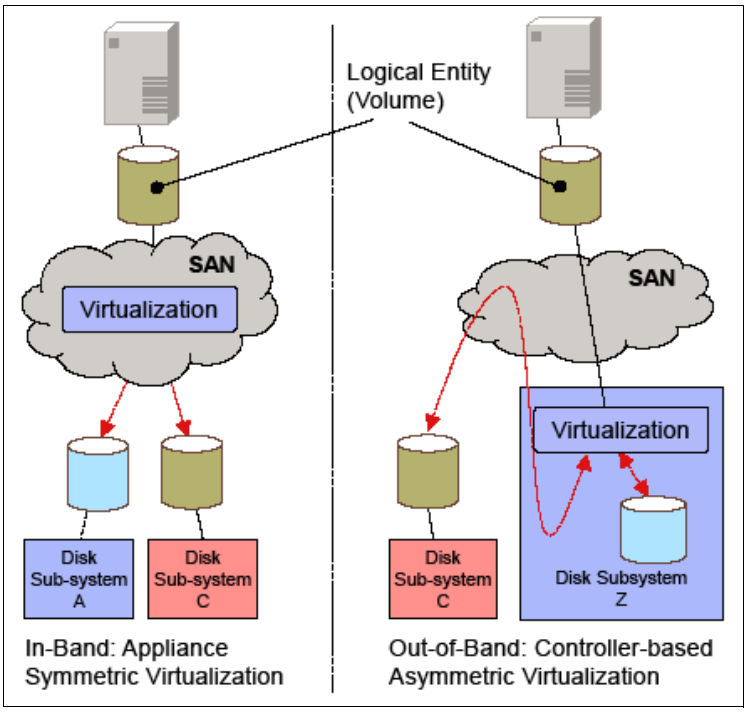


Figure 1-3 Overview of block-level virtualization architectures

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

The controller-based approach has high functionality, but it fails in terms of scalability or upgradeability. Because of the nature of its design, no true decoupling occurs with 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.

With 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.

With a 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. There are no other costs 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 Spectrum Virtualize.

IBM SAN Volume Controller includes the following key characteristics:

- It is highly scalable, which provides an easy growth path to two-*n* nodes (grow in a pair of nodes due to the cluster function).
- It is SAN interface-independent. It supports Fibre Channel (FC), FC-NVMe, iWARP, RoCE, Fibre Channel over Ethernet (FCoE), and internet Small Computer Systems Interface (iSCSI). It also is open for future enhancements.
- 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 Spectrum Virtualize V8.4 also provides these 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 IBM SAN Volume Controller to interact with cloud service providers (CSPs).
- Metro Mirror (MM) (synchronous copy).
- Global Mirror (GM) (asynchronous copy).
- Safeguarded Copy (SGC) at 8.4.2 only
- Data migration.
- 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.
- Supporting IBM HyperSwap.
- Supporting 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 functionality with IBM Spectrum Virtualize for Public Cloud.

1.3.2 IBM Spectrum Virtualize

IBM Spectrum Virtualize is a software-defined storage solution and a key member of the IBM Spectrum Storage portfolio. Built with IBM Spectrum Virtualize software, IBM SAN Volume Controller simplifies infrastructure and eliminates differences in management, function, and even hybrid multicloud support.

For more information, see the IBM Spectrum Storage portfolio web page.

Naming: With the introduction of the IBM Spectrum Storage family, the *software* that runs on IBM SAN Volume Controller and on IBM FlashSystem products is called IBM Spectrum Virtualize. The name of the underlying *hardware* platform remains intact.

IBM SAN Volume Controller systems helps organizations achieve better data economics by supporting these new workloads that are critical to their success. It can handle the massive 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.

IBM Spectrum Virtualize is the core software engine of the entire family of IBM FlashSystem products. The contents of this book are intentionally related to the deployment considerations of IBM SAN Volume Controller.

Benefits of IBM Spectrum Virtualize

IBM Spectrum Virtualize delivers leading benefits that improve storage infrastructures in many ways, including:

- Cost reduction of storing data by increasing utilization and accelerating applications to speed business insights. To achieve this, the solution:
  - Uses data reduction technologies to increase the amount of data you can store in the same space
  - Enables rapid deployment of cloud storage for DR along with the ability to store copies of local data
  - Moves data to the most suitable type of storage based on policies you define by using IBM Spectrum Control to optimize storage
  - Improves storage performance so you can get more done with your data
- Data protection from theft or inappropriate disclosure while enabling a high-availability strategy that includes protection for data and application mobility and DR. To achieve this, the solution:
  - Uses software-based encryption to improve data security
  - Provides fully duplexed copies of data and automatic switch over across data centers to improve data availability
  - Eliminates storage downtime with nondisruptive movement of data from one type of storage to another

- Provides a data strategy that is independent of your choice of infrastructure, and delivers tightly integrated functionality and consistent management across heterogeneous storage. To achieve this, the solution:
  - 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 hardware vendor for simplified management, consistent functionality, and greater efficiency
  - Supports common capabilities across storage types, and provides flexibility in storage acquisition by allowing a mix of vendors in the storage infrastructure

Note: These benefits are not a complete list of features and functions that are available with IBM Spectrum Virtualize software; it is only a subset of its total capabilities.

1.3.3 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 the external storage system and hosts.

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

Typically, the hosts cannot see or operate on the same physical storage (LUN) from the storage system that is assigned to IBM SAN Volume Controller. If the same LUNs are not shared, storage systems can be shared between the IBM SAN Volume Controller and direct host access.

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, FCoE, iSCSI over Ethernet, or possible future types.

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 practical implementations that have HA requirements (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 counterpart SANs, which provide alternative paths for each SAN-attached device.

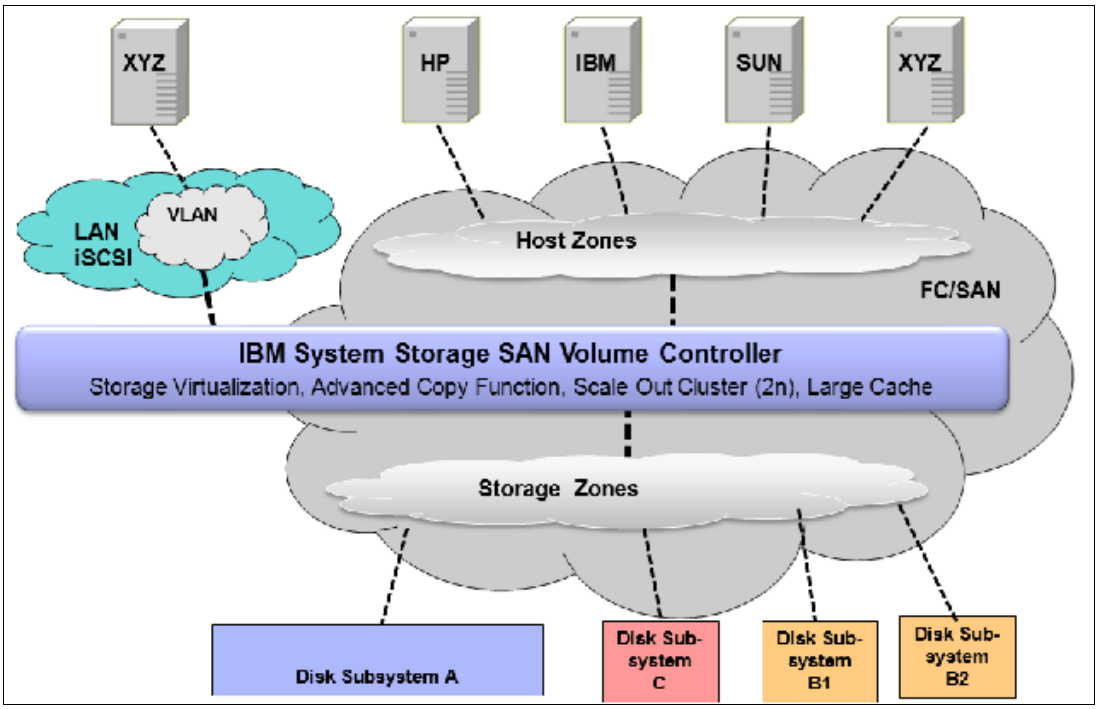


Figure 1-4 IBM SAN Volume Controller conceptual and topology overview

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 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 1.3.1, “IBM SAN Volume Controller architectural overview” on page 15, 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 (VLAN) path for storage traffic prevents any IP interface, switch, or target port failure from compromising the iSCSI connectivity across servers and storage controllers.

1.4 IBM SAN Volume Controller models

The following IBM SAN Volume Controller engine models are supported at the IBM Spectrum Virtualize V8.4 code level:

- IBM SAN Volume Controller Model SV2
- IBM SAN Volume Controller Model SA2
- IBM SAN Volume Controller Model SV1
- IBM SAN Volume Controller Model DH8

1.4.1 IBM SAN Volume Controller SV2 and SA2

Figure 1-5 shows the front view of the IBM SAN Volume Controller SV2 and SA2.



Figure 1-5 IBM SAN Volume Controller SV2 and SA2 front view

Figure 1-6 shows the rear view of the IBM SAN Volume Controller SV2 / SA2.

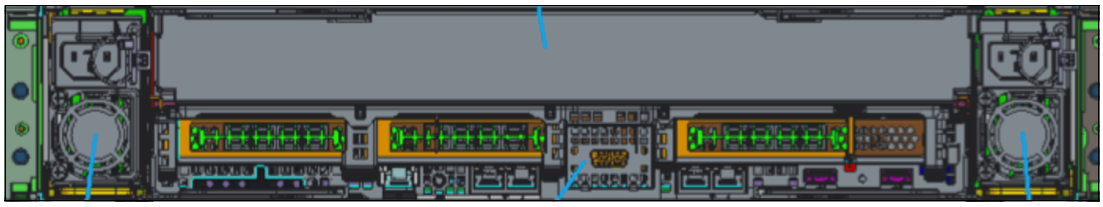


Figure 1-6 IBM SAN Volume Controller SV2 and SA2 rear view

Figure 1-7 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.

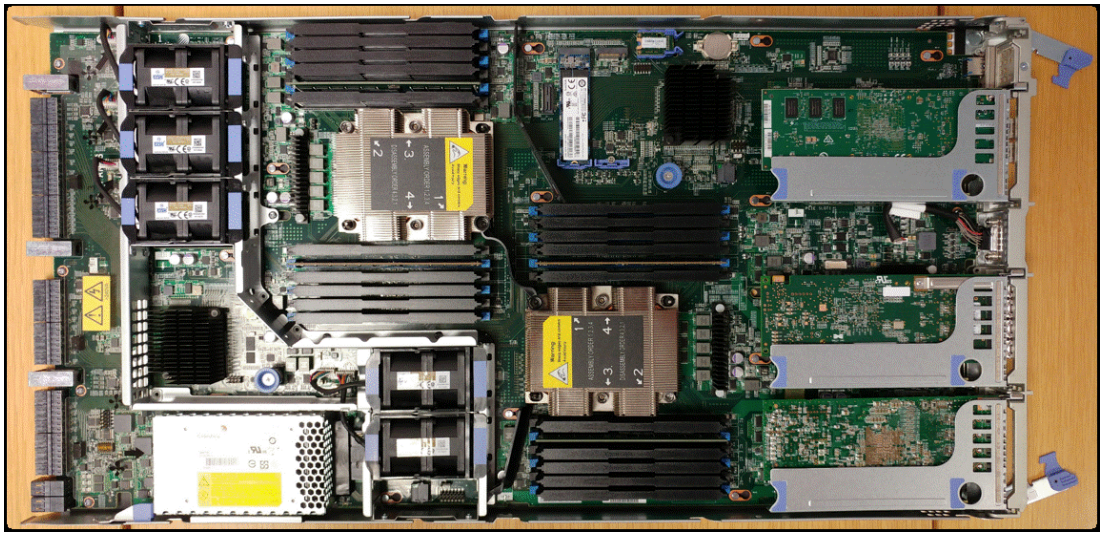


Figure 1-7 Internal hardware components

Figure 1-8 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.

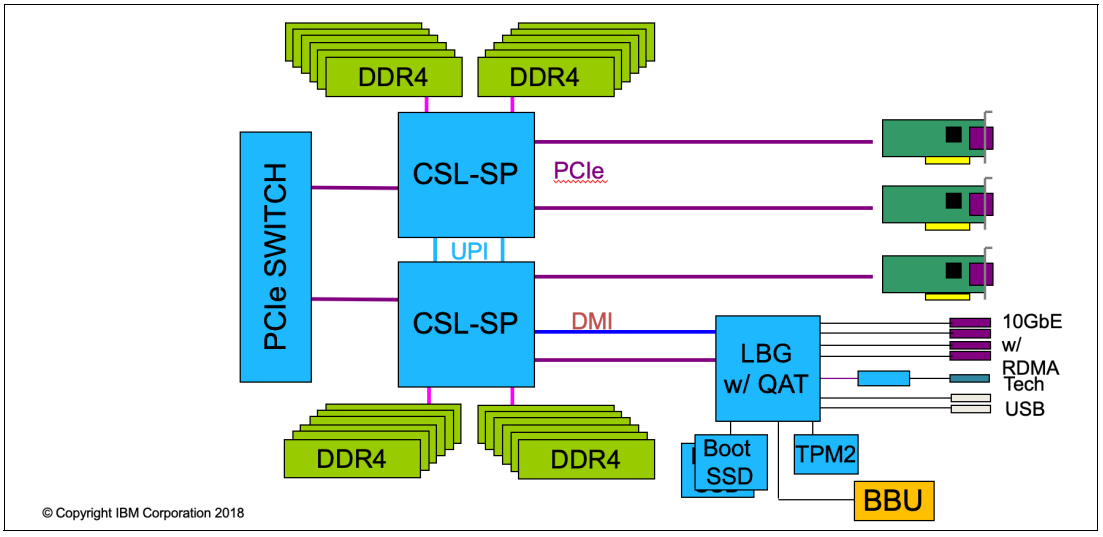


Figure 1-8 IBM SAN Volume Controller SV2 and SA2 internal architecture

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

1.4.2 IBM SAN Volume Controller SV1

Figure 1-9 shows the front view of the IBM SAN Volume Controller SV1.



Figure 1-9 IBM SAN Volume Controller SV1 front view

Figure 1.4.3 shows the rear view of the IBM SAN Volume Controller SV1.



1.4.3 IBM SAN Volume Controller model comparisons

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

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

The information in this book is valid at the time of this writing and covers IBM Spectrum Virtualize V8.4. However, as IBM SAN Volume Controller matures, expect to see new features and enhanced specifications.

Table 1-1 IBM SAN Volume Controller base models

Feature	2145/2147[a]-SV1	2145/2147-SV2	2145/2147-SA2
Processor	Two Intel Xeon E5 v4 Series, 8-cores, and 3.2 GHz	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)
Base cache memory	64 GB	128 GB	128 GB
I/O ports and management	Three 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	Four 10 Gb Ethernet ports for 10 Gb iSCSI connectivity and system management
Technician port	Single Gb Ethernet	Single 1 Gb Ethernet	Single 1 Gb Ethernet
Maximum host interface adapters slots	4	3	3
USB ports	4	2	2

Feature	2145/2147[a]-SV1	2145/2147-SV2	2145/2147-SA2
SAS chain	2	N/A	N/A
Max number of dense drawers per SAS chain	4	N/A	N/A
Integrated battery units	2	1	1
Power supplies and cooling units	2	2	2

a. Model 2147 is identical to 2145, but with an included enterprise support option from IBM.

The following optional features are available for IBM SAN Volume Controller SV1:

- A 256 GB cache upgrade fully unlocked with code Version 8.2 or later
- A 4-port 16 Gb FC adapter for 16 Gb FC connectivity
- A 4-port 10 Gb Ethernet adapter for 10 Gb iSCSI/FCoE connectivity
- Compression accelerator card for RtC
- A 4-port 12 Gb SAS expansion enclosure attachment card

Important: IBM SAN Volume Controller model 2145/2147-SV1 can contain a 16 Gb FC or a 10 Gb Ethernet adapter, but only one 10 Gbps Ethernet adapter is supported.

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)

Note: The 25 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.

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 comparison of current and previous models of IBM SAN Volume Controller is shown in Table 1-2. Expansion enclosures are not included in the list.

Table 1-2 Historical overview of IBM SAN Volume Controller models

Model	Cache (GB]	FC (Gbps)	iSCSI (Gbps)	Hardware base	Announced
2145-DH8	32 - 64	8 and 16	1, optional 10	x3550 M4	06 May 2014
2145-SV1	64 - 256	16	10	Xeon E5 v4	23 August 2016
2147-SV1	64 - 256	16	10	Xeon E5 v4	23 August 2016

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

The IBM SAN Volume Controller SV1 expansion enclosure consists of an enclosure and drives. Each enclosure contains two canisters that can be replaced and maintained independently. IBM SAN Volume Controller SV1 support three models of expansion enclosures: 12F, 24F, and 92F dense drawers.

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

Expansion Enclosure model 12F features two expansion canisters and holds up to 12 3.5-inch SAS drives in a 2U 19-inch rack mount enclosure.

Expansion Enclosure model 24F supports up to 24 internal flash drives, 2.5-inch SAS drives, or a combination of such drives. Expansion Enclosure 24F also features two expansion canisters in a 2U 19-inch rack-mounted enclosure.

Expansion Enclosure model 92F supports up to 92 3.5-inch drives in a 5U 19-inch rack-mounted enclosure. This model is known as dense expansion drawers, or *dense drawers*.

1.5 IBM SAN Volume Controller components

IBM SAN Volume Controller provides block-level aggregation and volume management for attached disk storage. In simpler terms, 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 (*target*) and to storage systems as one large host (*initiator*).

The SAN is zoned such that the application servers cannot “see” the back-end storage or controller. This configuration prevents any possible conflict between 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.

1.5.1 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 Spectrum 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.

1.5.2 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.

1.5.3 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 number 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 there is a disaster. 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, search for “Configuration Limits and Restrictions for IBM System Storage IBM SAN Volume Controller” at this IBM Support web page.

1.5.4 Expansion enclosures

Expansion enclosures are rack-mounted hardware that contains several components of the system: canisters, drives, and power supplies. Enclosures can be used to extend the capacity of the system. They are supported only on IBM SAN Volume Controller 2145 or 2147-SV1 nodes. For other models of the system, you must use external storage systems to provide capacity for data. The term *enclosure* is also used to describe the hardware and other parts that are plugged into the enclosure.

Expansion enclosures can be added to the IBM SAN Volume Controller 2145 or 2147-SV1 controller nodes to expand the available capacity of the system. (Other controller models do not support expansion enclosures.) Each system can have a maximum of four I/O groups, with two chains of expansion enclosures that are attached to each I/O group.

On each SAS chain, the systems can support up to a SAS chain weight of 10. Each 2145 / 2147-92F Expansion Enclosure adds a value of 2.5 to the SAS chain weight. Each 2145 / 2147-12F or 2145 / 2147-24F Expansion Enclosure adds a value of 1 to the SAS chain weight. For example, each of the following expansion enclosure configurations has a total SAS weight of 10:

- Four 2145 or 2147-92F enclosures per SAS chain
- Ten 2145 or 2147-12F enclosures per SAS chain
- Two 2145 or 2147-92F enclosures and five 2145 or 2147-24F enclosures per SAS chain

An expansion enclosure houses the following additional hardware: power supply units (PSUs), canisters, and drives. Enclosure objects report the connectivity of the enclosure. Each expansion enclosure is assigned a unique ID. It is possible to change the enclosure ID later.

Figure 1-10 on page 28 shows the front view of the 2145 or 2147 model 12F Expansion Enclosure.



Figure 1-10 IBM SAN Volume Controller Expansion Enclosure front view 2145 / 2147 model 12F

The drives are positioned in four columns of three horizontally mounted drive assemblies. The drive slots are numbered 1 - 12, starting at the upper left and moving left to right, top to bottom.

Figure 1-11 shows the front view of the 2145 or 2147 model 24F Expansion Enclosures.

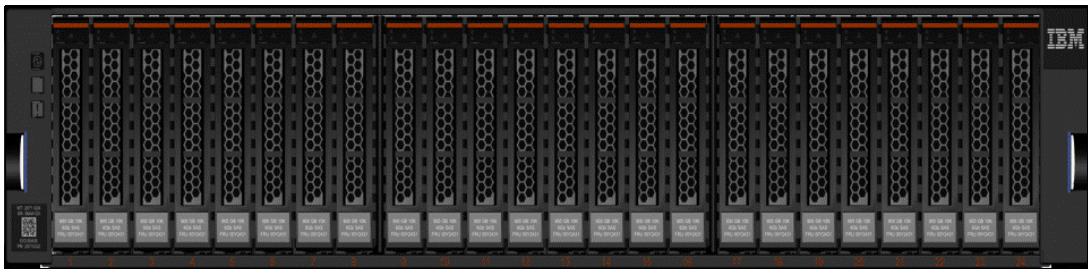


Figure 1-11 IBM SAN Volume Controller Expansion Enclosure front view 2145 / 2147 model 24F

The drives are positioned in one row of 24 vertically mounted drive assemblies. The drive slots are numbered 1 - 24, starting from the left. A vertical center drive bay molding is between slots 8 and 9 and another between slots 16 and 17.

Note: Any references to internal drives or internal drive functions in this book do not apply to IBM SAN Volume Controller Models SV2 and SA2 because these models do not support internal drives nor any type of expansion enclosure. These functions might apply to externally virtualized systems, and therefore be available.

Dense expansion drawers

Dense expansion drawers, also known as *dense drawers*, are optional 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. The two Secondary Expander Modules (SEMs) are centrally located in the chassis. One SEM addresses 54 drive ports, and the other addresses 38 drive ports. Dense drawers can support many different drive sizes and types.

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

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

Figure 1-12 shows a dense expansion drawer.



Figure 1-12 Dense expansion drawer

1.5.5 Flash drives

Flash drives can be used to overcome a growing problem that is known as the *memory bottleneck* or *storage bottleneck* specifically in single-layer cell (SLC) or multi-level cell (MLC) NAND flash-based disks.

Storage bottleneck problem

The memory or storage bottleneck describes the steadily growing gap between the time that is required for a CPU to access data that is in its cache memory (typically in nanoseconds) and data that is on external storage (typically in milliseconds).

Although CPUs and cache and memory devices continually improve their performance, mechanical disks that are used as external storage generally do not improve their performance.

Figure 1-13 shows these access time differences.

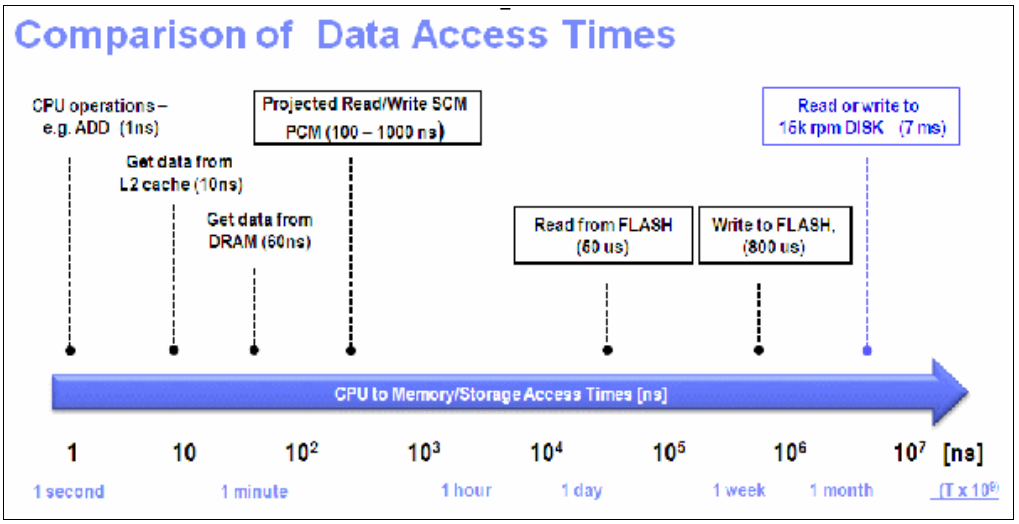


Figure 1-13 The memory or storage bottleneck

The actual times that are shown are not that important, but a noticeable difference exists between accessing data that is in cache and data that is on an external disk.

In the example that is shown in Figure 1-13 on page 29, we added a scale that gives you an idea of how long it takes to access the data in a scenario where a single CPU cycle takes 1 second. This scale shows the importance of future storage technologies closing or reducing the gap between access times for data that is stored in cache or memory versus access times for data that is stored on an external medium.

Since magnetic disks were first introduced by IBM in 1956 (the Random Access Memory Accounting System, also known as the *IBM 305 RAMAC*), they showed remarkable performance regarding capacity growth, form factor, and size reduction; price savings (cost per GB); and reliability.

However, the number of I/Os that a disk can handle and the response time that it takes to process a single I/O did not improve at the same rate, although they certainly did improve. In actual environments, you can expect from today's enterprise-class FC SAS disk up to 200 input/output operations per second (IOPS) per disk with an average response time (latency) of approximately 6 ms per I/O.

Table 1-3 shows a comparison of drive types and IOPS.

Table 1-3 Comparison of drive types to IOPS

Drive type	IOPS
NL - SAS	100
SAS 10,000 revolutions per minute (RPM)	150
SAS 15,000 RPM	250
Flash	> 500,000 read; 300,000 write

Today's spinning disks continue to advance in capacity, up to several terabytes, form factor/footprint (8.89 cm (3.5 inches), 6.35 cm (2.5 inches), and 4.57 cm (1.8 inches)), and price (cost per gigabyte), but they are not getting much faster.

The limiting factor is the number of RPM that a disk can perform (approximately 15,000). This factor defines the time that is required to access a specific data block on a rotating device. Small improvements likely will occur in the future. However, a significant step, such as doubling the RPM (if technically even possible), inevitably has an associated increase in power usage and price that is an inhibitor.

Flash drive solution

Flash drives can provide a solution for this dilemma, and no rotating parts means improved robustness and lower power usage. A remarkable improvement in I/O performance and a massive reduction in the average I/O response times (latency) are the compelling reasons to use flash drives in today's storage subsystems.

Enterprise-class flash drives typically deliver 500,000 read and 300,000 write IOPS with typical latencies of 50 µs for reads and 800 µs for writes. Their form factors of 4.57 cm (1.8 inches) / 6.35 cm (2.5 inches) / 8.89 cm (3.5 inches) and their interfaces (FC / SAS / Serial Advanced Technology Attachment (SATA)) make them easy to integrate into existing disk shelves. The IOPS metrics improve when flash drives are consolidated in storage arrays (flash arrays). In this case, the read and write IOPS are seen in millions for specific 4 KB data blocks.

Flash-drive market

The flash drive storage market is rapidly evolving. The key differentiator among today's flash drive products is not the storage medium but the logic in the disk internal controllers. The top priorities in today's controller development are optimally handling what is referred to as *wear-out leveling*, which defines the controller's capability to ensure a device's durability, and closing the remarkable gap between read and write I/O performance. Today's flash drive technology is only the first step into the world of high-performance persistent semiconductor storage.

IBM FlashCore Module and NVMe drives

Figure 1-14 shows an IBM FlashCore® Module (FCM) (NVMe) with a capacity of 19.2 TB that is built by using 64-layer TLC flash memory and an Everspin MRAM cache into a U.2 form factor.



Figure 1-14 FlashCore Module (NVMe)

FCM modules (NVMe) are designed for high parallelism and optimized for 3D TLC and updated FPGAs. IBM also enhanced the FCM modules by adding a read cache to reduce latency on highly compressed pages, and four-plane programming to lower the overall power during writes. FCM modules offer hardware-assisted compression up to 3:1, and are FIPS 140-2 compliant.

FCM modules carry the IBM patented Variable Stripe RAID at the FCM level, and use distributed RAID (DRAID) to protect data at the system level. VSR and DRAID together optimize RAID rebuilds by offloading rebuilds to DRAID and offers protection against FCM failures.

FCMs are available in the following capacity sizes: 4.8 TB, 9.6 TB, 19.2 TB, or 38.4 TB.

Note: At the time of this writing, FCM drives are available on the IBM FlashSystem family only. IBM SAN Volume Controller SV2 and SA2 do not support any internal drive types.

Storage-class memory

SCM promises a massive improvement in performance (in IOPS), and a real density, cost, and energy efficiency compared to today's flash-drive technology. IBM Research® is actively engaged in these new technologies.

For more information about nanoscale devices, see this web page.

For a comprehensive overview of the flash drive technology in a subset of the well-known Storage Networking Industry Association (SNIA) Technical Tutorials, see this web page.

When these technologies become a reality, they will fundamentally change the architecture of today's storage infrastructures. Figure 1-15 shows the different types of storage technologies versus the latency.

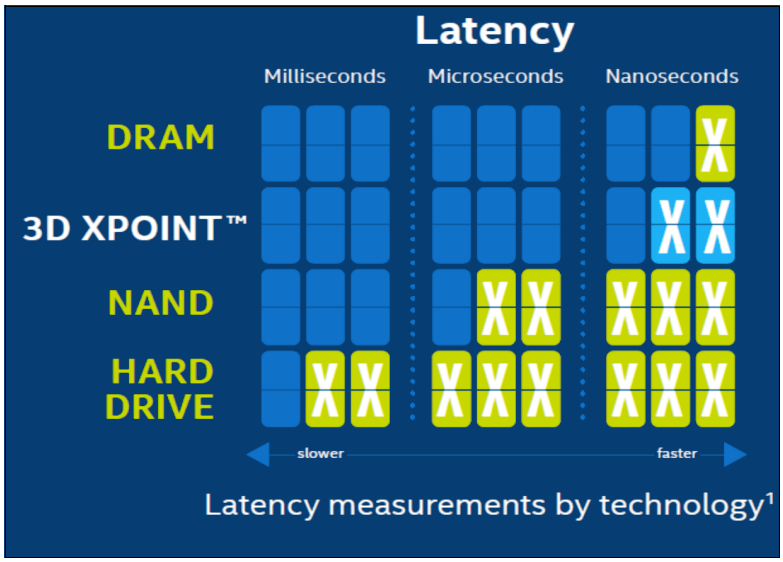


Figure 1-15 Storage technologies versus latency for Intel SCM drives

1.5.6 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 IBM SAN Volume Controller does not attempt to provide recovery from physical disk failures within the back-end storage system, an MDisk often is 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 *VDisks* or *volumes*. These disks are presented by the IBM SAN Volume Controller I/O groups through the SAN or LAN to the servers.

For information about the system limits and restrictions, search for "Configuration Limits and Restrictions for IBM System Storage IBM SAN Volume Controller" at this web page.

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

- 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 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 527.

    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 has an online path count that is the number of nodes that has access to 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 managed disks is not detected automatically and is set to *enterprise*. If the external managed disk is made up of flash drives or nearline Serial Attached SCSI (SAS) drives and you want to use Easy Tier, you must specify the tier when adding the managed disk to the storage pool or run the **chmdisk** command to modify the tier attribute.

1.5.7 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).

- The IBM SAN Volume Controller Model SV1 features 64 GB of memory with options for 256 GB of memory in a 2U 19-inch rack mount enclosure.
- 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 provides a flexible cache model. The node's memory can be used as read or write cache.

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 1-16 shows the separation of the upper and lower cache.

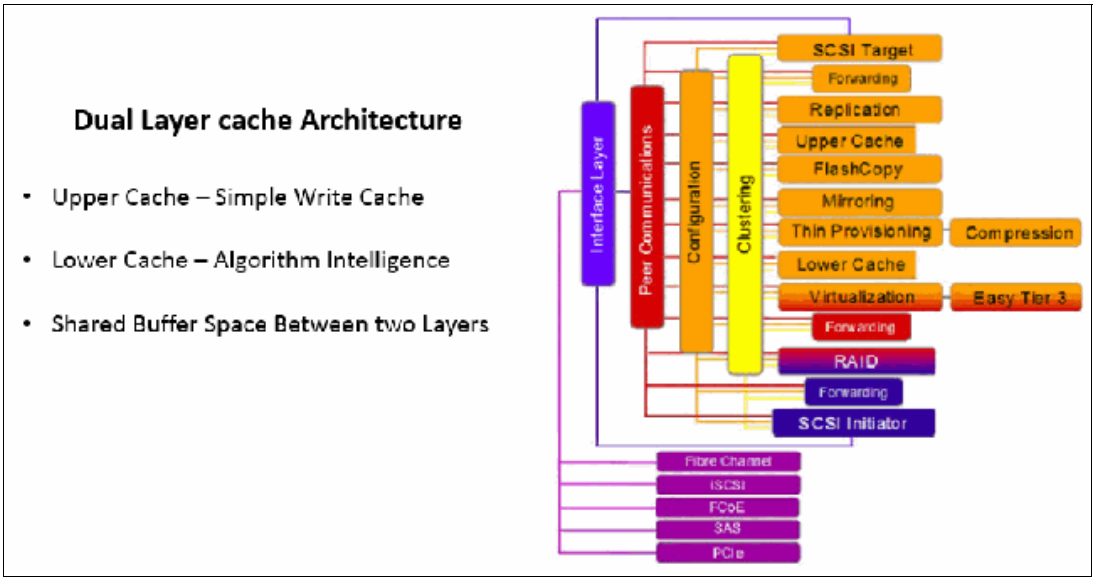


Figure 1-16 Separation of upper and lower cache

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

- Provides fast write response times to the host by being as high up in the I/O stack as possible.
- Provides partitioning.

The lower cache delivers the following additional functions:

- Ensures that the write cache between two nodes is in sync.
- Caches partitioning to ensure that a slow back-end 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 enhanced statistics for IBM Spectrum Control and IBM Storage Insights.
- Provides 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.
- Supports fast-write (two-way and one-way), flush-through, and write-through.
- Integrates with T3 recovery procedures.
- Supports two-way operation.
- Supports none, read-only, and read/write as user-exposed caching policies.
- Supports flush-when-idle.
- Supports expanding cache as more memory becomes available to the platform.
- Supports credit throttling to avoid I/O skew and offer fairness/balanced I/O between the two nodes of the I/O group.
- Enables switching of 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 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 shares memory with the main I/O process.

1.5.8 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 there is a problem 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 those 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, the owner continues to operate as the system and handles 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 there is a disaster. 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 IBM Documentation.

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, there are strict requirements on the IP network, and some disadvantages with using IP quorum applications.

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. The maximum number of applications that can be deployed is five.

1.5.9 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 1-17 shows the relationships of the IBM SAN Volume Controller entities to each other.

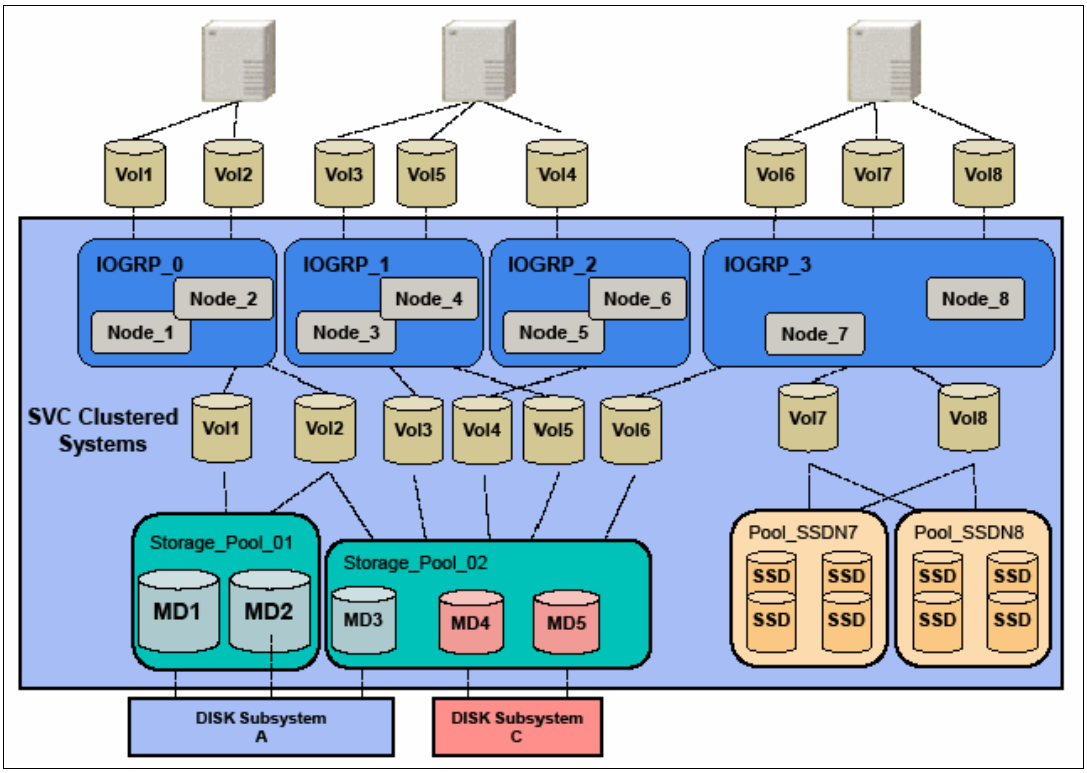


Figure 1-17 Overview of an IBM SAN Volume Controller clustered system with an I/O group

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.

1.5.10 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 1-18 shows how a striped volume is allocated, assuming that 10 extents are required.

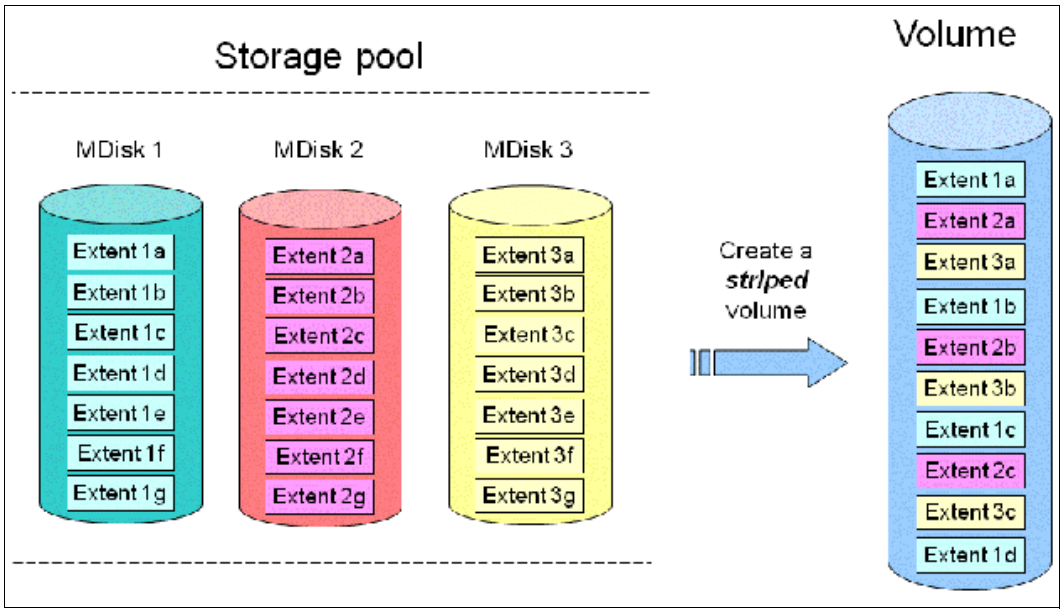


Figure 1-18 Striped volume

- Sequential

  A sequential volume is where the extents are allocated sequentially from one MDisk to the next MDisk (see Figure 1-19).

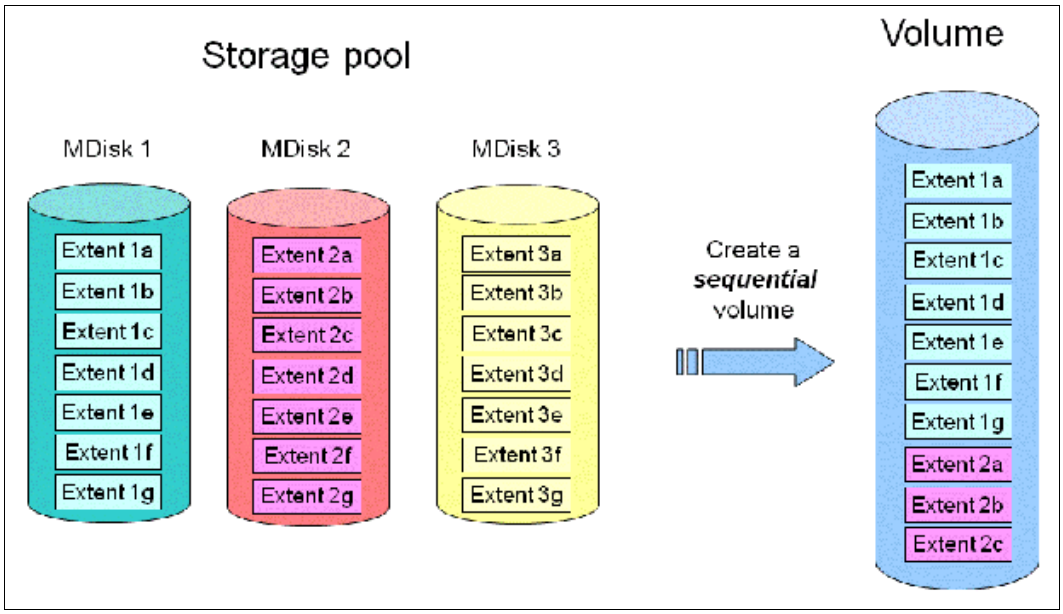


Figure 1-19 Sequential volume

- Image mode

  Image mode volumes (see Figure 1-20 on page 41) 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 non-virtualized) storage to the IBM SAN Volume Controller based virtualized infrastructure.

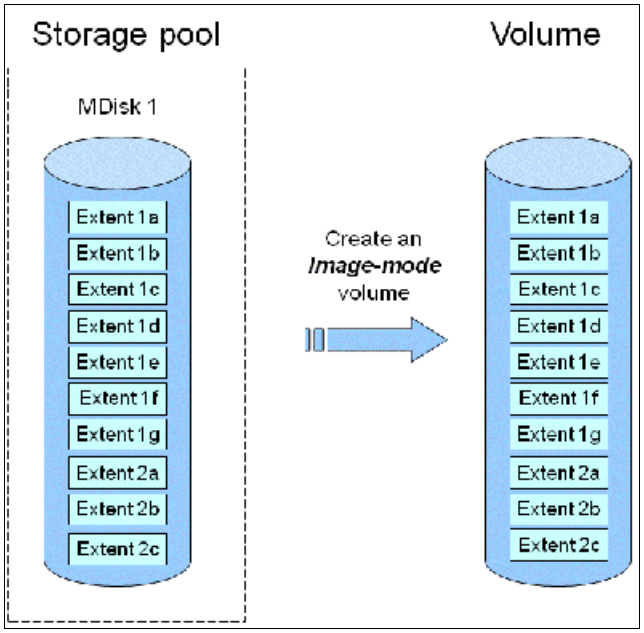


Figure 1-20 Image mode volume

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.

Some virtualization functions are not available for image mode volumes, so 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 containing 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.

Spectrum Virtualize code 8.4.2. increased the number of VDisks (VDisks + VDisk copies): from 10,000 to 15,864 for IBM SAN Volume Controller models DH8, SV1, SV2, and SA2. For more information, see “Volume scalability” on page 12.

1.5.11 Easy Tier

Easy Tier is a performance function that automatically migrate extents off a volume to or from one MDisk storage tier to another MDisk storage tier. Since V7.3, the IBM SAN Volume Controller code can support a three-tier implementation.

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. STAT can be downloaded for no initial cost from this web page.

For more information about Easy Tier, see Chapter 9, “Advanced features for storage efficiency” on page 559.

1.5.12 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 or Fibre Channel over Ethernet 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, ports do not come up until they are ready to service I/O, which improves host behavior around node unpends. In addition, path failures that occur because an offline node are 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.

1.5.13 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.

1.5.14 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, so 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 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.

In IBM Spectrum Virtualize V7.4, hardware encryption was introduced with the software encryption option that was introduced in Version 7.6. Encryption keys can be managed by an external key management system, such as the IBM Security™ Key Lifecycle Manager (SKLM), or stored on USB flash drives that are attached to a minimum of one of the nodes. Since Version 8.1, IBM Spectrum Virtualize provides a combination of external and USB key repositories.

SKLM is an IBM solution to provide 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 SKLM is used as a key manager for the IBM SAN Volume Controller encryption, you can run into 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. Up to four SKLM servers are supported.

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. That data encryption key is protected by a 256-bit AES key wrap when it is stored in non-volatile form.

Another data security enhancement, which is delivered with the Spectrum Virtualize 8.4.2 code, is the new SafeGuarded Copy function that can provide protected read-only air gap copies of volumes. this enhancement gives the customer effective data protection against cyber attacks.

For more information, see *Implementation Guide for SpecV/FlashSystem Safeguarded Copy*, REDP-5654.

1.5.15 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 Spectrum Virtualize software, not hardware. In Version 7.7, IBM introduced software capabilities to enable the underlying virtualized storage to attach to IBM SAN Volume Controller by using the iSCSI protocol.

The iSCSI protocol enables the transport 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 existing 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 an iSCSI name and address:

- 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 host name 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 way of Dynamic Host Configuration Protocol (DHCP). An IBM SAN Volume Controller node represents an iSCSI node and provides statically allocated IP addresses.

IBM SAN Volume Controller models SV2 and SA2 supports 25 Gbps Ethernet adapters that provide both iSCSI and iSCSI Extensions over RDMA (iSER) connections.

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 current hardware. Currently, 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.

1.5.16 IBM Real-time Compression

RtC is an attractive solution to address the increasing requirements for data storage, power, cooling, and floor space. When applied, RtC can save storage space so more data can be stored, and fewer storage enclosures are required to store a data set.

Note: IBM SAN Volume Controller models SV2 and SA2 do not support RtC software compression. They support only the newer DRP software compression.

RtC provides the following benefits:

- Compression for active primary data. RtC can be used with active primary data.
- Compression for replicated/mirrored data. Remote volume copies can be compressed in addition to the volumes at the primary storage tier. This process also reduces storage requirements in MM and GM destination volumes.
- No changes to the existing environment are required. RtC is part of the storage system.
- Overall savings in operational expenses. More data is stored and fewer storage expansion enclosures are required. Reducing rack space has the following benefits:
  - Reduced power and cooling requirements. More data is stored in a system, requiring less power and cooling per gigabyte or used capacity.
  - Reduced software licensing for more functions in the system. More data is stored per enclosure, which reduces the overall spending on licensing.
- Disk space savings are immediate. The space reduction occurs when the host writes the data. This process is unlike other compression solutions in which some or all the reduction is realized only after a post-process compression batch job is run.

When compression is applied, it is a best practice to monitor the overall performance and CPU utilization. Compression can be implemented without any impact to the environment, and it can be used with storage processes running.

1.5.17 Data Reduction Pools

Data Reduction Pools (DRP) represent a significant 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.

DRP is a new type of storage pool, implementing 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.

With 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 performance 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 existing 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, then pointers are shifted to reference a single copy of the chunk, and the duplicate data sets are then released.

To estimate potential capacity savings that data reduction can provide on the system, use the Data Reduction Estimation Tool (DRET). 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.

DRET is available for download at this IBM Support web page.

1.5.18 IP replication

IP replication was introduced in Version 7.2, and it enables data replication between IBM Spectrum Virtualize family members. IP replication uses the IP-based ports of the cluster nodes.

The IP replication function is transparent to servers and applications like traditional FC-based mirroring is. All remote mirroring modes (MM, GM, and GMCV) are supported.

The configuration of the system is straightforward, and IBM Spectrum Virtualize family systems normally “find” 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. Remember, remote mirror is a chargeable option, but the price does not change with IP replication. Existing remote mirror users have access to the function at no additional charge.

IP connections that are used for replication can have long latency (the time to transmit a signal from one end to the other), 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 utilization as low as 20% (based on IBM measurements). In addition, this scenario gets worse the longer the latency.

Bridgeworks SANSlide technology, which is integrated with the IBM Spectrum 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.

Other enhancements were made to the IP stack with Spectrum Virtualize code level 8.4.2. with the introduction of Multiple IPs/VLANs per port and Multiple IP Partnerships. For more information, see “IP improvements” on page 9.

1.5.19 IBM Spectrum Virtualize copy services

IBM Spectrum Virtualize supports the following copy services functions:

- Remote copy (synchronous or asynchronous)
- FlashCopy (PiT copy) and Transparent Cloud Tiering (TCT)
- HyperSwap

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

1.5.20 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 Spectrum Virtualize, Metro Mirror, and Global Mirror are the IBM branded terms for the functions that are synchronous remote copy and asynchronous remote copy.

Synchronous remote copy ensures that updates are committed at both the primary and the secondary volumes before the application considers the updates complete. Therefore, the secondary volume is fully up to date 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 Spectrum Virtualize 6.3. It uses change volumes that are associated with the primary and secondary volumes. These point 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 1-21 shows an example of this function where you can see the relationship between volumes and change volumes.

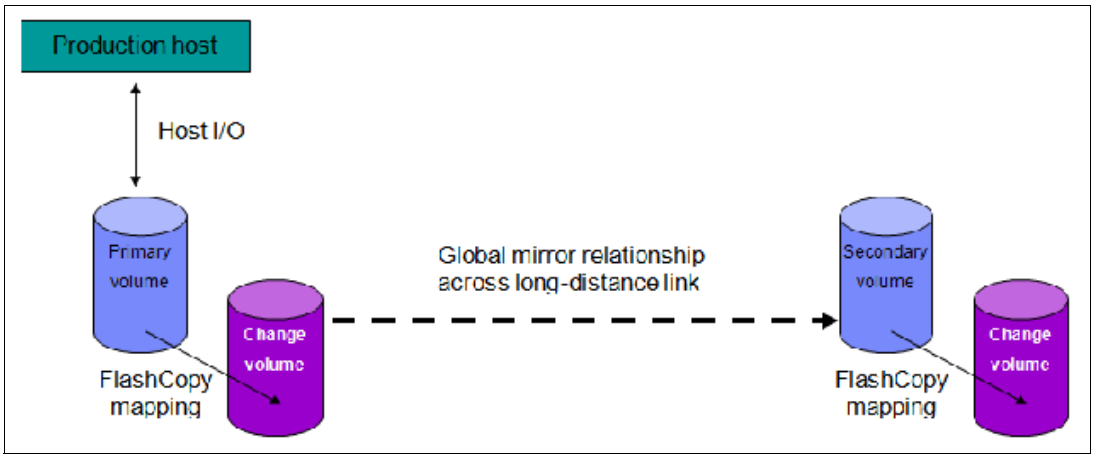


Figure 1-21 Global Mirror with change volumes

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.

1.5.21 FlashCopy and Transparent Cloud Tiering

FlashCopy and Transparent Cloud Tiering (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 (PiT). 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 (PiT) 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.

With IBM Spectrum 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 Spectrum Virtualize supports multiple targets and multiple rollback points.

Most clients aim to integrate the FlashCopy feature for PiT copies and quick recovery of their applications and databases. An IBM solution for this goal is provided by IBM Spectrum Protect. For more information, see this web page.

Transparent Cloud Tiering

IBM Spectrum Virtualize Transparent Cloud Tiering (TCT) is an alternative solution for data protection, backup, and restore 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 cloud, the volume snapshot is stored as object storage.

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

The management GUI provides an easy-to-use initial setup, advanced security settings, and audit logs that records all backup and restore to cloud.

To learn more about IBM Cloud Object Storage, see this web page.

HyperSwap

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

When a 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 you are using ownership groups 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 Spectrum Virtualize 3-Site Orchestrator.

1.6 Business continuity

In simple terms, 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.

1.6.1 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 there are power failures or site-wide outages.

Stretched Clusters are considered HA solutions because both sites work as instances of the production environment (there is no standby location). Combined with application and infrastructure layers of redundancy, Stretched Clusters can provide enough protection for data that requires availability and resiliency.

In Stretched Clusters 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.

1.6.2 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, no FC connectivity is required. 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 details and implementation guidelines about deploying Stretched Cluster or ESC, see *IBM Spectrum Virtualize and SAN Volume Controller Enhanced Stretched Cluster with VMware*, SG24-8211.

1.6.3 Business continuity with HyperSwap

The HyperSwap HA feature in the IBM Spectrum Virtualize software enables business continuity during hardware failure, power failure, connectivity failure, 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 Spectrum Virtualize software.

The HyperSwap feature provides HA volumes that are accessible through two sites at up to 300 km 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 itself 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 Spectrum Virtualize software.
- Uses intra-cluster synchronous remote copy (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 extra 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

1.6.4 Business continuity with three-site replication

Solutions were available for some time by using the IBM SAN Volume Controller stretched cluster to provide HA across two sites combined with Global Mirror or Global Mirror with Change Volumes to replicate to a third site. However this type of implementation requires manual intervention and custom scripts that increases management complexity.

Another possibility to implement a three-site replication solution was made available in limited deployments on code version 8.3.1 where data is replicated from the primary site to two alternative sites and in addition, the remaining two sites are aware of the difference between themselves.

This configuration ensures that if a disaster occurs at any one of the sites, the remaining two sites can establish a consistent_synchronized remote copy relationship among themselves with minimal data transfer; that is, within the expected RPO.

Spectrum Virtualize version 8.4 expands the three-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 alternate site. HyperSwap allows active-active configurations to maintain data availability, eliminating the need to failover if communications should be disrupted. This provides a more robust environment, allowing up to 100% uptime for data, and recovery options inherent to DR solutions.

To better assist with three-site replication solutions, IBM Spectrum Virtualize 3-Site Orchestrator coordinates replication of data for DR and HA scenarios between systems.

IBM Spectrum 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 Spectrum Virtualize products.

Figure 1-22 shows the two supported topologies for the three-site replication co-ordinated solutions.

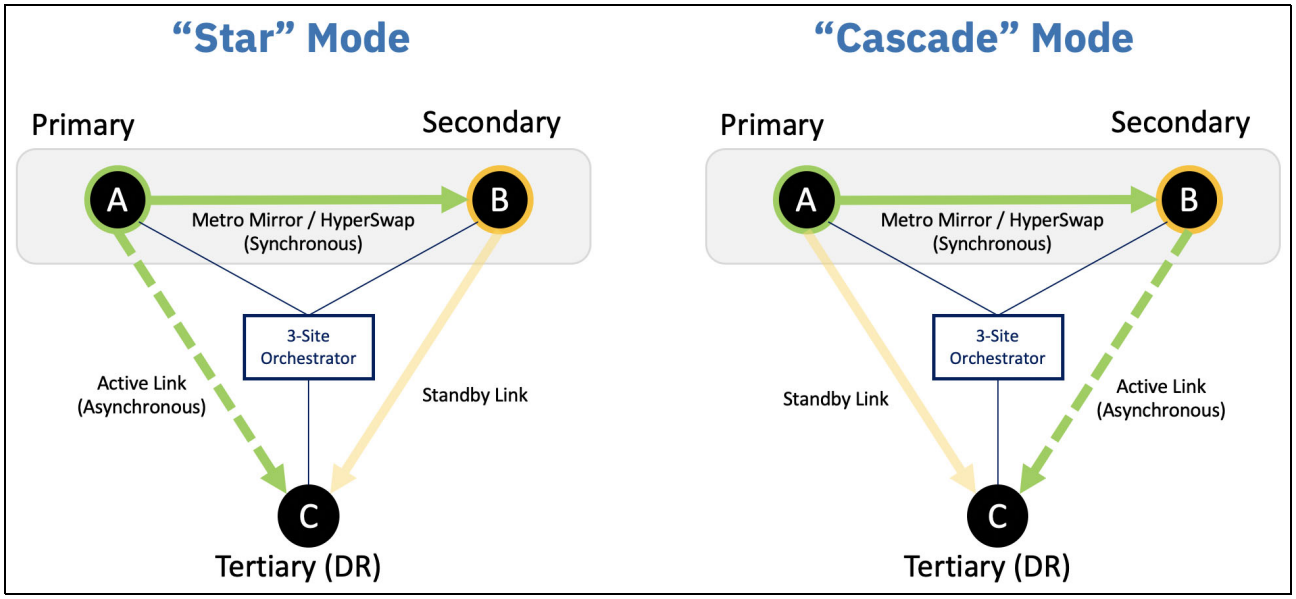


Figure 1-22 "Star" and "Cascade" modes in a three-site solution

1.6.5 Automatic hot spare nodes

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. Starting in version 8.2, the system can automatically take on the spare node to replace a failed node in a cluster or to keep the whole 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 and 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 is used by the system to determine suitable hot-spare nodes:

- Requires an exact match:
  - Memory capacity
  - Fibre Channel port ID
  - Compression support
  - Site
- Recommended to match, 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 Spectrum Virtualize virtualized storage ports to take over the job of the failed node 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 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 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, so the upgrades should be seamless for administrators 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 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.

1.7 Management and support tools

The IBM Spectrum Virtualize system can be managed through the included management software that runs on the IBM SAN Volume Controller hardware.

1.7.1 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 Spectrum 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 either 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 also can 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 SAN Volume Controller V8.4 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 SAN Volume Controller status and assists administrators with troubleshooting and fix activities. Remote Support Assistance is available at no extra charge, and no extra license is needed.

With Spectrum Virtualize code 8.4.2, some changes were made in the way Remote Support Assistance and the code download function of Remote Code Load upgrades are used and configured. For more information, see “Remote support interface improvements” on page 13.

For more information about setting up Remote Support Assistance, see this IBM Support web page.

1.7.2 Event notifications

IBM SAN Volume Controller can use SNMP traps, syslog messages, and a 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. Each event that IBM SAN Volume Controller detects is assigned a notification type of `Error`, `Warning`, or `Information`. You can configure the IBM SAN Volume 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 Spectrum Virtualize can send SNMP messages that notify personnel about an event. You can use an SNMP manager to view the SNMP messages that IBM Spectrum Virtualize sends. You can use the management GUI or the CLI to configure and modify your SNMP settings.

You can use the MIB file for SNMP to configure a network management program to receive SNMP messages that are sent by the IBM Spectrum 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 either Internet Protocol Version 4 (IPv4) or Internet Protocol Version 6 (IPv6).

IBM SAN Volume Controller can send syslog messages that notify personnel about an event. The event messages can be sent in either expanded or concise format. You can use a syslog manager to view the syslog messages that IBM SAN Volume Controller sends.

IBM Spectrum 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

Call Home notification improves 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 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 the IBM SAN Volume Controller is within the warranty period.

To send email, you must configure at least one SMTP server. You can specify as many as five more 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 use 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. You must 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 not using DNS but you have a firewall to protect your internal network from outside traffic, you need to enable certain IP addresses and ports to establish a connection to the support center. Ensure that your network firewall allows outbound traffic to the following IP addresses on port 443:

- 129.42.56.189
- 129.42.54.189
- 129.42.60.189

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 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 about the features and functions of Call Home methods, see this IBM Support web page.

1.8 Useful IBM SAN Volume Controller web links

For more information about IBM SAN Volume Controller-related topics, see the following web pages:

- IBM SAN Volume Controller Support
- IBM SAN Volume Controller Home
- IBM SAN Volume Controller Documentation

2

Planning

This chapter describes steps that are required to plan the installation and configuration of an IBM System Storage SAN Volume Controller in your storage network.

This chapter is *not* intended to provide in-depth information about the described topics; it provides only general guidelines. For more information, see *IBM System Storage SAN Volume Controller, IBM Storwize V7000,* and *IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

Note: Make sure that the planned configuration is reviewed by IBM or an IBM Business Partner before implementation. Such a review can both increase the quality of the final solution and prevent configuration errors that might impact the solution delivery.

This chapter includes the following topics:

- 2.1, “General planning rules” on page 60
- 2.2, “Planning for availability” on page 61
- 2.3, “Physical installation planning” on page 62
- 2.4, “Planning for system management” on page 63
- 2.5, “Connectivity planning” on page 64
- 2.6, “Fibre Channel SAN configuration planning” on page 65
- 2.7, “IP SAN configuration planning” on page 71
- 2.8, “Planning topology” on page 77
- 2.9, “Back-end storage configuration” on page 79
- 2.10, “Internal storage configuration” on page 80
- 2.11, “Storage pool configuration” on page 80
- 2.12, “Volume configuration” on page 86
- 2.13, “Host attachment planning” on page 89
- 2.14, “Planning copy services” on page 90
- 2.16, “Data migration” on page 93
- 2.17, “Ansible Automation for IBM SAN Volume Controller” on page 95
- 2.21, “Configuration backup procedure” on page 101

© Copyright IBM Corp. 2021. All rights reserved.

2.1 General planning rules

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 that makes you able to reach these goals.

Consider the following points when planning a system:

- Collect and document 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, percentage of sequential/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:
  - Assess the capacity and performance capabilities of the existing back-end storage that is present in the environment and intended to be virtualized by the SAN Volume Controller.
  - Plan and install new back-end storage if it will be deployed with the SAN Volume Controller.
  - Account 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)/RTO requirements and plan for high availability (HA) and Remote Copy (RC) functions. Decide whether you require a dual-site deployment with Enhanced Stretched Cluster (ESC) or HyperSwap feature, or a three-site deployment. In addition, decide 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 cyber attack.

    Note: Contact your IBM sales representative or IBM Business Partner to perform these calculations.

  - Ensure sufficient licensing is scoped to allow for data encryption or the implementation of Copy Service Manager software that is required by Safeguarded Copy.
- Define the number of input/output (I/O) groups (nodes). The number of I/O groups 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 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 of the volumes.
- Define a naming convention for the system nodes, volumes, and other storage objects.
- Plan a management Internet Protocol (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 Configuration Limits and Restrictions.

- 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 SAN Volume Controller 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 also are supported.

- Using spare nodes.

  You can purchase and configure a hot spare node to minimize the impact of hardware failures.

- 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

  The SAN Volume Controller uses three managed disks (MDisks) (external storage 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.

- Failure domain sizes

  A failure of a single Managed Disk (MDisk) within a storage pool, can take the whole storage pool offline. To reduce the impact 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. This configuration however 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 one of the SAN Volume Controller nodes, investment in a dual-rack configuration for placement of the second node is not justified. Any incident affecting the rack that holds the critical switch brings down the whole system no matter where the second SAN Volume Controller node is placed.

2.3 Physical installation planning

You must consider several key factors when you plan the physical site of a system. The physical site must have the following characteristics:

- Sufficient rack space must exist to install controller and disk enclosures.
- The site must meet the power, cooling, and environmental requirements.

For more information about the power and environmental requirements, see this IBM Documentation web page.

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, plan for a separate independent power source for each of the two redundant power supplies of a system enclosure.

When planning SAN cabling, make sure that your physical topology adheres to zoning rules and recommendations.

SAN Volume Controller physical installation and initial setup is performed by an IBM System Services Representative (IBM SSR).

2.4 Planning for system management

Each system's node has a *technician port*. It is a dedicated 1 gigabits per second (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.

For management, each system node requires at least one Ethernet connection. The cable must be connected to port 1, which is a 10 Gbps Ethernet port (it does not negotiate speeds that are less than 1 Gbps) on IBM SAN Volume Controller 2145-SV2 and 2145-SA2 nodes and a 1 Gbps Ethernet port on an IBM SAN Volume Controller 2145-SV1 node. For increased availability, an optional management connection can be configured over Ethernet port 2.

For configuration and management, you must allocate an IP address to each node, which is referred to as the *service IP address*. Internet Protocol Version 4 (IPv4) and Internet Protocol Version 6 (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 SAN Volume Controller I/O groups requires a minimum of five unique IP addresses: one for each node and one for the system as a whole.

Ethernet ports 1 and 2 are not reserved only for management. They can be also used for internet Small Computer Systems Interface (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, point a web browser to the cluster management IP address. To access the management CLI, point a Secure Shell (SSH) client to a cluster management IP 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 678.

2.4.1 User password creation options

IBM Spectrum Virtualize includes a password policy support feature with which system administrators can set security requirements. These requirements are related to password creation and expiration, timeout for inactivity, and actions after failed log on attempts.

Password policy support allows administrators to set security rules that are based on your organization's security guidelines and restrictions. Next, we describe the password and security-related rules that the system supports.

Password creation rules

Administrator 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.
- Require passwords to contain special characters.
- Prevent users from reusing recent passwords.
- Require users to change password on next login under any of the following conditions:
  - Their password expired.
  - An administrator created accounts with temporary passwords.
- Password History checking can be enabled
- The minimum required password age can be set to prevent bypassing the password history restriction by rapidly changing passwords multiple times.

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 password expiration limit.
- Set a password to expire immediately.
- Set number of failed login attempts before the account is locked.
- Set time 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 implementation of these features, see Chapter 4, “IBM Spectrum Virtualize GUI” on page 155.

2.5 Connectivity planning

An IBM SAN Volume Controller 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 (NVMe) over Fibre Channel (FC-NVMe) (also known as NVMe over Fabric (NVMe-oF))), and Fibre Channel over Ethernet (FCoE), and IP network technologies (iSCSI and iSCSI Extensions for Remote Direct Memory Access (RDMA) (iSER)). The connection options 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. 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**
NVMe over FC	**Yes**	No	No	No
FCoE	**Yes**	**Yes**	**Yes**	**Yes**
iSCSI	**Yes**	**Yes**	No	No[a]
iSER	**Yes**	No	**Yes**	No[a]

a. Replication traffic can be sent over an IP network with native IP replication, which can be configured on both onboard 10-Gigabit Ethernet (GbE) ports and optional 25 GbE ports.

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

2.6 Fibre Channel SAN configuration planning

Each IBM SAN Volume Controller 2145-SV1 node has four adapter slots, which can be used for the following FC adapters:

- 10 Gbps Ethernet FCoE adapter
- 16 Gbps FC adapter
- Two-port 32 Gbps FC adapter (A maximum of two adapters of this type are supported.)

Each IBM SAN Volume Controller 2145-SV2 and 2145-SA2 node has three adapter slots, which can be used for the following FC adapters:

- Four-port 16 Gbps FC adapter
- Four-port 32 Gbps FC adapter

16 Gbps and 32 Gbps adapters can be used for both SCSI FC and FC-NVMe attachment.

IBM SAN Volume Controller 2145-SV2 and 2145-SA2 nodes do not support FCoE attachment.

Note: It is recommended that a minimum of two adapters of any type be installed per IBM SAN Volume Controller node to use the CPU cores for optimal bandwidth and performance. This configuration also allows for the use of separate ports for dedicated traffic.

2.6.1 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 this IBM Documentation web page.

For more information about the planning and topology requirements for three-site replication configurations, see *Spectrum Virtualize 3-Site Replication*, SG24-8504.

2.6.2 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.

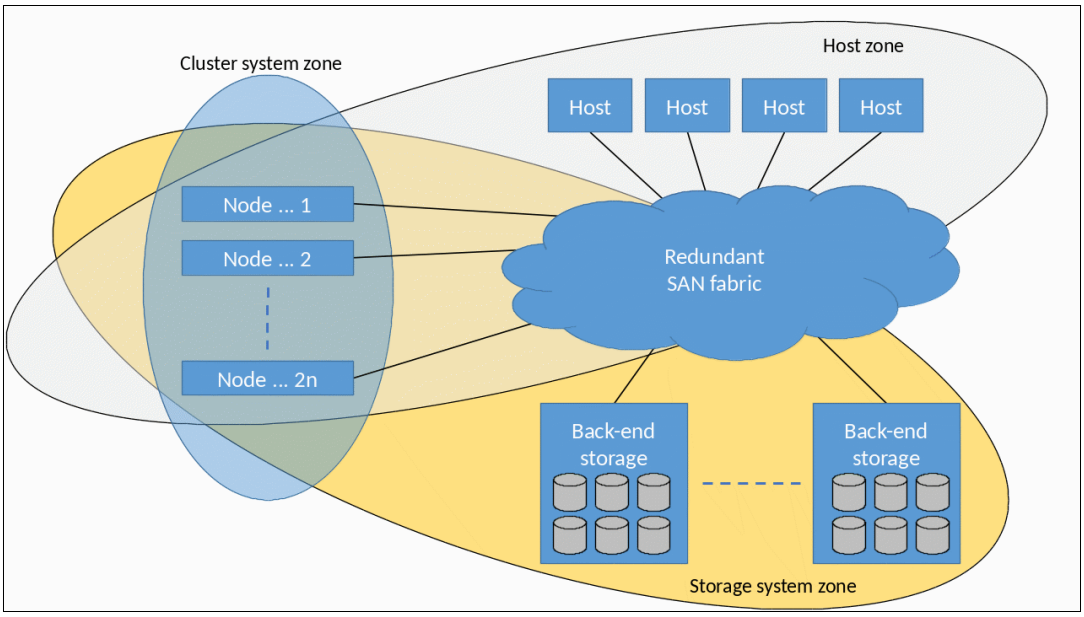


Figure 2-1 System zoning

The fundamental rules of system zoning are described in the rest of this section. However, you must review the latest zoning guidelines and requirements when designing zoning for the planned solution. For more information, see this IBM Documentation web page.

2.6.3 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 system physical FC port. This WWPN is available only for host connection. 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 7.4, “N_Port ID Virtualization support” on page 447.

Ensure that FC switches enable each physically connected system port with the ability to create four extra NPIV ports.

When performing zoning configuration, virtual WWPNs are used only for host communication, that is, “system to host” zones must include virtual WWPNs, and 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 this IBM Documentation web page.

2.6.4 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.

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. In each of them, place a single port per node 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 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.

2.6.5 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 put target ports and multiple initiator ports in a single zone for easy of management but act he same as though each initiator and target are configured in isolated zones. The use of these zoning techniques are supported for host attach and for storage virtualization. We recommend the use of normal zones when configuring ports for clustering or for replicating because these functions require the port to be an initiator and a target.

For more information about connecting back-end storage systems, see the following IBM Documentation web pages:

- External storage system configuration details (Fibre Channel)
- External storage system configuration with Fibre Channel connections

2.6.6 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 is mandatory.

Note: Cisco Smart Zoning and Brocade Peer Zoning are supported, with which you put target ports and multiple initiator ports in a single zone for easy of management but act exactly the same as though each initiator and target are configured in isolated zones. The use of these zoning techniques are supported for both host attach and for storage virtualization. We recommend the use of normal zones when configuring ports for clustering or for replication, as these functions require the port to be both an initiator and a target.

For smaller installations, you can have up to 40 FC ports (including both host HBA ports and the system's virtual WWPNs) in a host zone if the zone contains similar HBAs and operating systems (OSs). A valid zone can be 32 host ports plus eight system ports.

FC-NVMe applies more 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 either 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 3.4.5, "Configuring the local Fibre Channel port masking" on page 139.

- Balance the host load across the system's ports. For example, zone the first host with ports 1 and 3 of each node in 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 should 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.

2.6.7 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, exactly 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 at this IBM Documentation web page.

When designing zoning for a geographically dispersed solution, consider the effect of the cross-site links on the performance of the local system.

Using 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 2.6.8, "Port designation recommendations" on page 69.

For more information about zoning best practices, see *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

2.6.8 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 affect the system.

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. Nodes in one I/O group use it to synchronize write cache.

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.

To isolate both node-to-node and system-to-system traffic, use the port designations that are shown in Figure 2-2.

Adapter / Port	4 ports	8 ports	12 ports	16 ports	SAN Fabric
Adapter 1 Port 1	Host and Storage	Host and Storage	Host and Storage	Host and Storage	A
Adapter 1 Port 2	Host and Storage	Host and Storage	Host and Storage	Host and Storage	B
Adapter 1 Port 3	Intracluster and Replication	Intracluster	Intracluster	Intracluster	A
Adapter 1 Port 4	Intracluster and Replication	Intracluster	Intracluster	Intracluster	B
Adapter 2 Port 1	-	Host and Storage	Host and Storage	Host and Storage	A
Adapter 2 Port 2	-	Host and Storage	Host and Storage	Host and Storage	B
Adapter 2 Port 3	-	Intracluster or Replication	Replication or Host and Storage	Replication or Host and Storage	A
Adapter 2 Port 4	-	Intracluster or Replication	Replication or Host and Storage	Replication or Host and Storage	B
Adapter 3 Port 1	-	-	Host and Storage	Host and Storage	A
Adapter 3 Port 2	-	-	Host and Storage	Host and Storage	B
Adapter 3 Port 3	-	-	Intracluster	Intracluster	A
Adapter 3 Port 4	-	-	Intracluster	Intracluster	B
Adapter 4 Port 1	-	-	-	Host and Storage	A
Adapter 4 Port 2	-	-	-	Host and Storage	B
Adapter 4 Port 3	-	-	-	Replication or Host and Storage	A
Adapter 4 Port 4	-	-	-	Replication or Host and Storage	B
localfcportmask	1100	11001100 or 00001100	110000001100	0000110000001100	-
remotefcportmask	1100	00000000 or 11000000	000011000000	1100000011000000	-

- Host refers to host objects defined in the system.
- Replication refers to nodes that are part of a different cluster.
- Storage refers to controller objects defined in the system if external storage is being used.
- Intracluster refers to nodes within the same cluster.
- The word "and" indicates that both types are used.
- The word "or" indicates that one of the options must be selected. If using replication, preference should be given to replication.

Figure 2-2 Port masking configuration

To isolate traffic, 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 103.

For more information about Port Masking, see Spectrum Virtualize Port Masking Explained.

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 MTBF 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.

2.7 IP SAN configuration planning

IBM SAN Volume Controller 2145-SV1, 2145-SV2 and 2145-SA2 nodes are equipped with four on-board 10 Gbps Ethernet network interface ports. They can operate at link speeds of 1 Gbps and 10 Gbps. Any of those ports can be used for host I/O with iSCSI, 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 SV1, SV2, and SA2 nodes can also be configured with 2-port 25 Gbps Remote Direct Memory Access (RDMA)-capable Ethernet adapters. The maximum number of adapters depends on the system hardware type. Adapters can auto-negotiate link speeds of 1 - 25 Gbps. All their ports can be used for host I/O with iSCSI or iSER, external storage virtualization with iSCSI, node-to-node traffic, and for IP replication.

Note: It is recommended that a minimum of two adapters of any type be installed per IBM SAN Volume Controller node to use the CPU cores for optimal bandwidth and performance. This configuration also separates ports for dedicated traffic use.

IBM SAN Volume Controller supports the for 10 Gbps Finisar SFPs (Finisar FTLX8574D3BCL) that are 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 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.

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.

2.7.1 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.

Each physical Ethernet port can have a maximum of 64 IP addresses with each IP on a unique portset. However, each port 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 or iSER), 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 contains IP addresses from every I/O group. To access multiple nodes in a system, a portset must be configured with the IP address 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. The 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 have the following requirements:

- 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 IP address that can be routed (gateway specified). This 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 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.

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. The Portset 0, Portset 3, and replication portset always are globally owned and only global administrators can assign and modify IP addresses to the portsets.

In a typical configuration, a portset object is first 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 limitations of portsets, see this IBM Documentation web page.

2.7.2 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 either 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, providing faster connection and processing time than traditional iSCSI connections.

iSER requires optional 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. So, there are two types of 25 Gbps adapters that are available for a system, and they cannot be interchanged without a similar RDMA type change on the host side.

Either iSCSI or iSER works for standard iSCSI communications, that is, ones that do not use RDMA.

The 25 Gbps adapters come with 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 SAN Volume Controller supports the 10 Gbps Finisar SFPs (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.

2.7.3 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 certain classes of traffic on ports, thus providing better bandwidth for more important traffic. The system supports PFC on various supported Ethernet-based protocols on three types of traffic classes: system (node-to-node), host attachment, and back-end 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.

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 the all 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 these steps, check your switch documentation for details.
- 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 are planning on using the same port for different types of traffic, ensure that ETS settings are configured in the network.

For more information, see this IBM Documentation web page.

2.7.4 RDMA clustering

IBM SAN Volume Controller can use 25 Gbps adapters for node-to-node traffic. A dual-site HyperSwap configuration also can use the adapters for an inter-site link.

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.

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. Due to this limitation, do not exceed the maximum of four ports for node-to-node communications.
- 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 contain 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, use the **addnodediscoverysubnet**, **rmnodediscoverysubnet**, or **lsnodediscoverysubnet** commands.

For more information, see this IBM Documentation web page.

Note: Before you configure a system that uses RDMA-capable Ethernet ports for node-to-node communications in a standard, ESC, or HyperSwap topology system, contact your IBM representative.

2.7.5 iSCSI back-end storage attachment

IBM SAN Volume Controller supports the virtualization of external storage systems that are attached through iSCSI. On-board 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 using any RDMA capabilities.

Consider the following items when planning for iSCSI virtualization:

- A one-to-one mapping of source ports to target ports is required.
- Direct attachment between the system and external storage systems is not supported, and 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 both initiator and target nodes support more ports.

All planning and implementation aspects of external storage virtualization with iSCSI are described in detail in *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

2.7.6 IP network host attachment

You can attach the system to iSCSI or iSER hosts by using the Ethernet ports of the systems.

The same ports can be used for iSCSI and iSER host attachment concurrently, but a single host can establish either an iSCSI or iSER session, but not both.

iSCSI or iSER hosts connect to the system through the 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 both its own node-port IP addresses and also 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 also 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 reconfiguring all iSCSI hosts that communicate with the system.

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 in Spectrum Virtualize 8.4.2, 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 (iSCSI or iSER), back-end storage connectivity (iSCSI only), and IP replication. A host can access storage only through IP addresses 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.

Host portsets have following requirements:

- Portsets can have a maximum of four IP addresses per node.
- A single portset can contain IPv4, IPv6, or 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 nodes on that system.

For more information about iSCSI host attachment, see *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

ensure that iSCSI initiators, host iSER adapters, and Ethernet switches that are attached to the system are supported by using SSIC.

2.7.7 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 using FC configurations.

IP replication is supported on both onboard 10 Gbps Ethernet ports and optional 25 Gbps Ethernet ports. However, when configured over 25 Gbps ports, it does not use RDMA capabilities, and it does not provide a performance improvement 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.

With Spectrum Virtualize 8.4.2, a new portset feature is introduced that allows for multiple IP partnerships per IBM SAN Volume Controller. Up to three partnerships per IBM SAN Volume Controller system is allowed. A system can be part of only two IP partnerships.

IBM SAN Volume Controller systems with pre-v8.4.2.0 firmware are still limited to one IP partnership. Because partnerships on low memory platform nodes share memory resources, their performance might be denigrated.

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 each node). Similarly, for the remote system with four nodes, a portset on that system also can 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, 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.

Specific intersite link requirements must be met when you are planning to use IP partnership for Remote Copy. For more information about these requirements, see the following resources:

- This IBM Documentation web page
- Chapter 10, “Advanced Copy Services” on page 605.

2.7.8 Firewall planning

After you have your IP network planned, set up the suitable firewall rules for each data flow.

For more information about required network flows, see this IBM Documentation web page.

2.8 Planning topology

In this section, we discuss planning topology.

2.8.1 High availability

IBM SAN Volume Controller supports two dual-site topologies: enhanced stretched cluster (ECS) and HyperSwap. The key differences between these topologies are listed in Table 2-2.

Table 2-2 Differences between Enhanced Stretched Cluster and HyperSwap

Item	Enhanced stretched cluster	HyperSwap
Minimum number of I/O groups that are required.	1	2
Independent copies of data that are maintained.	2	2 (Four if volume mirroring to two pools in each site is configured.)
Cache that is retained if only one site is online.	No	Yes
Stale consistent data is retained during resynchronization for DR.	No	Yes
Ability to use MM, GM, or GM together with an HA solution.	Yes	Yes
Maximum HA volume count.	7932	2000
Licensing.	Included in base product.	Requires a Remote Mirroring license.

The topologies differ in how the nodes are distributed across the sites:

- 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 enables 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 Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317
- *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices,* REDP-5597
- *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

2.8.2 Three-Site replication

IBM SAN Volume Controller supports a three-site replication topology, which includes three-site replication with HyperSwap or three-site replication with Metro Mirror configurations.

With the three-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 operations with minimal data transfer. The remote-copy relationships are synchronous or asynchronous, depending on which site failed.

The three-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 at all times.

For more information, see *Spectrum Virtualize 3-Site Replication*, SG24-8504.

2.9 Back-end storage configuration

External back-end storage systems (also known as *controllers*) provide their logical volumes (LUs), which are detected by a system as managed disks (MDisks) and can be used in storage pools to provision their capacity to system's hosts.

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, 7000, and 9000 family) can be direct-attached by using FC.
- For migration purposes, all back-end storage systems are supported by 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 general practice is that:
  - EMC DMX/SYMM, all HDS, and SUN/HP HDS clones use one 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 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.
- The IBM SAN Volume Controller must have exclusive access to every LU that is provisioned to it from a back-end controller. Any specific LU cannot be presented to more than one system. Presenting the same back-end LU to a system and a host is not allowed.
- Data reduction (compression and deduplication) on the back-end controller is supported with only a limited set of IBM Storage systems.

In general, configure back-end controllers as though they are used as stand-alone systems. However, there might be specific requirements or limitations 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 this IBM Documentation web page.

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 (the system itself) and the lower (back-end controller) level 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.

2.10 Internal storage configuration

IBM SAN Volume Controller SV1 models can be equipped with an optional SAS-attached disk expansion enclosure. For general-purpose storage pools with various I/O applications, 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.

An array-level recommendation for all types of internal storage is distributed redundant array of independent disks (DRAID) 6 (DRAID 6). DRAID 6 outperforms other available redundant array of independent disks (RAID) levels in most applications while providing fault tolerance and high rebuild speeds.

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

For more information about internal storage configuration, see *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

2.11 Storage pool configuration

The storage pool is at the center of the many-to-many relationship between the internal drive arrays or externally virtualized logical unit numbers (LUNs), which are represented as Managed Disks (MDisks), and the volumes. It acts as a container of physical disk capacity from which chunks of MDisk space, which is known as *extents*, are allocated to form volumes that are presented to hosts.

The system supports two types of pools: data reduction pools (DRP) and standard pools. The type is configured when a pool is created and it cannot be changed later. The type of the pool determines the set of features that is available on the system.

The following features can be implemented only with standard pools:

- VMware vSphere integration with VMware vSphere Virtual Volumes (VVOLs)
- IBM SAN Volume Controller 2145-SV1 only: Real-time Compression (Random Access Compression Engine (RACE))

The following features can be implemented with DRPs only:

- Automatic capacity reclamation with SCSI UNMAP (This feature returns capacity that is marked as no longer used by a host back to storage pool. If you have back-end storage with data reduction features, you also benefit from the UNMAP, but you do not see it in the IBM SAN Volume Controller pool.)
- DRP compression (in-flight data compression)
- DRP deduplication
- FlashCopy with redirect-on-write

Note: FlashCopy with redirect-on-write is usable for volumes with supported deduplication only without mirroring relationships and within the same pool and I/O group. Automatic mode selection (RoW/CoW) is based on these conditions.

In addition to providing data reduction options, a DRP amplifies the I/O and CPU workload, which is accounted for during performance sizing and planning.

IBM Spectrum Virtualize provides a Comprestimator Always On feature, whereby the continuous comprestimation of all VDisks are provided so that compress-ability estimations are always available. This feature is enabled by default.

Another base storage pool parameter is the extent size. There are two implications of a storage pool extent size:

- 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 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 must 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 2-3 lists the minimum DRP capacity that is required to create a volume within the pool.

Table 2-3 Minimum overhead capacity requirements for DRPs

Extent size (in gigabytes)	Overhead capacity requirements (in terabytes)[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 Configuration Limits and Restrictions.

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, there is no way to encrypt it later. Your only option is to delete it and re-create it as encrypted.

When planning storage pool layout, consider the following aspects:

- 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. Using 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.
- Pool performance:
  - Do not mix same-tier arrays or MDisks with different performance characteristics in one 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 IBM 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.

2.11.1 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 VMware vSphere Virtual Volumes, sometimes referred to as VVols, that are used in VMware vCenter and 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 only the capacity details for child and migration pools.
- Child pools can be created and changed with the command-line or GUI interface.
- 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, standard child pools can be created to migrate volumes in a non-encrypted pool to encrypted child pools. 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 non-encrypted 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 to less than 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 from 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 Chapter 13, “Reliability, availability, serviceability, monitoring and troubleshooting” on page 853.

With Spectrum Virtualize 8.4.2, 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.

The system integrates with IBM Copy Services Manager to provide automated backup copies and data recovery. 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 pools 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 Administrator privileges or less cannot change or delete any pools that contain Safeguarded backups.

2.11.2 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 2-4 lists the limits of the write-cache data that can be used by a single storage pool.

Table 2-4 Limit 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.

2.11.3 Provisioning policies

With Spectrum Virtualize 8.4.2, you can configure provisioning policies to simplify managing capacity saving for volumes and pools. 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:

- Drive-based (`-capacitysaving drivebased`)

  Use this capacity saving option to indicate that volumes must 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 (`-capacitysaving thin`)

  Use this 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 (`-capacitysaving compressed`)

  Use this 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 (`-deduplicated`)

  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` flag when you create a provisioning policy to add deduplication to your capacity saving for a DRP.

2.12 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, “Creating and provisioning volumes” on page 331.

2.12.1 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.

2.12.2 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, the data is still reduced 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.

2.12.3 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.

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 using thin-provisioned volumes, 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.

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 back end when DRPs are implemented.

2.12.4 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, compressed volumes must belong to DRPs. Also, systems with IBM SAN Volume Controller 2145-SV1 nodes support RACE compression for volumes on standard pools.

Before implementing compressed volumes, analyze the data to know your average compression ratio and ensure that performance sizing was done for compression.

IBM Spectrum Virtualize provides a Compretimation Always On feature, which ensures that the continuous comprestimation of all VDisks is provided and always available. This feature is `On` by default.

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.

2.12.5 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 is 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 an existing signature that is stored on the system, then 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, and it might improve the performance of read operations to data that has an existing 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, then 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.

Before implementing deduplication, analyze the data to estimate deduplication savings and make sure 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 the following IBM Support web pages:

- IBM Data Reduction Estimator Tool (DRET) for SVC, Storwize and FlashSystem products
- 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 Host attachment planning

The system supports the attachment of a various host hardware types that are running different operating systems by using FC SAN or IP SAN. For more information that is specific to your host setup, see this IBM Documentation web page.

2.13.1 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 correct host queue depth for your environment, see this IBM Documentation web page.

2.13.2 SAN boot support

The system supports SAN boot or startup for selected configurations of hosts running AIX, Microsoft Windows, and other OSs. To check whether your configuration is supported for SAN boot, see the SSIC.

2.13.3 Planning for large deployments

Each I/O group can have up to 512 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 512 hosts to access the storage, you must divide them into groups of 512 hosts or less and map each group to a single I/O group only. With this approach, you can configure up to 2048 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 only to a host that is associated with the I/O group to which the volume belongs.

2.13.4 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.

The system supports 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.

UNMAP processing can be controlled with two separate settings:

- One setting enables hosts to send **UNMAP** commands,
- The other setting controls whether the system sends **UNMAP** commands to back-end storage (drives and external controllers).

By default, back-end UNMAP is on, and it is a best practice to keep it turned on for most use cases. Host UNMAP is off by default.

Enable host UNMAP if data reducing storage is used that is virtualized by the IBM SAN Volume Controller (for example) a FlashSystem 9200 system with self-compressing drives, to benefit from end-to-end UNMAP support.

Consider turning off host UNMAP if you do not use DRPs and use HDDs as a back end.

However, if the system has a virtualized slow back-end controller (for example, back-end storage running nearline (NL) SAS drives for the lowest tier), UNMAP requests that are sent from the host as large I/O chunks might overload the back end. Thorough planning is required if you plan to virtualize a mixture of slow back-end controllers and data reducing (compressing) back-end controllers.

2.14 Planning copy services

IBM SAN Volume Controller offers 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 605.

2.14.1 FlashCopy guidelines

With the IBM FlashCopy function of IBM Spectrum Virtualize, 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 that is less than the host operating system and cache, so those levels must be flushed by the OS for a FlashCopy copy to be consistent.

When you use the FlashCopy function, observe the following guidelines:

- Both 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 redirect-on-write uses the DRP’s 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 redirect-on-write is usable for volumes with supported deduplication only without mirroring relationships and within the same pool and I/O group. Automatic mode selection (RoW/CoW) is based on these conditions.

With Spectrum Virtualize 8.4.2, you can now 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 feature 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 *IBM System Storage SAN Volume Controller, IBM Storwize V7000,* and *IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

2.14.2 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. The systems can be up to 300 kilometers (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, if a failure occurs, you might lose data.

GM uses one of two methods to replicate data: a non-cycling mode, and a multi-cycling mode, which also is known as GM with change volumes. Multicycling GM is designed to replicate data while adjusting for bandwidth constraints. It is appropriate for environments where it is acceptable to lose a few minutes of data if a failure occurs.

For environments with higher bandwidth, non-cycling 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 you are going 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 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 using 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.

  Using 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 Throttles

Throttles is a mechanism that controls 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 can access the volume allows 100 IOPS for that volume. Any I/O operation that exceeds the throttle limit are 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 be used with virtual volumes.

You can set throttles for 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 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.

2.16 Data migration

Data migration is an important part of an implementation, so 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.
- Move a workload off old or failing storage ahead of decommissioning it.
- Move a workload to rebalance a changed load pattern.
- Migrate data from an older disk subsystem.
- Migrate data from one disk subsystem to another one.

Because multiple data migration methods are available, choose the method that best fits your environment, OS 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 OS, for example, by using the system's logical volume manager (LVM)
- Based on specialized data migration software
- Based on the system data migration features

For more information about system data migration tools, see Chapter 8, "Storage migration" on page 527, and 10.7, "Volume mirroring and migration options" on page 691.

With data migration, apply the following guidelines:

- Choose the data migration method that best fits your OS 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 ahead of time 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. Especially for a large or critical data migration, have the plan peer-reviewed and formally accepted by an appropriate technical design authority within your organization.
- 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 non-virtualized storage subsystem to another non-virtualized 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 are using slower disk subsystems for the secondary volumes for high-performance primary volumes because the system's cache might not be able to buffer all the writes. Flushing cache writes to slower back-end storage might impact performance of your hosts.
- Consider storage performance. The migration workload might be much higher than expected during normal operations of the system. If there is already application data on the system to which you are migrating, the application performance might suffer if the system is overloaded. Consider using host or volume level throttles when performing migration on a production environment.

With Spectrum Virtualize 8.4.2, you can now use a remote-copy relationship type, called *Non-Disruptive System Migration*, to migrate data from one IBM Spectrum 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 located 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. Non-Disruptive System Migration does not require a Remote Mirroring license, but 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 include 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.

- 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.

> **Note:** If ODX is disabled, it cannot be re-enabled on the 8.4.2 software. As of this writing, ODX support is not available on 8.4.2.

- 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.
- A Fibre Channel 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. Ensure that the partnership has 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 for nondisruptive system migration that are listed in Table 2-5.

Table 2-5 Supported hosts and connections types for nondisruptive system migration.

Supported hosts	Connection types
VMware ESXi 6.x	iSCSI and Fibre Channel
VMware ESXi 7.x	iSCSI, iSER, and Fibre Channel
RHEL 7.x, RHEL 8.x	iSCSI and Fibre Channel
SLES 12.x, SLES 15.x	iSCSI and Fibre Channel
Solaris 10, Solaris 11	Fibre Channel
HP-UX 11iV3	Fibre Channel

2.17 Ansible Automation for IBM SAN Volume Controller

IBM SAN Volume Controller includes integration with Red Hat Ansible Automation Platform, with which you can create an Ansible Playbook that automates the tasks that are repeated across your organization in a consistent way. This feature helps improve outcomes and reduces risk.

With IBM SAN Volume Controller and Red Hat Ansible Automation Platform, you can easily automate tasks, such as configuration management, provisioning, workflow orchestration, application deployment, and life-cycle management.

IBM as a Red Hat Certified Support Module Vendor provides simple management for the storage provisioning commands that are used in the IBM Spectrum Virtualize Ansible Collection:

- Collect facts: Gather array information, such as hosts, host groups, snapshots, consistency groups, and volumes.
- Manage:
  - Hosts: Create, delete, or modify hosts.
  - Volumes: Create, delete, or extend the capacity of volumes.
  - MDisks: Create or delete a managed disk.
  - Pools: Create or delete a pool (managed disk group).
  - Volume maps: Create or delete a volume map.
  - Consistency group snapshots: Create or delete consistency group snapshots.
  - Snapshots Create or delete snapshots.
  - Volume clones: Create or delete volume clones.
  - Replication: Create or delete replication relationships.

For more information about IBM Spectrum Virtualize Ansible modules, see Ansible Automation Hub.

For more information, see the *Automate and Orchestrate Your IBM FlashSystem Hybrid Cloud with RedHat Ansible,* REDP-5598.

2.18 Container Storage Integration

IBM SAN Volume Controller implements 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) that are running in Red Hat OpenShift Container Platform or Kubernetes clusters.

Storage provisioning can be fully automatized with support for cluster orchestration systems to automatically deploy, scale, and manage containerized applications.

The CSI driver requires Port 22 to be opened on the worker nodes operating system firewall.

Be sure that multipathing is installed and running.

For more information, see IBM block storage CSI driver.

For more information about implementing the IBM block storage CSI driver with IBM SAN Volume Controller, see 7.9, “Container Storage Interface drivers” on page 526.

2.19 Safeguarded Copy function

The Safeguarded Copy function isolates backup copies from production data. This isolation is necessary so that if a cyberattack occurs, you can quickly recover and restore data from Safeguarded copies.

The Safeguarded Copy function, which is available with IBM Spectrum Virtualize software 8.4.2, is supported by the following products:

- FlashSystem 9200
- FlashSystem 9100
- FlashSystem 7200
- FlashSystem 5200
- IBM SAN Volume Controller with FlashCopy license
- Storwize V7000 Gen3 with FlashCopy license
- FlashSystem 5100 with FlashCopy license
- Storwize V5100 with FlashCopy license

The Safeguarded Copy function uses IBM Copy Services Manager as an external scheduling application. IBM Copy Services Manager coordinates and automates Safeguarded Copy function across multiple systems.

IBM Copy Services Manager uses a Safeguarded policy to configure FlashCopy mapping and consistency groups automatically to create backup copies. When Safeguarded backups are created, IBM Copy Services Manager uses the retention time for the Safeguarded backups based on the settings in the Safeguarded policy. After copies expire, the IBM Spectrum Virtualize software deletes the expired copies from the Safeguarded backup location.

IBM Copy Services Manager queries the system every 5 minutes to process Safeguarded policies. The start time that is defined in the Safeguarded policy must factor in the possible 5-minute delay. When IBM Copy Services Manager detects a new Safeguarded policy for a volume group, it creates the session and scheduled task to create and manage the Safeguarded backups.

2.19.1 Prerequisites

Before you can configure the Safeguarded Copy function on your system, ensure that the following prerequisites are met:

- IBM Copy Services Manager:
  - If you do not have a IBM Copy Services Manager license, purchase the IBM Copy Manager for IBM Spectrum Virtualize license, which includes IBM Copy Services Manager version 6.3.0.1. This license option is available through iERP/AAS, Passport Advantage®, or your IBM Sales team.
  - If you have a license for IBM Copy Services Manager, download IBM Copy Services Manager version 6.3.0.1 at Latest Downloads for IBM Copy Services Manager.

> **Note:** If you use an existing license, ensure that the licensed capacity is adequate for use of the Safeguarded Copy function. If you need more capacity for Safeguarded Copy function, contact you IBM sales representative to update your licensed capacity for IBM Copy Services Manager.

  - After you download IBM Copy Services Manager, complete the instructions for your installation. IBM Copy Services Manager supports several installation options on different environments.
  - During installation, license files can be imported for IBM Copy Services Manager. If the license was not imported during the installation, you must apply the license to the installation.
- System:
  - All systems must be running the 8.4.2 or later release.
  - For existing systems, ensure that you completed capacity planning for Safeguarded Copy function.

For more information about the implementation of the Safeguarded Copy function, see *Implementation Guide for SpecV/FlashSystem Safeguarded Copy*, REDP-5654.

2.20 Performance monitoring with IBM Storage Insights

IBM Storage Insights is integral to monitoring and ensuring the continued availability of the system.

Available at no additional charge, the cloud-based IBM Storage Insights product provides a single dashboard that provides a clear view of all your IBM block storage. You can make better decisions by seeing trends in performance and capacity.

With storage health information, you can focus on areas needing attention and 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.

IBM Storage Insights provides a unified view of 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 occurring with your storage.

IBM Storage Insights provides advanced customer service with an event filter that provides the following functions:

- The ability for you and 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 starts looking into the problem. This feature can reduce the time to solve the case by as much as 50%.

Figure 2-3 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.

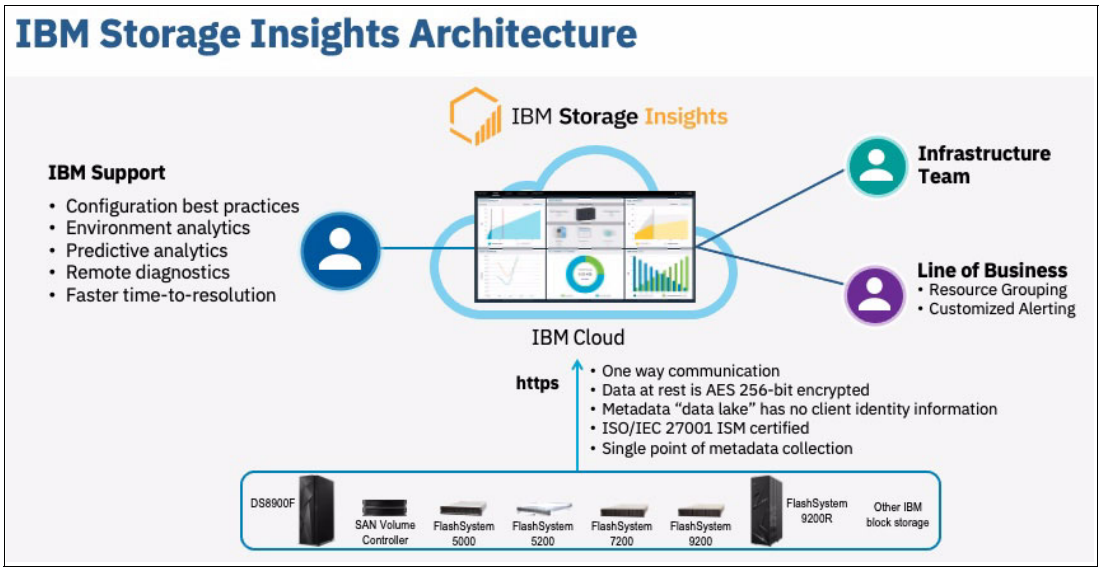


Figure 2-3 IBM Storage Insights architecture

IBM Storage Insights provides a lightweight data collector that is deployed on a Linux, Windows, or AIX server or a guest in a virtual machine (VM) (for example, a VMware guest).

The data collector streams performance, capacity, asset, and configuration metadata to your IBM Cloud instance.

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 Information Security Management certified.

To monitor storage systems, you must provide a user name and password to log in to the storage systems. The role or user group that is assigned to the user name must have the appropriate privileges to monitor the data that is collected. As of Spectrum Virtualize 8.3.1.2 and SI/ Spectrum Control 5.3.7 or later, data collection can be done with Monitor (least privileged) role.

Figure 2-4 shows the data flow from systems to the IBM Storage Insights cloud.

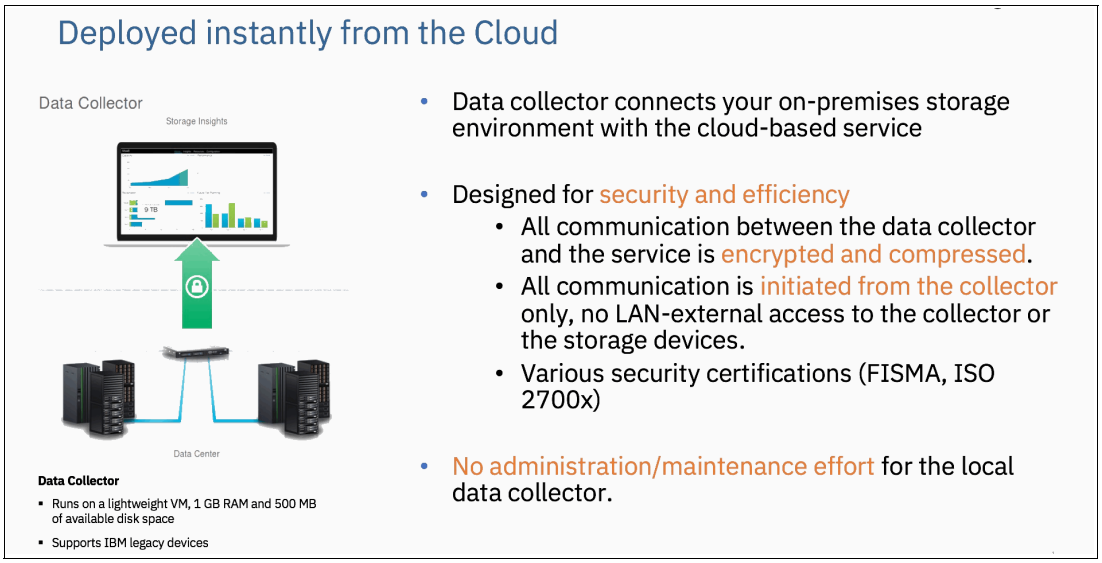


Figure 2-4 Data flow from the storage systems to the IBM Storage Insights cloud

Metadata about the configuration and operations of storage resources is collected, such as:

- Name, model, firmware, and type of storage system.
- Inventory and configuration metadata for the storage system's resources, such as volumes, pools, disks, and ports.
- Capacity values, such as capacity, unassigned space, used space, and the compression ratio.
- Performance metrics, such as read and write data rates, I/O rates, and response times.

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 day-to-day operation and maintenance of IBM Cloud instances.
- IBM Support for investigating and closing service tickets.

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 13.11, "IBM Storage Insights Monitoring" on page 920.

2.21 Configuration backup procedure

Save the configuration before and after any major configuration changes on 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. You must back up your application data by using the suitable backup methods.

For more information, see Chapter 13, “Reliability, availability, serviceability, monitoring and troubleshooting” on page 853.

3

Initial configuration

This chapter describes the initial configuration of the IBM SAN Volume Controller systems, provides step-by-step instructions about how to do the initial setup, 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 104
- 3.2, “System initialization” on page 105
- 3.3, “System setup” on page 109
- 3.4, “Base configuration” on page 124
- 3.5, “Configuring management access” on page 141

© Copyright IBM Corp. 2021. All rights reserved.

3.1 Prerequisites

Note: SAN Volume Controller is installed by an IBM System Services Representative (IBM SSR). You must provide all the necessary information to the IBM SSR by filling out the planning worksheets, which can be found at the IBM Documentation web page.

After the IBM SSR completes their task, continue the setup process by following the instructions in 3.3, “System setup” on page 109.

Before initializing and setting up the IBM SAN Volume Controller, ensure that the following prerequisites are met:

- The physical components fulfill all the requirements and are correctly installed, including:
  - The control enclosures 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 will use them.
  - The SAN Volume Controller nodes and optional expansion enclosures are powered on.

    Note: IBM SAN Volume Controller nodes need plenty of time to charge the batteries. How long it takes to recharge depends on how long it was waiting idle in stock and not 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 IBM Documentation.
- You have the required information, which can be found at IBM Documentation, 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 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 particular functions (depending on the system type):
    - Remote Copy (RC)
    - External Virtualization
    - IBM FlashCopy
    - Compression
    - Encryption
  - 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.
  - (optional) The Network Time Protocol (NTP) server IP address.

– (optional) 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.
– (optional) The IP addresses for Remote Support Proxy Servers, which are required only if you want to use them with the Remote Support Assistance feature.

3.2 System initialization

This section provides step-by-step instructions about how to create the SVC cluster.

To start the initialization procedure, connect a desktop PC or a notebook to the technician port. The *technician port* is a dedicated 1 Gb Ethernet port at the rear of each of the nodes in the control enclosure. It can be use 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.

The location of a technician port on the SAN Volume Controller 2145-SV2 and 2145-SA2 nodes is show on Figure 3-1.

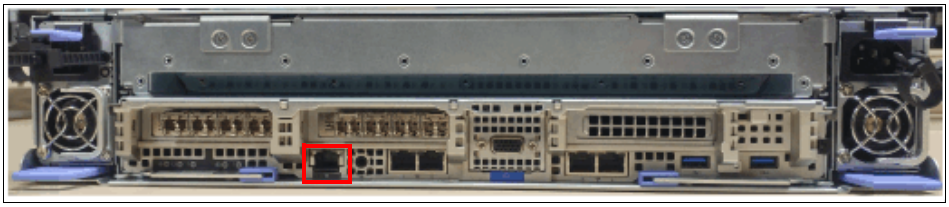


Figure 3-1 Location of technician port on SAN Volume Controller 2145-SV2 and 2145-SA2 nodes

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 PC or notebook 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 PC or notebook.

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 the SAN Volume Controller is initialized, each node of a new system remains in the *candidate* state and cannot process I/O. During initialization, a *cluster* is created, which at that moment consists only of one node. All other nodes except the first one must not be initialized, and they are added to the cluster by using a cluster management interface (GUI or CLI) after first one 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 in a system. During initialization, candidate nodes are added automatically. You also can use the management GUI or CLI to add nodes to the system.

Complete the following steps to initialize a new system:

1. Connect your PC or notebook to a technician port of any canister of the control enclosure. Ensure that you obtained a valid IPv4 address with DHCP.
2. Open a supported web browser and go 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 are prompted to accept untrusted certificates because the system certificates are self-signed. If you are directly connected to the service interface, there is no doubt about the identity of the certificate issuer, so you can safely accept the certificates.

   If the system is not in a state that allows initialization, you are redirected to the Service Assistant interface. Use the displayed error codes to troubleshoot the problem.
3. The login to Service Assistant window appears as shown in Figure 3-2. This step is the first step of initializing the system. Enter the default superuser password (`passw0rd`) and then, click **Log in**.

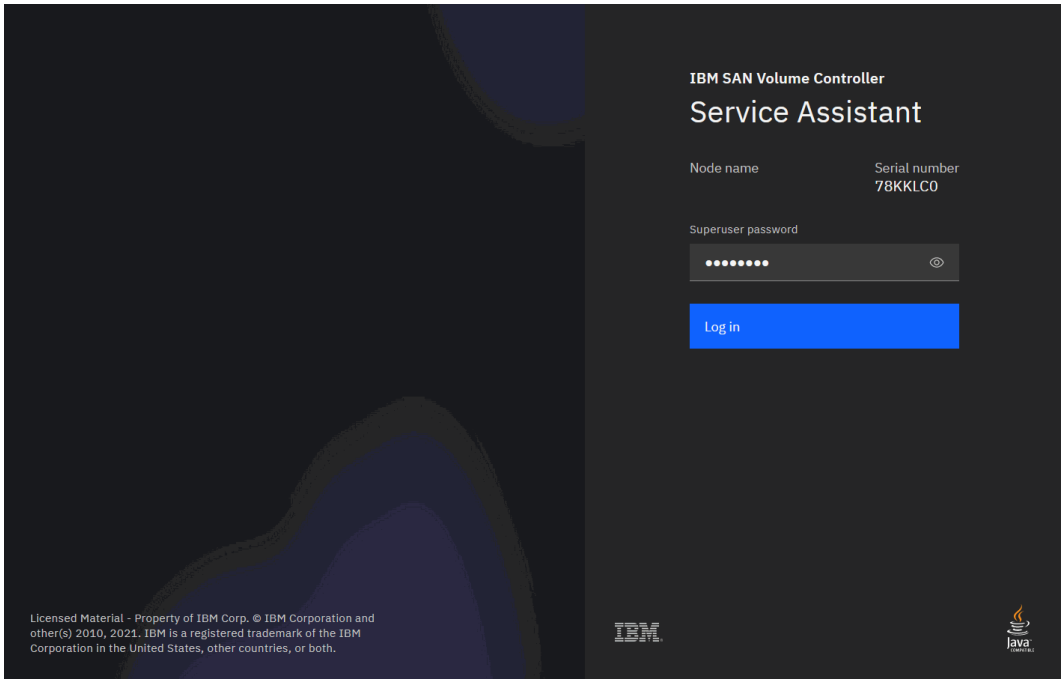


Figure 3-2 Login to Service Assistant via the technician port

4. The System Initialization wizard prompts whether the node is the first node of a new cluster or another node in a system, as shown in Figure 3-3. Click **Next** to continue.

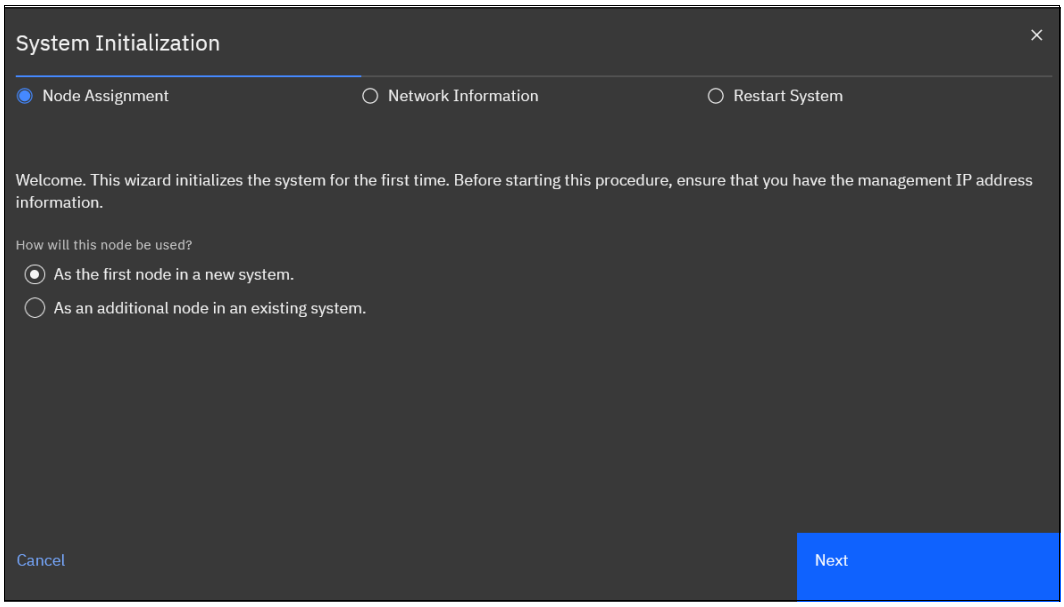


Figure 3-3 System Initialization: Node assignment

If you select **As an additional node in an existing system**, you are prompted to disconnect from the technician port and use the GUI of the existing system to add nodes.

5. Enter the management IP address information for the new system, as shown in Figure 3-4. Set the IP address, network mask, and gateway. Then, click **Next**.

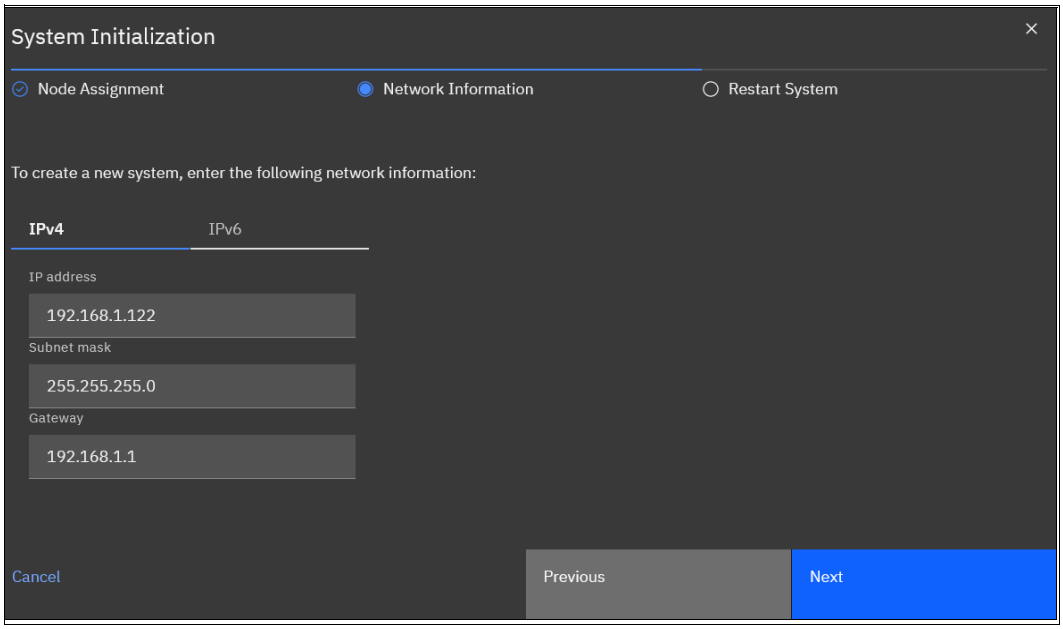


Figure 3-4 System Initialization: Enter Management IP

6. A window with restart timer opens. 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 SAN Volume Controller web-server. The timeout counter window is shown in Figure 3-5.

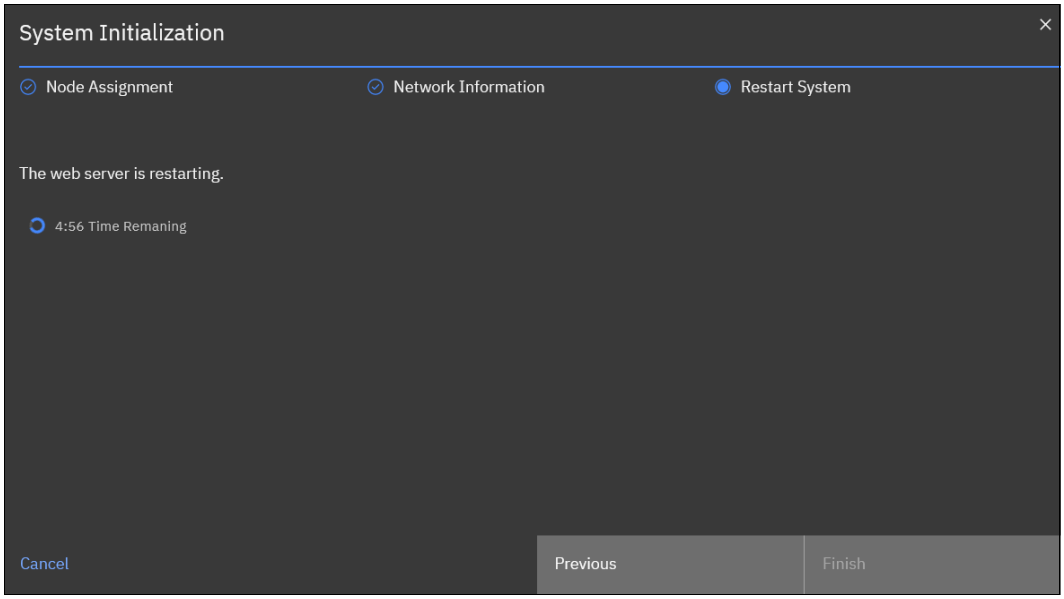


Figure 3-5 System Initialization: Web-server restart timer counting down from five minutes

7. The System Initialization completed wizard is shown in Figure 3-6. Click **Finish**.

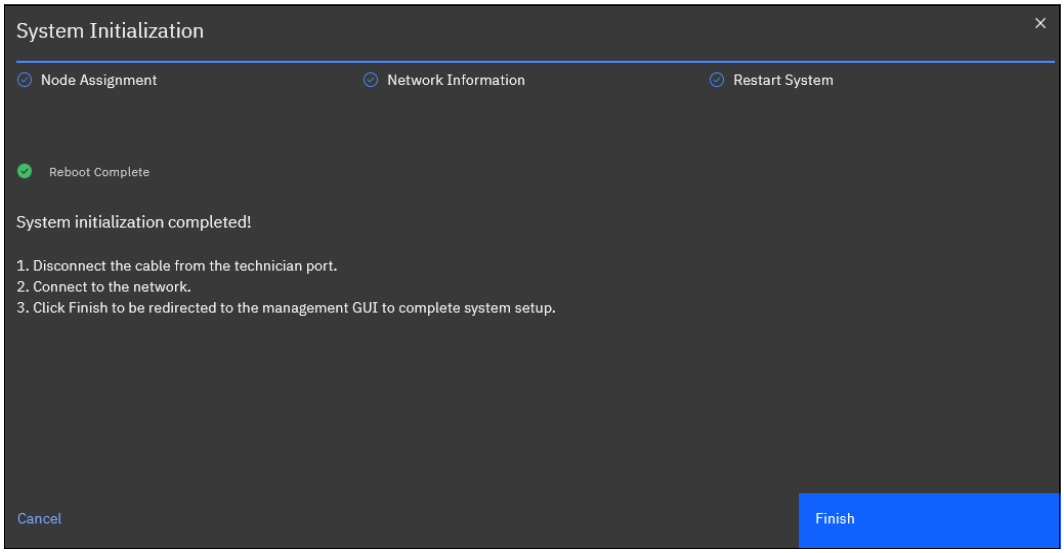


Figure 3-6 System Initialization completed

8. 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 also is available directly from the technician port. The System Setup wizard is available both ways.

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

After the initialization is complete and you are redirected to a management GUI from your PC or notebook, or you browse to the management IP address of a freshly initialized system from another workstation, you must complete the system setup wizard to define the basic settings of the system.

Notes: Consider the following points:

- Experienced users can disable the system setup wizard and complete the configuration manually. However, this method is *not* recommended for most use cases.

  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 are 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 3.5.1, “Configuring secure communications” on page 141.

To complete 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 3-7 on page 110. 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.

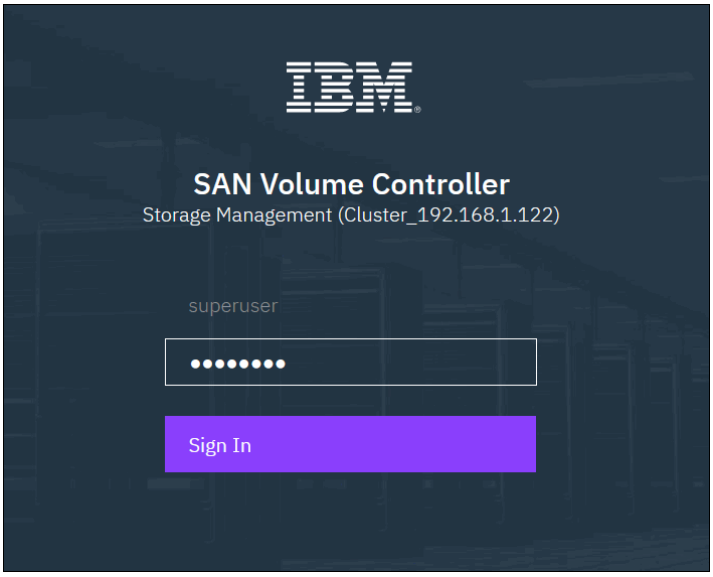


Figure 3-7 Logging in for the first time

2. The welcome window opens, as shown in Figure 3-8. Click **Next**.

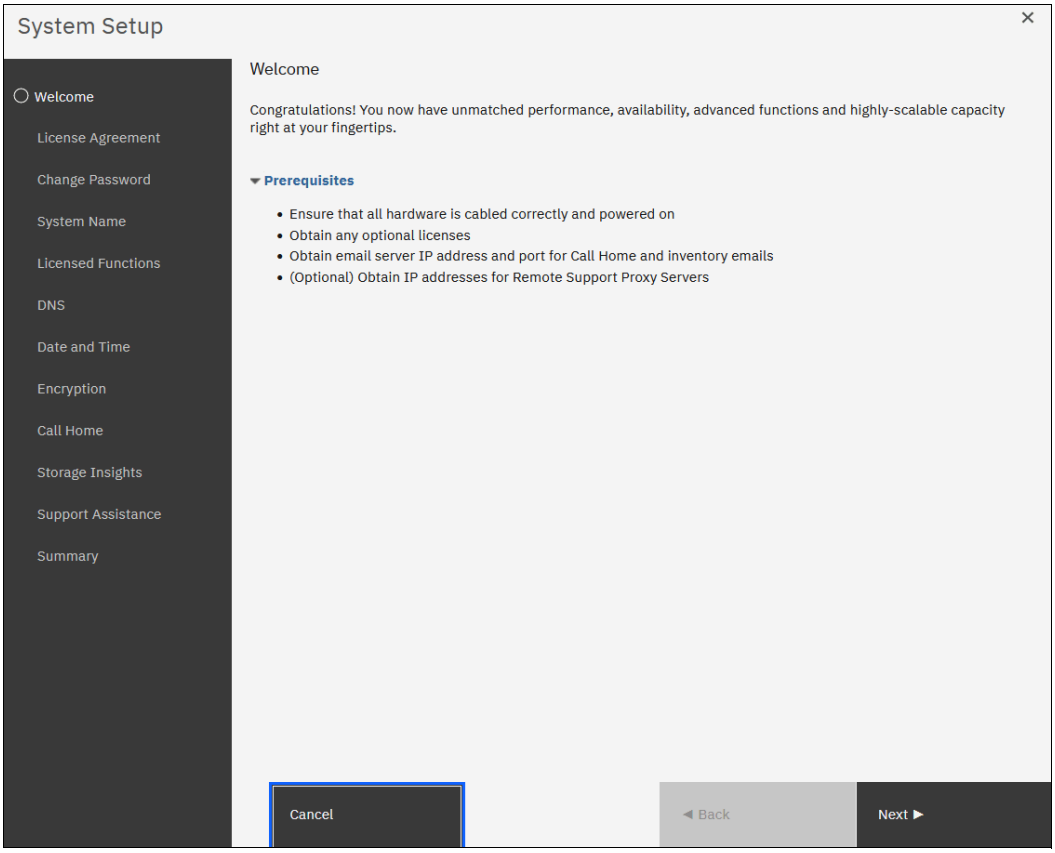


Figure 3-8 System Setup welcome window

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 3-9. Click **Next**.

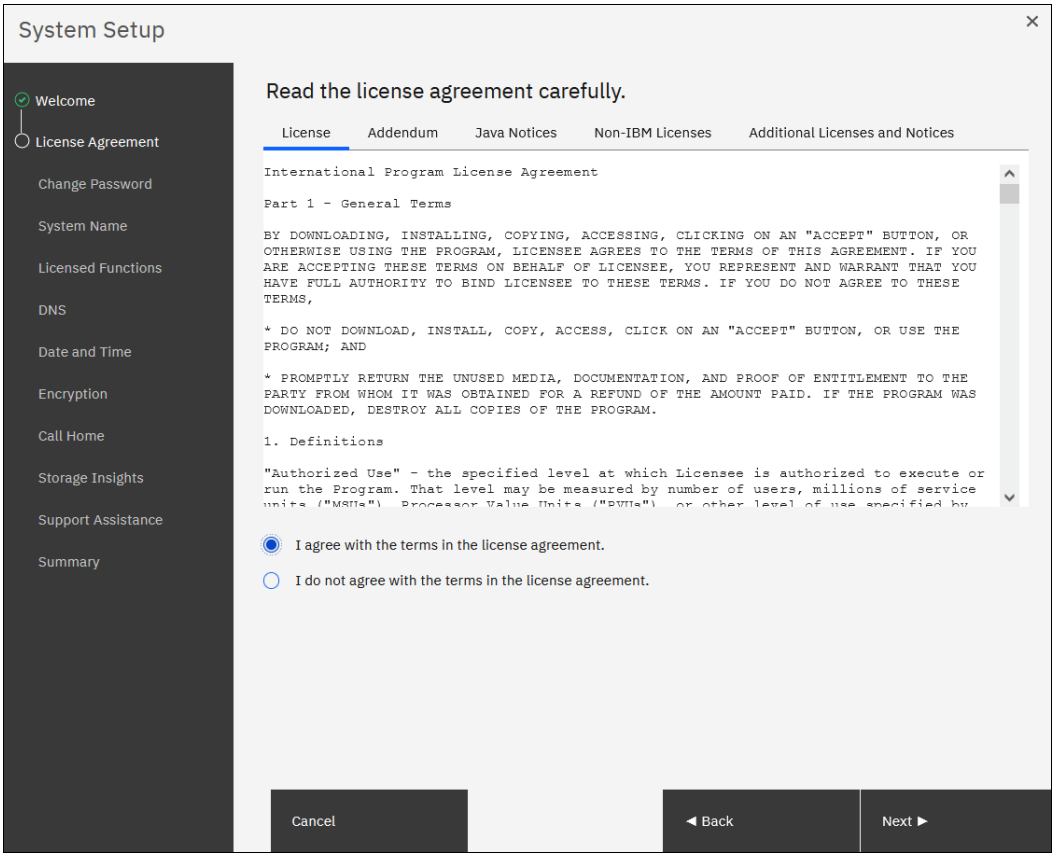


Figure 3-9 License agreement

4. Enter a new password for the superuser, as shown in Figure 3-10.

   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 3.5.2, “Configuring password policies” on page 144. Click **Apply and Next**.

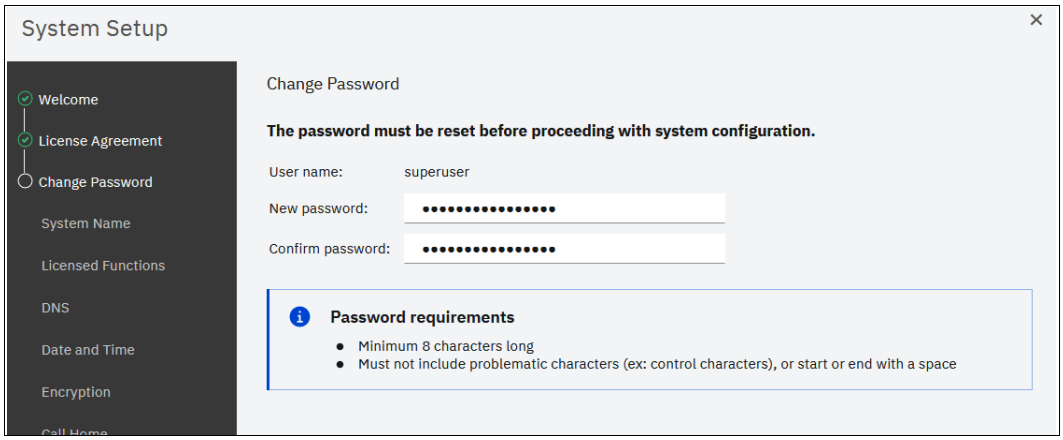


Figure 3-10 Change superuser password

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 the commands during the System Setup wizard.

5. Enter the name that you want to give the new system, as shown in Figure 3-11. Click **Apply and Next**.

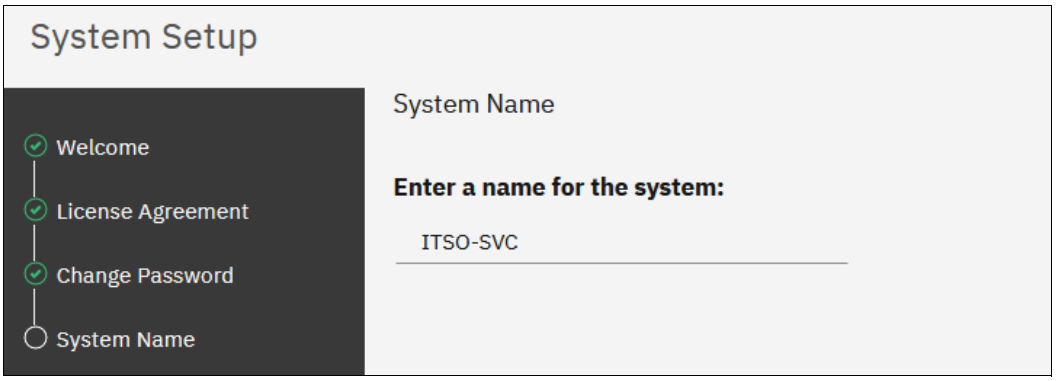


Figure 3-11 Entering system name

Avoid the use of an underscore (_) in a system name. Although permitted, it cannot be used in domain name server (DNS) short names 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: To prepare the IBM Spectrum Virtualize clusters at Master, AuxNear, and AuxFar sites to work in a 3-Site replication solution, ensure that the system name is unique for all three clusters. The system names must remain different through the life of the 3-Site configuration.

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 and clicking **Monitoring** → **System Hardware** and then, selecting **System Actions** → **Rename System**.

6. Enter the licensed storage capacity units (SCU) and capacity for each function, as shown in Figure 3-12.

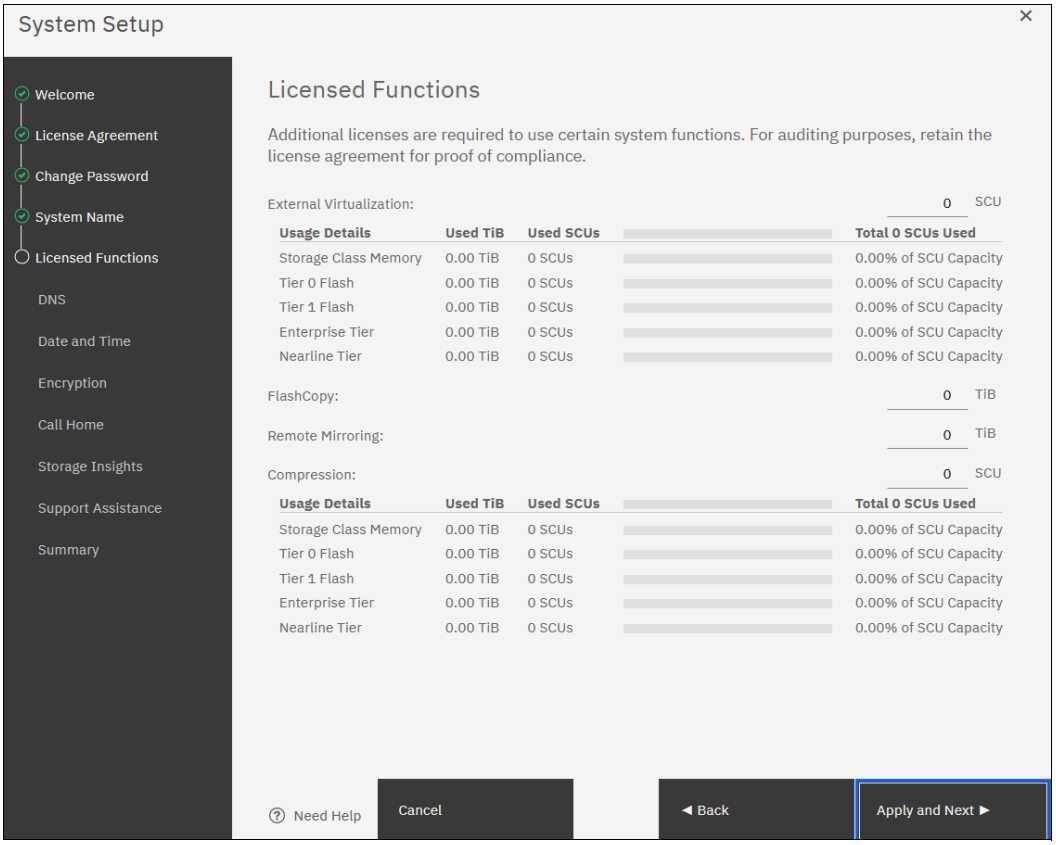


Figure 3-12 Licensed functions

For external virtualization and compression, differential licensing offers different pricing rates for different types of storage and is based on the number of SCUs that are purchased.

Ensure that the numbers you enter here match the numbers in your license authorization paperwork. For more information about licensing, see Chapter 1, “Introduction and system overview” on page 1.

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 3-13. DNS helps the system to resolve the names of the computer resources that are in the external network if these names are not indicated by an IP-address. Click **Apply and Next** when done.

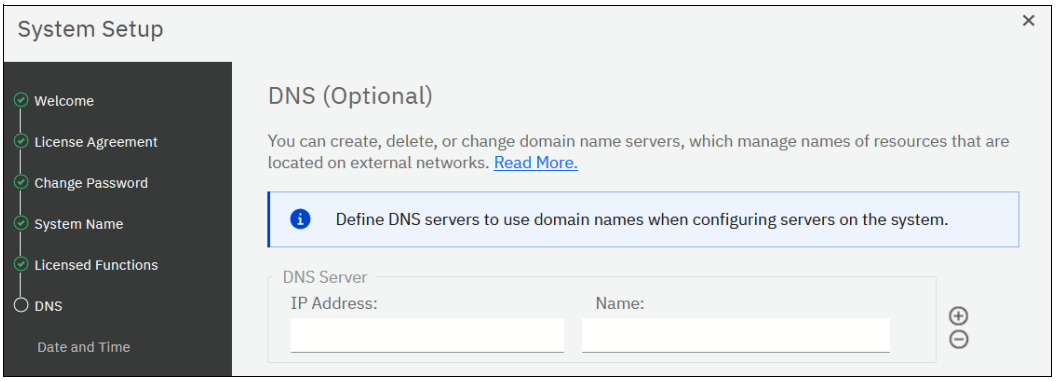


Figure 3-13 DNS server set up window (optional)

9. Enter the date and time settings. In the example that is shown in Figure 3-14, the date and time are set by using an NTP server. Generally, use an NTP server so that all of your SAN and storage devices feature a common timestamp. This practice facilitates troubleshooting and prevents time stamp-related errors. When done, click **Apply and Next** (see Figure 3-14).

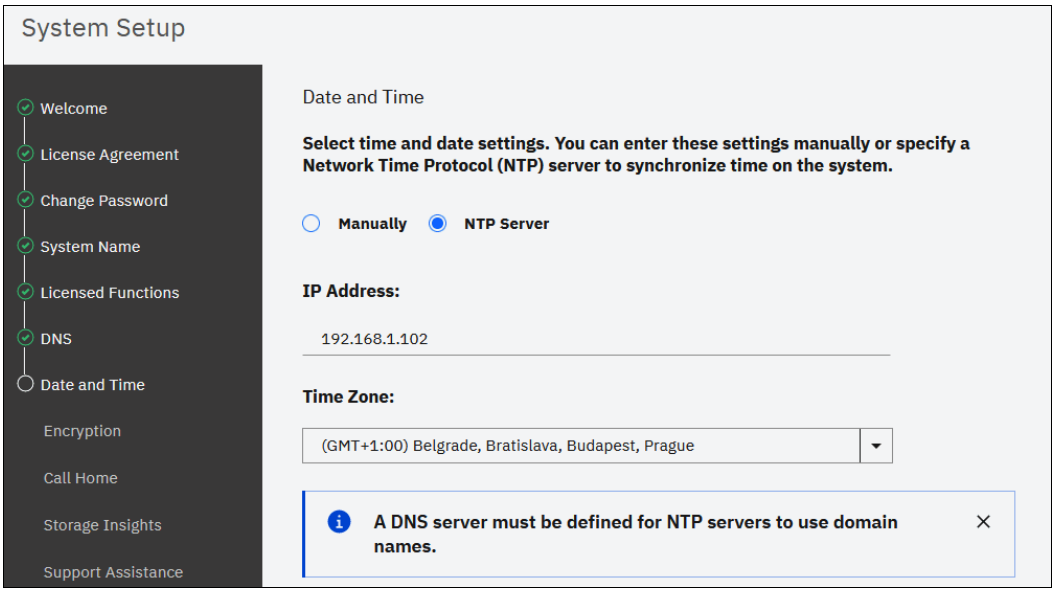


Figure 3-14 Data and Time set up window

10. The System Setup wizard prompts for enablement of encryption licensing. Select whether the encryption feature was purchased for this system, as shown in Figure 3-15.

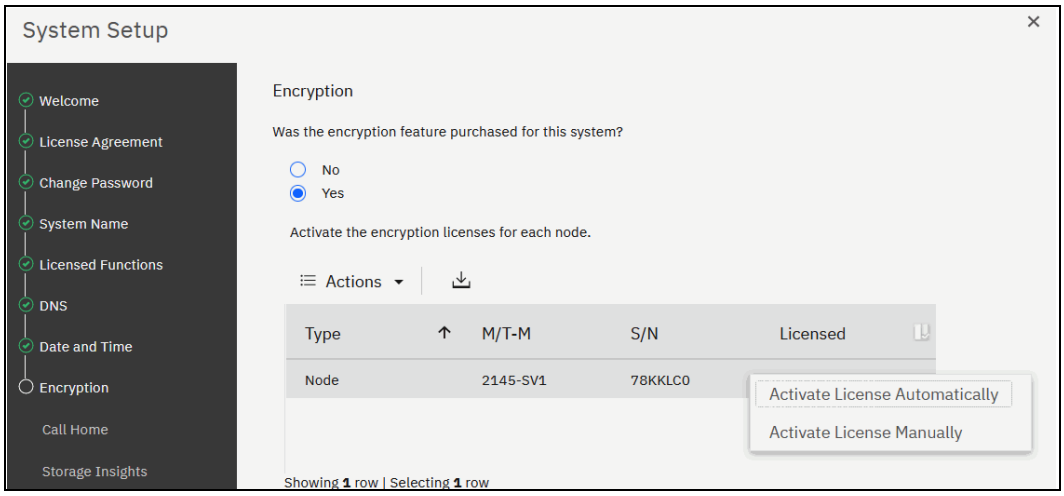


Figure 3-15 Activating encryption license

If encryption is not planned now, select **No** and then, click **Next**. You can enable this feature later, as described in Chapter 12, “Encryption” on page 793.

Note: When encryption is enabled on the system, encrypted storage pools can be created. If the system is a single control enclosure system in which all FCM-drives must 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 this 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 PC or notebook that you use to connect to the GUI and run the System Setup wizard includes internet access. If no internet connection is available, use the manual activation option and follow the instructions. For more information, see Chapter 12, “Encryption” on page 793.

After the encryption license is activated, you see a green check mark for each enclosure, as shown in Figure 3-16 on page 116. After all of the control enclosures show that encryption is licensed, click **Next**.

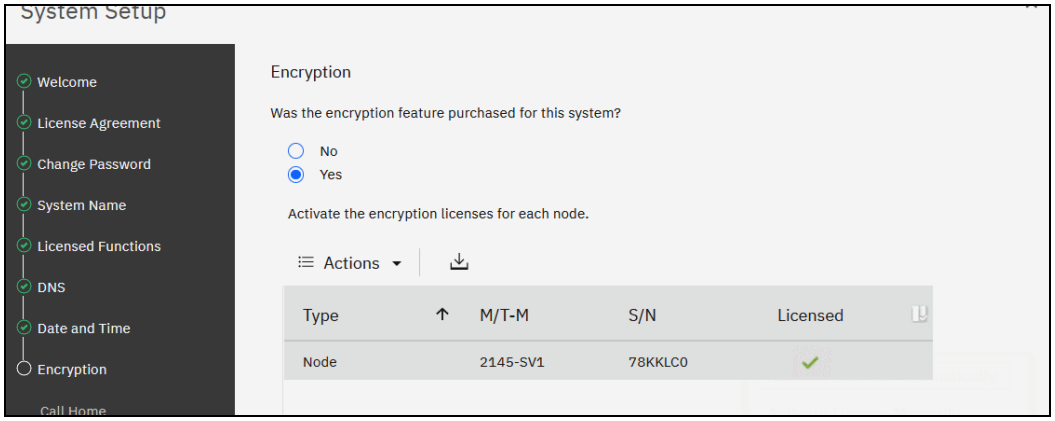


Figure 3-16 Encryption licensed

11. Call Home provides multiple benefits. It enables automatic ticket creation at IBM when errors occur, which improves the speed and efficiency by which calls are handled. Call Home also enables Storage Insights and Remote Support. Click **Next** in the Call Home window, as shown in Figure 3-17.

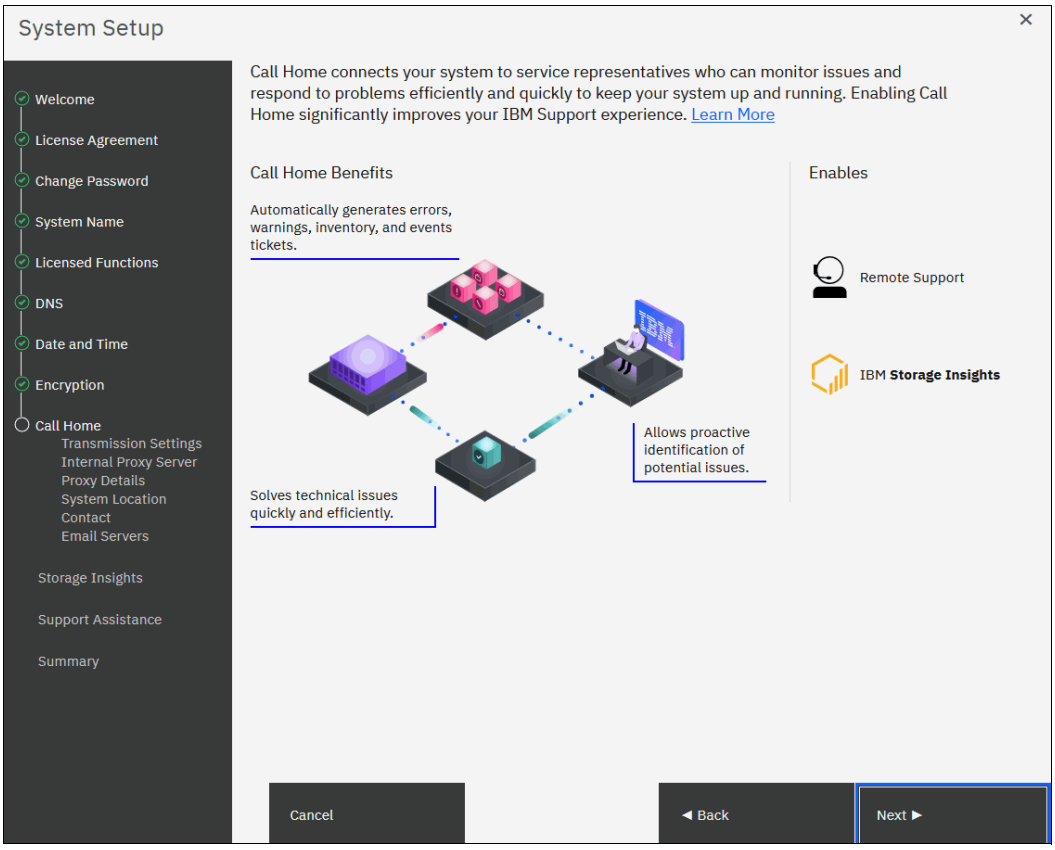


Figure 3-17 Call Home information window

On IBM SAN Volume Controller systems, an IBM SSR configures Call Home during installation. You must check only whether all the entered data is correct.

All IBM SAN Volume Controller systems support two 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 also can 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 Chapter 13, “Reliability, availability, serviceability, monitoring and troubleshooting” on page 853.

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 populated. In this step, the system also verifies that it can contact the Cloud Call Home servers.

12. System Setup prompts for which transmission types are used for Call Home. Select both options and click **Apply and Next**, as shown in Figure 3-18.

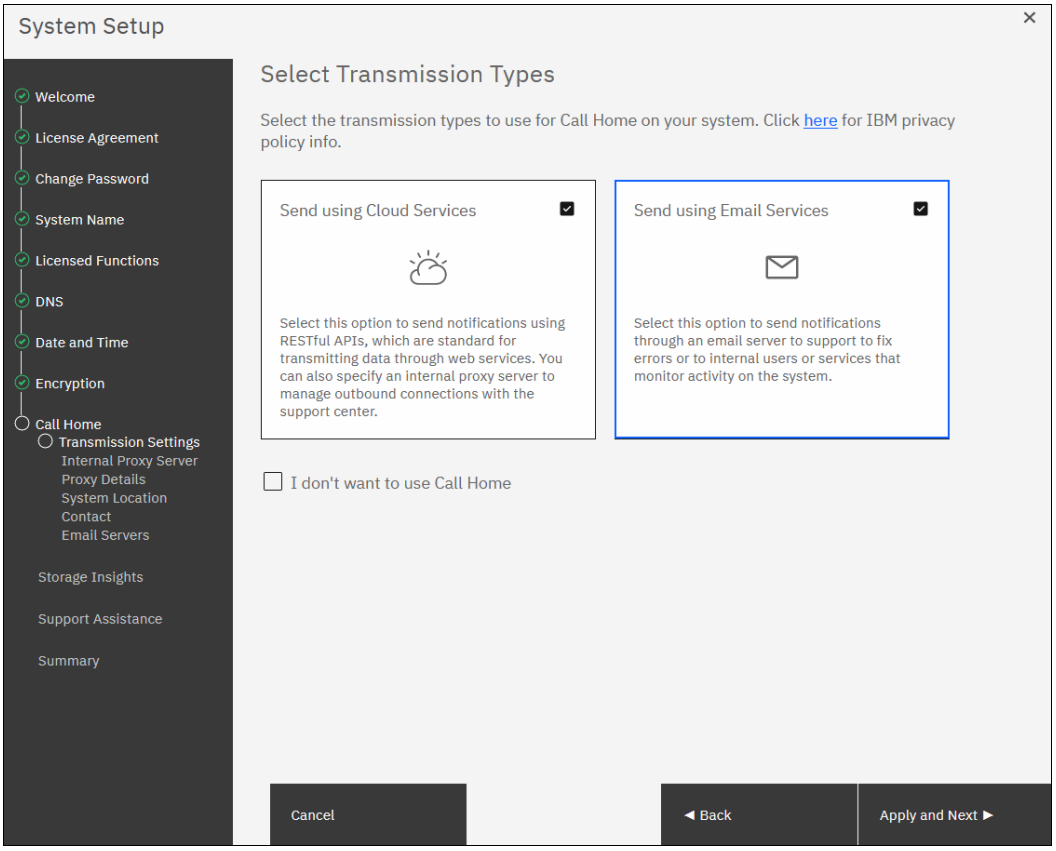


Figure 3-18 Call Home transmission type

Note: It is not recommended to select the **I don't want to use Call Home** option.

13. System Setup gives you the opportunity to enter proxy server details (if required) to allow Call Home connectivity (see Figure 3-19). Click **Next**.

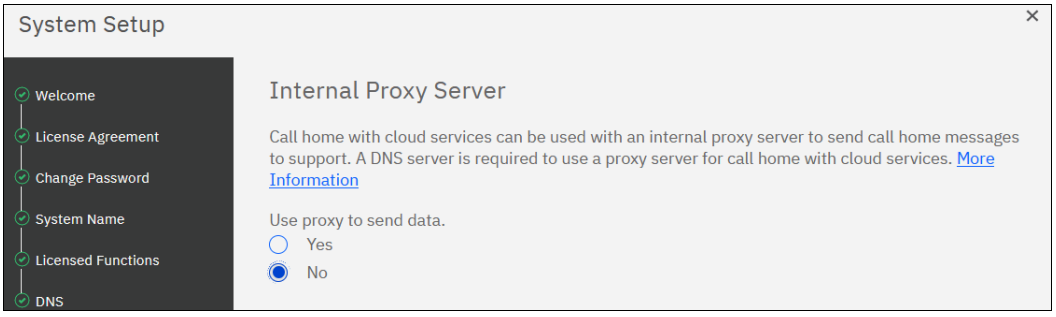


Figure 3-19 Proxy server (optional)

14. System Setup now prompts for the system location, as shown in Figure 3-20. Enter all information and click **Next**.

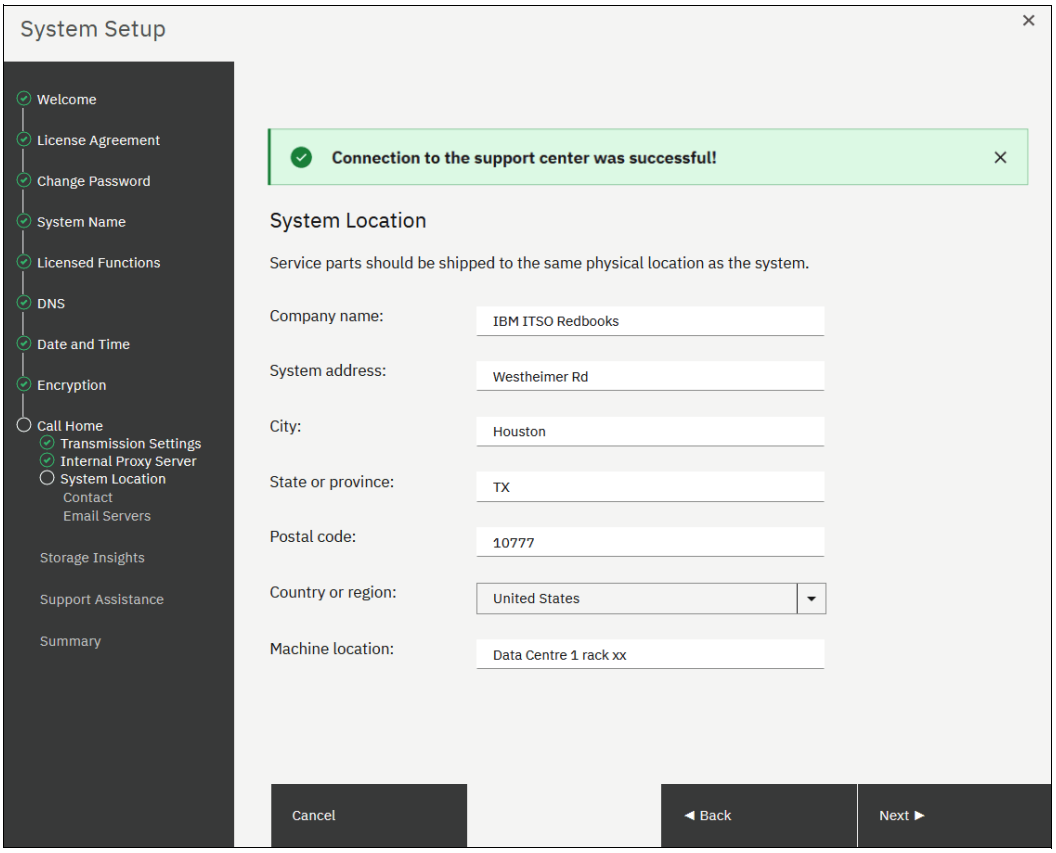


Figure 3-20 System location

15. Contact information is entered at the Contact window, as shown in Figure 3-21. Enter all information and then, click **Apply and Next**.

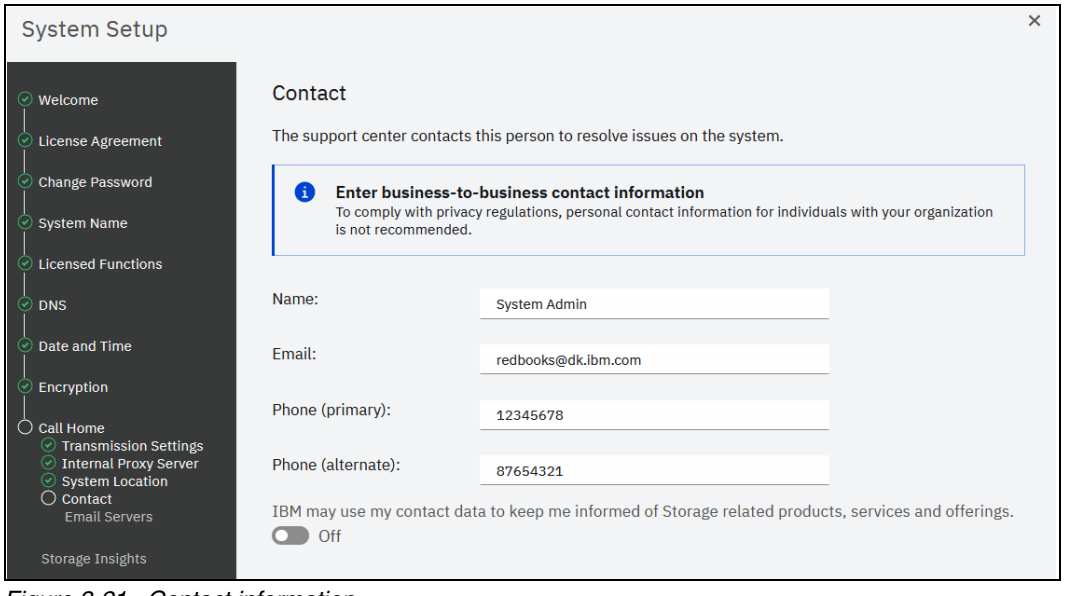


Figure 3-21 Contact information

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 3-22. Enabling email for Call Home also enables the possibility for storage administrators to receive email if warnings or errors occur. Enter the IP address of one or more servers and then, click **Apply and Next**.

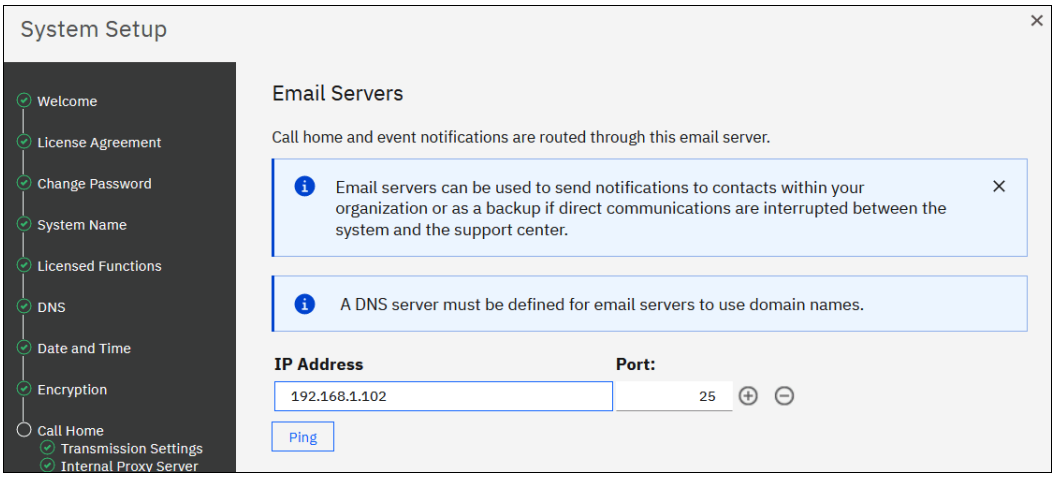


Figure 3-22 System Setup: Email servers

17. 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 the service is available, you are prompted to sign up, as shown in Figure 3-23. System Setup now provides information about how to enable Storage Insights. Click **Next**.

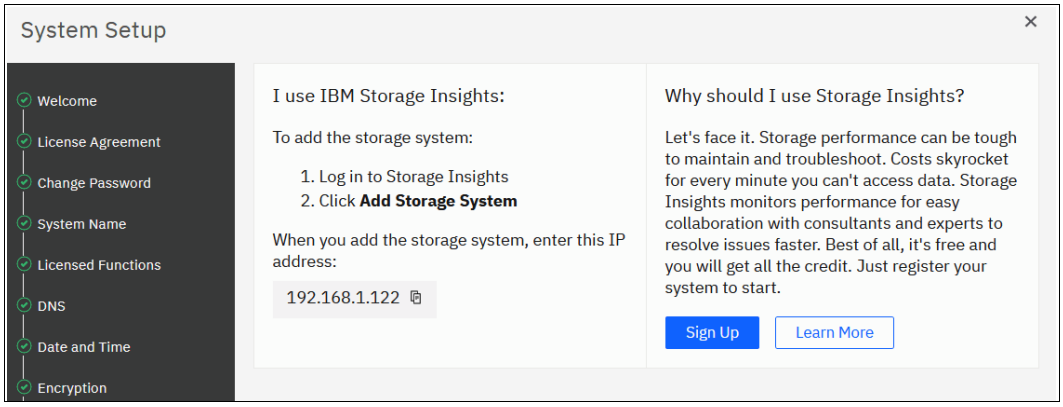


Figure 3-23 Storage Insights information window

18. If you enabled at least one Call Home method, the Support Assistance configuration window opens, as shown in Figure 3-24. The Support Assistance function requires Call Home. Therefore, if it is disabled, Support Assistance cannot be used. Click **Next**.

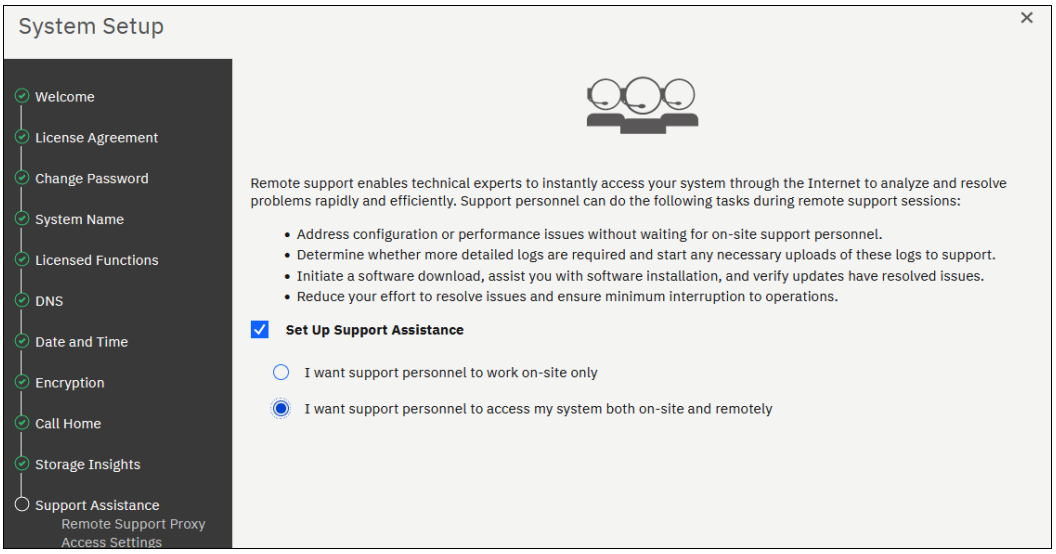


Figure 3-24 Support assistance

By using the Support Assistance feature, you allow IBM Support to perform maintenance tasks on your system with support personnel onsite or remote.

If an IBM service support representative (SSR) is onsite, the SSR can log in locally by using 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 also can 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 Chapter 13, “Reliability, availability, serviceability, monitoring and troubleshooting” on page 853.

19. If you allow remote support, you are provided with the IP addresses and ports of the remote support centers. You also can provide proxy server information (if required) to allow the connectivity, as shown in Figure 3-25. Click **Apply and Next**.

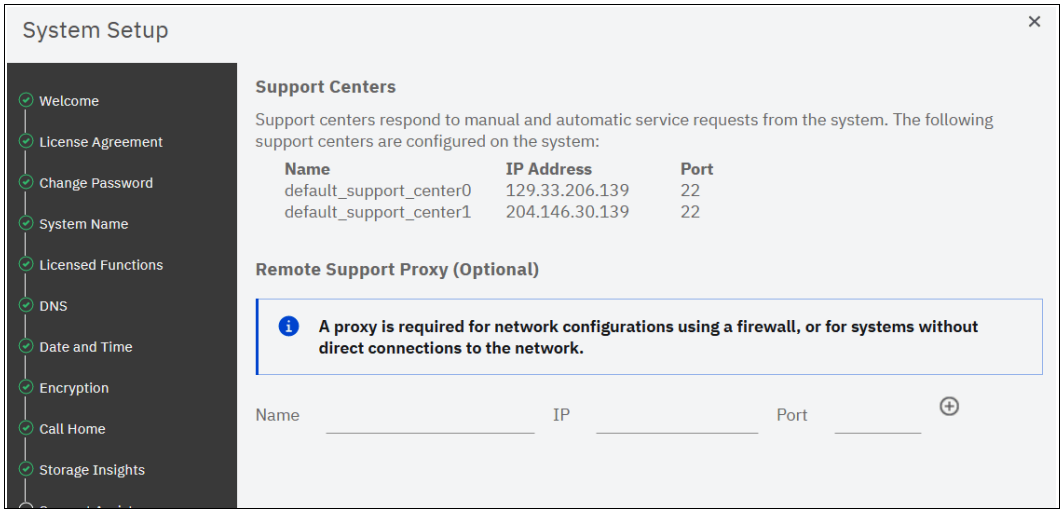


Figure 3-25 System communicating with named IBM Support servers

20. You can allow remote connectivity at any time or only after obtaining permission from the storage administrator (see Figure 3-26). Click **Apply and Next**.

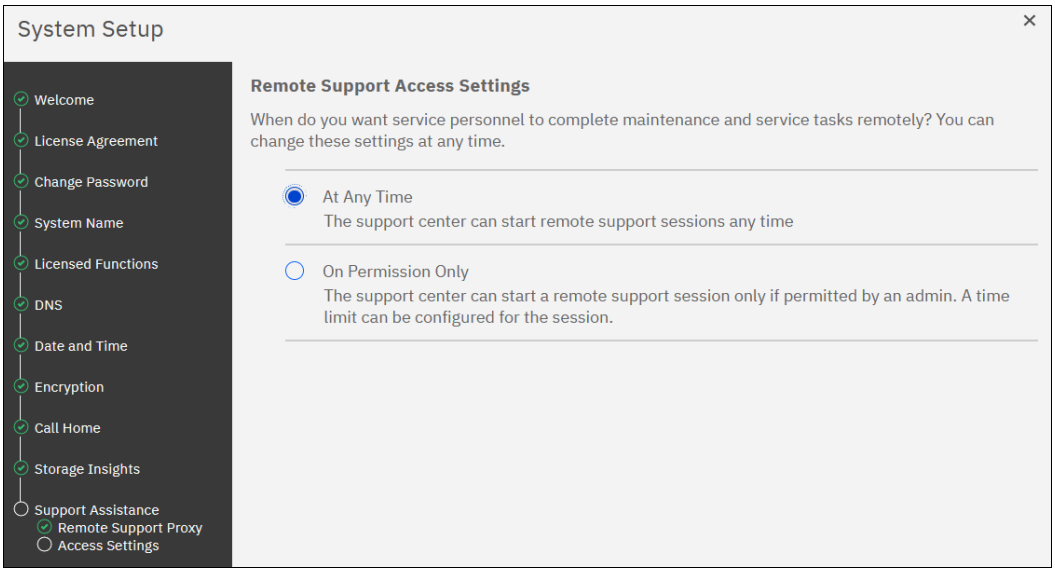


Figure 3-26 Remote support access settings

21. On the Summary page, the settings that were set by the System Setup wizard are shown. If corrections are needed, return to a previous step by clicking **Back**. Otherwise, click **Finish** to complete the system setup wizard (see Figure 3-27).

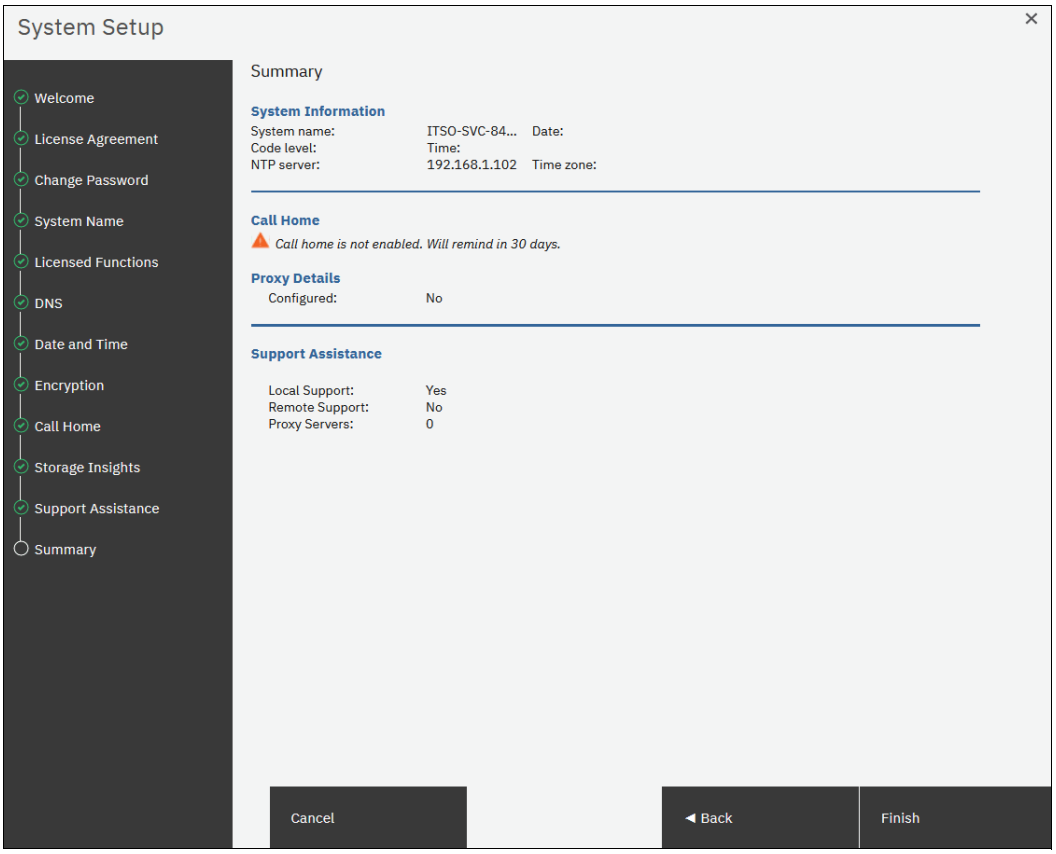


Figure 3-27 System Setup: summary page

Note: If the system does not have access to send data to IBM Support, Call Home and Support Assistance cannot be enabled.

After the wizard completes, your system consists of only the nodes that were available as candidates, connected to SAN, and zoned together. If more nodes exist, you must add them to complete system setup. For more information about how to add a node, see 3.4.2, “Adding a node or hot spare node” on page 128.

If no other nodes must be added to the system, the setup process is complete. All of the mandatory steps of the initial configuration are done. If required, 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.

3.4 Base configuration

Tasks that are listed in this section are used to define global system configuration settings. Often, they are performed during system setup. However, they also can be performed any time later, such as when the system is expanded or the system environment is reconfigured.

3.4.1 Configuring Remote Direct Memory Access clustering

Up to four control enclosures can be joined in an IBM HyperSwap or a standard topology cluster. This subsection describes the configuration steps that must be performed if a system is designed for IP-based Remote Direct Memory Access (RDMA) node-to-node traffic. For FC SAN clustering, no special configuration is required on the system, but the SAN must be set up as described in Chapter 2, “Planning” on page 59.

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 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 the 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 will 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 IP addresses on each port of each node

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 going to `https://<node_service_IP>/service` and then, click **Change Node IP**, as shown in Figure 3-28.

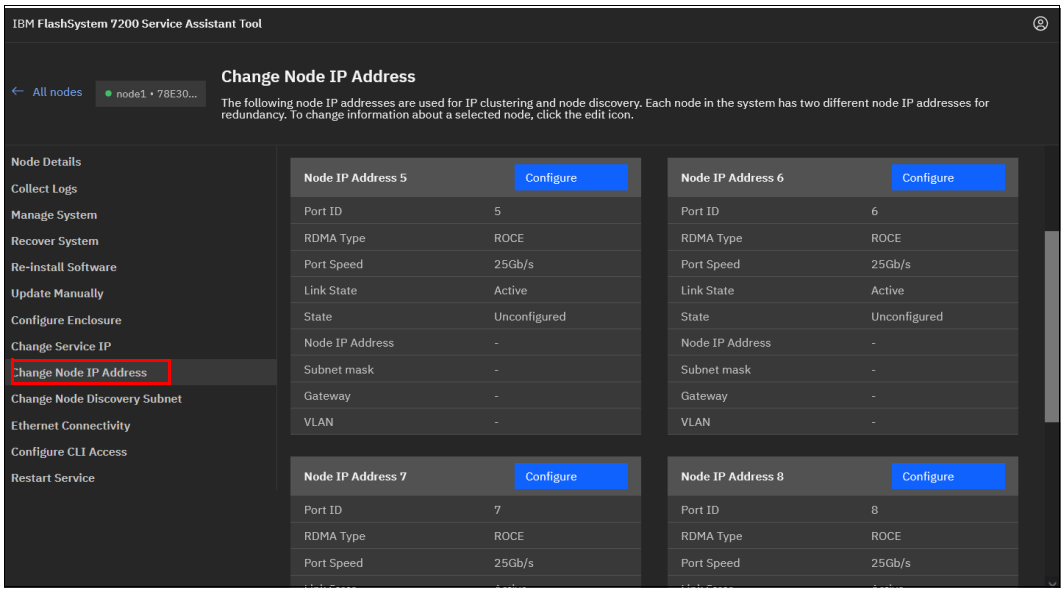


Figure 3-28 Node IP address setup for RDMA clustering

Figure 3-28 shows that ports 1 - 3 do not show any RDMA type; therefore, they cannot be used for node-to-node traffic. Because Ports 4 and 5 show RDMA type RoCE, they can be used.

2. Hover your cursor over a tile with a port and click **Modify** 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 completed successfully, as shown in Figure 3-29. 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.

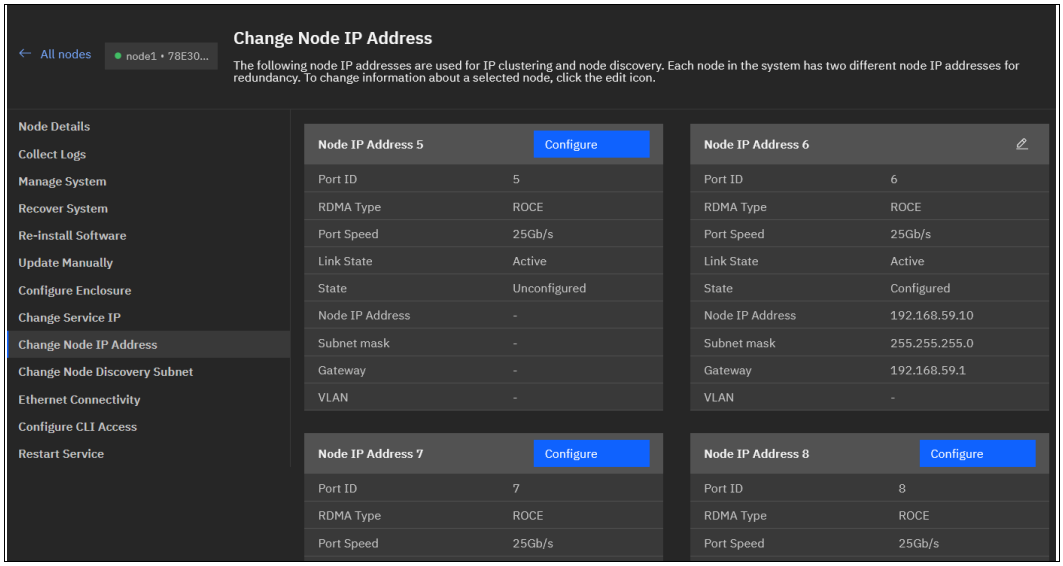


Figure 3-29 Node IP addresses configured

To list the node IP configuration by using the CLI, run the **`sainfo lsnodeip`**, as shown in Example 3-1.

Example 3-1 Listing node IPs currently not set

```
IBM_2145:ITSO-SVC: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 3-2.

Example 3-2 Executing 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 3-3.

Example 3-3 Changed node IP (output shortened for clarity)

```
IBM_2145:ITSO-SVC: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.

   To configure unicast discovery, see the information about the **satask addnodediscoverysubnet**, **satask rmnodediscoverysubnet**, or **sainfo lsnodediscoverysubnet** commands in IBM Documentation. You can also configure discovery subnets by using the Service Assistant interface menu option **Change Node Discovery Subnet**, as shown in Figure 3-30.

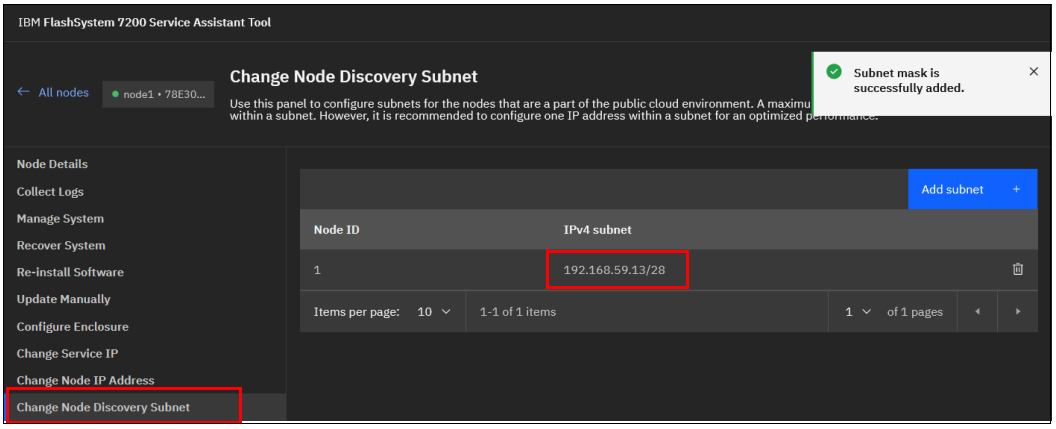


Figure 3-30 Setting the node discovery subnet

4. 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, the enclosure can be added to the cluster.

3.4.2 Adding a node or hot spare node

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 existing system. For FC node-to-node communication, verify that correct the 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 3-31.

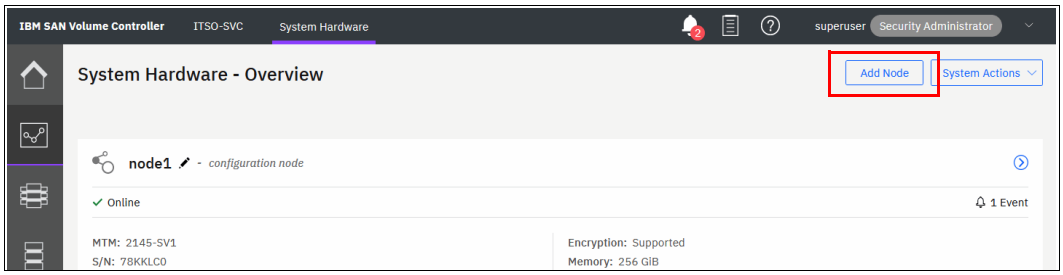


Figure 3-31 Add Node button

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 3-32. 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**.

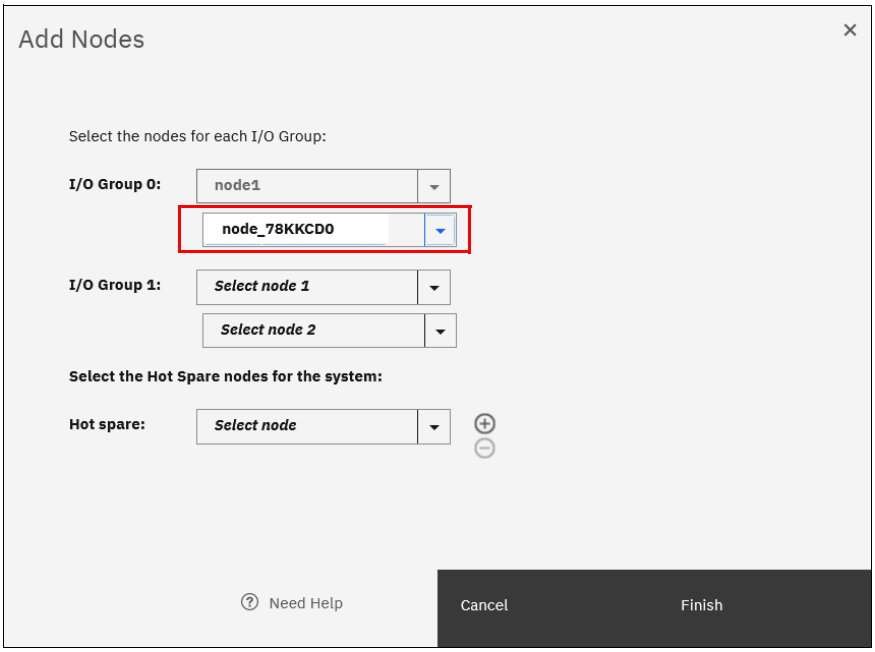


Figure 3-32 Adding a node

The **Monitoring** → **Systems Hardware** window now changes and shows the node is added, as shown in Figure 3-33.

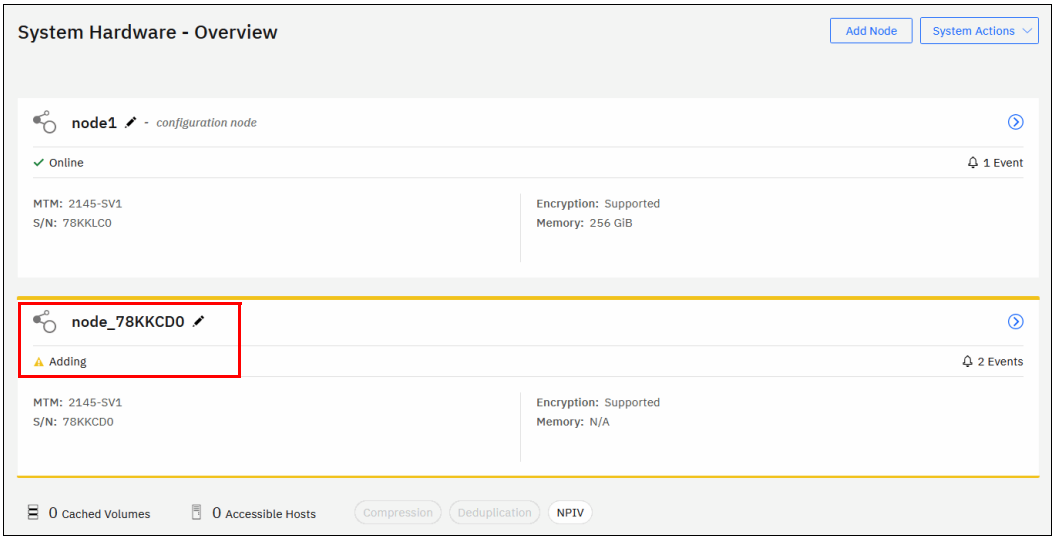


Figure 3-33 IBM SAN Volume Controller is adding node to cluster

The node is added as shown in Figure 3-34 on page 130.

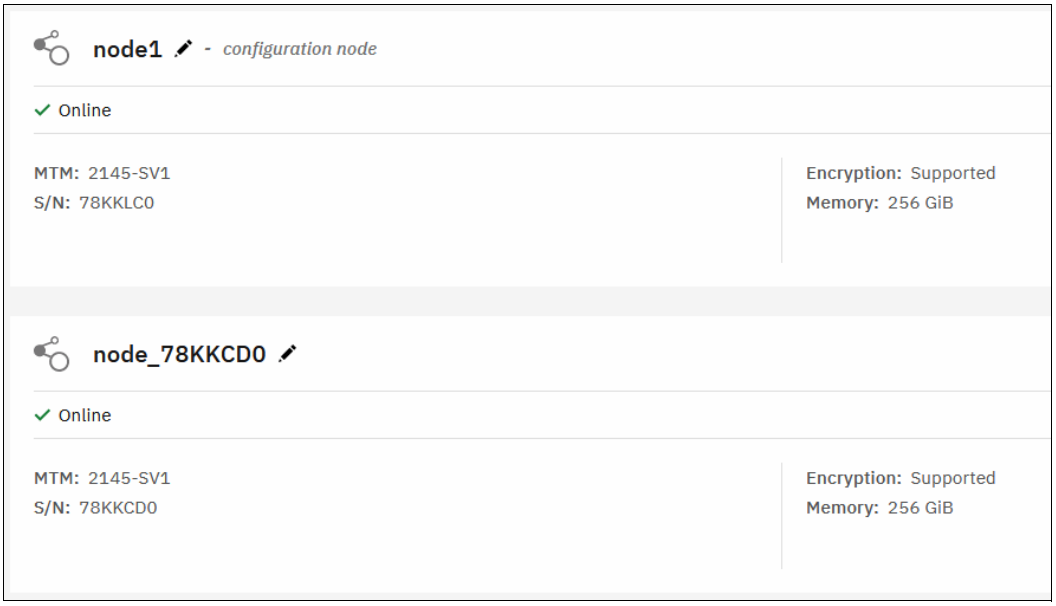


Figure 3-34 Node added

Note: When a node is added, the software version that is running is upgraded or rolled backed to match the cluster software version. This process can take 30 minutes or more to complete, and 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 IBM Documentation:

1. When adding nodes, check for unpopulated I/O groups by running **`lsiogrp`**. Each complete I/O group has two nodes. Example 3-4 shows that only `io_grp0` has nodes; therefore, a new control enclosure can be added to `io_grp1`.

Example 3-4 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 3-5.

Example 3-5 Listing the candidate nodes

```
IBM_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 3-6. The command triggers only the process, which starts in background and can take 30 minutes or more.

Example 3-6 Adding a node as a spare

```
IBM_2145:ITSO-SVC:superuser>addnode -panelname 78KKLD0 -spare
Node, id [3], successfully added
```

Example 3-7 shows same command, but used to add a node to an I/O group `io_grp1`.

Example 3-7 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 3-8, 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 3-8 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
```

Shown in Example 3-9 is a two-IO-group system with no spare.

Example 3-9 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**.

3.4.3 Changing the system topology

IBM SAN Volume Controller supports two multi-site topologies: HyperSwap and Enhanced Stretched Cluster (ESC). 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 HyperSwap topology places both nodes of an I/O group at the same site. Both topologies enable full configuration of the highly available (HA) volumes through a single point of configuration.

If your solution is designed for ESC or HyperSwap, 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
- HyperSwap system configuration details

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 then, select **Modify System Topology**, as shown in Figure 3-35.

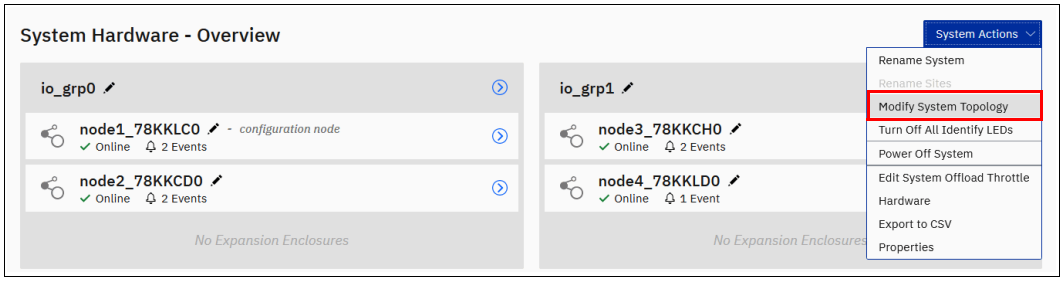


Figure 3-35 Starting the Modify System Topology wizard

2. The Modify Topology wizard welcome window opens. Click **Next**. You are prompted to change the default site names, as shown in Figure 3-36. The site names can indicate, for example, building locations for each site, or other descriptive information.

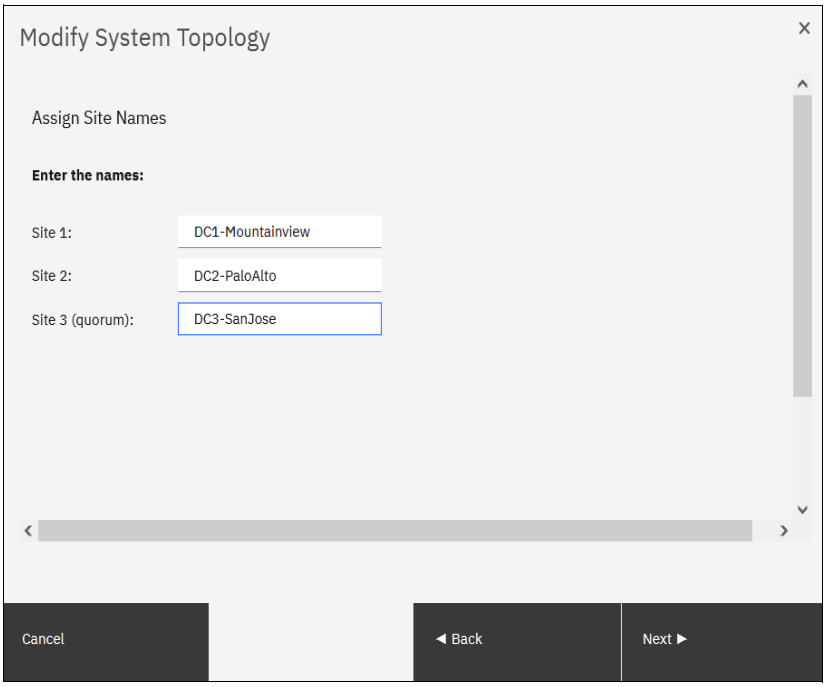


Figure 3-36 Assigning site names

3. Select **HyperSwap System** or **Stretched System** for the topology, and assign I/O groups to the sites. Click the marked icons in the center of the window to swap site assignments, as shown in Figure 3-37 on page 134. After checking serial numbers and site assignments, select HyperSwap and click **Next**.

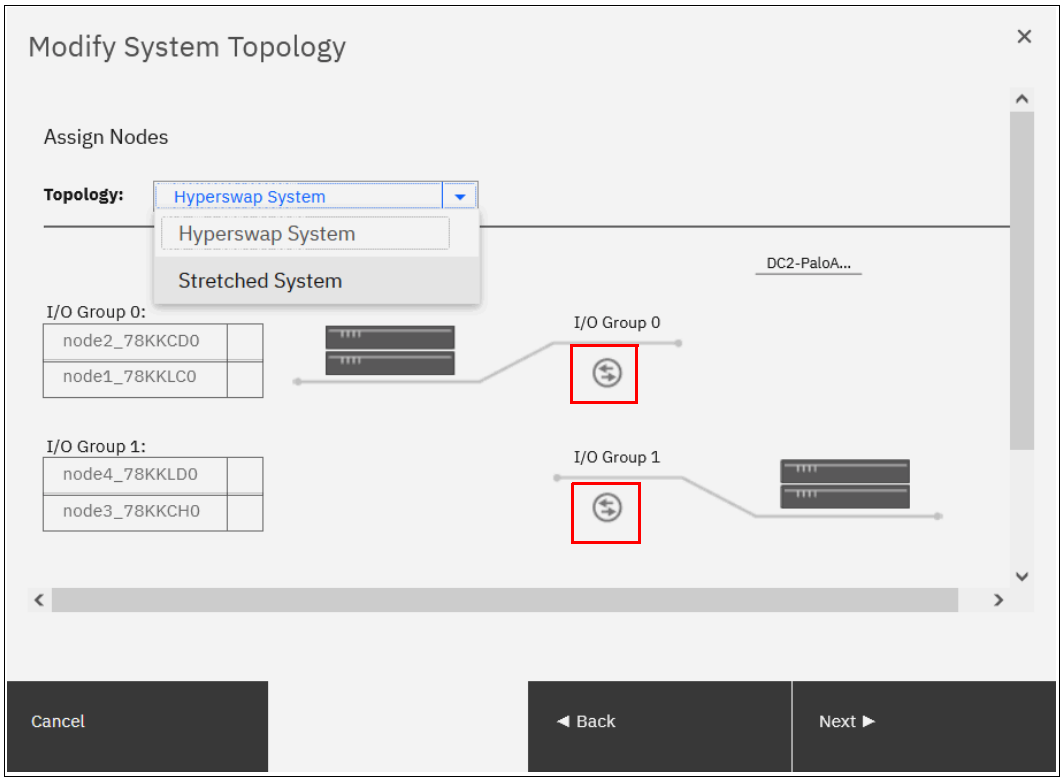


Figure 3-37 Specifying the system topology

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**, as shown in Figure 3-38. Click **Next** when done.

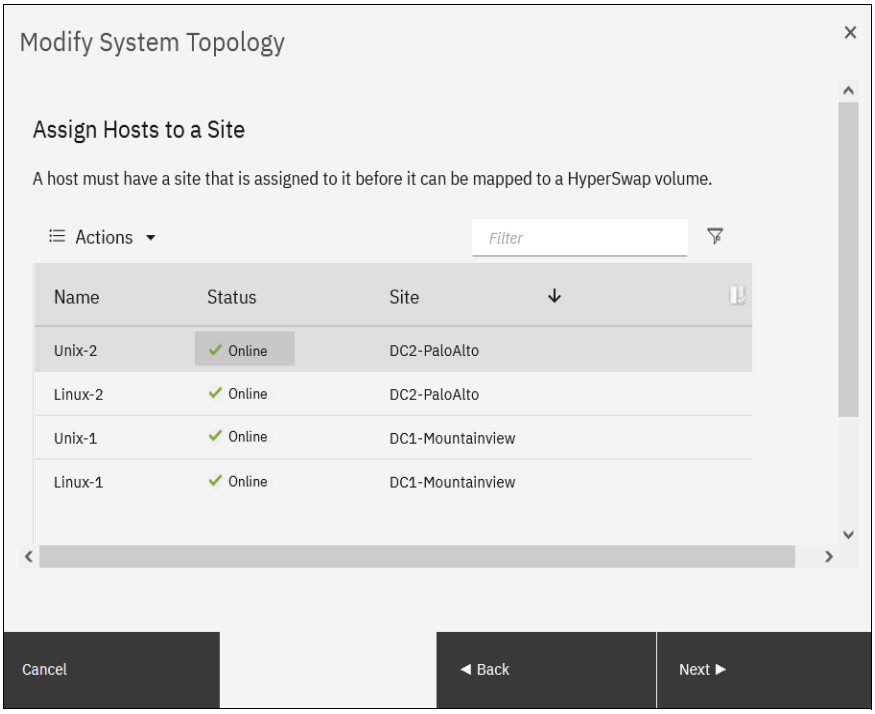


Figure 3-38 Assigning controllers to the sites

5. If you are configuring HyperSwap topology, set the maximum background copy operations bandwidth between the sites. *Background copy* is the initial synchronization and any subsequent resynchronization traffic for HyperSwap volumes. Use this setting to limit the impact of volume synchronization to host operations. You can also set it higher during the initial setup (when there are no host operations on the volumes yet), and set it lower when the system is in production.

   As shown in Figure 3-39, 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 Enhanced Stretched Cluster (ESC) topology system does not require this setting.

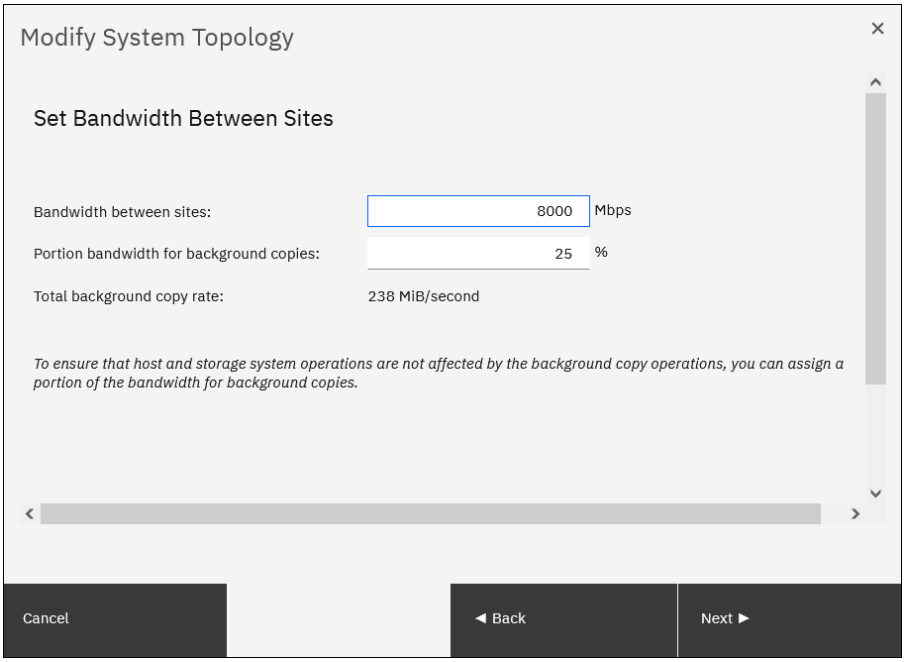


Figure 3-39 Setting the bandwidth between the sites for a HyperSwap topology

6. Review the summary and click **Finish**. The wizard starts implementing changes to migrate the system to the topology.

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.

3.4.4 Configuring quorum disks or applications

Quorum devices are required for a system to hold a copy of important system configuration data. 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 and to avoid a “split brain” condition, which happens when both halves of the system continue I/O processing independently of each other.

For systems with a standard topology, quorum devices are automatically assigned from a managed MDisk. No special configuration actions are required. Optionally, an IP quorum device can be configured to provide extra redundancy.

For ESC and HyperSwap topology systems, an active quorum device must be on a third, independent site. Due to 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.

A stretched or HyperSwap system can be configured without a quorum device at a third site. If there is no third site, then quorum must be configured to select a site to always win a tie-breaker. If there is a loss of connectivity between the sites, then 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. If there is a site outage at the wining site, then the system stops processing I/O requests until this site is recovered or the manual quorum override procedure is used.

Creating and installing an IP quorum application

To create and install an IP quorum application, complete the following steps:

1. Select **Settings** → **IP Quorum** to download the IP quorum application, as shown in Figure 3-40. If you use IPv6 for management IP addresses, the Download IPv6 Application button is available and the IPv4 option is disabled.

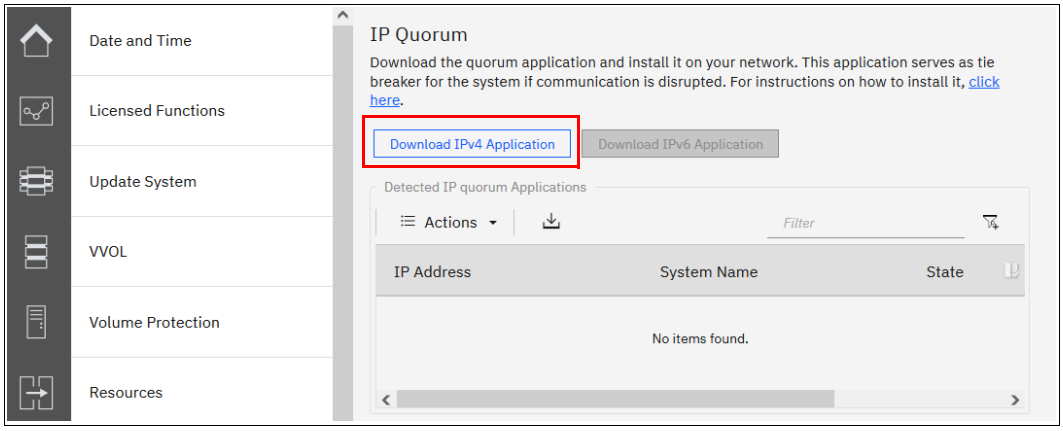


Figure 3-40 Download IPv4 quorum button

2. Click **Download...**, and a window opens, as shown in Figure 3-41 on page 137. This window 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.

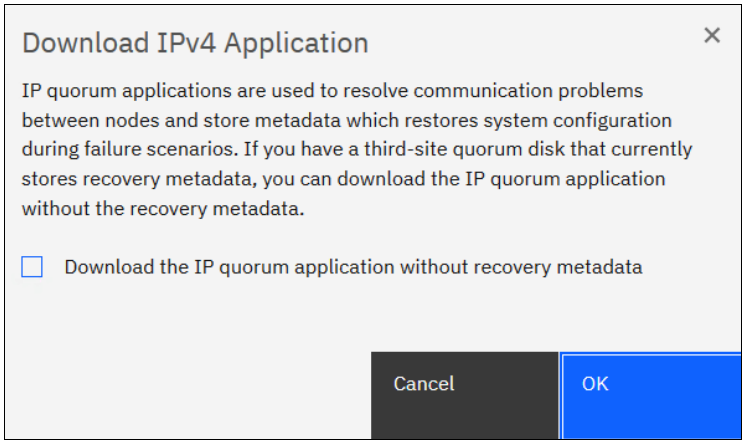


Figure 3-41 Download IP quorum application window

An application that does not store recovery metadata requires less channel bandwidth for a link between the system and the quorum application, which can be a decision-making factor for using a multi-site HyperSwap system.

For a full list of IP quorum app requirements, see this IBM Documentation web page.

3. After you 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 app, as shown in Example 3-10.

Example 3-10 Starting the IP Quorum app 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 which runs the IP Quorum must be in the same subnet as the IBM SAN Volume Controller. You can have a total number of five IP Quorums

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 3-42.

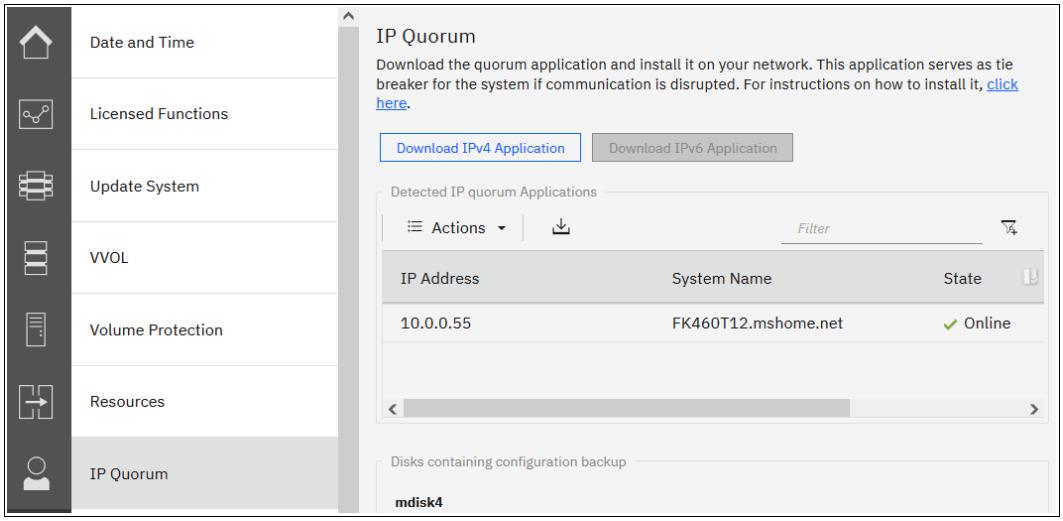


Figure 3-42 IP quorum application that is deployed and connected

Configuring IP quorum mode

On a standard topology system, only the Standard quorum mode is supported. No additional configuration is required. On an ESC or HyperSwap topology, you can configure different tie-breaker scenarios (a tie occurs when exactly half of the nodes that were previously 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, leaving more time for the preferred site to win. If during an extended period a preferred site cannot contact the IP quorum app (for example, if it is destroyed), a non-preferred site continues working.
- If the quorum mode is set to **Winner**, the selected site always is the tie-breaker winner. If the winner site is destroyed, the remaining site can continue operating only after manual intervention.

The Preferred and Winner quorum modes are supported only with an IP quorum. For a FC-attached active quorum MDisk, only Standard mode is possible.

To set a quorum mode, select **Settings** → **IP Quorum** and click **Quorum Setting**. The Quorum Setting window opens, as shown in Figure 3-43.

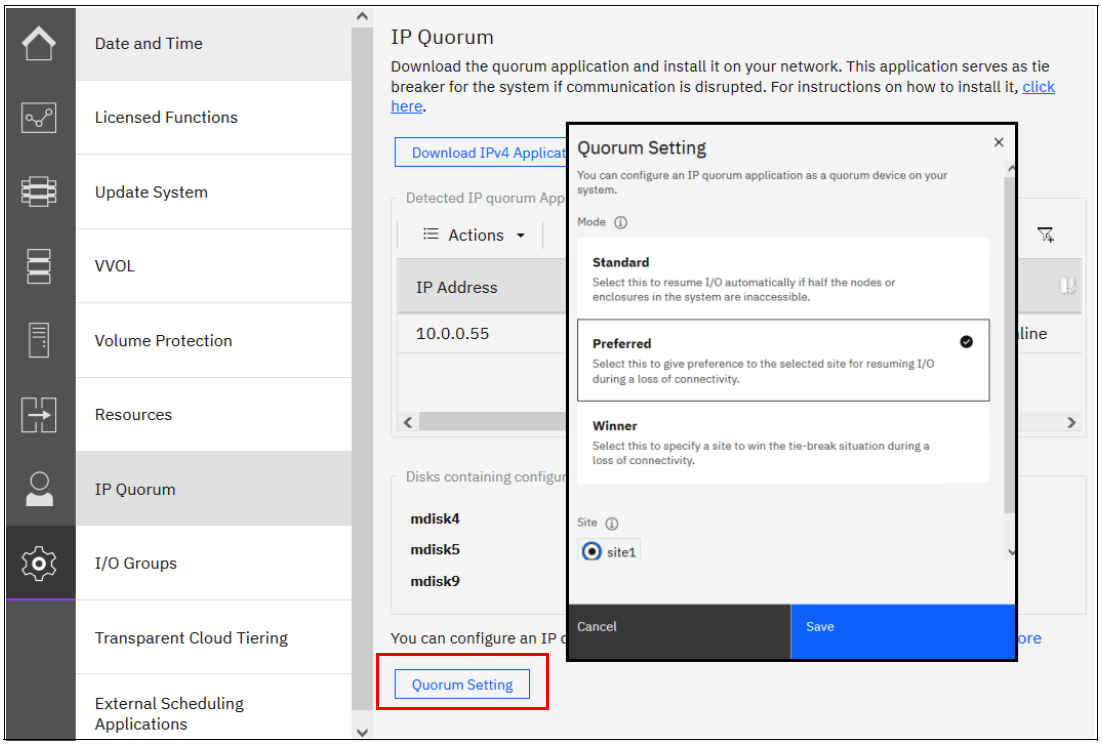


Figure 3-43 Changing the quorum mode

3.4.5 Configuring the local Fibre Channel port masking

With FC port masking, you control the usage of FC ports. By applying a mask, you restrict node-to-node communication or FC RC traffic on selected ports.

To decide whether your system must have port masks configured, see 2.6.8, “Port designation recommendations” on page 69.

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 3-44.

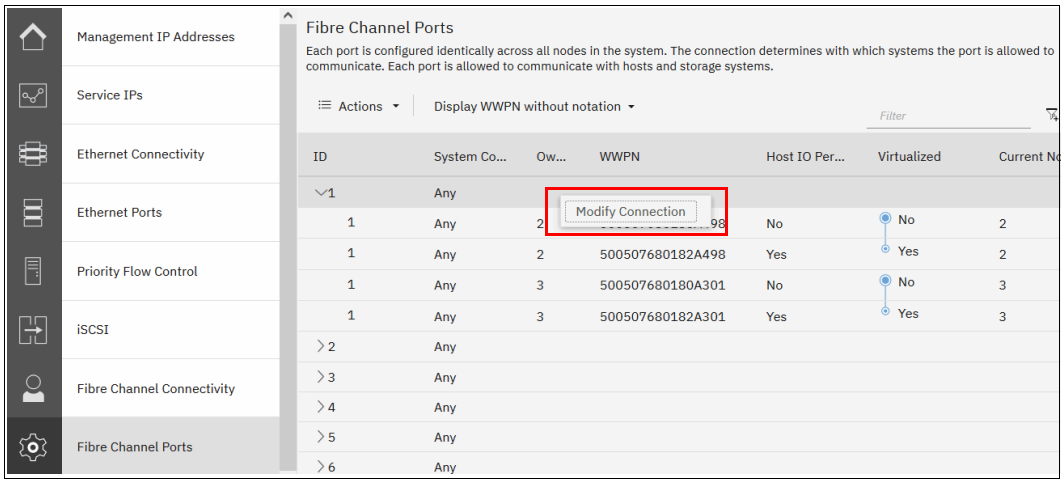


Figure 3-44 Applying a port mask by using a GUI

By default, all system ports can send and receive traffic of any kind:

- 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 3-45), you can choose which type of traffic that a port can send. Make your selection and click **Modify**.

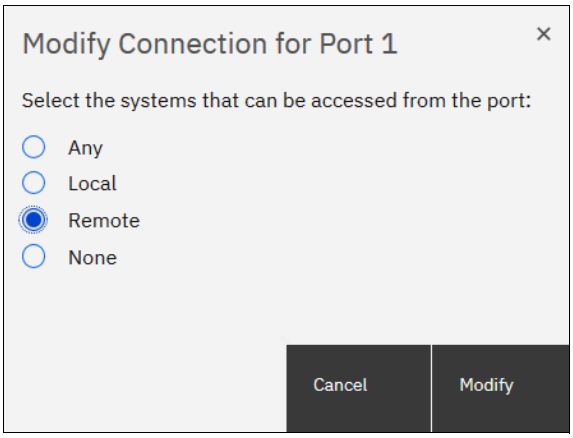


Figure 3-45 Modify Connection dialog box

Any	A port can work with all four types.
Local	Remote replication traffic is blocked on this port.
Remote	Blocks local node-to-node traffic.
None	Both local and remote systems traffic, but there is still system to host and system to back-end storage communication.

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 3-11. The output shows that all system ports allow all kinds of traffic.

Example 3-11 Viewing the local port mask

```
IBM_2145:ITSO-SVC:superuser>lssystem |grep mask
local_fc_port_mask 1111111111111111111111111111111111111111111111111111111111111111
partner_fc_port_mask 1111111111111111111111111111111111111111111111111111111111111111
```

To set the local or remote port mask, run the **chsystem** command. Example 3-12 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 3-12 Setting a local port mask by running the chsystem command

```
IBM_2145:ITSO-SVC:superuser>chsystem -localfcportmask 0011
IBM_2145:ITSO-SVC:superuser>chsystem -partnerfcportmask 1100
IBM_2145:ITSO-SVC: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 explicitly 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.

3.5 Configuring management access

The system can be managed by using the GUI and the CLI. Access to the system management interfaces require user authentication. User authentication and the secure communication implementation steps are described in this section.

3.5.1 Configuring secure communications

During system initialization, a *self-signed* Secure Sockets Layer (SSL) certificate is automatically generated by the system to encrypt communications between the browser and the system. Self-signed certificates generate web browser security warnings and might not comply with organizational security guidelines.

Signed SSL certificates are issued by a trusted CA. A browser maintains a list of trusted CAs that are identified by their *root* certificate. The root certificate must be included in this list in order for the signed certificate to be trusted.

To see the details of your system certificate, select **Settings** → **Security** and click **Secure Communications** (see Figure 3-46) or run the **`lssystemcert`** command.

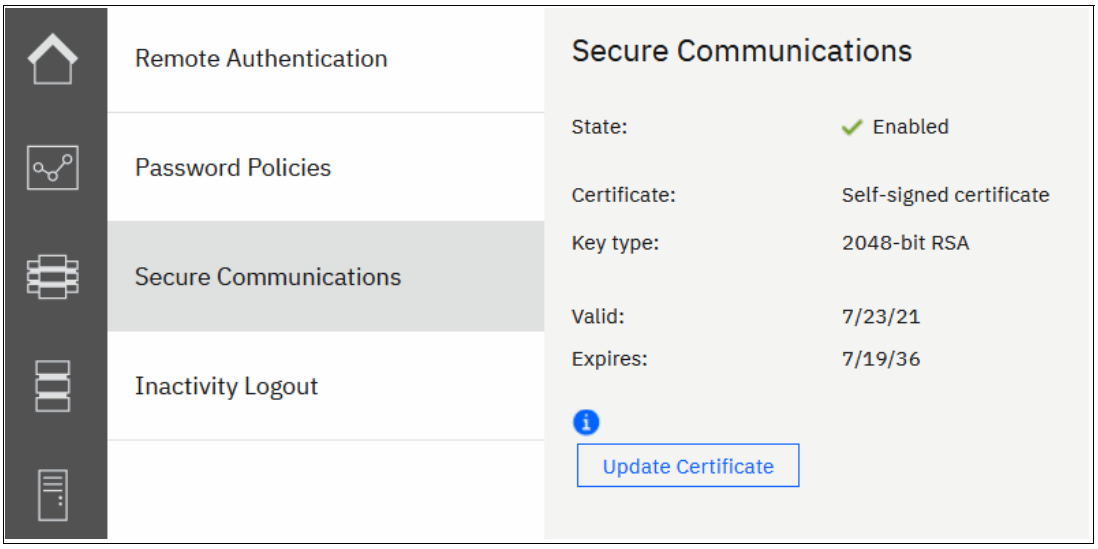


Figure 3-46 Accessing the Secure Communications window

Based on the security requirements for your system, you can create a self-signed certificate or install a signed certificate that was created by a third-party CA.

Generating a self-signed certificate

If a self-signed certificate is expired or its key type does not comply with your company's security policy, you can regenerate it. To renew a self-signed certificate, complete the following steps:

1. Select **Update Certificate** on the Secure Communications window, as shown in Figure 3-46.
2. Select **Self-signed certificate** and enter the details for the new certificate. The "Key type" and "Validity days" are the only mandatory fields.

> **Note:** Before re-creating a self-signed certificate, ensure that your browser supports the type of keys that you are going to use for a certificate. See your organization's security policy to ensure what key type is required.

3. Click **Update**.

You are prompted to confirm the action. Click **Yes** to proceed. Close the browser, wait approximately 2 minutes, and reconnect to the management GUI.

To regenerate an SSL certificate by using a CLI, run the **`chsystemcert`** command, as shown in Example 3-13. Valid values for **`-keytype`** are `rsa2048`, `ecdsa384`, or `ecdsa521`.

Example 3-13 Regenerating a self-signed certificate

```
IBM_2145:ITSO-SVC:superuser>chsystemcert -mkselfsigned -keytype ecdsa521 -validity
365
```

Configuring a signed certificate

If your company's security policy requests certificates to be signed by a trusted authority, complete the following steps to configure a signed certificate:

1. Select **Update Certificate** in the Secure Communications window.
2. Select **Signed certificate** and enter the details for the new certificate signing request, as shown in Figure 3-47. All fields are mandatory except for the Subject Alternative Name. For the "Country" field, use a two-letter country code. Click **Generate Request**.

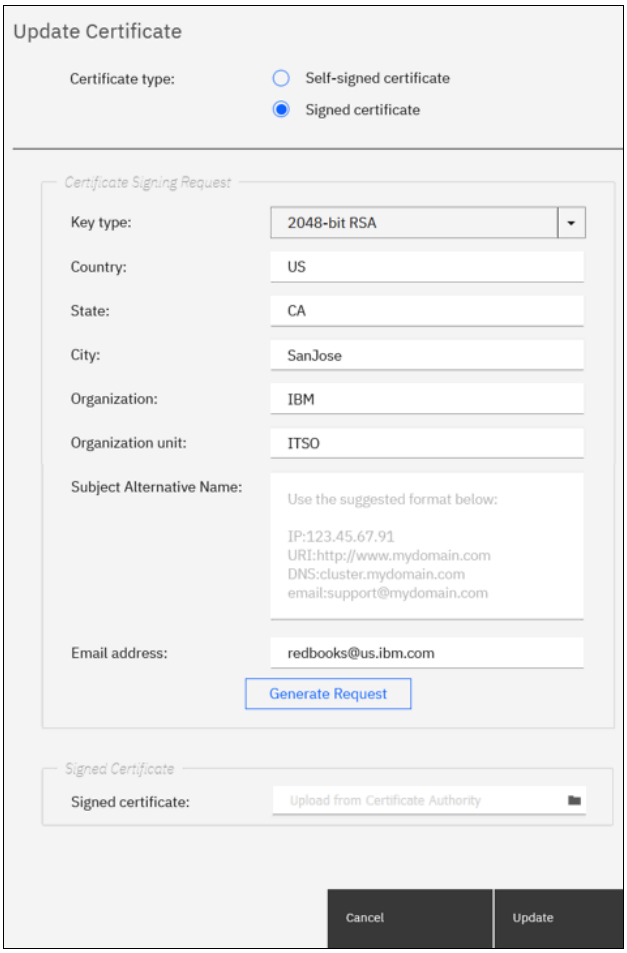


Figure 3-47 Generating a certificate request

3. When prompted, save the `certificate.csr` file that contains the certificate signing request.

   Until the signed certificate is installed, the Secure Communications window shows that an outstanding certificate request exists.

   Attention: If you must update a field in the certificate request, generate a new request and submit it for signing by the proper CA. However, this process invalidates the previous certificate request and prevents the installation of the signed certificate that is associated with the original request.

4. Submit the request to the CA to receive a signed certificate. Notify the CA that you need a certificate (or certificate chain) in base64-encoded Privacy Enhanced Mail (PEM) format.

5. When you receive the signed certificate, select **Update Certificate** in the Secure Communications window again.
6. Select **Signed Certificate** and click the folder icon that is next to the **Signed Certificate** input field of the Update Certificate window, as shown in Figure 3-47 on page 143. Click **Update**.
7. You are prompted to confirm the action. Click **Yes** to proceed. After your certificate is installed, the GUI session disconnects. Close the browser window and wait approximately 2 minutes before reconnecting to the management GUI.
8. Reconnect to the GUI and select **Settings** → **Security** → **Secure Communications**. The window that opens should show that you are using a signed certificate, as shown in Figure 3-48.

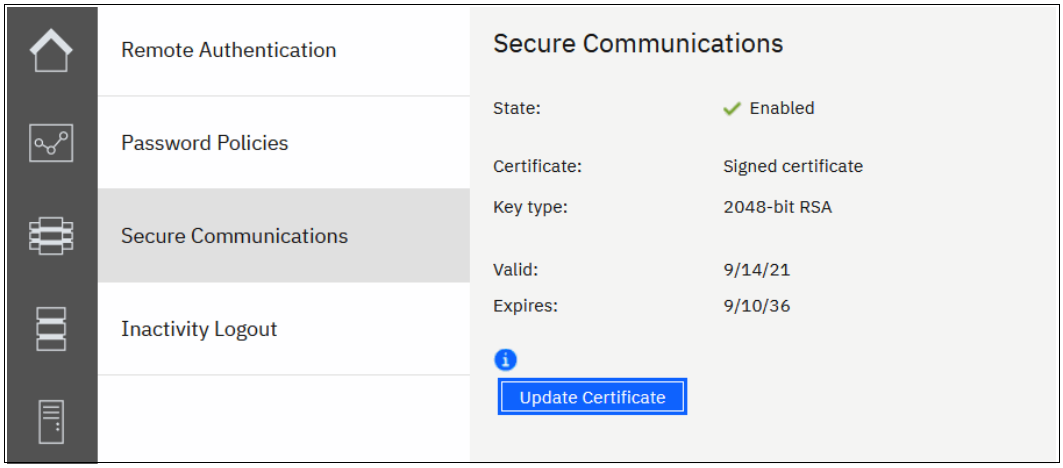


Figure 3-48 Signed certificate installed

3.5.2 Configuring password policies

A set of options were added to IBM Spectrum Virtualize v8.4.0 and later that allow a securityadmin to create policies for passwords, account lockout, and session timeout. A single system-wide policy applies to all local accounts (session timeouts also apply to remote accounts).

Password creation options

Password creation options can be customized to use the following policies:

Note: A new policy does not apply retroactively to existing passwords. However, any new passwords must meet the current policy setting.

- Minimum password length (6 - 64 characters)
- Minimum number of uppercase characters (1 - 3)
- Minimum number of lowercase characters (1 - 3)
- Minimum number of special characters (1 - 3)
- Minimum number of digits (1 - 3)
- Password History checking can be enabled.
- 0 - 10 previous passwords can be checked.
- Stores the previous password hashes only (for example, no plaintext).

- 0: Compare the current password only.
- 10: Check that the new password does not match the current password or the 10 passwords that were used before the current password.
- The minimum required password age can be set (0 – 365 days).

  A minimum age of 1 means that a user can change a password only once per day (which prevents a user from cycling through the history).

> **Note:** The history is not checked when a security administrator changes another user's password.

To set password creation options and password creation rules policies from the GUI, select **Settings** → **Security** and then, click **Password Policies**, as shown in Figure 3-49.

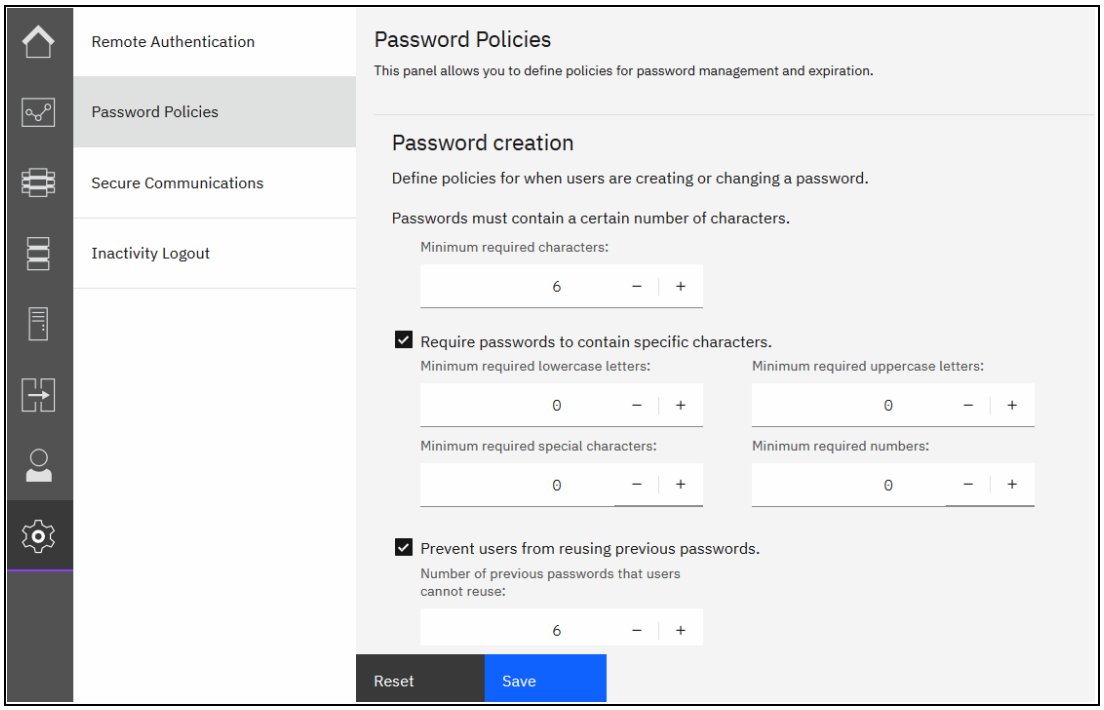


Figure 3-49 Creating password policies for password creation

Password expiration and account lockout

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 will be set to expire in *X* days when the setting is first enabled.
- A user with an expired password can log into the system, but cannot run any **`svctask`** commands until they change their password.
- An expiry warning can be enabled (0 - 30 days), which warns the user on login that their password expires in *X* days (only on the CLI on Spectrum Virtualize v8.4).

The security admin can force a user to change their password at any time. The password expires immediately (CLI – individual users, GUI –button to reset all user password):

- Can be used when creating a new user to require a password change on first login.
- Can be used after changing password policy settings.

The following methods can be used to force account locking:

- Manual

  The security admin can manually lock and unlock user accounts on the CLI, as shown in Example 3-14.

  Example 3-14 Manually locking and unlocking a user account

```
IBM_2145:IBM Redbook SVC:superuser>svctask chuser -lock bill
IBM_2145:IBM Redbook SVC:superuser>svctask chuser -unlock ted
```

  Note: A locked account is not allowed to log in to the system.

- Automatic

  Accounts can also be locked automatically by setting the maximum number of failed login attempts (0 - 10).

  Note: The counter is reset on a successful login.

  Accounts also can be locked automatically by setting the maximum number of failed login attempts (0 - 10) and by setting the length of time a user is locked out of the system (0 - 10080 minutes, which is 7 days, in which 0 = indefinite).

  Disabling the superuser account can be done from the GUI or CLI.

- The specific option to enable superuser locking is shown in Example 3-15.

  Example 3-15 To manually enable superuser account locking option

```
IBM_2145:IBM Redbook SVC:admin>svctask chsecurity -superuserlocking enabled
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
IBM_2145:IBM Redbook SVC:admin>
```

- The superuser can then be locked, as shown in Example 3-16.

  Example 3-16 To manually lock superuser account

```
IBM_2145:IBM Redbook SVC:admin>svctask chuser -lock superuser
IBM_2145:IBM Redbook SVC:admin>
```

A good use case is assuming some enterprises have policies that all systems should use remote authority. Therefore, configure remote authority, create a remote security admin, 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; for example, T3/T4 recovery. It is automatically unlocked for recovery and must be manually locked again later.

Session timeouts can be configured for CLI and GUI sessions:

- Configurable CLI timeout of 5 - 240 minutes
- Separate configurable GUI timeout of 5 - 240 minutes

The set of options to govern password expiry, require a password change, account lockouts, disable the superuser account, and session timeouts can be done from the GUI by selecting **Settings** → **Security** and clicking **Password Policies**, as shown in Figure 3-50.

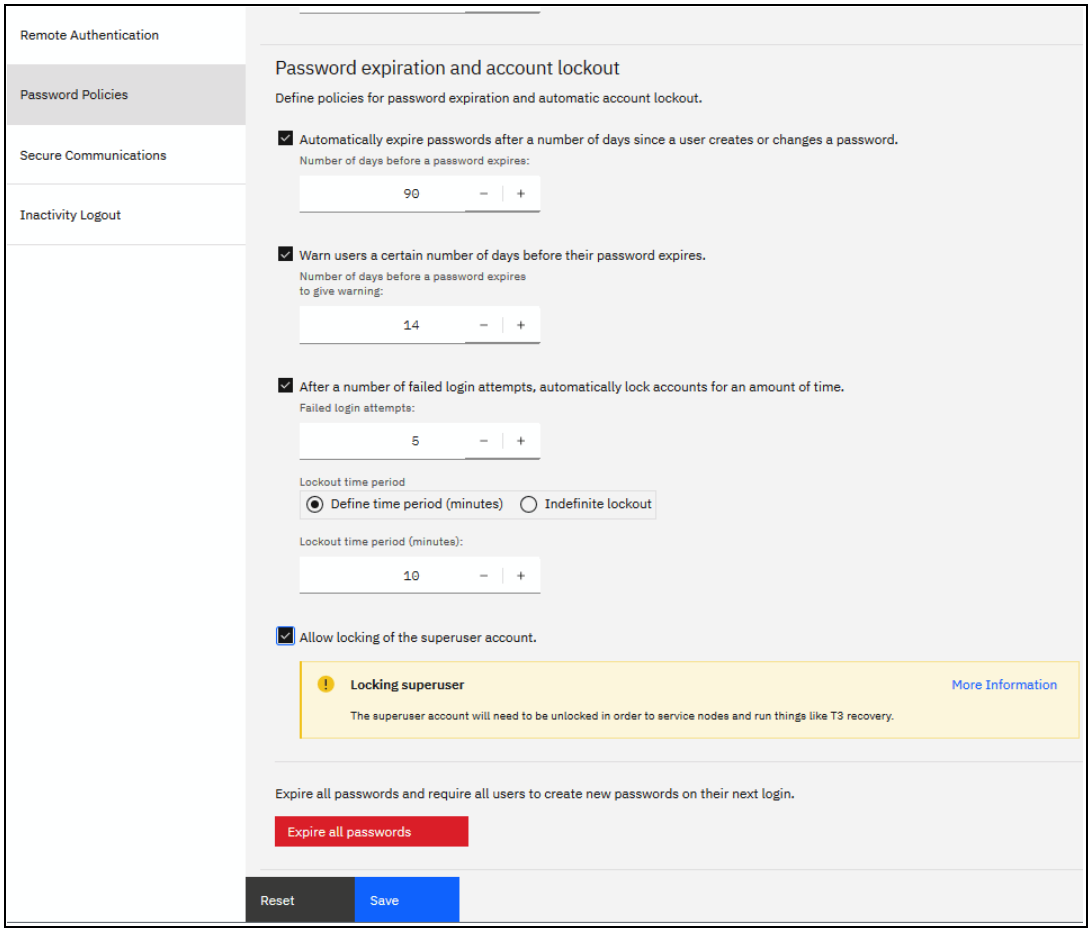


Figure 3-50 Creating password polices for password expiration and account lockout

3.5.3 Configuring user authentication

The following user authentication methods can be used to control access to the GUI and CLI:

- *Local user authentication* is performed within the SAN Volume Controller system. GUI users authenticate with user name and password. CLI users must provide a user name and a Secure Shell (SSH) public key or a password.
- *Remote user authentication* allows users to authenticate to the system by using credentials that are stored on an external authentication service. With this feature, you use user credentials 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.

Locally administered users can coexist with remote authentication.

User roles and groups

User groups are used to determine what tasks the user is authorized to perform. Each user group is associated with a single role. Roles apply to both local and remote users on the system and are based on the user group to which the user belongs. A local user can belong only to a single group, so the role of a local user is defined by the single group to which that user belongs.

For a list of user roles and their tasks, and a description of a preconfigured user group, see this IBM Documentation web page.

Superuser account

Every system has a default user that is called the *superuser*. It cannot be deleted or modified, except for changing the password and SSH key. The superuser is a *local* user and cannot be authenticated remotely. The superuser has a *SecurityAdmin* user role, which has the most privileges within the system.

Note: The superuser is the only user that can log in to the Service Assistant interface. It is also the only user that can run **`sainfo`** and **`satask`** commands through the CLI.

The password for superuser is set during the system setup. The superuser password can be reset to its default value of `passw0rd` by using a procedure that is described in IBM Documentation by expanding **Troubleshooting** → **Resolving a problem** → **Procedure: Resetting the superuser password.**

Note: The superuser password reset procedure uses system internal USB ports. Systems with IBM SAN Volume Controller 2145-SV2 and 2145-SA2 nodes can be configured to disable those ports. No workaround is available if the USB ports are disabled and no users with the SecurityAdmin role and a known password exist. In this case, the superuser password cannot be reset without replacing the system hardware and deleting the system configuration.

Local authentication

A *local user* is a user whose account is managed entirely on the system. A local user belongs to one user group only, and it must have a password, an SSH public key, or both. Each user has a name, which must be unique across all users in one system.

User names can contain up to 256 printable American Standard Code for Information Interchange (ASCII) characters. Forbidden characters are the single quotation mark ('), colon (:), percent symbol (%), asterisk (*), comma (,), and double quotation marks (“). A user name 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 connecting to the CLI, encryption key authentication is attempted first with the user name and password combination available as a fallback. The SSH key authentication method is available for CLI and file transfer access only. For GUI access, only the password is used.

To add a user that is authenticated without a password by using only an SSH key, select **Access** → **Users by Group**, click **Add user**, and then, click **Browse** to select the SSH public key for that user, as shown in Figure 3-51 on page 149. The Password field can be left blank. The system accepts public keys that are generated by PuTTY (SSH2), OpenSSH, and RFC 4716-compliant keys that are generated by other clients. Click **Create** when done.

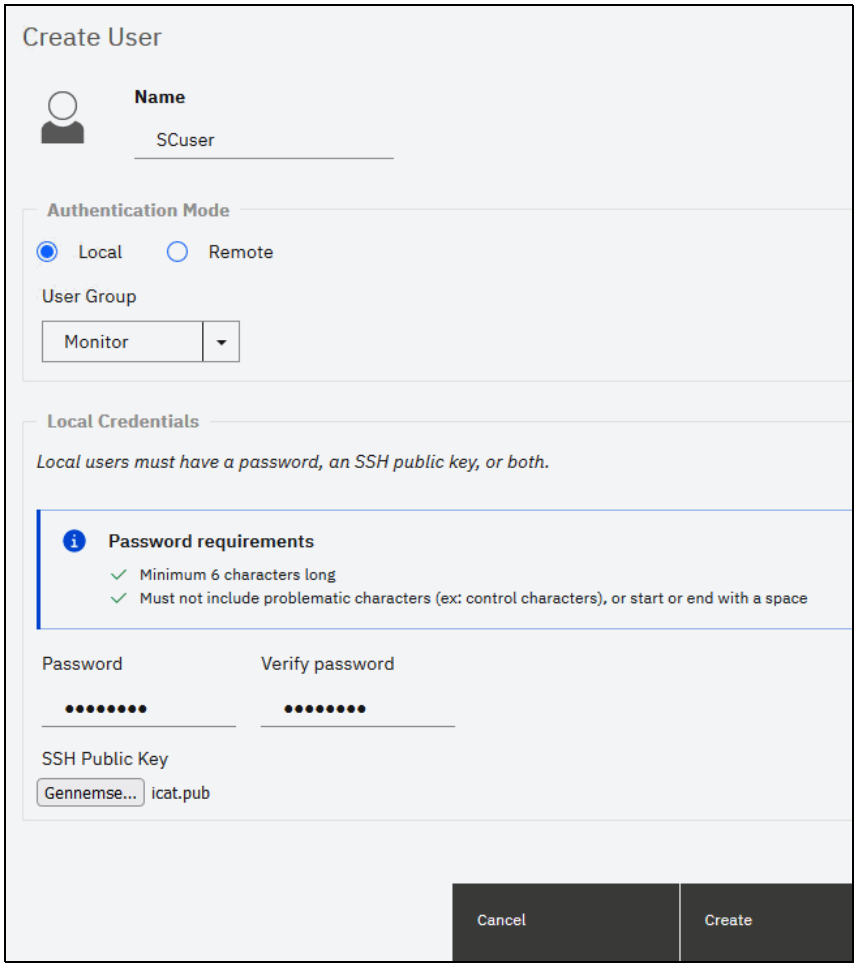


Figure 3-51 Creating a user authenticated by a SSH key

If local authentication is used, user accounts must be created for each system. If you want access for a user on multiple systems, you must define the user in each system.

Remote authentication

A *remote user* is authenticated by using identity information that is accessible by using the Lightweight Directory Access Protocol (LDAP). The LDAP server must be available for the users to log in to the system. Remote users have their groups defined by the remote authentication service.

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 more than one LDAP server to improve resiliency. Authentication requests are processed by those LDAP servers that are marked as preferred 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 the same system must be of the same type.

If users that are part of a group on the LDAP server are to be authenticated remotely, a user group with an identical name must exist on the system. The user group name is *case-sensitive*. The user group must also be enabled for remote authentication on the system.

A user who is authenticated remotely is granted permissions according to the role that is assigned to the user group of which the user is a member.

To configure remote authentication by using LDAP, start by enabling remote authentication by completing the following steps:

1. Select **Settings** → **Security**, click **Remote Authentication**, and then, click **Configure Remote Authentication**, as shown in Figure 3-52.

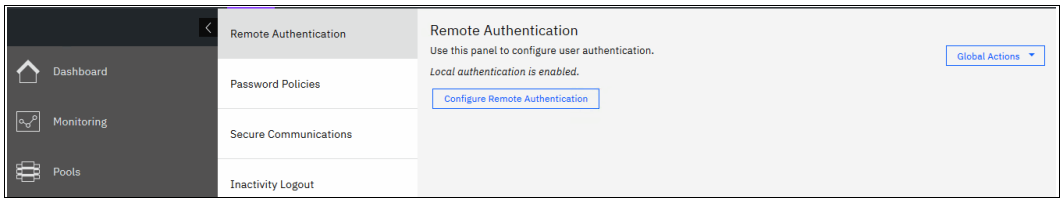


Figure 3-52 Configuring remote authentication

2. Enter the LDAP settings. These settings are not server-specific. They are applied to all LDAP servers that are configured in the system. Extra optional settings are available by clicking **Advanced Settings**, as shown in Figure 3-53.

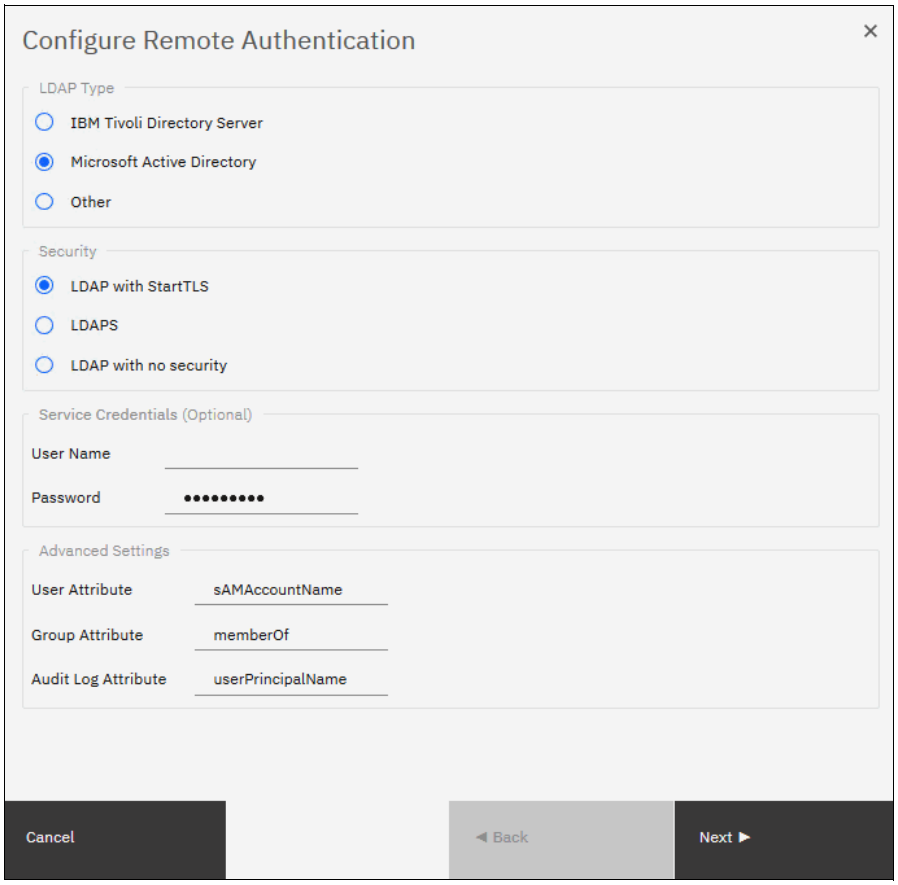


Figure 3-53 Configure Remote Authentication settings

The following settings are available:

- LDAP type:
  - IBM Security Directory Server (for IBM Security Directory Server).
  - Microsoft Active Directory (AD).
  - Other (other LDAP v3-capable directory servers, for example, OpenLDAP).
- Security:
  - LDAP with StartTLS: Select this option to use the StartTLS extension (RFC 2830). It works by establishing a non-encrypted connection with an LDAP server on a standard LDAP port (389), and then performing a TLS handshake over an existing connection.
  - LDAPS: Select to use LDAP over SSL and establish secure connections by using port 636.
  - None: Select to transport data in clear text format without encryption.
- Service Credentials: Sets a user name and password for administrative binding (the credentials of a user that has the authority to query the LDAP directory). Leave it empty if your LDAP server is configured to support anonymous bind.

  For AD, a user name must be in User Principal Name (UPN) format.
- Advanced settings

  Speak to the administrator of the LDAP server to ensure that these fields are completed correctly:

  - User Attribute

    This LDAP attribute is used to determine the user name of remote users. The attribute must exist in your LDAP schema and must be unique for each of your users.

    This advanced setting defaults to `sAMAaccountName` for AD and to `uid` for IBM Security Directory Server and Other.
  - Group Attribute

    This LDAP attribute is used to determine the user group memberships of remote users. The attribute must contain either the distinguished name of a group or a colon-separated list of group names.

    This advanced setting defaults to `memberOf` for AD 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

    This LDAP is an attribute that is used to determine the identity of remote users. When an LDAP user performs an audited action, this identity is recorded in the audit log. This advanced setting defaults to `userPrincipalName` for AD and to `uid` for IBM Security Directory Server and the Other type.

3. Enter the server settings for one or more LDAP servers, as shown in Figure 3-54.

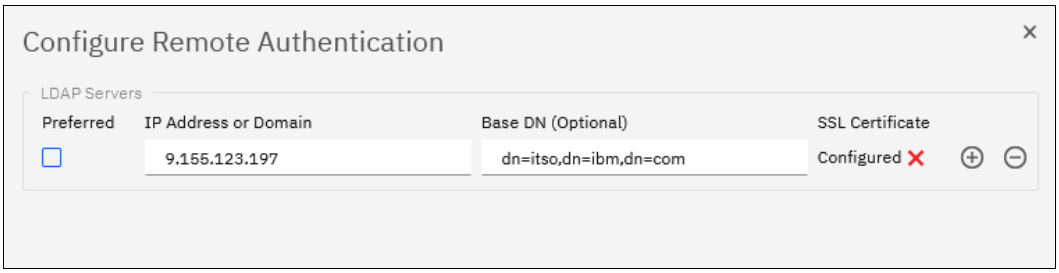


Figure 3-54 Configure Remote Authentication: Creating an LDAP server

The following settings are available:

– Preferred

  One or more configured LDAP servers can be marked as Preferred. Requests are distributed among these servers, and use only non-preferred servers if all the preferred servers failed.

– IP Address

  The IP address of the server.

– Base DN

  The distinguished name to use as a starting point for searching for users on the server (for example, `dc=itso,dc=ibm,dc=com`).

– SSL Certificate

  The SSL certificate that is used to securely connect to the LDAP server. This certificate is required only if you chose to use SSL or Transport Layer Security as a security method earlier.

  Note: If your organization is using a tiered CA hierarchy, a server certificate that is exported for use on a system must include all the certificates in a chain. To accomplish this task, export the certificate in MS Windows in `.P7B` format and use third-party tools (OpenSSL) to convert it to PEM format. Otherwise, the exported certificate will not contain all certificates in the certification path.

  If you set a certificate and you want to remove it, click the red cross that is next to **Configured**.

a. Click the plus (+) and minus (-) signs to add or remove LDAP server records. You can define up to six servers.

b. Click **Finish** to save the settings.

4. To verify that LDAP is enabled, select **Settings** → **Security** → **Remote Authentication**, as shown in Figure 3-55. You also can test the server connection by selecting **Global Actions** → **Test LDAP connections** and verifying that all servers return "`CMMVC7075I The LDAP task completed successfully`".

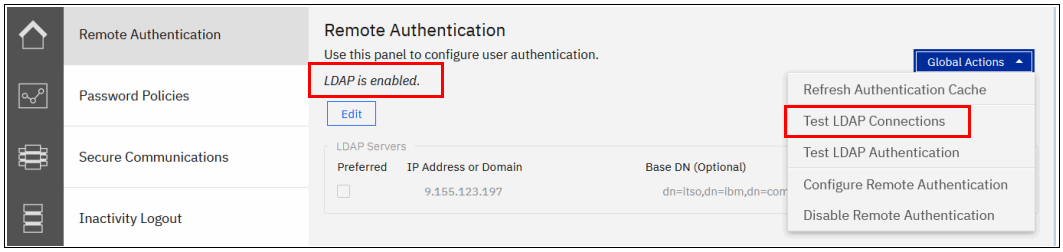


Figure 3-55 Verifying that LDAP is enabled

Use the **Global Actions** menu to disable remote authentication and switch to local authentication only.

After remote authentication is enabled, the remote user groups must be configured. You can use the default built-in user groups for remote authentication. However, the name of the default user groups cannot be changed. If the LDAP server contains a group that you want to use and you do not want to create this group on the storage system, the name of the group must be changed on the server side to match the default name. Any user group, whether default or self-defined, must be enabled for remote authentication before LDAP authentication can be used for that group.

To create a user group with remote authentication enabled, complete the following steps:

1. Select **Access** → **Users by Group** and click **Create User Group**. Enter the name for the new group, select the **LDAP** option, and then, choose a role for the users in the group, as shown in Figure 3-56.

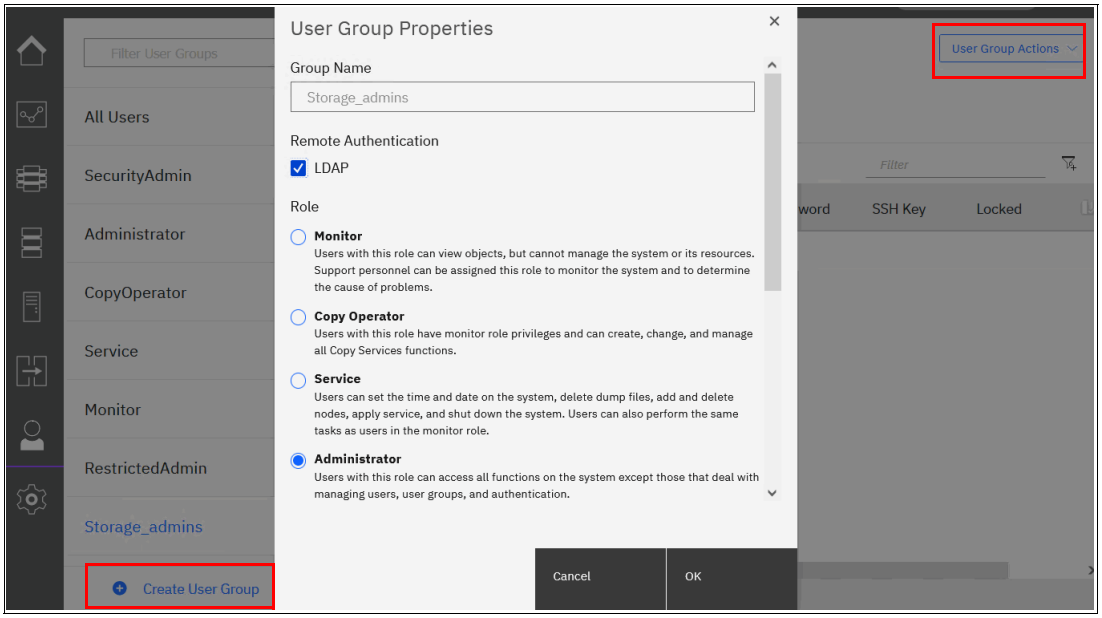


Figure 3-56 Creating a user group with remote authentication enabled

2. To enable LDAP for one of the existing groups, select it in the list, select **User Group Actions** → **Properties** in the upper right corner, and then, select the LDAP option.

3. When you have at least one user group that is enabled for remote authentication, verify that you set up your user group on the LDAP server correctly by checking whether 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 system role.
4. To test the user authentication, select **Settings** → **Security** → **Remote Authentication**, and then, select **Global Actions** → **Test LDAP Authentication** (for an example, see Figure 3-55 on page 153). Enter the user credentials of a user that is defined on the LDAP server and click **Test**. A successful test returns the following message:

   `CMMVC7075I The LDAP task completed successfully.`

A user can log in with their short name (that is, without the domain component) or with the fully qualified user name in the UPN format (`user@domain`).

4

IBM Spectrum Virtualize GUI

This chapter describes an overview of the IBM Spectrum Virtualize GUI. The management GUI is a tool that is enabled and provided by IBM Spectrum 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 IBM SAN Volume Controller environment running as quickly as possible by using the GUI.

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).

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 more in detail in other chapters of the book only are referred to here. For example, Storage pools (Chapter 5, “Using storage pools” on page 271), Volumes (Chapter 6, “Creating and provisioning volumes” on page 331), Hosts (Chapter 7, “Hosts” on page 443), and Copy Services (Chapter 10, “Advanced Copy Services” on page 605) are described in separate chapters.

This chapter includes the following topics:

- 4.1, “Normal operations by using the GUI” on page 156
- 4.2, “GUI introduction” on page 160
- 4.3, “System Hardware - Overview window” on page 167
- 4.4, “Monitoring menu” on page 171
- 4.5, “Pools” on page 178
- 4.6, “Volumes” on page 179
- 4.7, “Hosts” on page 180
- 4.8, “Copy Services” on page 180
- 4.9, “Access” on page 181
- 4.10, “Settings” on page 198
- 4.11, “Other frequent tasks in the GUI” on page 257

© Copyright IBM Corp. 2021. All rights reserved.

4.1 Normal operations by using the GUI

This section describes the graphical icons and the indicators that you use to manage IBM Spectrum Virtualize. For the example in this book, we configure the system in a standard topology and not in a Enhanced Stretched- or Split Cluster configuration.

Multiple users can be logged in to the GUI. However, no locking mechanism exists, so be aware that if two users change the same object simultaneously, the last action that is entered from the GUI is the action that takes effect.

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 IBM SAN Volume Controller Documentation.

4.1.1 Accessing the GUI

To access the IBM SAN Volume Controller GUI, enter the Internet Protocol (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 (see Figure 4-1).

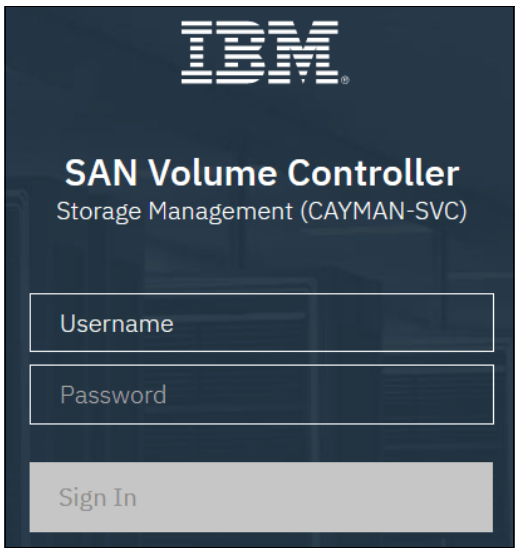


Figure 4-1 Login window of IBM SAN Volume Controller

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 on page 157.

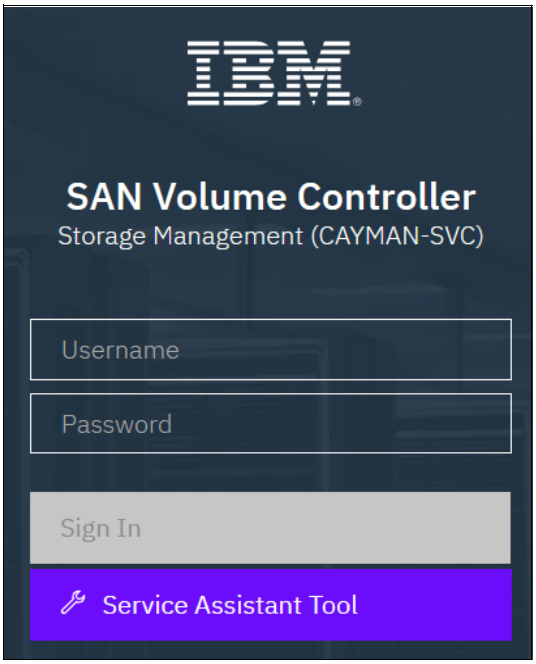


Figure 4-2 Login window of IBM SAN Volume Controller when connected to the configuration node

It is a best practice for each user to have their own unique account. The default user accounts should be disabled for use or their passwords changed and kept secured for emergency purposes only. This approach helps to identify personnel working on the systems and track all important changes that are done by them. The *superuser* account should be used for initial configuration only.

After a successful login, the Version 8.4 Welcome window opens and displays the new system dashboard (see Figure 4-3).

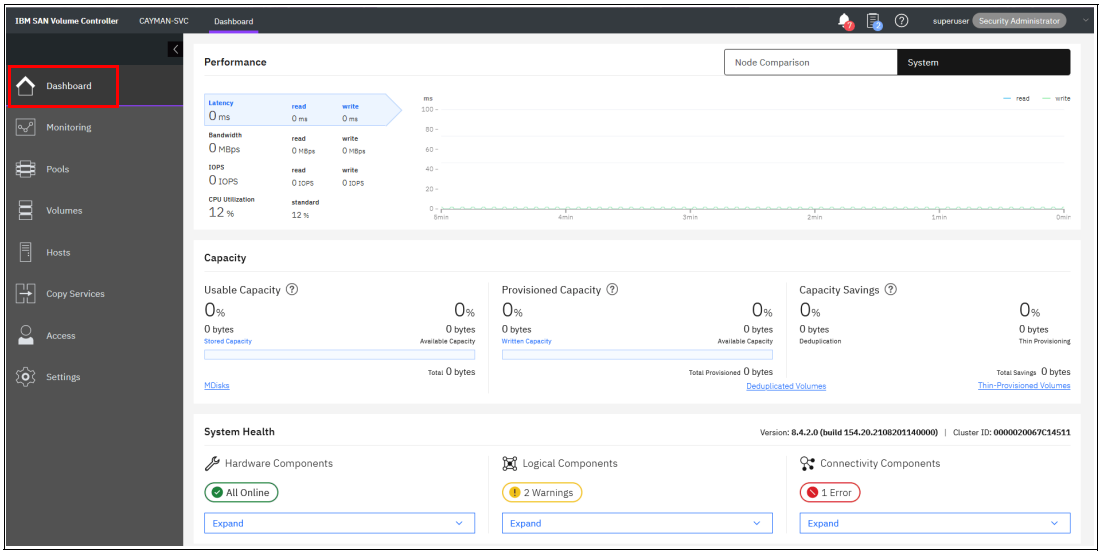


Figure 4-3 Welcome page with new dashboard

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 either the IBM SAN Volume Controller system level or node level. A “Node comparison” view shows the differences in characteristics of each node (see Figure 4-4). The performance graph is updated with new data every 5 seconds.

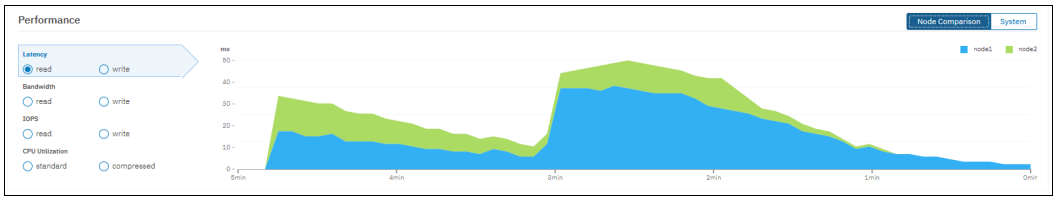


Figure 4-4 Performance statistics

- Capacity

  This section shows the current utilization of attached storage and its usage. Apart from the usable capacity, it also shows provisioned capacity and capacity savings. You can select the **Compressed Volumes**, **Deduplicated Volumes**, or **Thin Provisioned Volumes** options to display a complete list of the options in the Volumes tab.

  The “Overprovisioned External Storage” section (highlighted in the red box in Figure 4-5), appears only when attached storage systems that are overprovisioned are present.

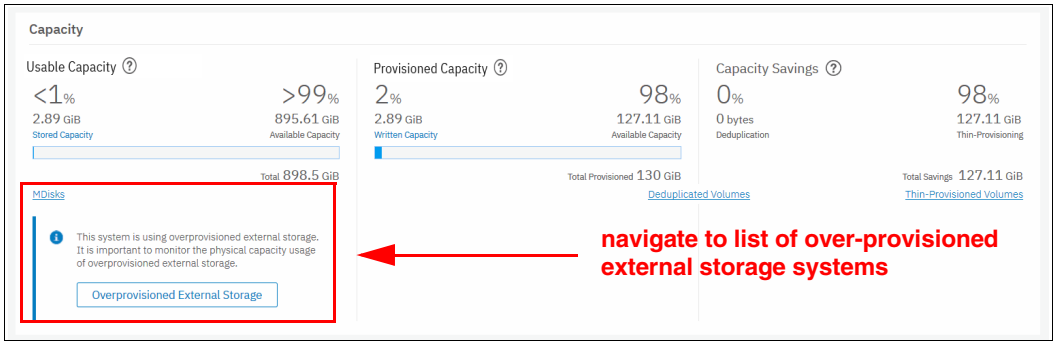


Figure 4-5 Capacity overview

Selecting this option provides a list of Overprovisioned External Systems that you can select to see a list of related managed disks (MDisks) and pools, as shown in Figure 4-6.

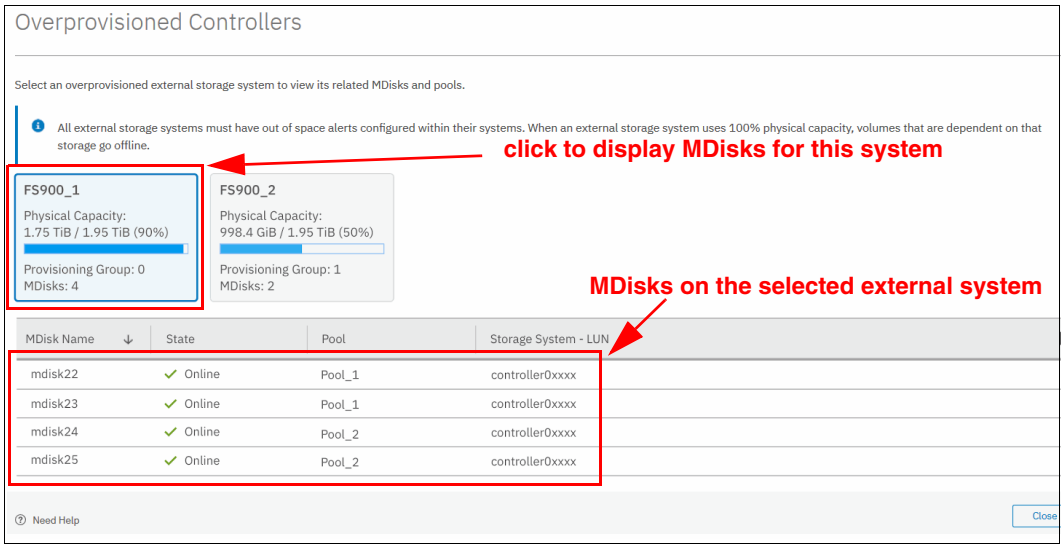


Figure 4-6 List showing overprovisioned external storage

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 4-7. 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.

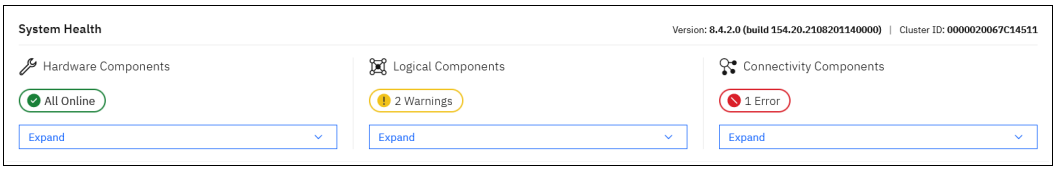


Figure 4-7 System health overview

Figure 4-8 shows the expanded view.

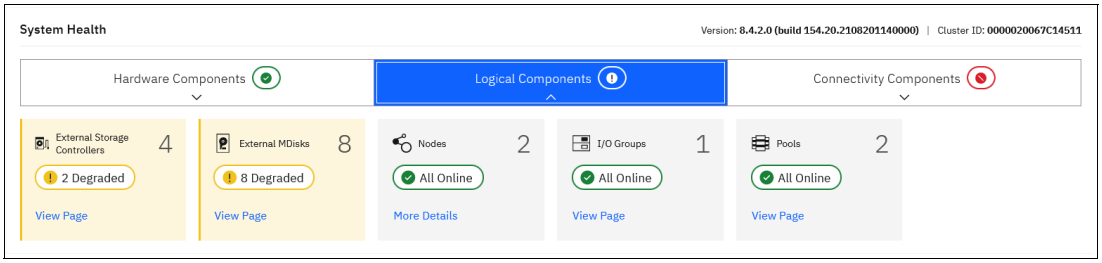


Figure 4-8 Expanded System health view

The dashboard displays as a welcome page instead of the system pane as in previous versions. This system overview was moved to the **Monitoring** → **System Hardware** menu.

Although the Dashboard pane provides key information about system behavior, the System Hardware menu is a preferred starting point to obtain the necessary details about your system components.

4.2 GUI introduction

As shown in Figure 4-9, the former IBN SAN Volume Controller GUI System pane was relocated to **Monitoring** → **System Hardware**.

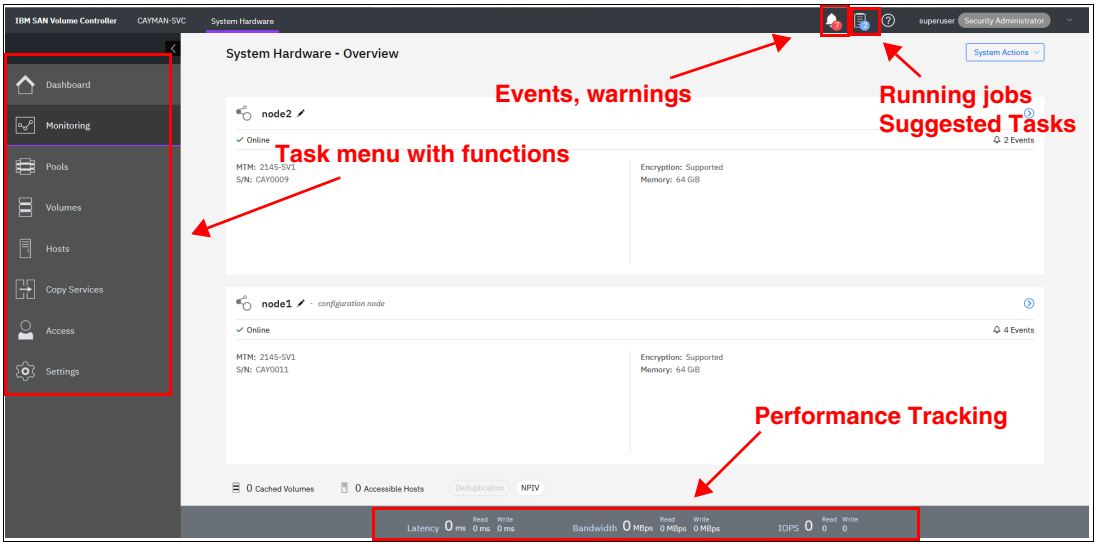


Figure 4-9 IBM SAN Volume Controller System window

4.2.1 Task menu

The IBM Spectrum 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 4-10.

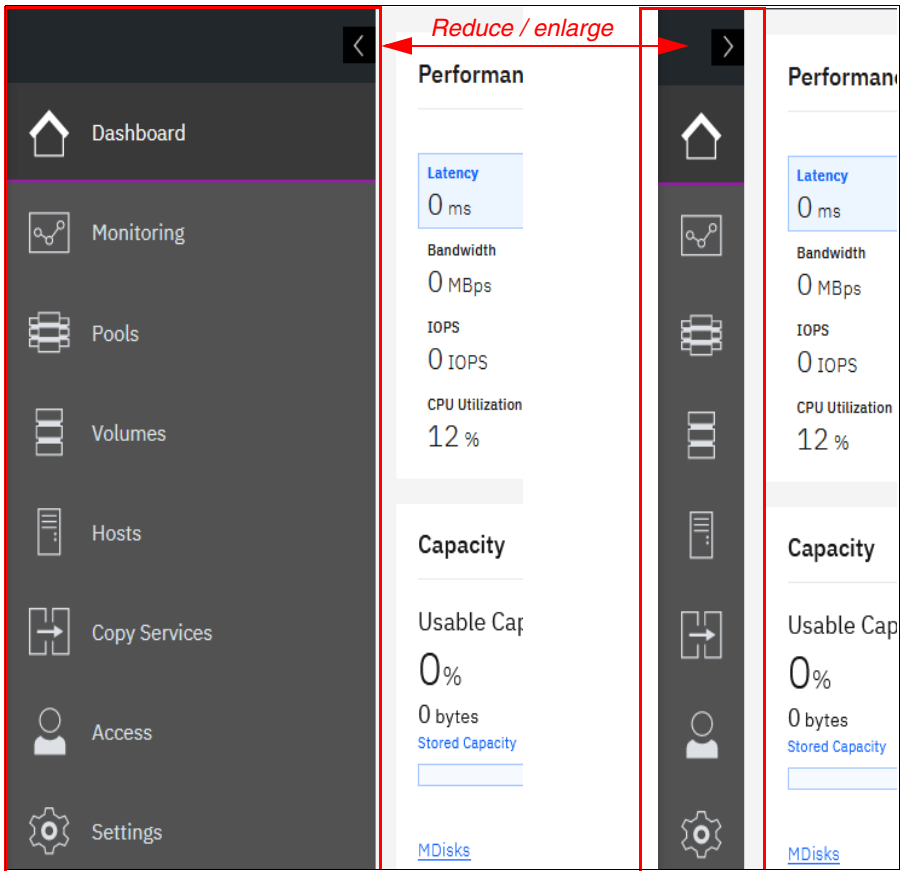


Figure 4-10 The task menu on the left side of the GUI

By reducing the horizontal size of your browser window, the wide task menu shrinks to the icons only.

4.2.2 Suggested tasks

After the initial configuration process is complete, IBM Spectrum Virtualize shows the information about suggested tasks to notify the administrator that several key functions are not yet configured. If necessary, this indicator can be closed and these tasks can be performed later. Figure 4-11 shows the suggested tasks in the System pane.

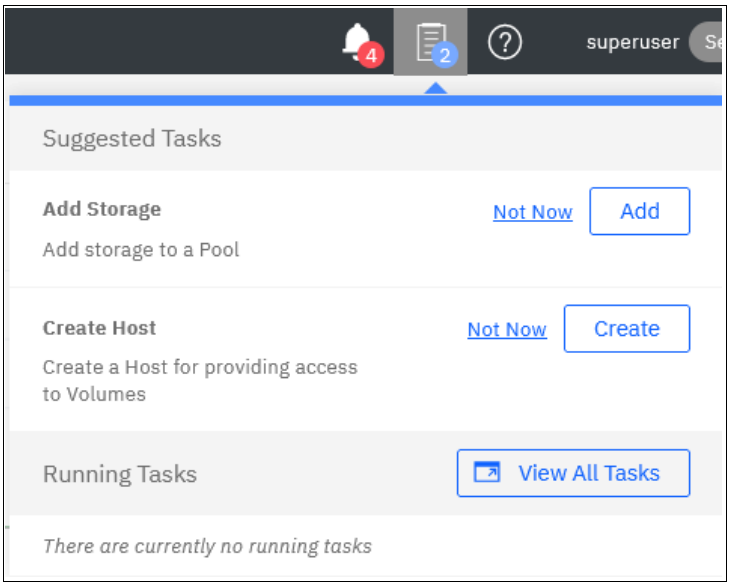


Figure 4-11 Suggested tasks

In this case, the GUI has two suggested tasks that help with the general administration of the system: You can directly perform the tasks from this window, or cancel them and run the procedure later. Other suggested tasks that typically appear after the initial system configuration are to create a volume and configure a storage pool.

The dynamic IBM Spectrum Virtualize menu contains the following panes:

- Dashboard
- Monitoring
- Pools
- Volumes
- Hosts
- Copy Services
- Access
- Settings

4.2.3 Notification icons and help

Three notification icons are in the upper navigation area of the GUI (see Figure 4-12). 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 4-12 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 4-13).

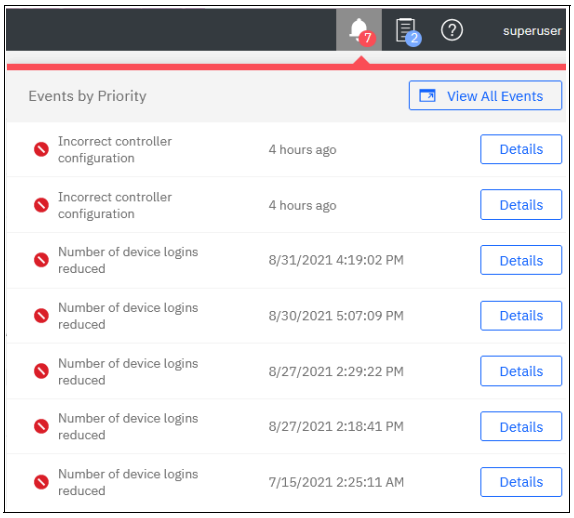


Figure 4-13 System alerts

You can go directly to the Events menu by clicking the **View All Events** option, as shown in Figure 4-14.

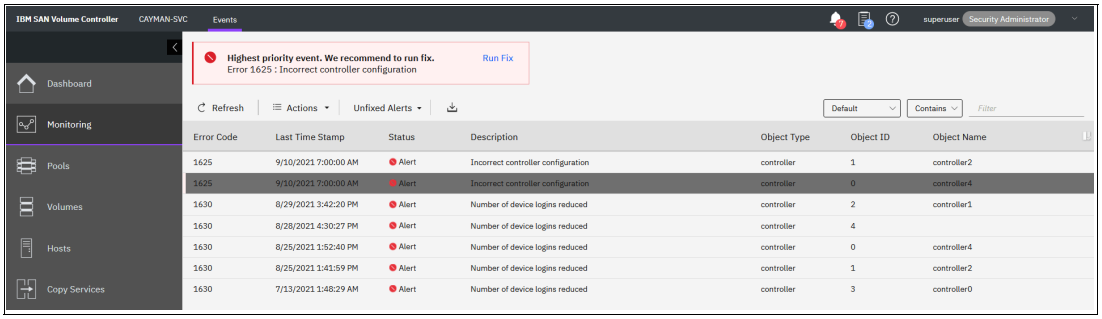


Figure 4-14 View all events

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 run the suggested fix procedure, as shown in Figure 4-15.

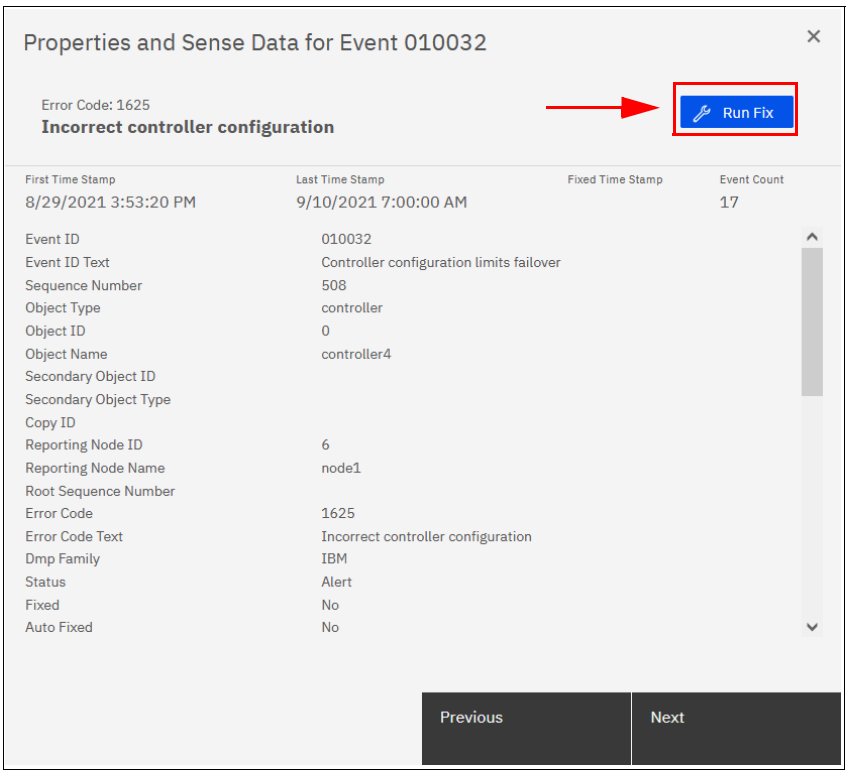


Figure 4-15 Incorrect controller configuration

Running tasks and suggested tasks

The middle icon in the notification area provides an overview of currently running tasks that are triggered by administrator. It also includes the suggested tasks that recommend that users perform specific configuration actions.

In the example that is shown in Figure 4-16, a Storage Pool is not yet defined in the system. Therefore, the system suggests that we do so and offers us direct access to the associated Add Storage menu. As a second task, you can create a Host. Click **Create** to create a Host, as described in Chapter 7, “Hosts” on page 443. If you do not want to create a host now, click **Not Now** and the suggestion message disappears.

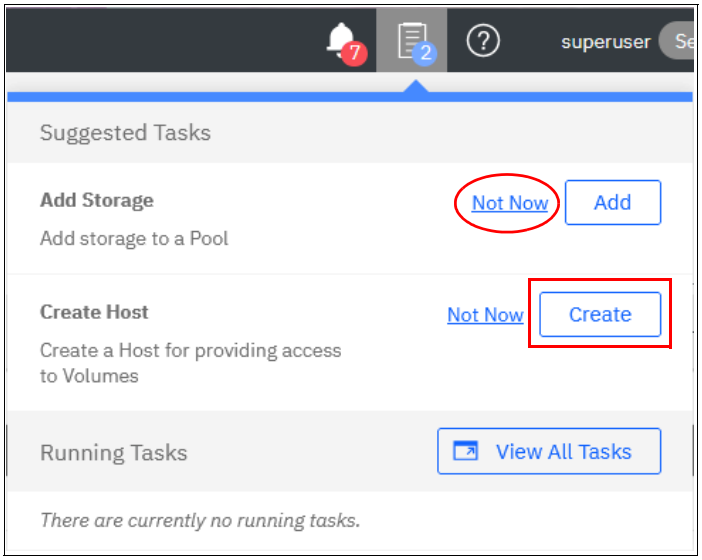


Figure 4-16 Suggested Tasks

Similarly, you can analyze the details of running tasks (all of them together in one window or of a single task). Click **View** to open the Array Initialization job, as shown in Figure 4-17.

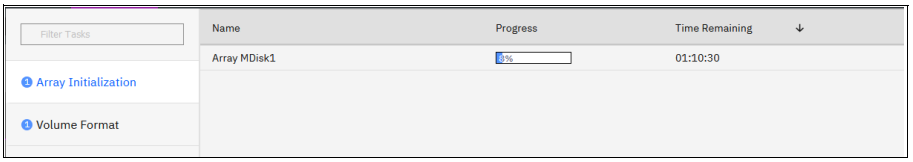


Figure 4-17 Details of running task

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

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.

While pressing and holding the Shift key, 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 4-18.

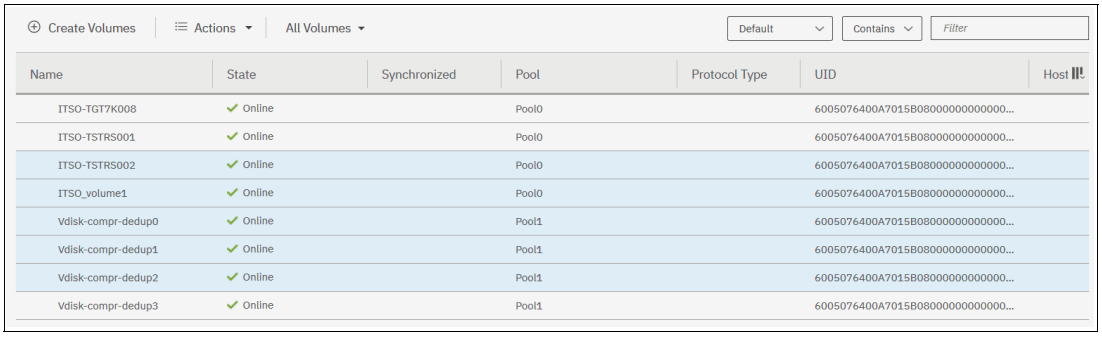


Figure 4-18 Selecting items by using the Shift key

While pressing and holding the Ctrl key, select any items from the entire list. You can select items that do not appear in sequential order, as shown in Figure 4-19.

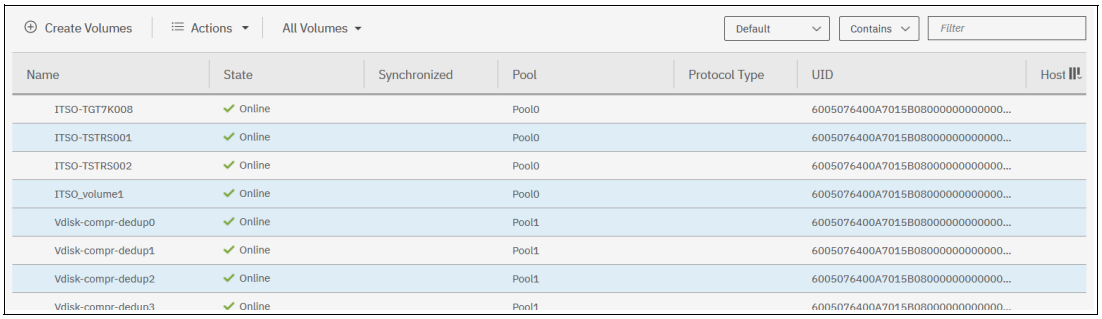


Figure 4-19 Selecting items by using the Ctrl key

You also can select items by using the built-in filtering function. For more information, see 4.3.1, “Content-based organization” on page 168.

Help

If you need help, select the (**?**) button, as shown in Figure 4-20. The following options are available:

- The first option opens a new tab with plain text information about the pane you are on 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 IBM SAN Volume Controller.

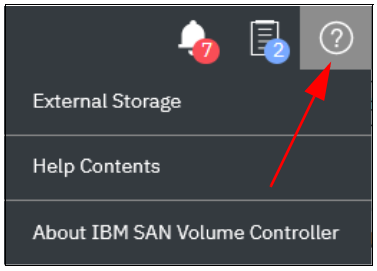


Figure 4-20 Access help menu

For example, in the System Hardware pane, you can open help that is related to the system in general, as shown in Figure 4-21.

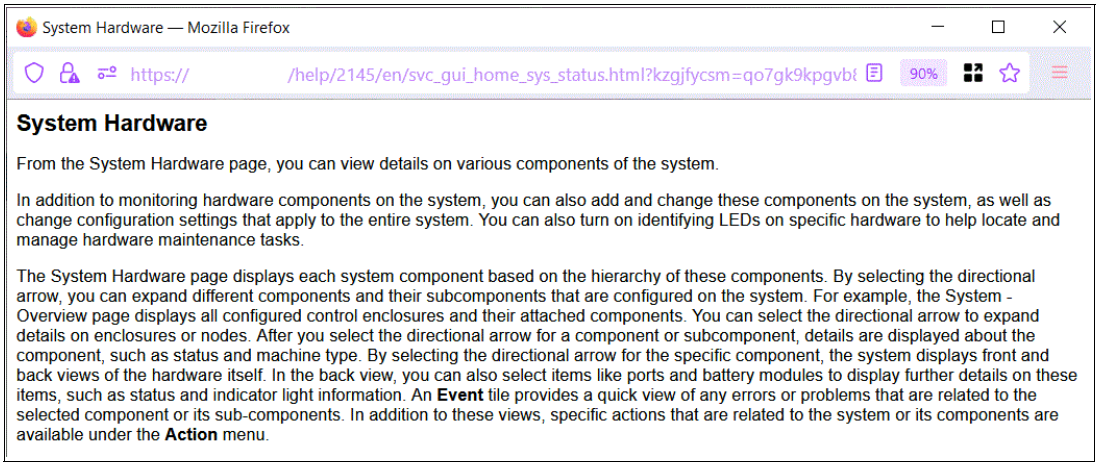


Figure 4-21 Example of System Hardware content

4.3 System Hardware - Overview window

The System Hardware - Overview window is shown in Figure 4-22.

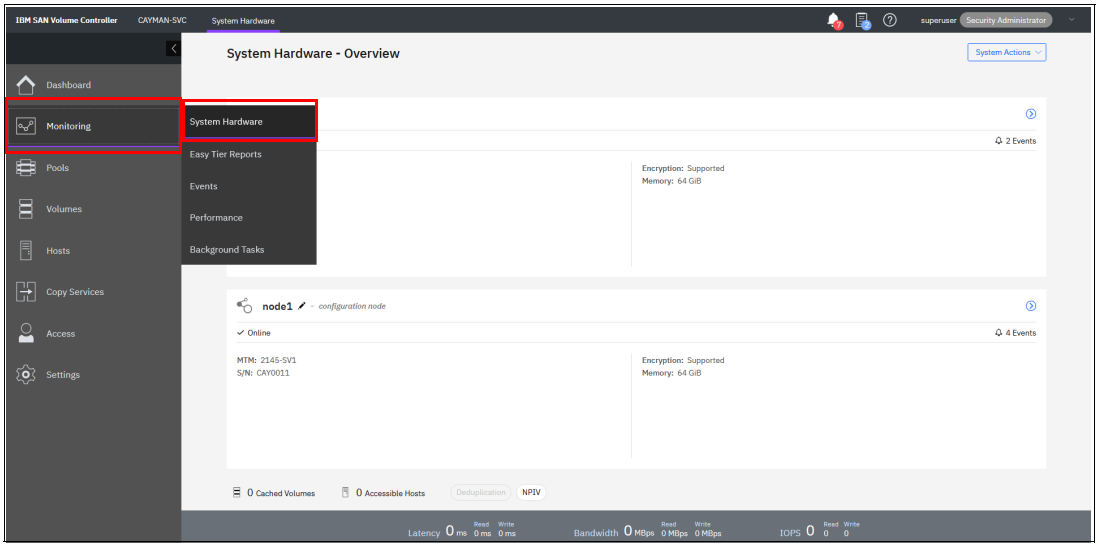


Figure 4-22 The System Hardware - Overview window

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 side 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. As shown in the Filter box that is shown in Figure 4-23, 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.

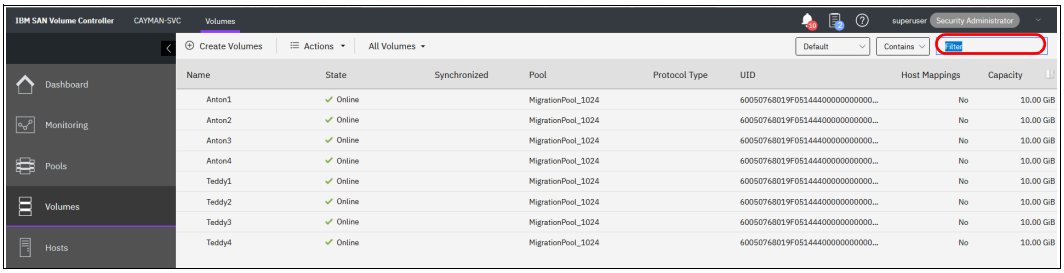


Figure 4-23 Filter search box

2. Enter the text string that you want to filter and press **Enter**.

   By using this function, you can filter your table that is based on column names. In our example, a volume list is displayed that contains the names that include **Anton** somewhere in the name. **Anton** is highlighted in amber, as are any columns that contain this information, as shown in Figure 4-24. The search option is not case-sensitive.

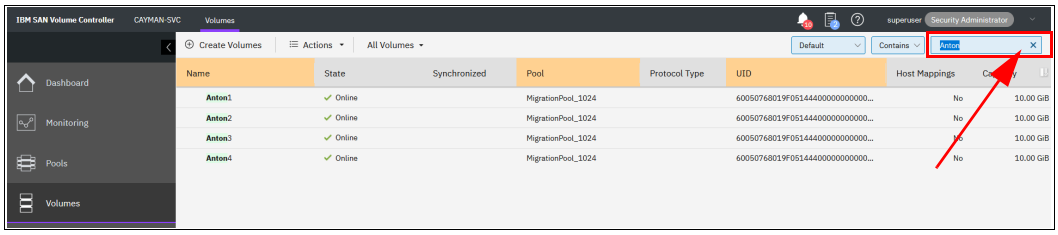


Figure 4-24 Show filtered rows

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.

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 pane, 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 left corner of the table header. A list of all of the available columns displays, as shown in Figure 4-25.

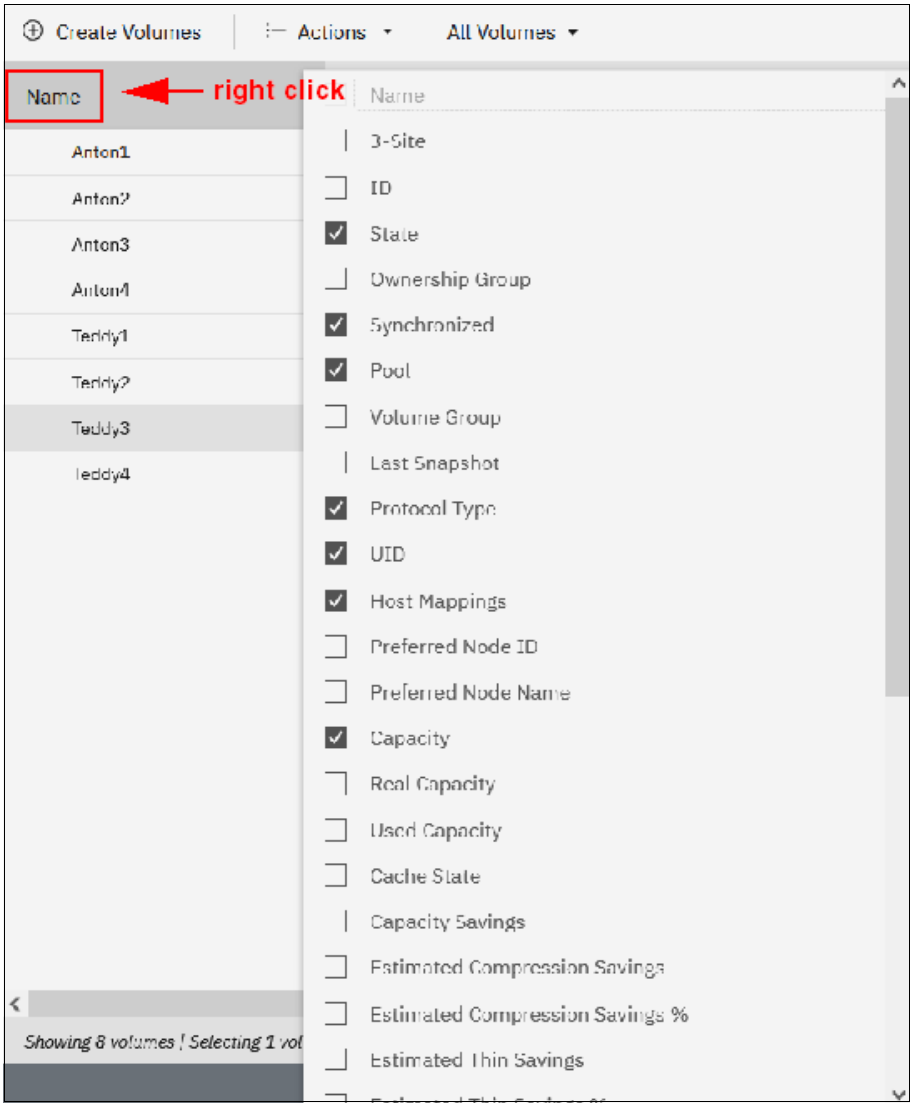


Figure 4-25 Add or remove details in a table

2. Select the column that you want to add or remove from this table. In our example, we added the volume Real Capacity column, as shown in Figure 4-26.

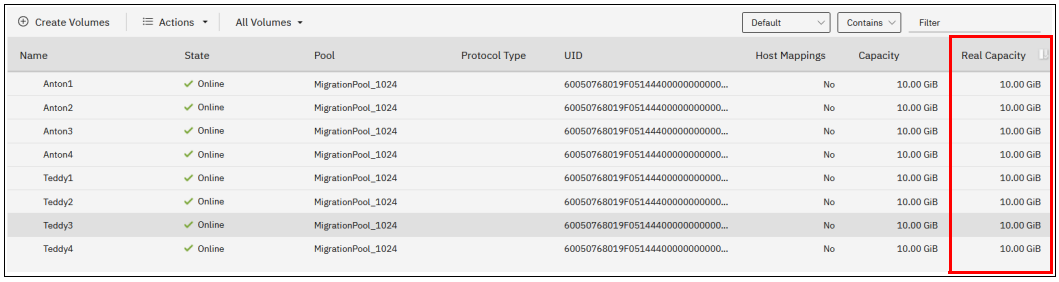


Figure 4-26 Table with an added Real Capacity column

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 4-27.

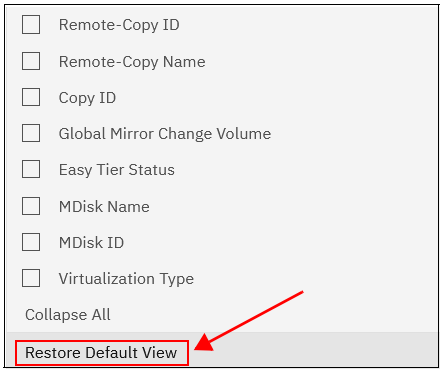


Figure 4-27 Table Restore Default View

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 4-28. In this example, we attempt to move the Capacity column before the State column.

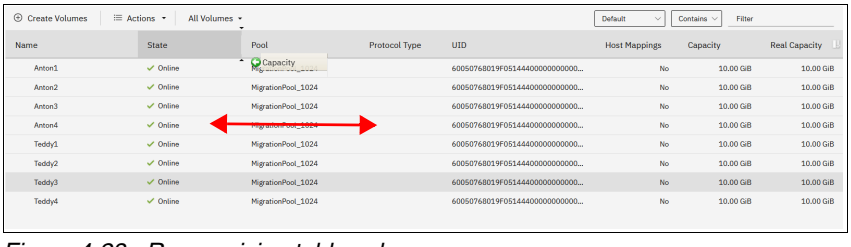


Figure 4-28 Reorganizing table columns

4.4 Monitoring menu

Click the **Monitoring** icon in left pane to open the Monitoring menu (see Figure 4-29). The Monitoring menu offers these 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 the nodes is displayed, including node status, number of events against each node, and key node information, such as cache size and serial numbers. For more information, see 4.4.1, “System Hardware overview” on page 172.
- 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 goes into balancing mode. When the pool contains multiple types of MDisks, Easy Tier is automatically turned on. For more information, see 4.4.2, “IBM Easy Tier Reports” on page 174
- 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” on page 175.
- 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 Figure 4.4.4 on page 177.
- Background Tasks: The option shows the progress of all tasks running in the background as listed in “Running tasks and suggested tasks” on page 164.

The following sections describe each option of the Monitoring menu (see Figure 4-29).

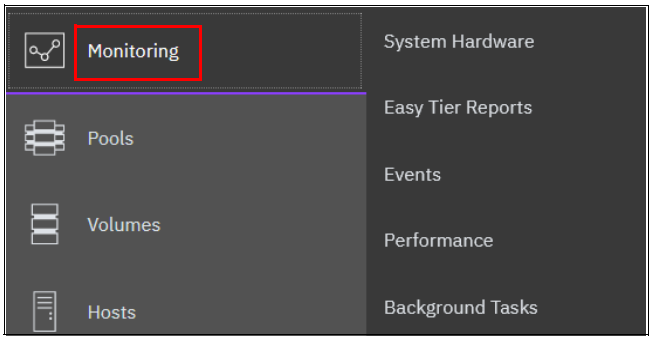


Figure 4-29 Monitoring menu

4.4.1 System Hardware overview

The System Hardware option of the Monitoring menu provides a general overview of the IBM SAN Volume Controller and key information. If you have more than one I/O group, this view is presented by I/O group, with nodes displayed within their own I/O group section, as shown in Figure 4-30.

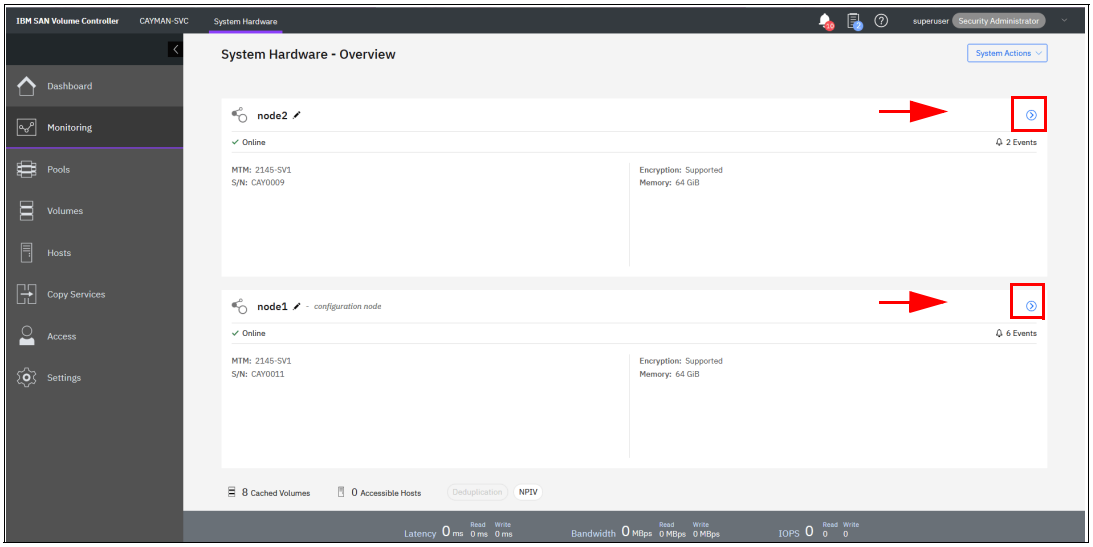


Figure 4-30 System overview window showing both nodes

You can select a component to view more details about it by using the arrow that is highlighted in the red box in Figure 4-30. Clicking this arrow shows detailed technical attributes, such as ports that are in use, memory, serial number, node name, encryption status, configuration node and node status (online or offline). Selecting an individual component on the image of your IBM SAN Volume Controller displays its details in the Component Details area on the right side, as shown in Figure 4-31.

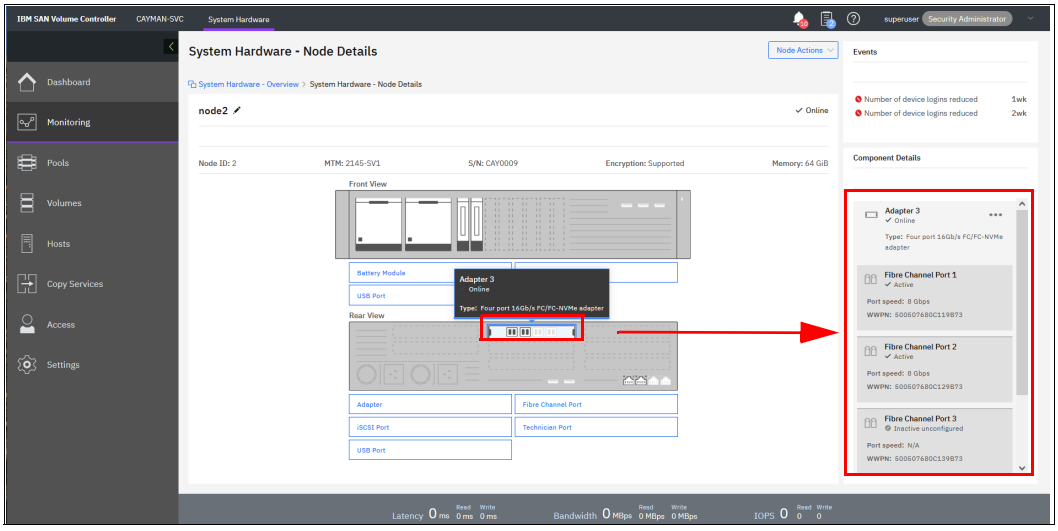


Figure 4-31 Component details

In an environment with multiple 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 pane by completing the following steps:

1. Click **Identify** in the menu that is shown in Figure 4-32.

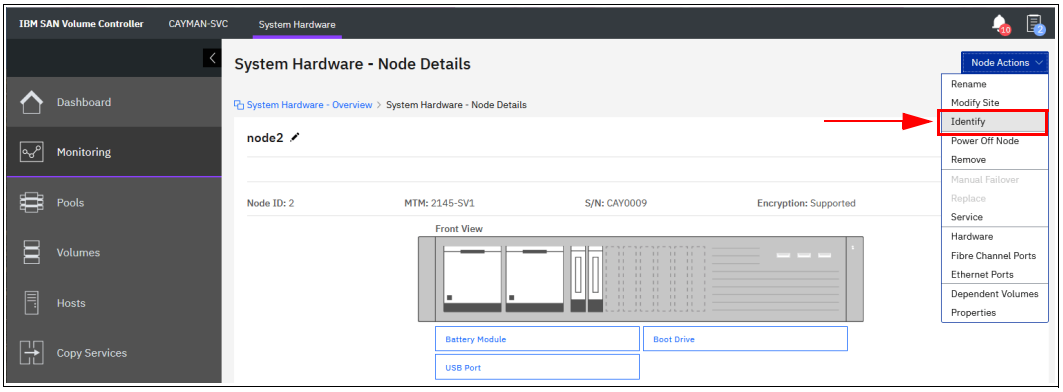


Figure 4-32 Turning on the identification LED

2. Wait for confirmation from the technician that the device in the data center was correctly identified.
3. After the confirmation, click **Turn LED Off** (see Figure 4-33).

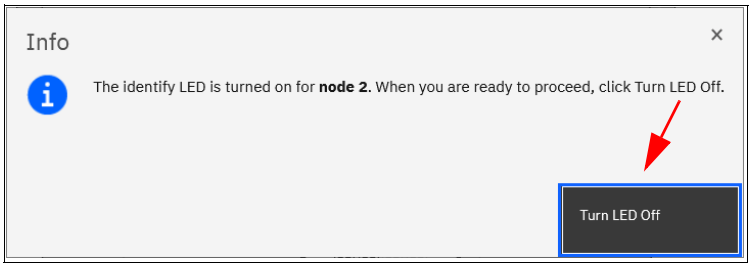


Figure 4-33 Turning off the identification LED

Alternatively, you can use the IBM SAN Volume Controller command-line interface (CLI) to get the same results. Run the following commands in the sequence:

1. **`svctask chnode -identify yes 1`** (or **`chnode -identify yes 1`**)
2. **`svctask chnode -identify no 1`** (or **`chnode -identify no 1`**).

4.4.2 IBM Easy Tier Reports

The management GUI supports monitoring Easy Tier data movement in graphical reports to help you understand what is happening with 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 can be 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, select **Monitoring** → **Easy Tier Reports** (see Figure 4-34).

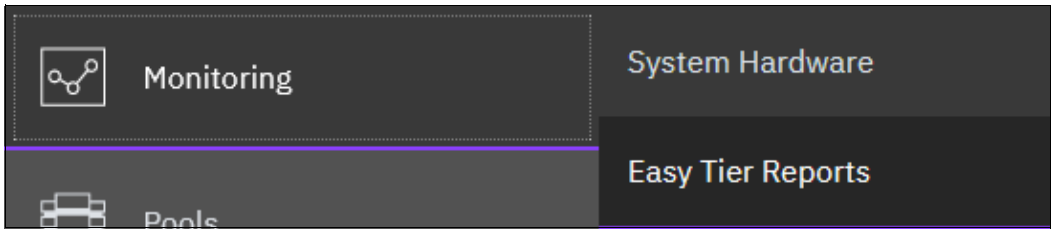


Figure 4-34 Selecting Easy Tier Reports

Data Movement statistics

The Data Movement chart displays the migration actions that are triggered by Easy Tier.

Tier composition statistics

The Tier Composition 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 Workload Skew Comparison chart displays the percentage of I/O workload compared to the total capacity.

Figure 4-35 shows you how to export the Easy Tier Reports.

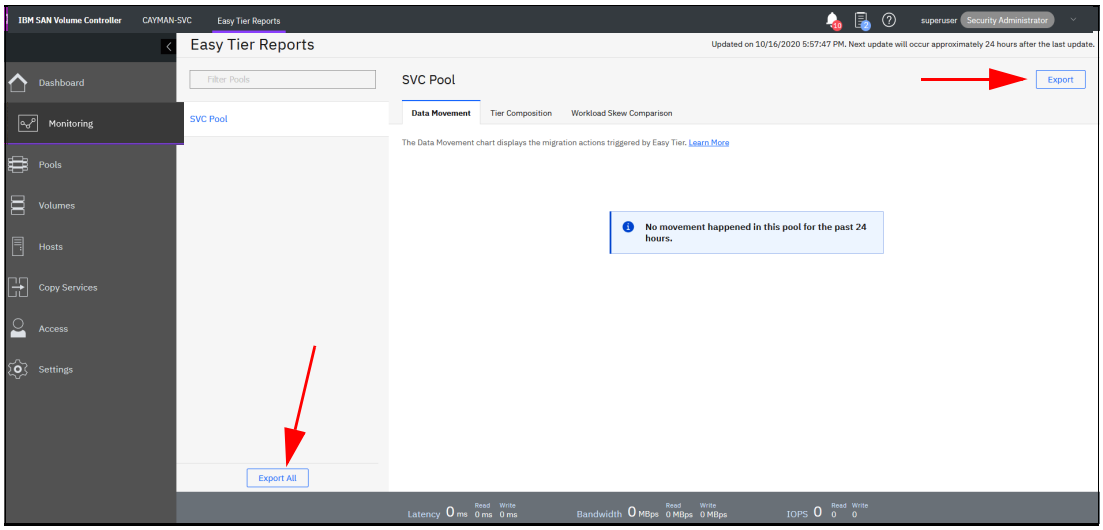


Figure 4-35 Easy Tier Reports

You can export your Easy Tier statistics to a CSV file for further analysis. For more information about Easy Tier Reports, see Chapter 9, “Advanced features for storage efficiency” on page 559.

4.4.3 Events

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 4-36 provides an example of records in the system Event log.

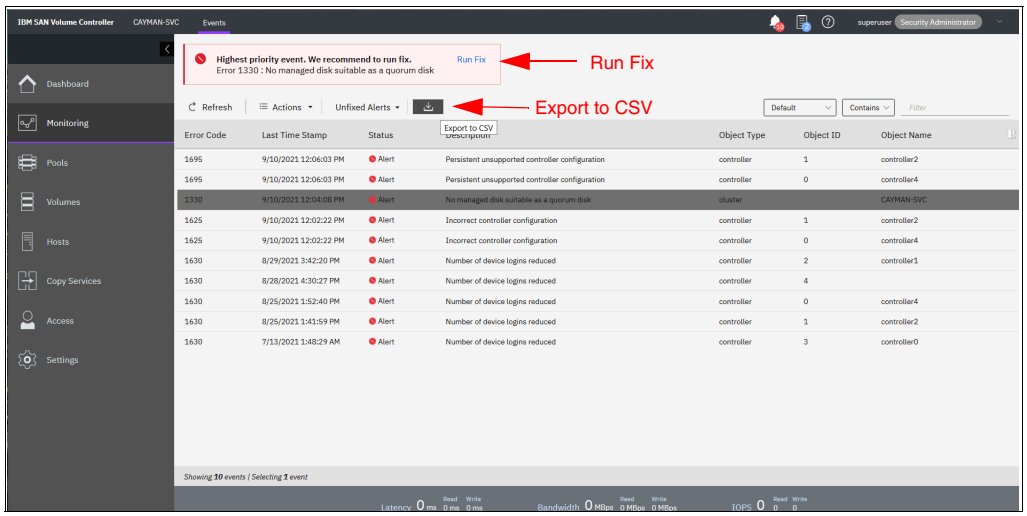


Figure 4-36 Event log list

For the error messages with the highest internal priority, perform corrective actions by running fix procedures. Click **Run Fix** (see Figure 4-36 on page 175), and the fix procedure wizard opens, as shown in Figure 4-37.

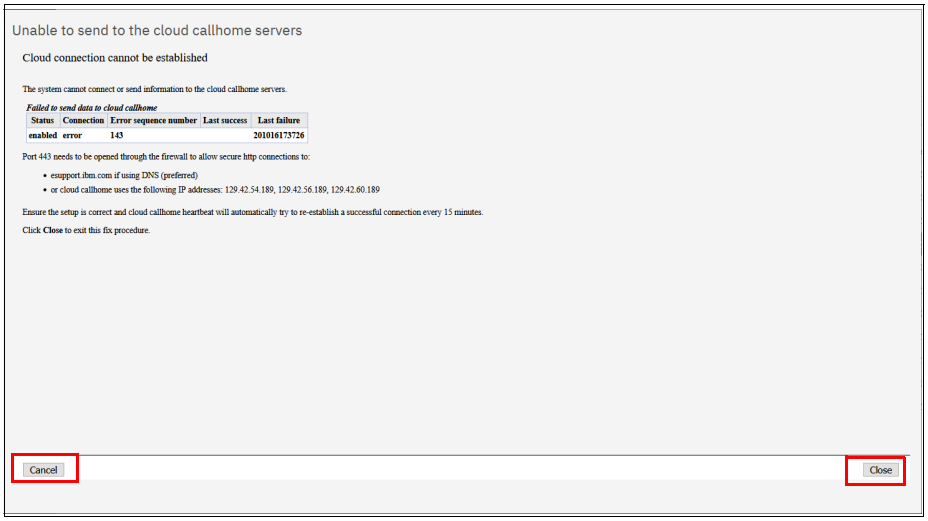


Figure 4-37 Performing a fix procedure

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 13, "Reliability, availability, serviceability, monitoring and troubleshooting" on page 853.

4.4.4 Performance

The Performance pane 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 node in the system if problems exist after a node failover occurs (see Figure 4-38).



Figure 4-38 Performance statistics of the IBM SAN Volume Controller

The performance statistics in the GUI show, by default, the latest 5 minutes of data. To see details of each sample, click the graph and select the time stamp, as shown in Figure 4-39.

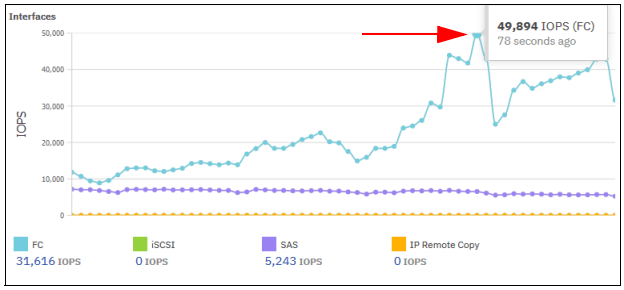


Figure 4-39 Sample details

The charts that are shown in Figure 4-39 represent 5 minutes of the data stream. For in-depth storage monitoring and performance statistics with historical data about your IBM SAN Volume Controller system, use IBM Spectrum Control (enabled by the former IBM Tivoli® Storage Productivity Center for Disk and IBM Virtual Storage Center) or IBM Storage Insights.

You can also obtain a no-charge unsupported version of the Quick Performance Overview (**qperf**) for the IBM SAN Volume Controller and Spectrum Virtualize systems from this IBM Support web page.

4.4.5 Background Tasks

Use the Background Tasks pane to view and manage current tasks that are running on the system (see Figure 4-40).

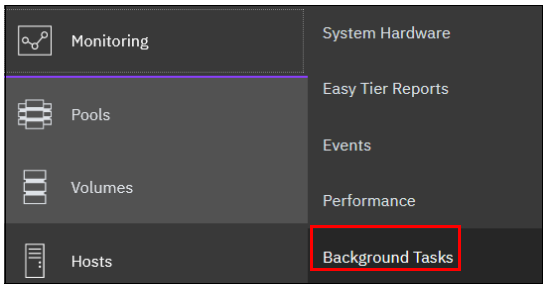


Figure 4-40 Selecting Background Tasks

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 4-41.

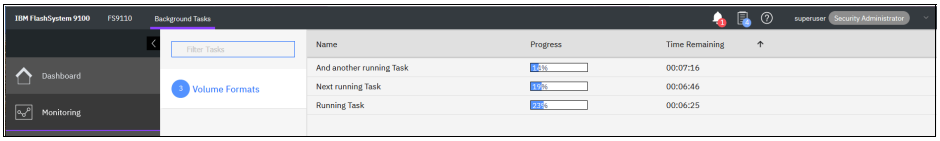


Figure 4-41 List of running tasks

4.5 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 GUI (see Figure 4-42):

- Pools
- Volumes by Pool
- Internal Storage
- External Storage
- MDisks by Pool
- Import External Storage

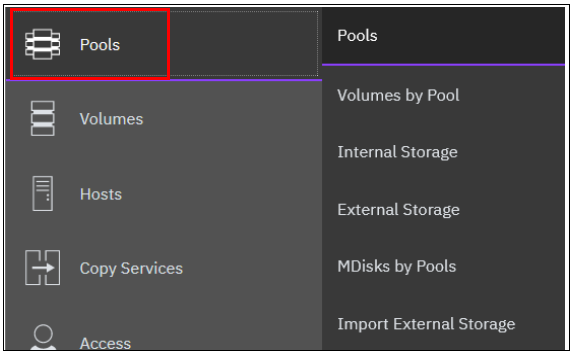


Figure 4-42 Pools menu

For more information about storage pool configuration and management, see Chapter 5, “Using storage pools” on page 271.

4.6 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 items, as shown in Figure 4-43 on page 179:

- Volumes
- Volumes by Pool
- Volumes by Host and Cluster
- Cloud Volumes
- Volume Groups

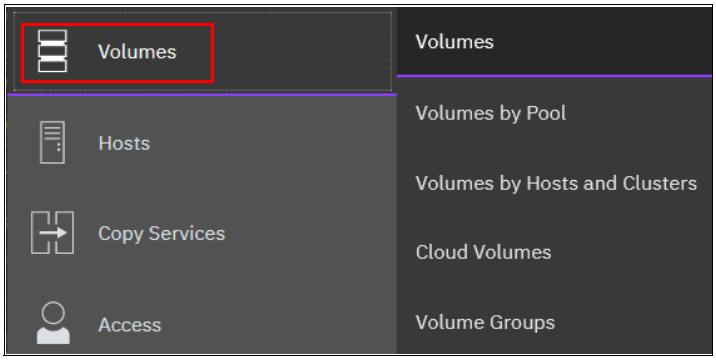


Figure 4-43 Volumes menu

For more information about these tasks and configuration and management process guidance, see Chapter 6, “Creating and provisioning volumes” on page 331.

4.7 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 the IBM SAN Volume Controller. Both FC and serial-attached Small Computer System Interface (SCSI) (SAS) connections use WWPNs to identify the host interfaces to the systems.

The Hosts menu consists of the following choices, as shown in Figure 4-44:

- Hosts
- Host Clusters
- Mappings
- Volumes by Host and Clusters

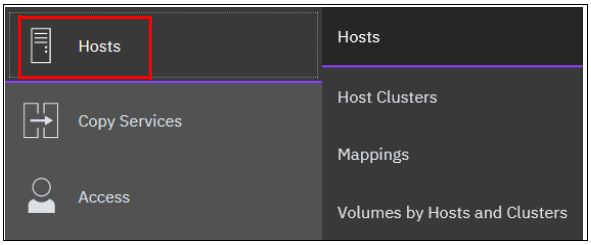


Figure 4-44 Hosts menu

For more information about configuration and management of hosts by using the GUI, see Chapter 7, “Hosts” on page 443.

4.8 Copy Services

The IBM Spectrum Virtualize Copy Services and Volumes Copy operations are based on the FlashCopy function. In its basic mode, the function creates copies of content on a source volume to a target volume. Any data on the target volume is lost and is replaced by the copied data.

More advanced functions allow FlashCopy operations to occur on multiple source and target volumes. Management operations are coordinated to provide a common, single point-in-time (PiT) for copying target volumes from their respective source volumes. This technique creates a consistent copy of data that spans multiple volumes.

The IBM SAN Volume Controller Copy Services menu offers the following operations in the GUI, as shown in Figure 4-45:

- FlashCopy
- FlashCopy Consistency Groups
- FlashCopy Mappings
- Remote Copy

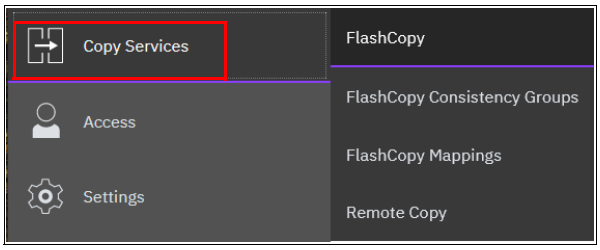


Figure 4-45 Copy Services in GUI

Because Copy Services is one of the most important features for resiliency solutions, see Chapter 10, “Advanced Copy Services” on page 605.

4.9 Access

The Access menu in the GUI maintains who can log in to the system, defines the access rights to the user, and tracks what was done by each privileged user to the system. It is logically split into three categories:

- Ownership groups
- Users by group
- Audit log

In this section, we explain 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 pane, as shown in Figure 4-46.

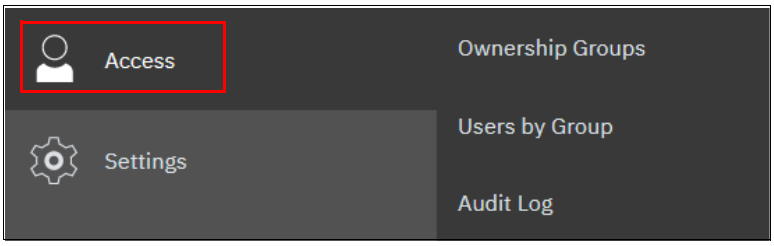


Figure 4-46 Access menu

4.9.1 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.

When you start the Ownership Group task for the first time, you see the window that is shown in Figure 4-47.

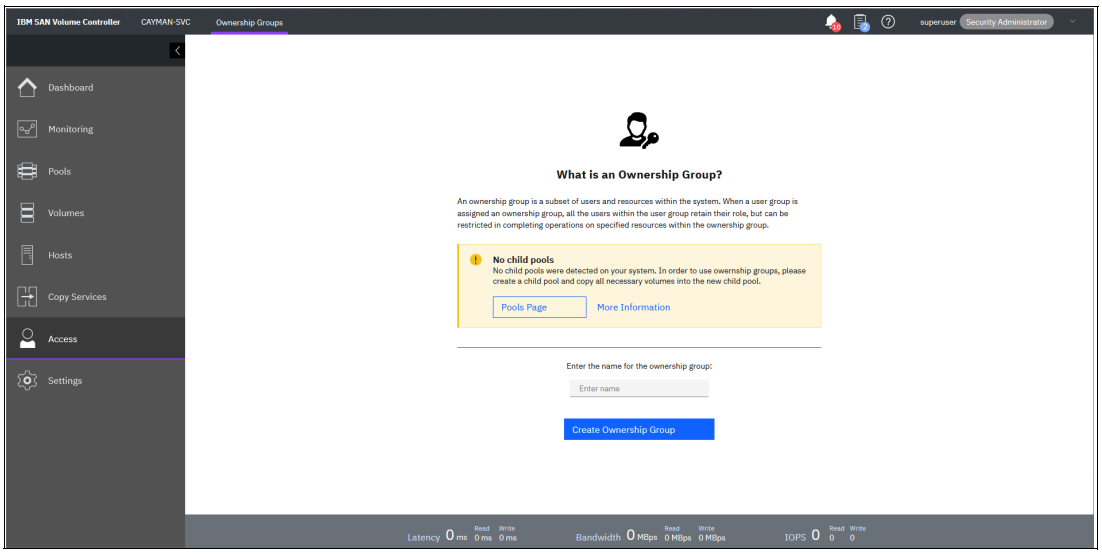


Figure 4-47 Start of an Ownership Group

In our example, no child pool exist, so the GUI guide you to the Pools page to create child pools.

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

Two basic use cases can be applied to using ownership groups on the system. The first use case for using ownership groups is on a system where objects are created within the ownership group. Other objects can exist on the system that are not in the ownership group. The second use case for using ownership groups is 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 RC 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 have access to through the ownership group are displayed. Additionally, only events and commands that are related to the ownership group in which a user belongs are viewable by those users.

Inheriting ownership

Depending on the type of resource, ownership can be defined explicitly or ownership can be inherited from the user, user group, or from other parent resources. Objects inherit their ownership group from other objects whenever possible:

- 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 they belong to, if applicable.
- Host mappings inherit the ownership group from both the host and the volume to which the host is mapped.

These objects cannot be explicitly 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), then the owner must choose the group when the object is created.

Figure 4-48 on page 184 shows how different objects inherit ownership from ownership groups.

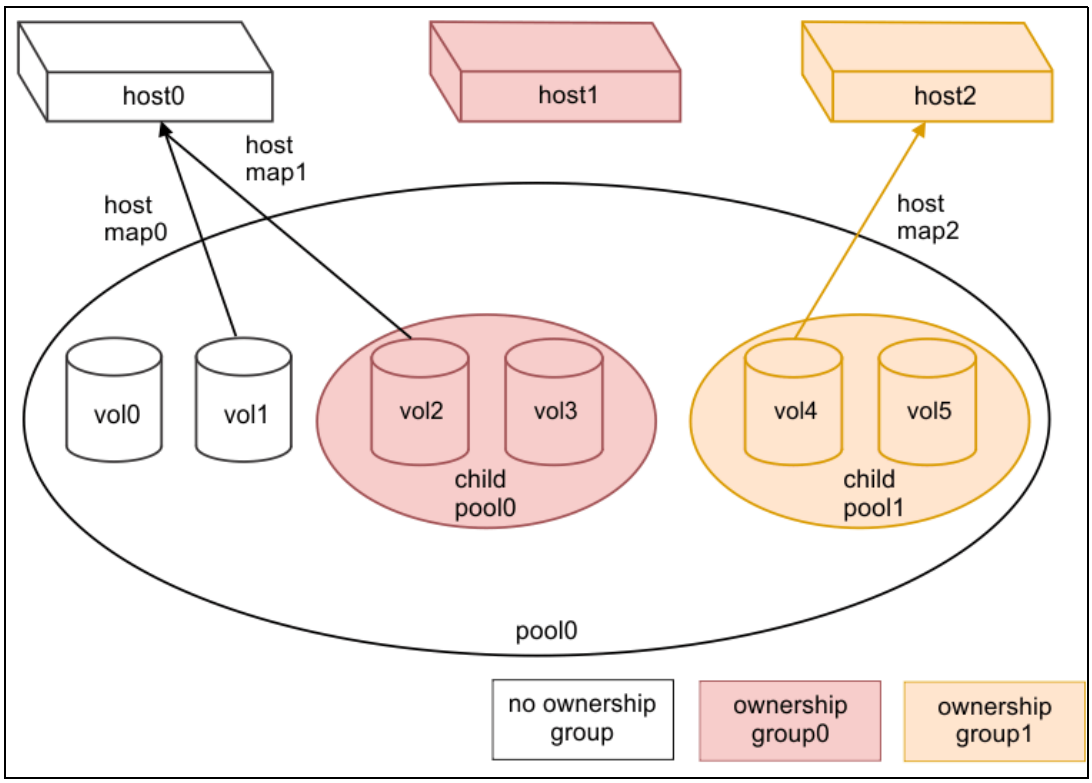


Figure 4-48 Ownership group inheritance

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 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 an existing 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 new volume group or change an existing volume group.
- Volumes can be added to a volume group if both the volume and the volume group are within the same ownership group or if both are not in an ownership group. There are situations where 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, then 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 itself and indirectly change the volume, such as deleting a volume from the volume group. Neither the ownership group of the child pools or the owner of the volume group can 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 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.
- There are situations where 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 impact 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. A user cannot change the ownership group of the resource, but can change the ownership group of the parent object.

Portsets

The following rules apply to portsets:

- Restricted users can view and create IP addresses or assign hosts to portset that are in their ownership group.
- Restricted users can assign hosts to portsets that are 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 can view all IP addresses in portsets that they own and portsets that are not assigned to an ownership group.
- Restricted users cannot modify portsets that are associated with other restricted users.
- Restricted users cannot view portsets that were assigned to a different ownership group.
- Restricted users cannot view IP addresses that are associated with a portset that is assigned to a different ownership group.
- Restricted users cannot create or delete IP addresses from a portset that is part of a different ownership group.
- Only unrestricted users can own remote copy and storage portsets.

- Unrestricted users can view, create, and manage all portsets on the system.
- Unrestricted users can view and create IP addresses to any portsets on the system.
- Unrestricted users can assign hosts in an ownership group to portsets that are not assigned to an ownership group.
- Unrestricted users can assign hosts to portsets that are assigned the same ownership group.
- Unrestricted users cannot assign hosts to portsets if the host and portsets are in different ownership groups.
- Unrestricted users cannot change the ownership of remote copy and storage portsets.

Other objects inherit the ownership group from these specifically assigned objects. Often, the ownership group is defined when an object is created. Ownership also can be changed when the specifically 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 command-line interface, run the **`-inconsistentownershipgroup`** command to allow for inconsistent ownership groups. However, it is not recommended to leave volumes or volume copies in different ownership groups. After the migration, the user with Security Administrator role needs to ensure 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:

- Volume-to-host mappings inherit the ownership group of the host or host cluster and volume in the mapping.
- If host or host cluster and volume are within different ownership groups, then 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, then the mapping cannot be assigned to an ownership group.
- Like 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 has no impact on the ownership of the mappings that it contains.

Configuring ownership groups

You can configure ownership groups to manage access to resources on the system. 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.

Migrating to ownership groups

If you updated your system to a software level that supports ownership groups, you must reconfigure certain resources if you want to configure 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.

To create an ownership group, select **Create Ownership Group**, as shown in Figure 4-49.

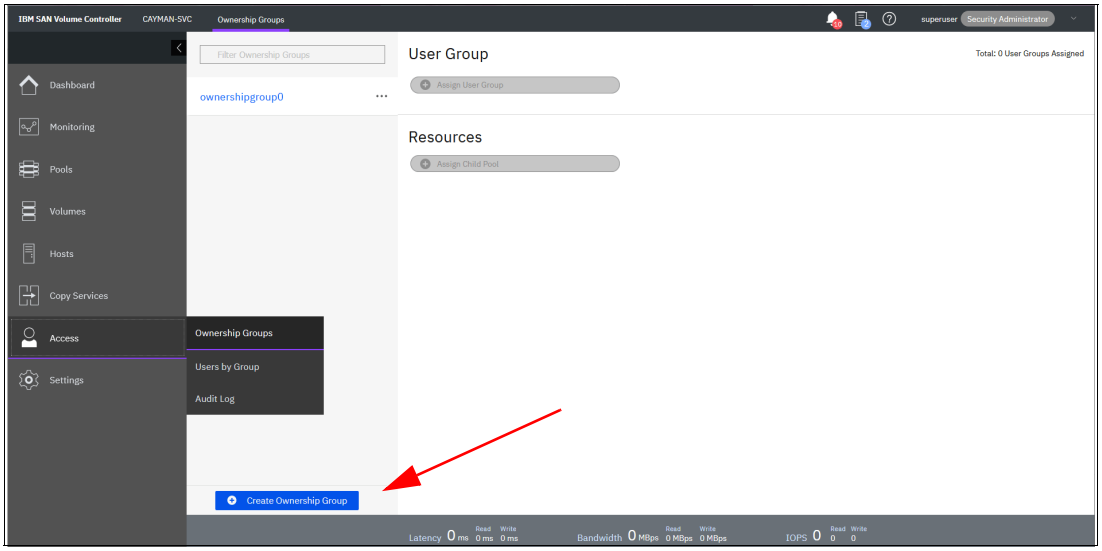


Figure 4-49 Ownership by Groups

4.9.2 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.

To define your user group in the IBM SAN Volume Controller, select **Access** → **Users by Group**, as shown in Figure 4-50.

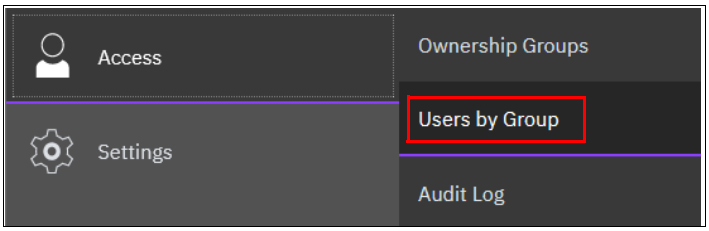


Figure 4-50 Access Users by Group

Select **Create User Group**, as shown in Figure 4-51.

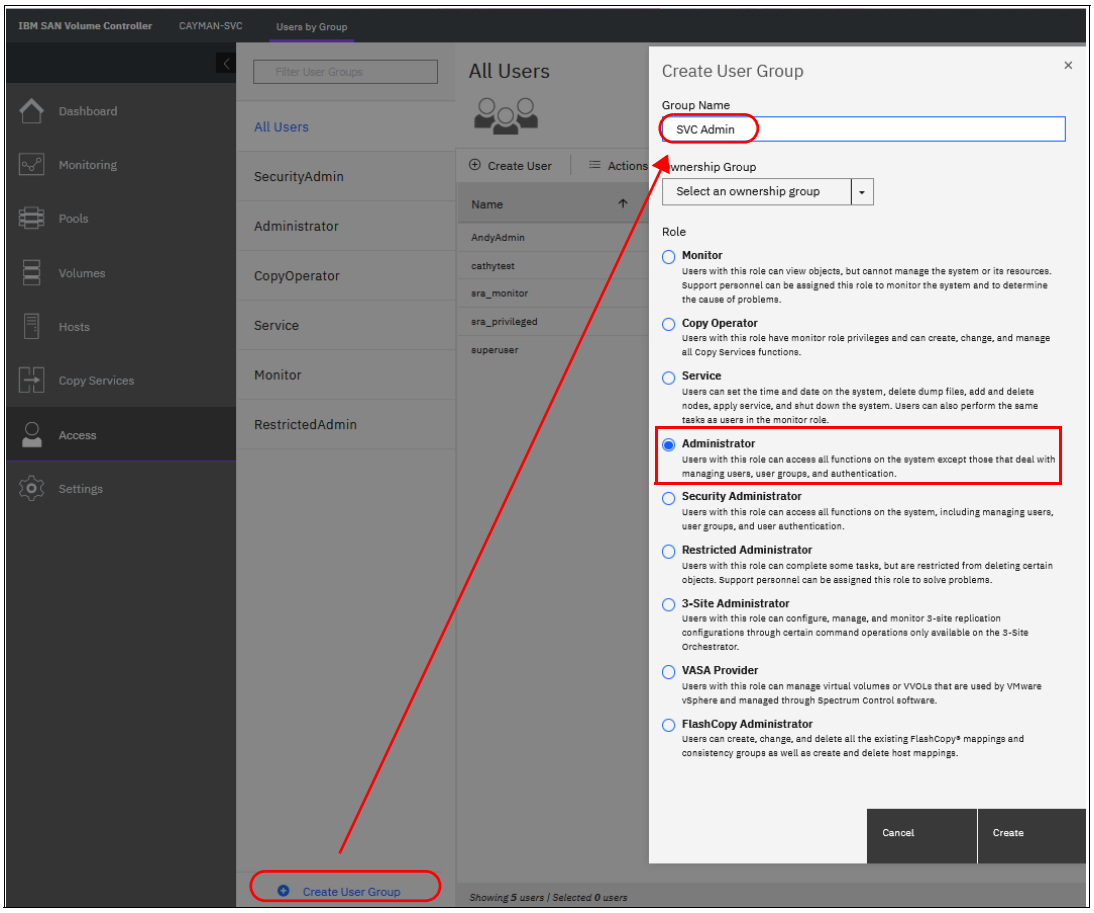


Figure 4-51 Creating a user group in IBM SAN Volume Controller

Figure 4-52 shows the newly created User Group.

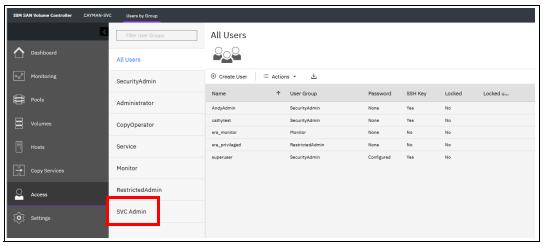


Figure 4-52 User Group

The following privileged user group roles exist in IBM Spectrum 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.

- CopyOperator

  These users can start and stop all existing 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.

- Service

  These users can set the time and date on the system, delete dump files, add and delete nodes, apply service, and shut down the system. Users can also complete the same tasks as users in the monitor role.

- 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.

- 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 through certain 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.

- vSphere APIs for Storage Awareness (VASA) Provider

  Users with this role can manage virtual volumes or VVOLs that are used by VMware vSphere and managed through Spectrum Control software.

Registering a user

After you define your group (in our example, `SVC_Administrator` with Administrator privileges), you can register a user (`SVC_Cluster_Admin`) within this group by clicking **Create User** and then, selecting **SVC Admin** (see Figure 4-53).

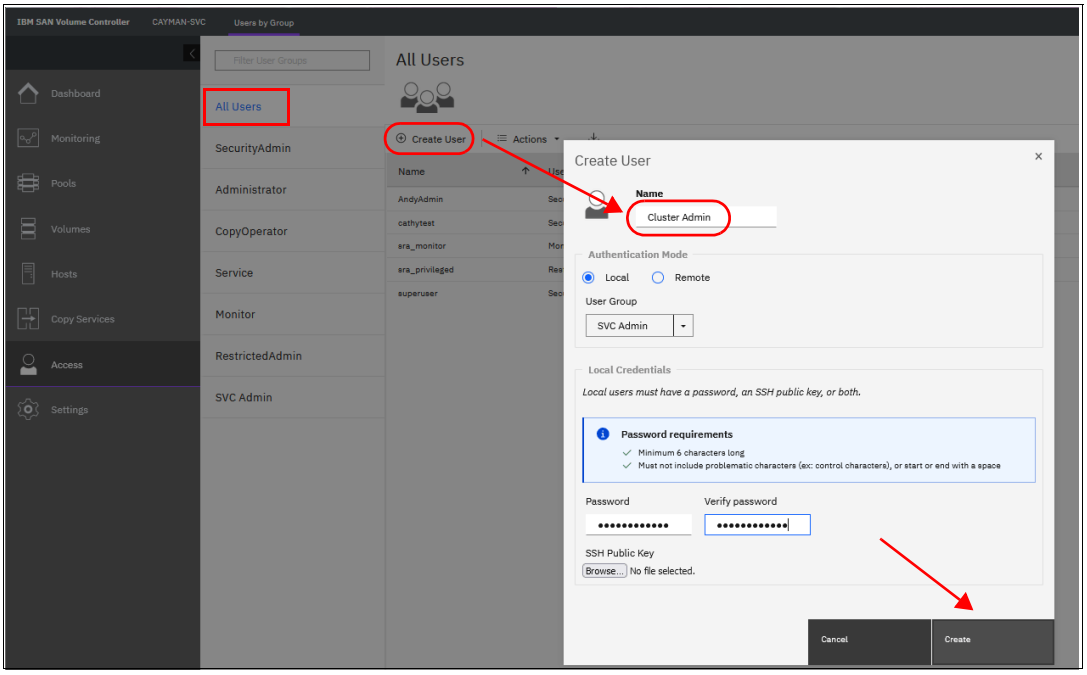


Figure 4-53 Registering a user account

Deleting a user

To remove a user account, right-click the user in the **All Users** list and select **Delete**, as shown in Figure 4-54.

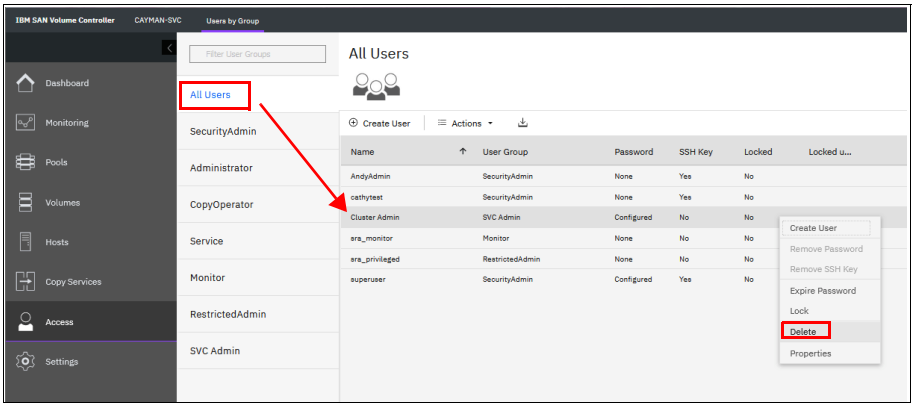


Figure 4-54 Deleting a user account

Attention: When you click **Delete**, the user account is directly deleted. No other confirmation request is presented.

Setting a new 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 4-55).

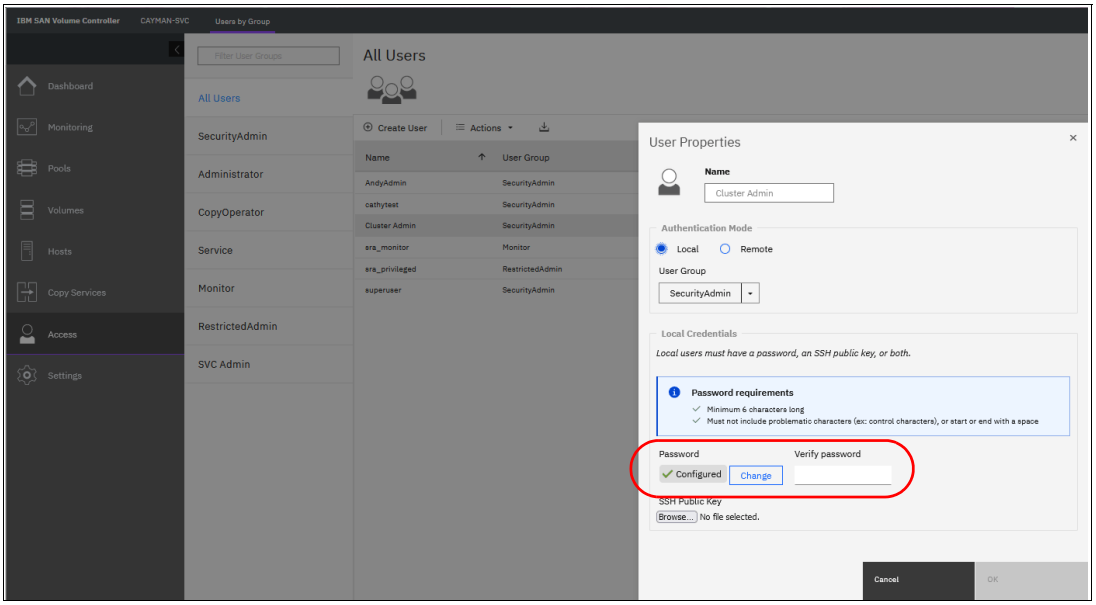


Figure 4-55 Setting a new password

New functions with 8.4.2 are Lock User and Expire Password.

Lock Password

To lock a user, right-click the user name and select **Lock**, as shown in Figure 4-56.

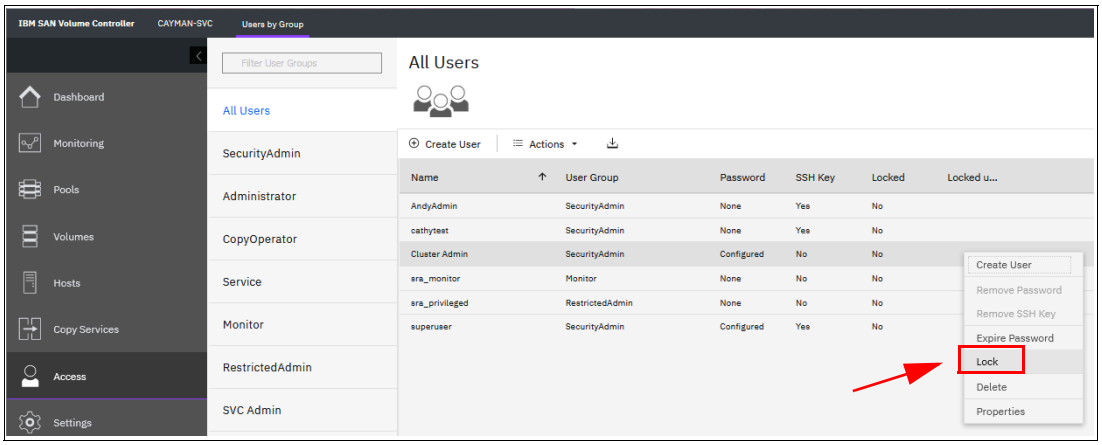


Figure 4-56 Lock User

A verification window opens (see Figure 4-57).

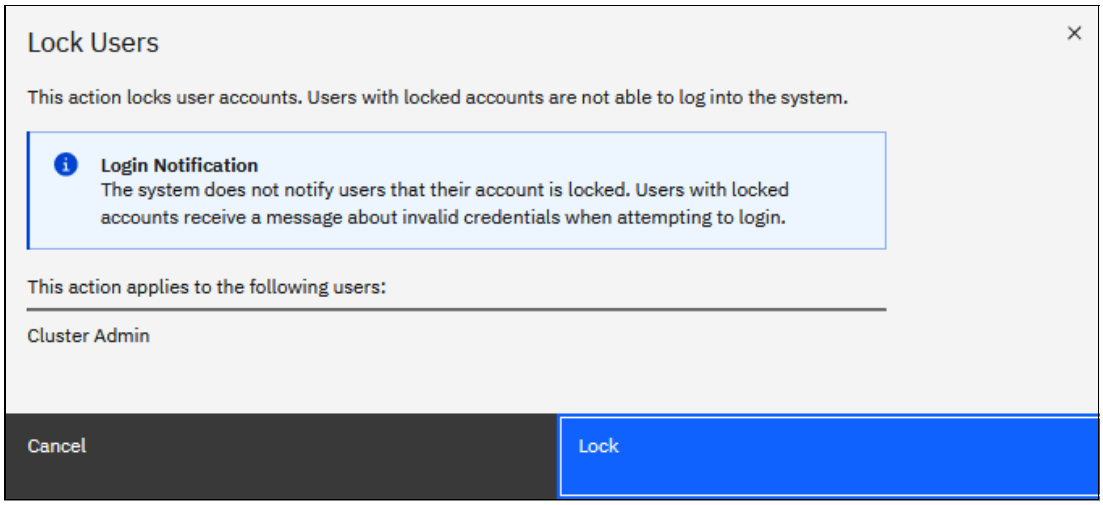


Figure 4-57 Lock verification

Note: Remember that users do not receive a notification that their account was locked. Users with locked accounts receive a message only about invalid credentials when attempting to log in.

Figure 4-58 shows a locked user.

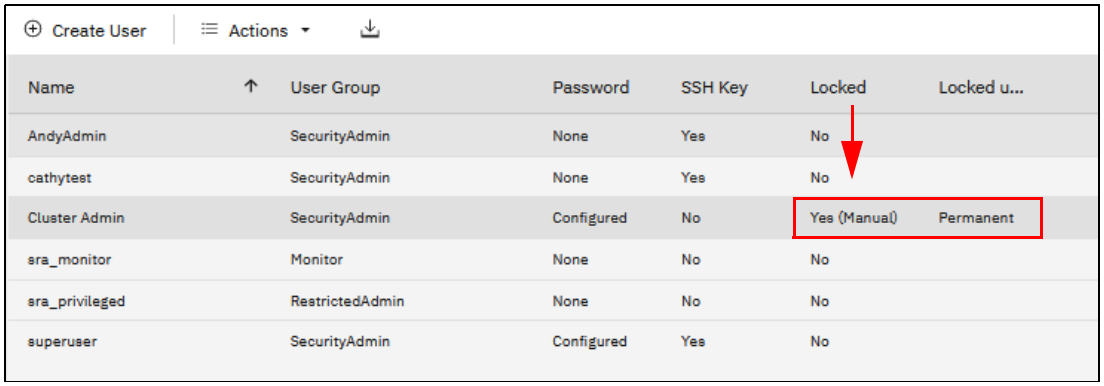


Figure 4-58 Locked user

To unlock the user, right-click the user and select **Unlock**, as shown in Figure 4-59.

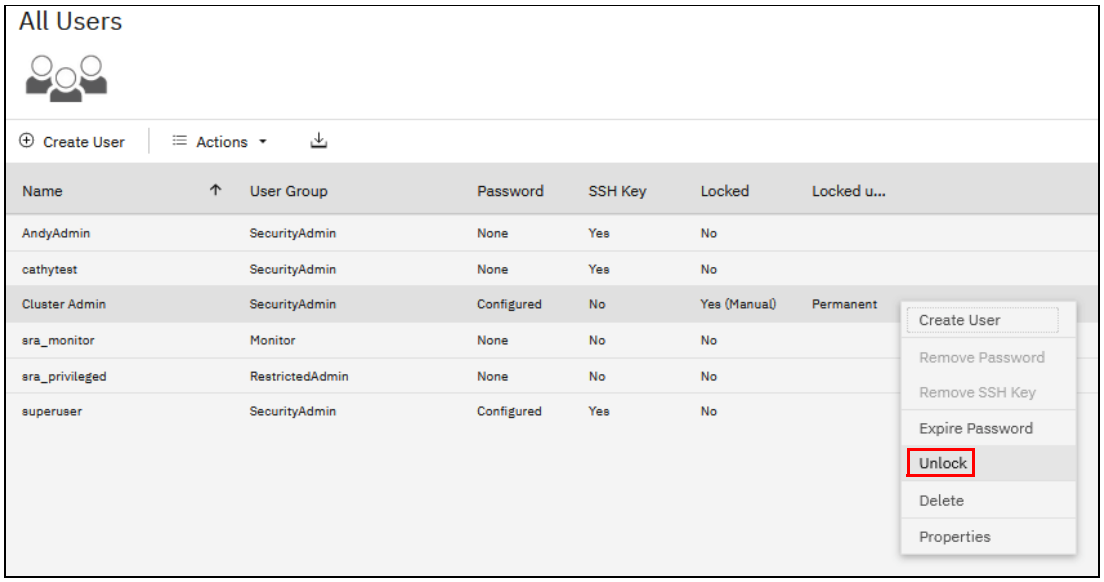


Figure 4-59 Unlock User

A confirmation window opens (see Figure 4-60).

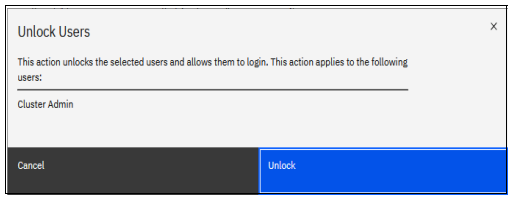


Figure 4-60 Unlock

Expiring a password

To force a password change for a dedicated user, you can use the Expire Password function (see Figure 4-61).

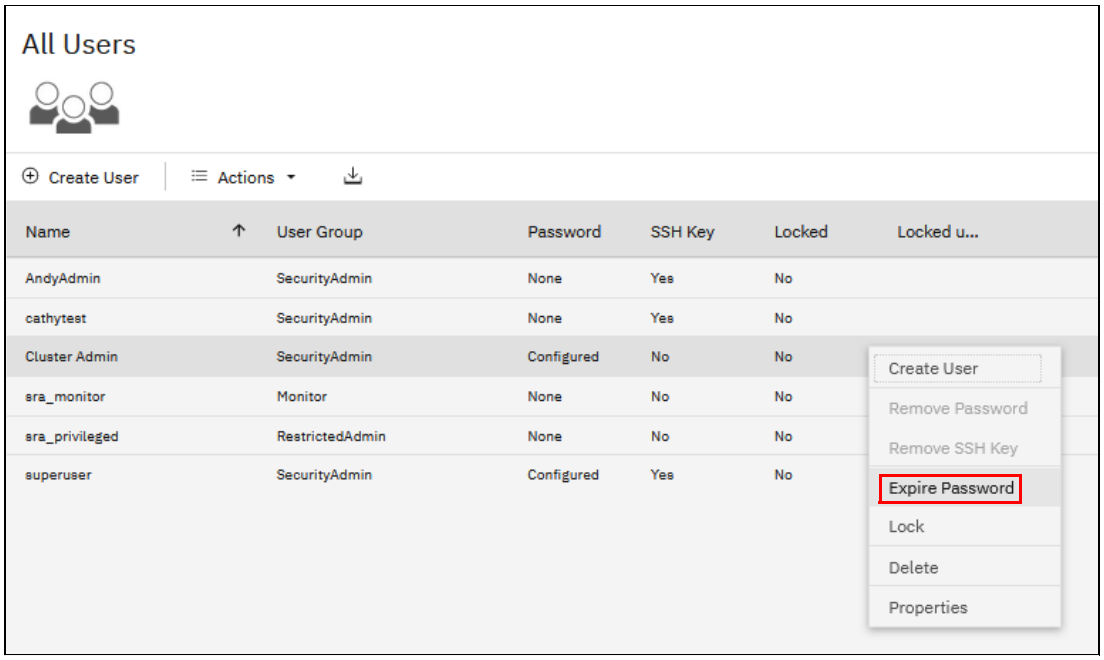


Figure 4-61 Expire Password

After you select **Expire Password**, a verification windows opens, as shown in Figure 4-62.

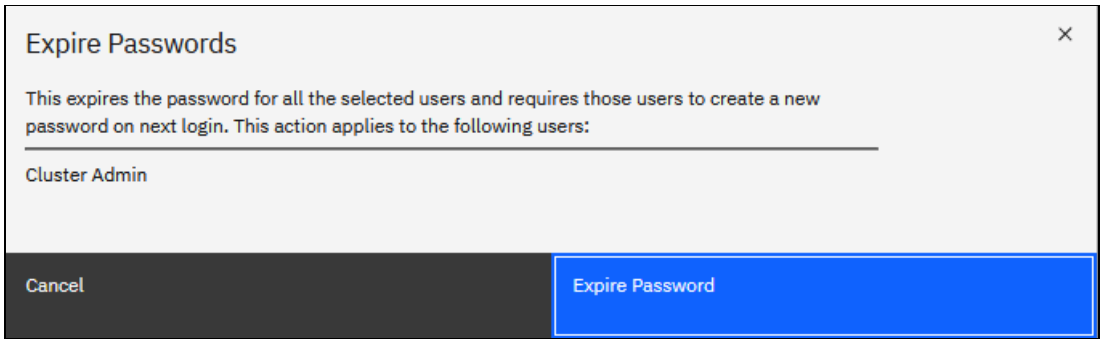


Figure 4-62 Expire Password verification

Note: This process expires the password for all of the selected users and requires those users to create a new password on next log in.

4.9.3 Audit log

An *audit log* documents actions that are 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 command-line interface, the time stamps for the logs in the CLI are the system time. However, when logs are displayed in the management GUI, the time stamps 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 information about a challenge.

  This action indicates the challenge with which the support user 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.

An example of the audit log is shown in Figure 4-63.

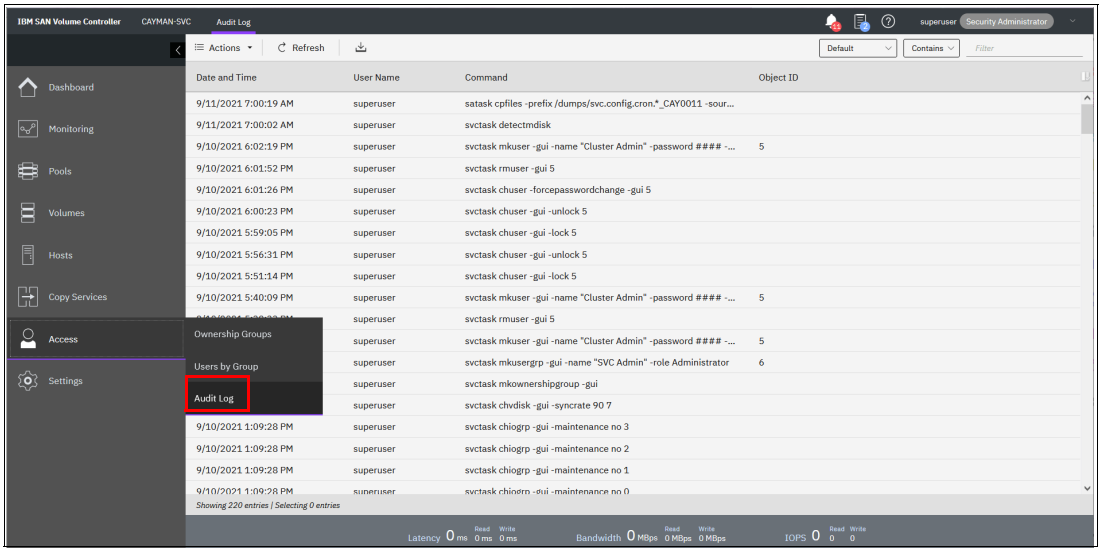


Figure 4-63 Audit log

The following commands are not documented in the audit log:

- `dumpconfig`
- `cpdumps`
- `finder`
- `dumperrlog`

The following items are also 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.

Important: Failed commands are not recorded in the audit log. Commands that are triggered by IBM Support personnel are recorded with the **`Challenge`** flag because they use challenge-response authentication.

4.10 Settings

Use the Settings pane to configure system options for notifications, security, IP addresses, and preferences that are related to display options in the management GUI (see Figure 4-64).

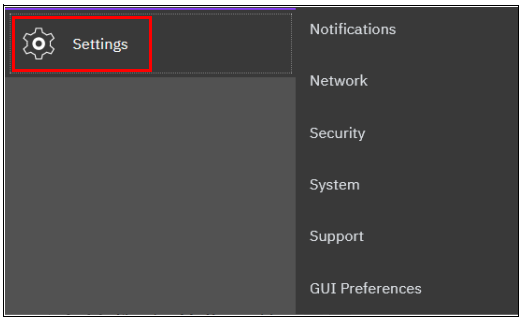


Figure 4-64 Settings menu

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 pane 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 pane 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 welcome message after login, and refresh internals and GUI logout timeouts.

These options are described next.

4.10.1 Notifications menu

The IBM SAN Volume Controller 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

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 IBM SAN Volume Controller.

To view the SNMP configuration, click the **Settings** icon and then, select **Notification** → **SNMP** (see Figure 4-65).

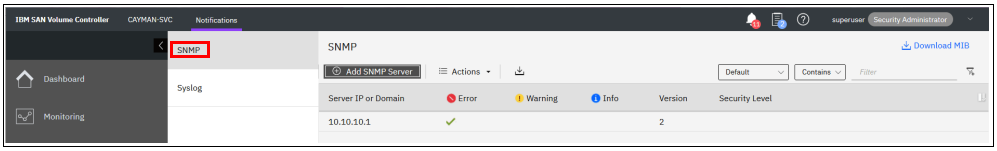


Figure 4-65 Setting SNMP server and traps

As shown in Figure 4-65, 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 by only devices that support SNMPv1 and SNMPv2c protocol. SNMPv3 uses username/password authentication, along with an encryption key. By convention, most SNMPv1-v2c equipment ships from the factory with a read-only community string set to “public”.

- Server Port

  The remote port number for the SNMP server. The remote port number must be a value 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 and select **Remove**. To add another 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 at least five octets in length and cannot be more than 32 octets long. Each octet can contain a value from 0 up to 255 or hexadecimal values from ‘00’H up to ‘FF’H. The engine ID translates to an even number of hexadecimal between 10 and 64 characters long.

- Security Name

  User name 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 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 certain 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) known by sender and receiver
  - Multiple levels of compliances with respect to DES due to problems associated with international use
  - 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 pane to view the syslog messages that are sent by the IBM SAN Volume Controller.

To view the Syslog configuration, go to the System pane and click **Settings** and then, select **Notification** → **Syslog** (see Figure 4-67). A domain name server (DNS) server is required to use domain names in syslog.

To create a Syslog server select: **Create Syslog Server** as shown in Figure 4-66.

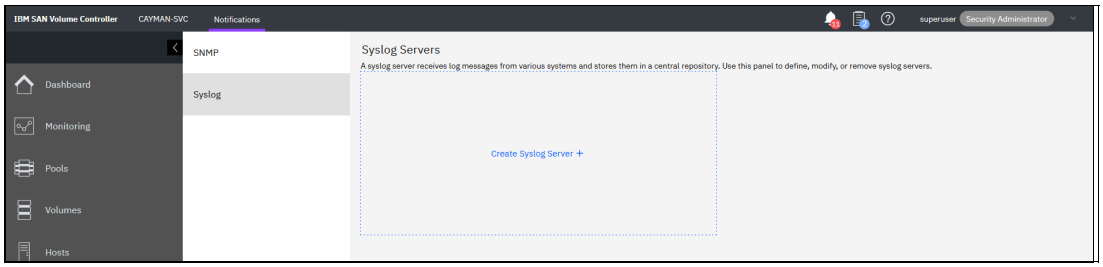


Figure 4-66 Creating Syslog server

In Figure Figure 4-67 you see what input is needed to create a Syslog Server.

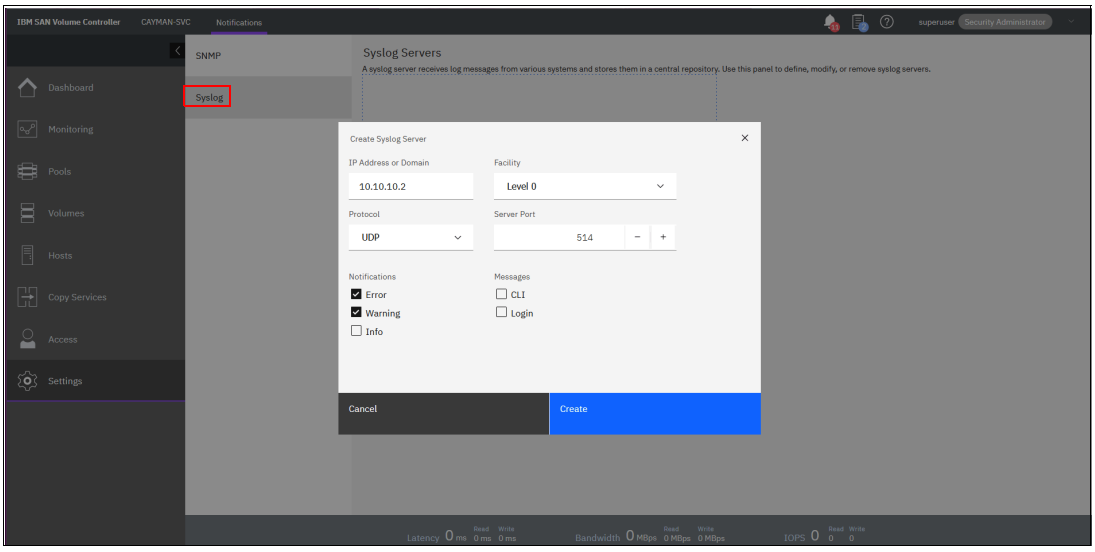


Figure 4-67 Setting the syslog messages

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

  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**.

- 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:

  - The concise message format provides standard details about the event.
  - The expanded format provides more details about the event.

To remove a syslog server, select **Delete** from the drop-down menu, as shown in Figure 4-68.

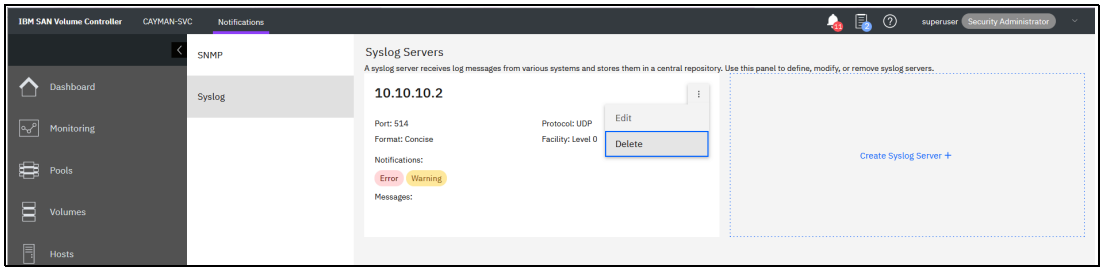


Figure 4-68 Delete Syslog Server

To create another Syslog Server, select **Create Syslog Server** (see Figure 4-68).

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=CAYMAN-SVC #Timestamp=Wed Oct 16 08:00:59 2021 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=CAYMAN-SVC #Timestamp=Wed Oct 16 08:00:59 2021 BST
#ObjectType=Node #ObjectName=Node1 #CopyID=0 #ErrorSequenceNumber=100 #ObjectID=2
#NodeID=2 #MachineType=2145DH8#SerialNumber=1234567 #SoftwareVersion=8.4.2.0
(build 154.20.2108201140000)#FRU=fan 24P1118, system board 24P1234
#AdditionalData(0->63)=0000000021000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000#Additional
Data(64-127)=0000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000
```

4.10.2 Network

This section describes how to view the network properties of the IBM SAN Volume Controller system.

The network information can be obtained by clicking **Network**, as shown in Figure 4-69.

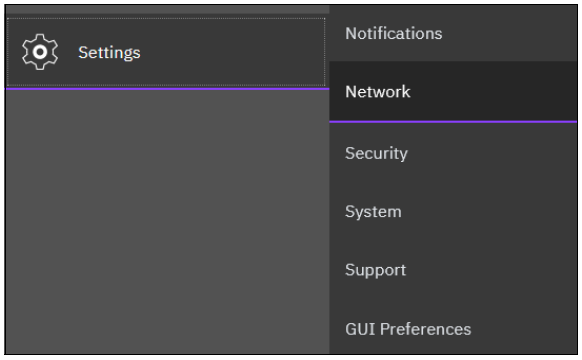


Figure 4-69 Accessing network information

Configuring the network

The procedure to set up and configure IBM SAN Volume Controller network interfaces is described in Chapter 3, “Initial configuration” on page 103.

Management IP addresses

To view the management IP addresses of IBM Spectrum Virtualize, select **Settings** → **Network**, and click **Management IP Addresses**. The GUI shows the management IP address by pointing to the network ports, as shown in Figure 4-70.

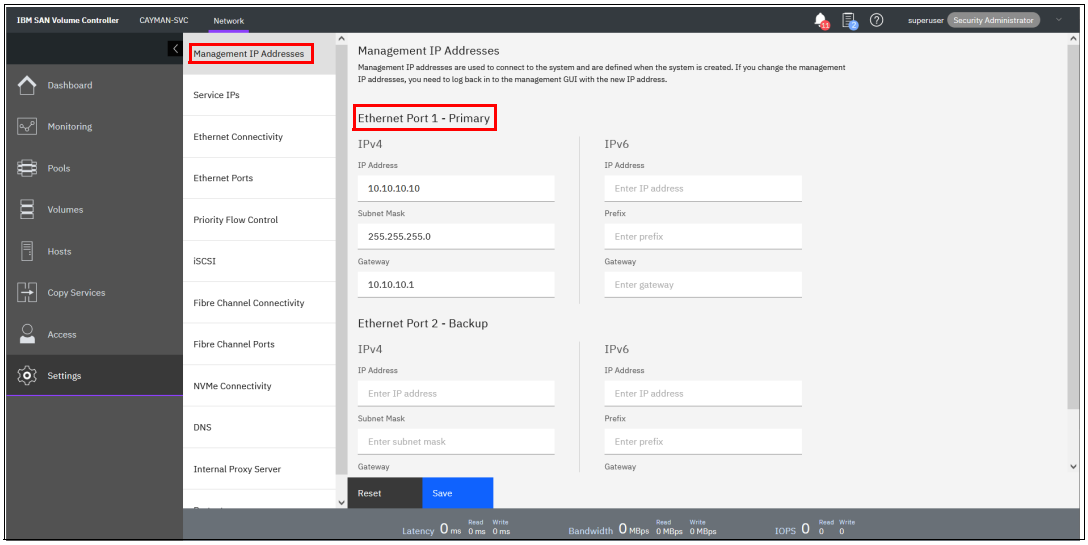


Figure 4-70 Viewing the management IP addresses

Service IP information

To view the Service IP information of your IBM Spectrum Virtualize installation, select **Settings** → **Network**, as shown in Figure 4-69. Click the **Service IPs** option to view the properties, as shown in Figure 4-71.

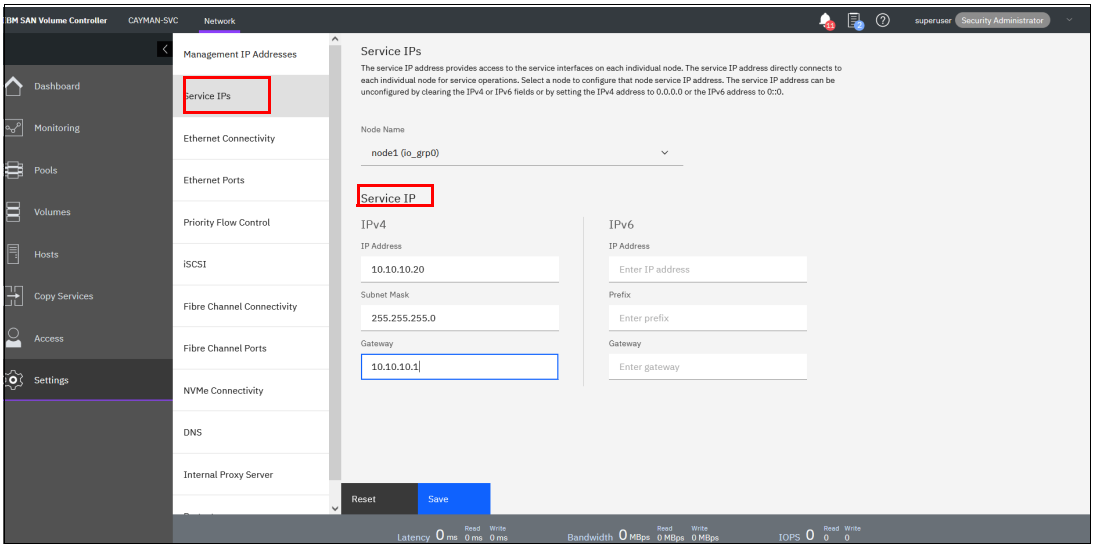


Figure 4-71 Viewing service IP addresses

The service IP address is commonly used to provide access to the network interfaces on each individual node.

Instead of reaching the management IP address, the service IP address directly connects to each individual node for service operations. You can select a node 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 window 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.

RDMA technologies, such as RoCE and iWARP, enable the RDMA-capable adapter to transfer data directly between nodes, bypassing CPU and caches and making 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 the connection 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 being 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 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 IBM Documentation.

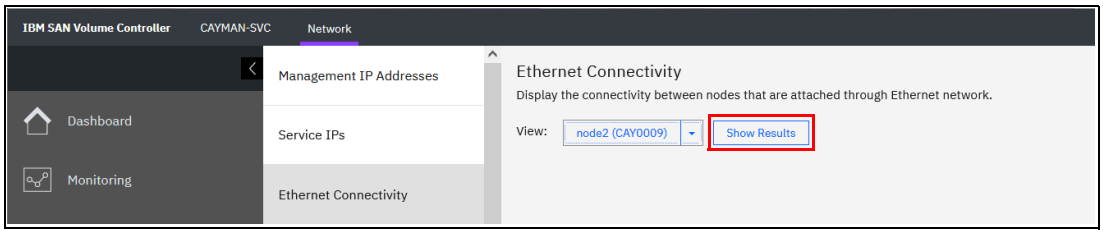


Figure 4-72 Ethernet Connectivity

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 either the RDMA over Converged Ethernet (RoCE) protocol or the internet Wide Area RDMA Protocol (iWARP). The panel 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**. Select one of the following options to change the use of the port:

- Modify VLAN
- Modify IP Settings
- 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 right-click the node and select **IP Login Information**. This information can be used to detect connectivity issues between the system and hosts and to improve the configuration of iSCSI host to optimize performance. Select **Ethernet Ports** for an overview from the menu, as shown in Figure 4-73 on page 207.

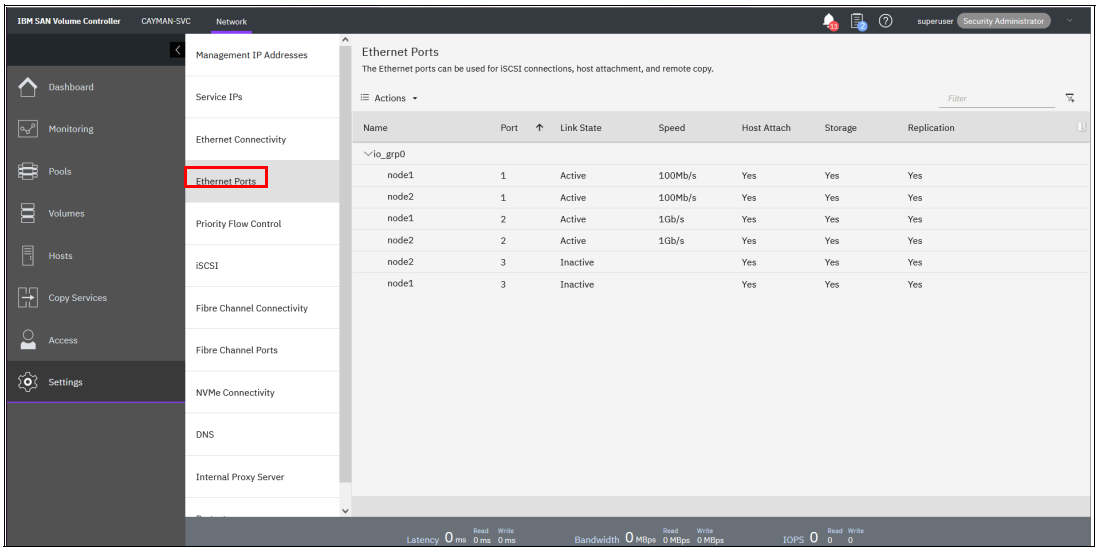


Figure 4-73 Ethernet Ports

For more information about planning, see 2.5, “Connectivity planning” on page 64.

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 the guidelines and examples that are presented here.

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 details.

- 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 are planning on using the same port for different types of traffic, ensure that the ETS settings are configured in the network.

4.10.3 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 4-74.

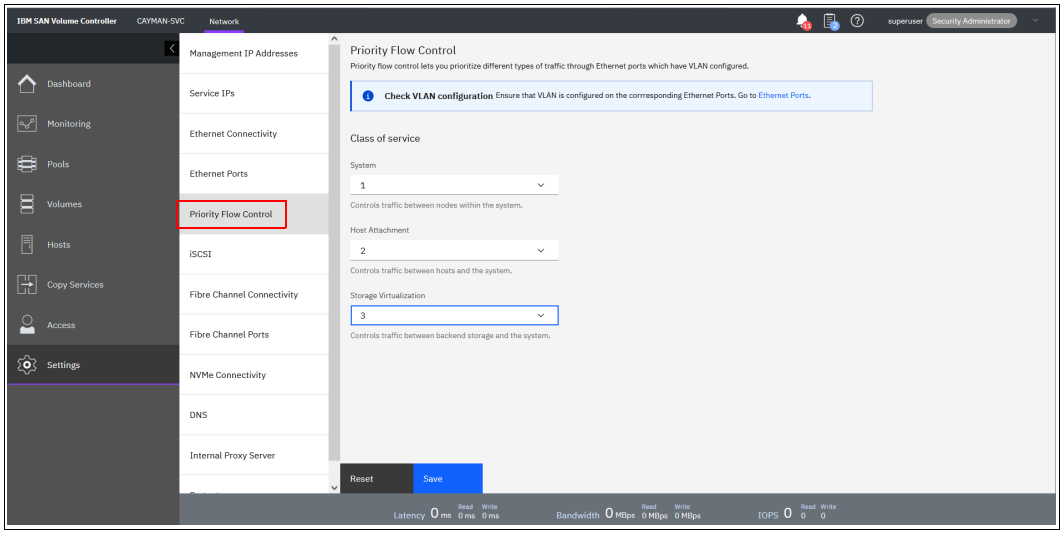


Figure 4-74 Priority flow control

2. For each of 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 on 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 on 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.

     Make sure that IP is configured with VLAN.

iSCSI information

From the iSCSI pane in the **Settings** menu, you can display and configure parameters for the system to connect to iSCSI-attached hosts, as shown in Figure 4-75.

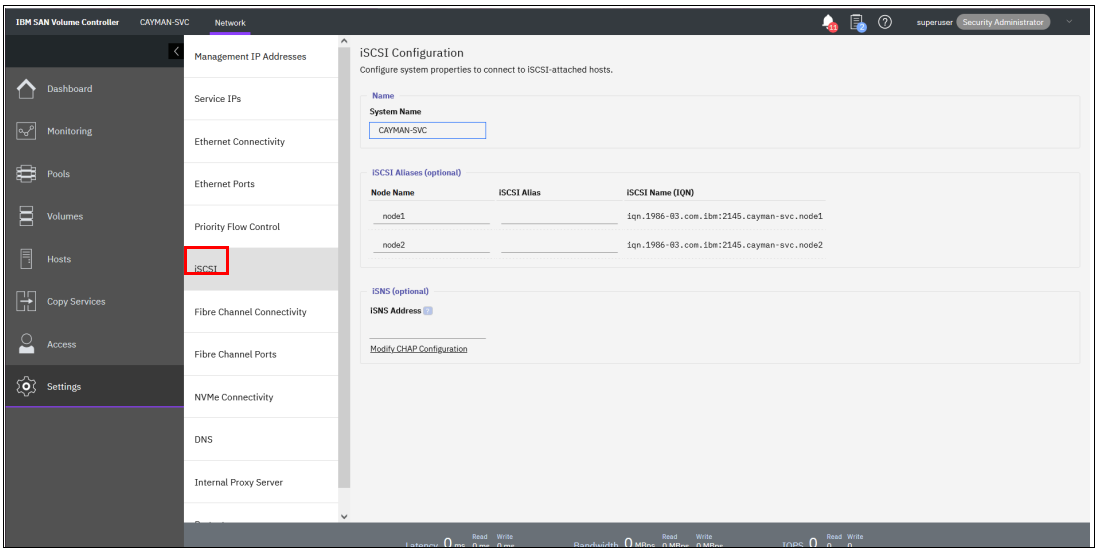


Figure 4-75 iSCSI Configuration pane

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. Complete the following steps to change an iSCSI alias: Click an 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 whole 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 using CHAP.

Fibre Channel information

As shown in Figure 4-76, you can use the Fibre Channel (FC) Connectivity pane to display the FC connectivity 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

You can view Fibre Channel Connectivity as shown in Figure 4-76.

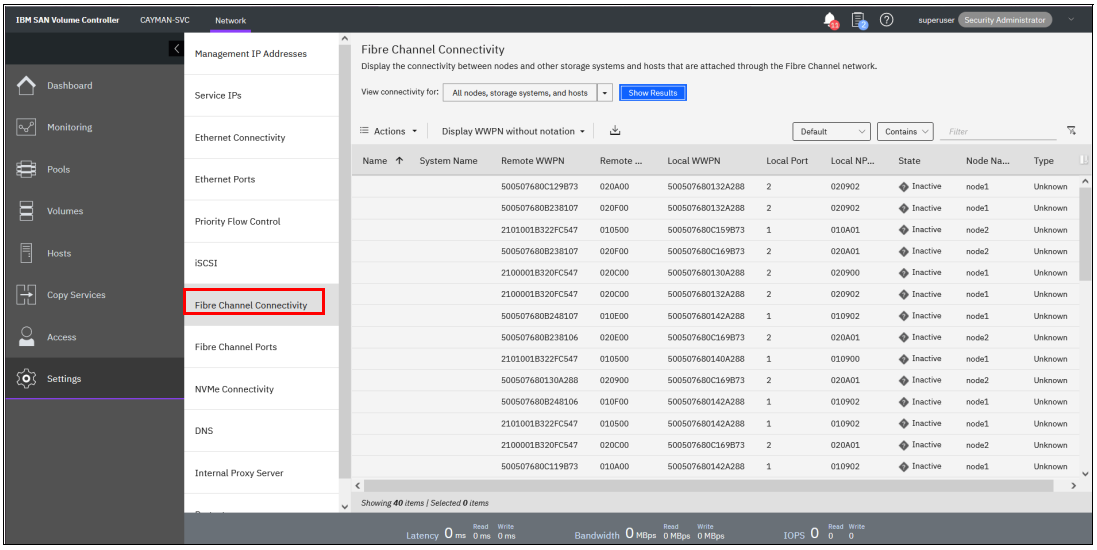


Figure 4-76 Fibre Channel Connectivity

In the Fibre Channel Ports pane, you can use this view to display how the FC port is configured across all control node canisters in the system. For example, this view helps 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 4-77).

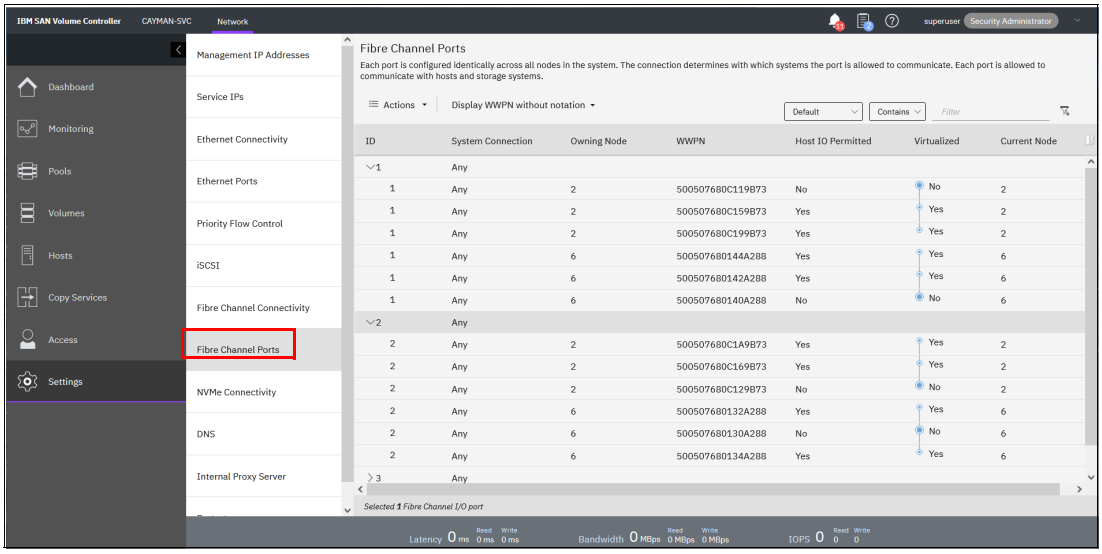


Figure 4-77 Viewing Fibre Channel Port properties

You can use the FC ports panel 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 FC port masking is used to filter out the data that is not intended on a specific FC port. In this way, you can choose the type of the traffic (service) that is required on each port. Four services normally use FC 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 either **`mkhost`** or **`chhost`** commands. External storage system communication also is 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 FC 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 connectivity

Use the Non-Volatile Memory Express (NVMe) Connectivity page to display connections between nodes and hosts that use a NVMe over FC connection. For other specific information about NVMe over Fibre Channel (FC-NVMe), such as interoperability requirements, see this IBM Support web page.

If your system supports an FC-NVMe connection between nodes and hosts, you can display details about each side of the connection. To display node details, select the node from the drop-down menu and 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. Since the system sends 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. If a connection problem persists between the host and the node, a reduced node login count or the status of the host indicates it is degraded, which you can view by selecting **Hosts** → **Ports by Host** in the management GUI. Verify these values in the management GUI, and view the messages by selecting **Monitoring** → **Events**.

Figure 4-78 shows the NVMe connectivity menu.

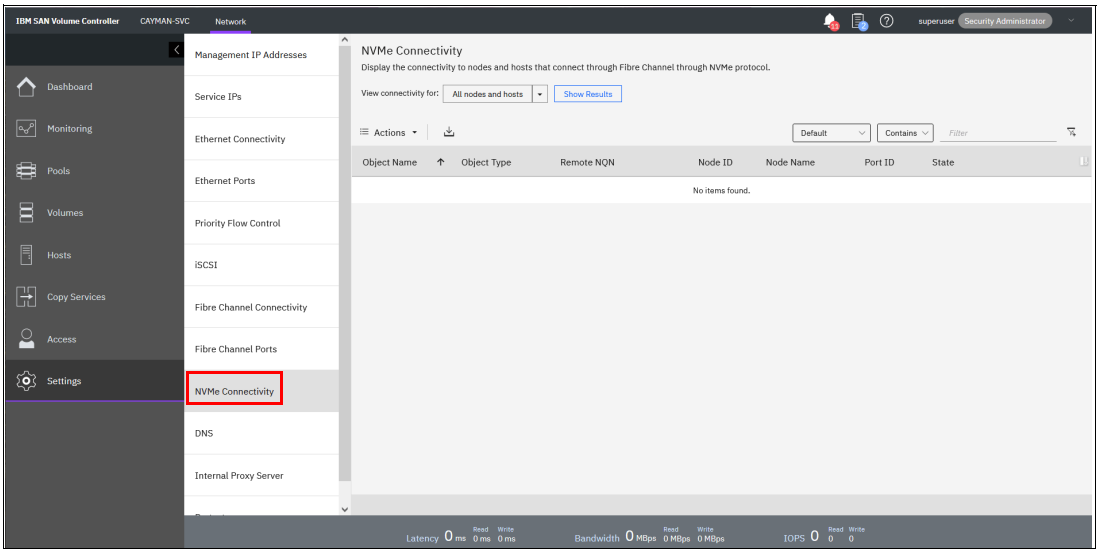


Figure 4-78 NVMe Connectivity window

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 Linux specific guidelines.

Domain name server

IBM Spectrum Virtualize allows DNS entries to be manually set up in the system. The information about the DNS servers helps the system to access the DNS servers to resolve the names of the computer resources that are in the external network.

To view and configure DNS server information in IBM Spectrum Virtualize, complete the following steps:

1. In the left pane, click **DNS** and select **Add DNS server**. Enter the IP address and Name of each DNS server. IBM Spectrum Virtualize supports up two DNS servers for IPv4 or IPv6 (see Figure 4-79).

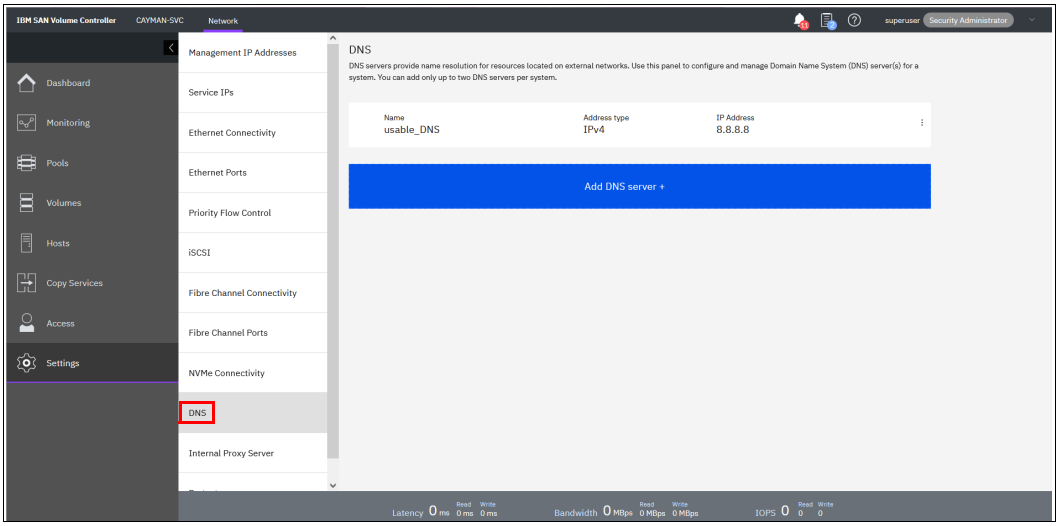


Figure 4-79 DNS information

2. Click **Save** after you finish entering the DNS server information.

Internal Proxy Server

You can configure an internal proxy server to manage incoming and outgoing connections to the system (see Figure 4-80).

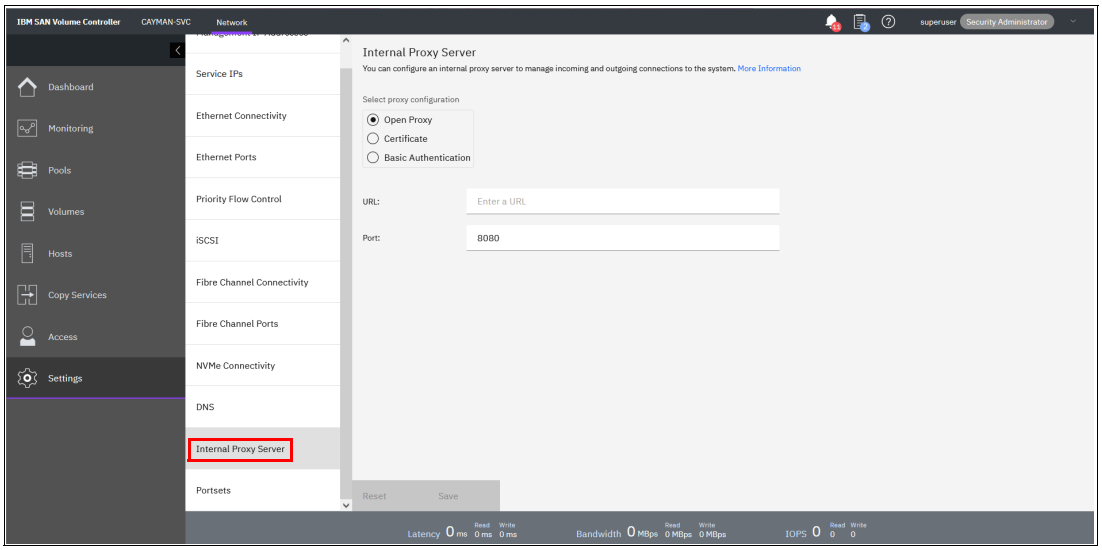


Figure 4-80 Internal Proxy Server

As part of IBM 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 outside of your network.

Call Home with cloud services or remote support assistance requires connections to the IBM 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 also is 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. command. To define an internal proxy server, complete the following steps:

1. In the management GUI, select **Settings** → **Network** → **Internal Proxy Server**.
2. Select one of the following types of proxy configurations 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 have 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 (see Figure 4-81).

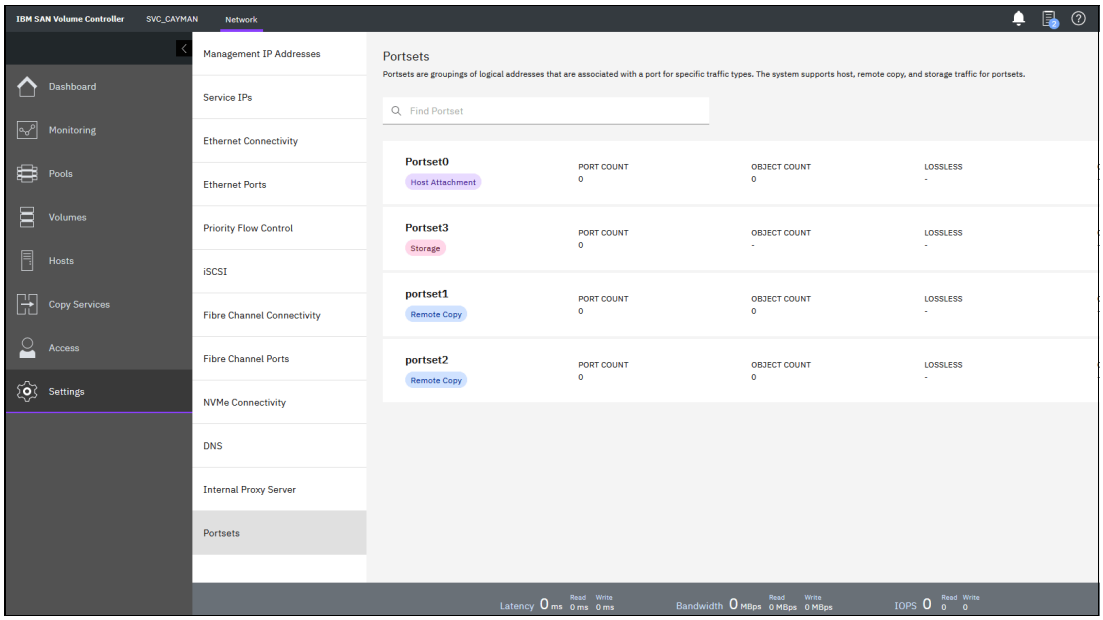


Figure 4-81 Portsets window

A port supports one IP 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 port 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 requirements:

- 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.
- Multiple portsets can share IP addresses across different portsets that allow the same IP address to be used for host, storage, and remote-copy traffic. All duplicate 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 is 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 mean that not all 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 need to configure VLAN settings on the switches and hosts before you can create portsets and IP addresses. For more information, see IBM Documentation.

- 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.
- 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 addresses configured for storage traffic (in previous releases), it is automatically mapped to portset 3 after upgrade.

Requirements for host portsets

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 four IP addresses per node.
- A single portset can contain IPv4 or IPv6 IP or a mix of IPv4 and IPv6 addresses.
- Portsets can have IP addresses that are configured on ports with different capabilities (support iWARP+RoCE capable hardware).

Requirements for replication portsets

You can create more portsets for IP partnerships. The following requirements are specific to replication portsets:

- Replication portset can have maximum of 1 IP address per node.
- All IP addresses in replication portsets must be either IPv4 or IPv6 addresses. You cannot mix IP protocol versions on replication portsets. The protocol for 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 Link 2 field.
- 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 located 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:

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 (see Figure 4-82).

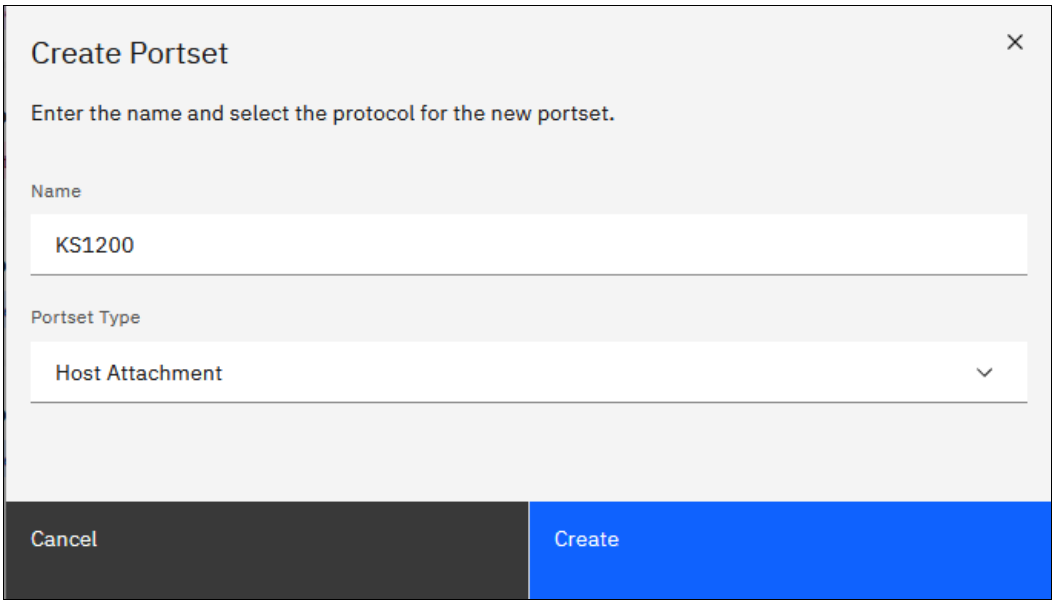


Figure 4-82 Create Portset

4. 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: Enter the IP address to associate with the selected port.
   – Type: Select the IP protocol version of the IP address.
   – Subnet Mask or Prefix: Enter the subnet mask for the IPv4 addresses or enter the prefix for IPv6 addresses.
   – VLAN: Enter the corresponding VLAN tag to which this IP address belongs.
   – Portset: Select the name of the portset and ensure that 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 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.

4.10.4 Security

Click **Settings** → **Security** to view and change security settings, authenticate users, and manage secure connections (see Figure 4-83)).

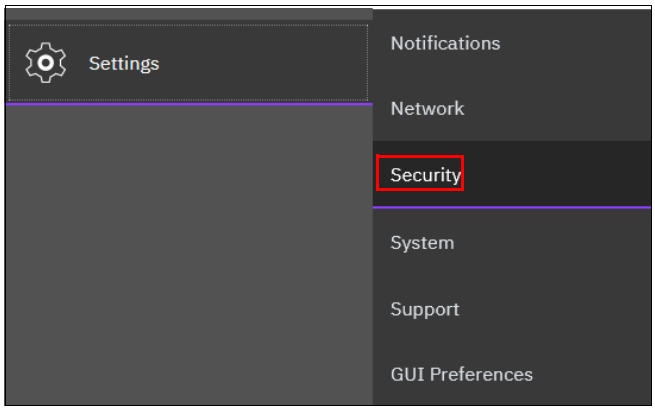


Figure 4-83 Security settings

Remote Authentication

In the Remote Authentication pane, you can configure remote authentication with LDAP, as shown in Figure 4-84. By default, the system has local authentication that is enabled.

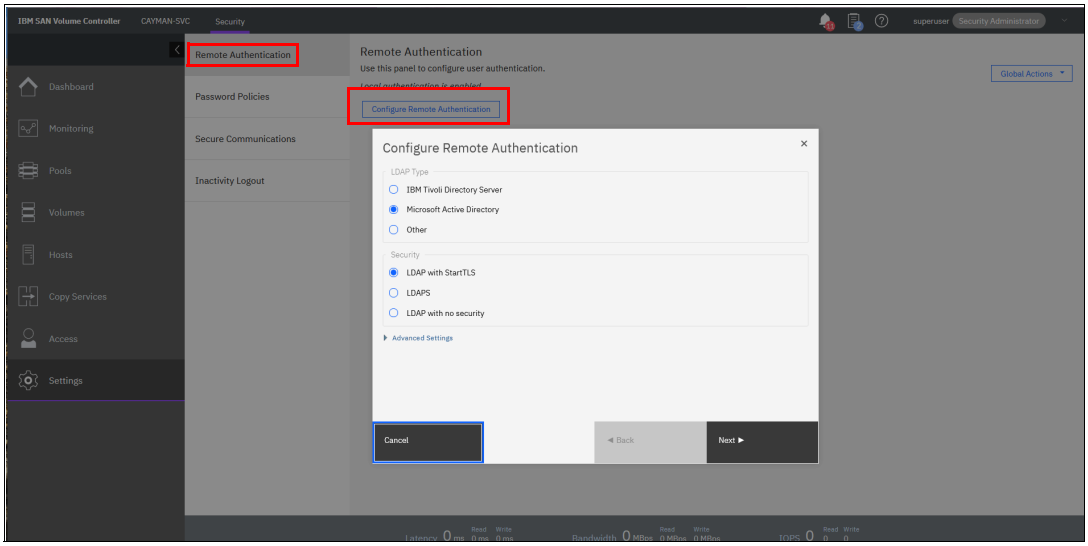


Figure 4-84 Configuring Remote Authentication

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 about how to configure remote log on, see this IBM Documentation web page.

Encryption

As shown in Figure 4-85, you can enable or disable the encryption function on an IBM SAN Volume Controller System. For more information, see Chapter 12, “Encryption” on page 793 for more information.

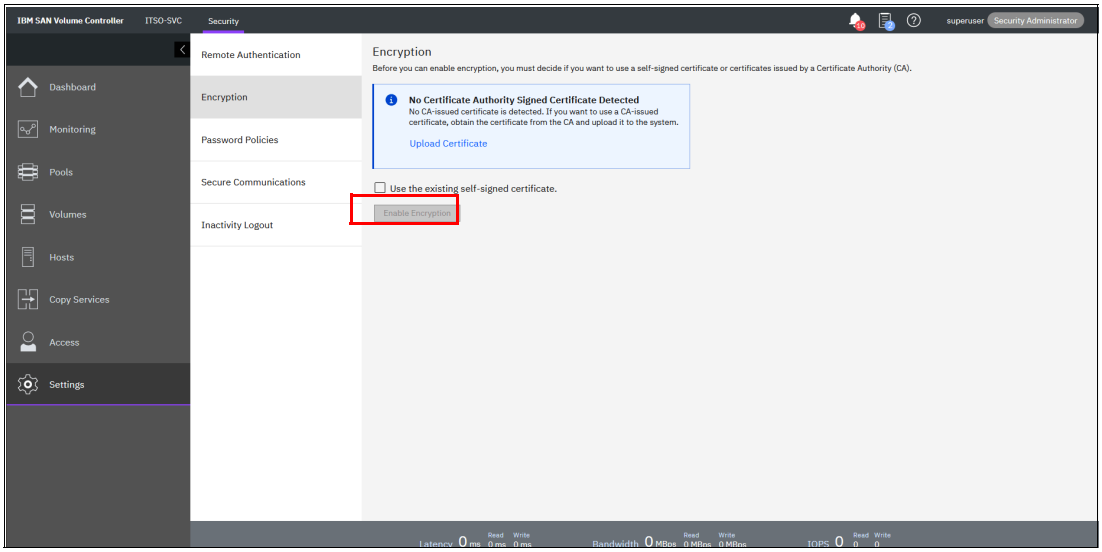


Figure 4-85 Enable Encryption

Password Policies

By using the Password Policies panel, you can define policies for password management and expiration (see Figure 4-86).

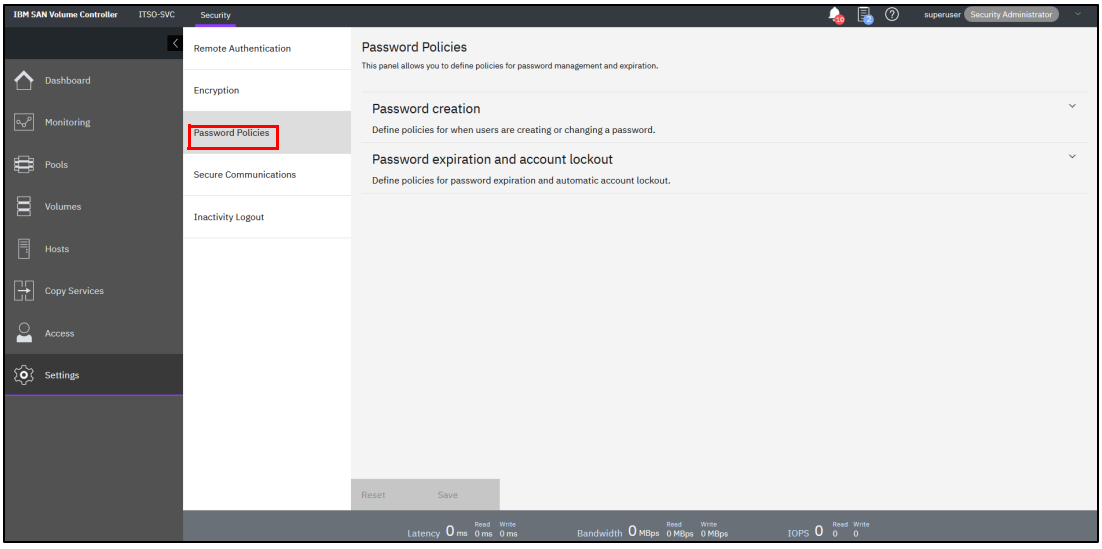


Figure 4-86 Password Policies panel

With password policy support, system administrators can set security requirements that are related to password creation and expiration, time-out for inactivity, and actions after failed log in 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 with this support:

- Password creation rules

  Administrator can set and manage the following rules for all passwords that are created in the system:

  - Specify password length requirements for all users.
  - Require passwords to use uppercase and lowercase characters.
  - Require passwords to contain special characters.
  - Prevent users from reusing recent passwords.
  - Require users to change password on next login under any of these conditions:
    - Their password expired.
    - An administrator created new accounts with temporary passwords.

- Password expiration and account locking rules

  The administrator can create the following rules for password expiration:

  - Set password expiration limit.
  - Set a password to expire immediately.
  - Set number of failed login attempts before the account is locked.
  - Set time 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.

Secure Communications

To enable or manage secure communications, select the **Secure Communications** pane, as shown in Figure 4-87. Before you create a request for either type of certificate, ensure that your current browser does not have restrictions about the type of keys that are used for certificates.

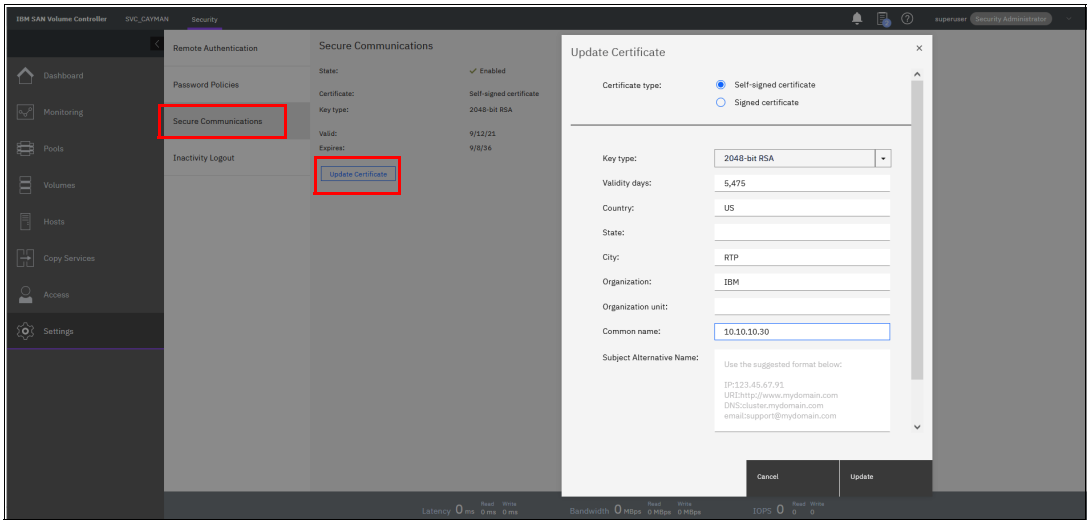


Figure 4-87 Configuring secure communications and updating certificates

Some browsers limit the use of specific key-types for security and compatibility issues. Select **Update Certificate** to add certificate details, including certificates that were created and signed by a third-party certificate authority (CA).

If your system supports key server-based encryption, the system also uses certificates to secure communications between key servers that are used to distribute and manage encryption keys to the system. If encryption is used on the system, certificates also secure communications between key servers that are used to distribute and manage encryption keys to the system. If a certificate is changed, the certificate must also be updated on all configured key servers or access to encrypted data can be lost.

During system setup, an initial certificate is created to use for secure connections between web browsers. Based on the security requirements for your system, you can create a self-signed certificate or install a signed certificate that is created by a third-party CA. Self-signed certificates are generated automatically by the system and encrypt communications between the browser and the system. Self-signed certificates can generate web browser security warnings and might not comply with organizational security guidelines.

Signed certificates are created by a third-party CA. These certificate authorities ensure that certificates include the required security level for an organization based on purchase agreements. Signed certificates often feature higher security controls for encryption of data and do not cause browser security warnings.

To use a signed certificate, complete the following steps:

1. Generate and download a request for a certificate that is based on the values that are specified on the Secure Communication page.
2. Submit this request to the CA to receive a signed certificate.
3. Install the certificate by using the **Settings** → **Security** → **Secure Communications** → **Update Certificate** page.

The Chrome browser and other browsers, require a Subject Alternative Name, which is an extension to the internet standard for public key certificates. The Subject Alternative Name extension is used to match the domain name and site certificate and can be an email address, IP address, URI, or DNS name. A certificate can contain a collection of these values so that the certificate can be used on multiple sites.

For example, if the system has a single DNS server, enter `DNS:dns.myco.com` in the Subject Alternate Name field. For multiple values, list each value on a separate line within the box of the Subject Alternate Name field:

- DNS: `dns.myco.com`
- IP: `1.2.3.20`
- URI: `http:\\www.myco.com`
- Email: `support@myco.com`

Inactivity Logout

By using this panel, you can set the inactivity time that is allowed before the system logs out users (see Figure 4-88). You can set values for command-line access and management GUI access.

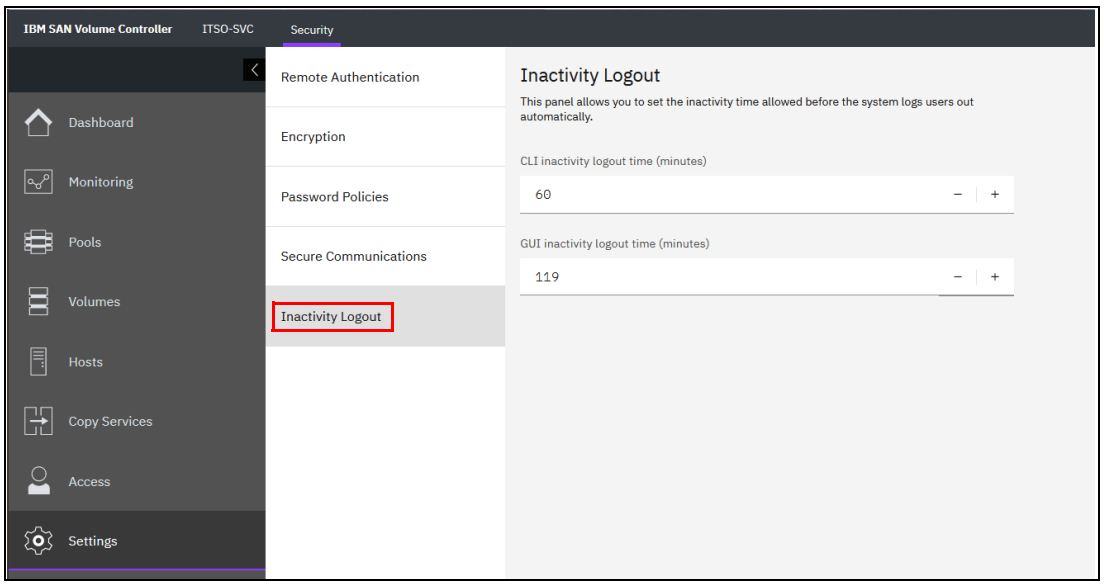


Figure 4-88 Inactivity Logout panel

4.10.5 System menus

Click the **System** option from the **Settings** menu (see Figure 4-89) to view and change the time and date settings, work with licensing options, download configuration settings, work with VMware VVOLs and IP Quorum, or download software upgrade packages.

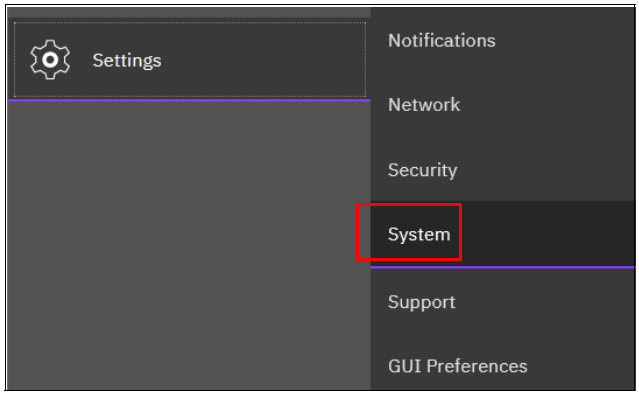


Figure 4-89 System option

Date and time

Complete the following steps to view or configure the date and time settings:

1. From the IBM SAN Volume Controller System pane, click **Settings** and click **System**.
2. In the left column, select **Date and Time**, as shown in Figure 4-90 on page 223.

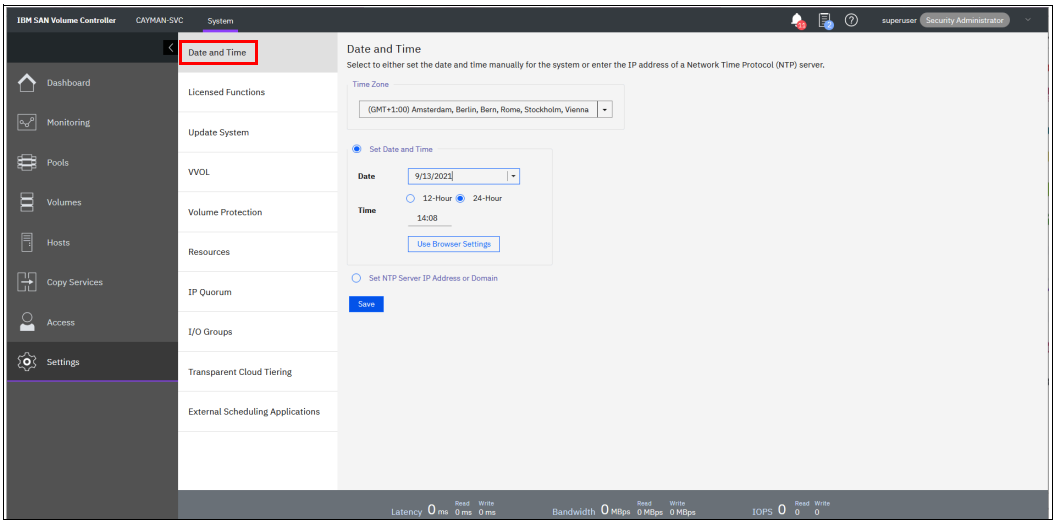


Figure 4-90 Date and Time window

3. From this pane, 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, select **Set Date and Time**, and then, manually enter the date and time for your system, as shown in Figure 4-91. You can click **Use Browser Settings** to automatically adjust the date and time of your IBM SAN Volume Controller system with your local workstation date and time.

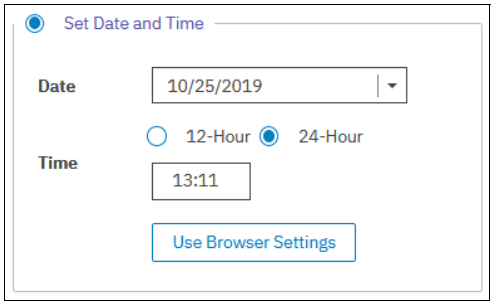


Figure 4-91 Set Date and Time window

     - If you are using 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 4-92.

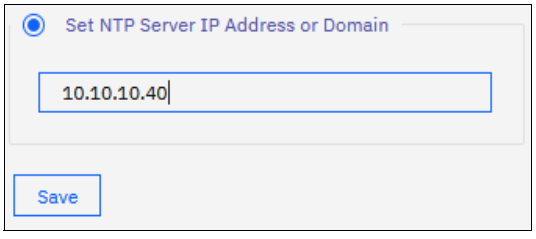


Figure 4-92 Set NTP Server IP Address window

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.

The IBM SAN Volume Controller supports both 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 are based on the number of Storage Capacity Units (SCUs) that are purchased. With other functions, like 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 based on the type of storage.

Table 4-1 lists the different storage types and the associated SCU ratios.

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 usable of Category 1 storage
Flash	All flash devices, other than SCM drives	One SCU equates to 1 TiB usable of Category 1 storage
Enterprise	10K or 15K RPM drives	One SCU equates to 1.18 TiB usable of Category 2 storage
Nearline (NL)	Nearline Serial ATA (SATA) drives	One SCU equates to 4.00 TiB usable of Category 3 storage

License settings are initially entered into a system initialization wizard. They can also be changed later.

To change the license settings or view the current license usage, select **Settings** → **System** → **Licensed Functions**, as shown in Figure 4-93.

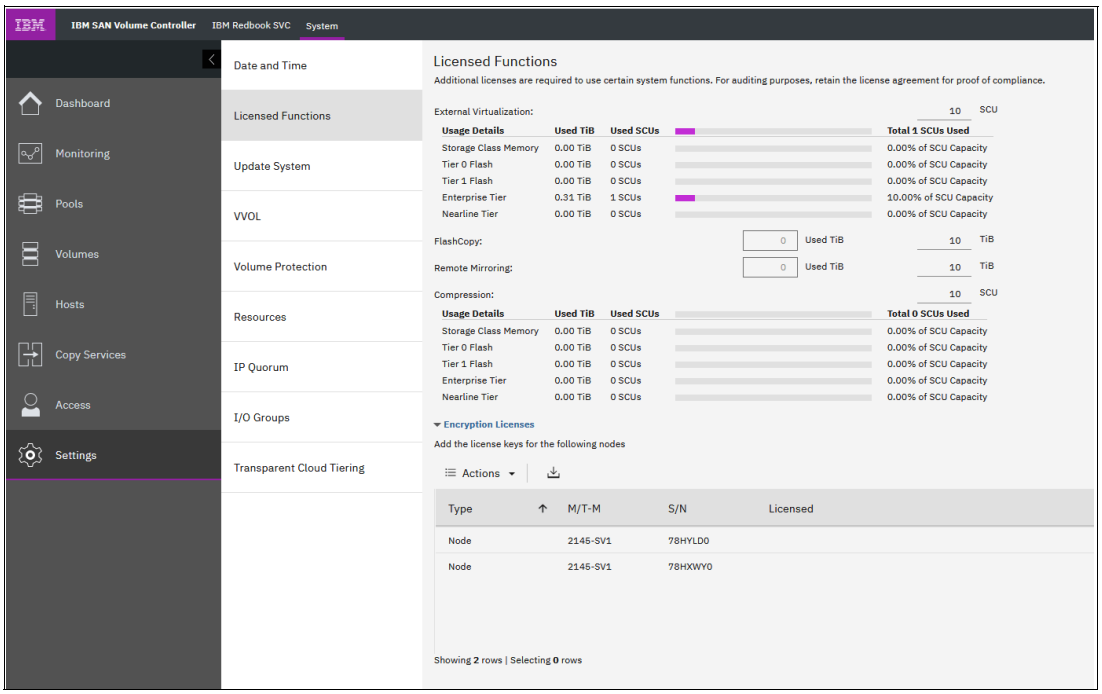


Figure 4-93 Licensed functions pane

The following elements are displayed:

- External Virtualization

  You can enter the number of SCU units that are licensed for External Virtualization for your IBM SAN Volume Controller environment.

  When monitoring External Virtualization license usage, consider the following items:

  - The license accounts for usable MDisk capacity. For example, one SCU is used for one 1 TB Tier0 MDisk when it is assigned to a storage pool, independently of the amount of the actual data written on the MDisk.
  - SCU is used for complete and incomplete chunks of MDisk capacity. For example, if a combined capacity of all NL Tier MDisks in your system is 5 TB, two SCUs are needed: one SCU for 4 TB of NL storage, and one SCU for another 4 TB (even if 1 TB is used).

The system supports virtualization licensing for different tiers of storage. A license can be purchased for a specific quantity of storage capacity units (SCU) that can be used to virtualize a designated number of tebibytes (TiB) of storage. The used capacity for each tier can be monitored to learn 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 endurance and reduces latency 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.

- Tier 0 Flash

  Tier 0 flash contains high-performance tier 0 flash 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.

- Tier 1 Flash

  Tier 1 flash contains high performance tier 1 flash 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. 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. 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 Value

  The FlashCopy function copies the contents of a source volume to a target volume. It also is used to create cloud snapshots of volumes in systems that include TCT enabled. FlashCopy is licensed in terabytes.

  Important: The used capacity for FlashCopy mappings is the sum of all the volumes that are the source volumes of a FlashCopy mapping and volumes with cloud snapshots.

- Remote Mirroring Value

  Enter the capacity that is available for MM and GM relationships. 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. Remote mirroring is licensed in terabytes.

  Important: The used capacity for GM and MM is the sum of the capacities of all of the volumes that are in an MM or GM relationship. Both master and auxiliary volumes are counted.

- Compression Value

  With the compression function, data is compressed as it is written to disk, saving 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. But for a Compression License, *used capacity* (and not *usable capacity*) is calculated.

  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 SCUs, 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 following tiers of storage:

- Storage Class Memory

  Storage Class Memory storage uses persistent memory technologies that improves 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.

- Tier 0 Flash

  Tier 0 flash contains high-performance tier 0 flash 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.

- Tier 1 Flash

  Tier 1 Flash contains tier 1 flash 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. 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. 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.

Encryption licenses

In addition to the enclosure-based licensed functions, the system 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 system 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.

Encryption is activated on a per system basis, and an active license is required for each node that uses encryption. During system setup, the system detects the nodes that support encryption, and a license should be applied to each one. If more nodes are added and require encryption, more encryption licenses must be purchased and activated.

Note: To monitor license usage, you can run the **`lslicense`** CLI command. For more information, see this IBM Documentation web page.

Updating your system

Use the Update System panel to install new versions of the software. Before you start a system update, ensure that no problems exist on the system that might interfere with a successful update of the system.

Each update requires that you run the update test utility and then, download the correct software package. The update test utility indicates whether your current system includes issues that must be resolved before you update to the next level. The test utility is run as part of the system update process.

After you run the update test utility and verify that your system is ready to be updated, the latest code is downloaded. Contact your support center for the most recent update for the system.

After the update package is downloaded, 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 *must* be completed in a specific order or the update fails.

For more information about the update procedure that uses the GUI, see Chapter 13, “Reliability, availability, serviceability, monitoring and troubleshooting” on page 853.

VMware vSphere Virtual Volumes (VVOL)

IBM Spectrum Virtualize can manage VVOLs directly in cooperation with VMware. It enables VMware virtual machines (VMs) to get the assigned disk capacity directly IBM SAN Volume Controller rather than from the ESXi data store. This technique enables storage administrators to control the appropriate usage of storage capacity, and to enable enhanced features of storage virtualization directly to the VM (such as replication, thin-provisioning, compression, and encryption).

VVOL management is enabled in the IBM SAN Volume Controller in the System section, as shown in Figure 4-94. The NTP server must be configured before enabling VVOLs management. As a best practice, use the same NTP server for ESXi and your IBM SAN Volume Controller.

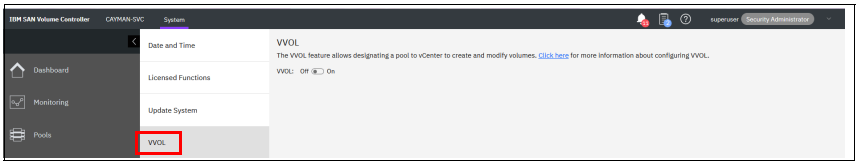


Figure 4-94 Enabling VVOLs management

Restriction: You cannot enable VVOLs support until the NTP server is configured in IBM SAN Volume Controller.

For more information about VVOLs, see the following publications:

- *Quick-start Guide to Configuring VMware Virtual Volumes for Systems Powered by IBM Spectrum Virtualize*, REDP-5321.
- *Configuring VMware Virtual Volumes for Systems Powered by IBM Spectrum Virtualize*, SG24-8328.

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 either 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 you force the deletion.

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 remote-copy 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 has 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
- Deleting a volume copy
- Deleting a host or a host cluster mapping
- Deleting a storage pool
- Deleting a host from an I/O group
- Deleting a host or host cluster
- Deleting a defined host port
- Creating a remote-copy relationship

Figure 4-95 shows the menu for volume protection.

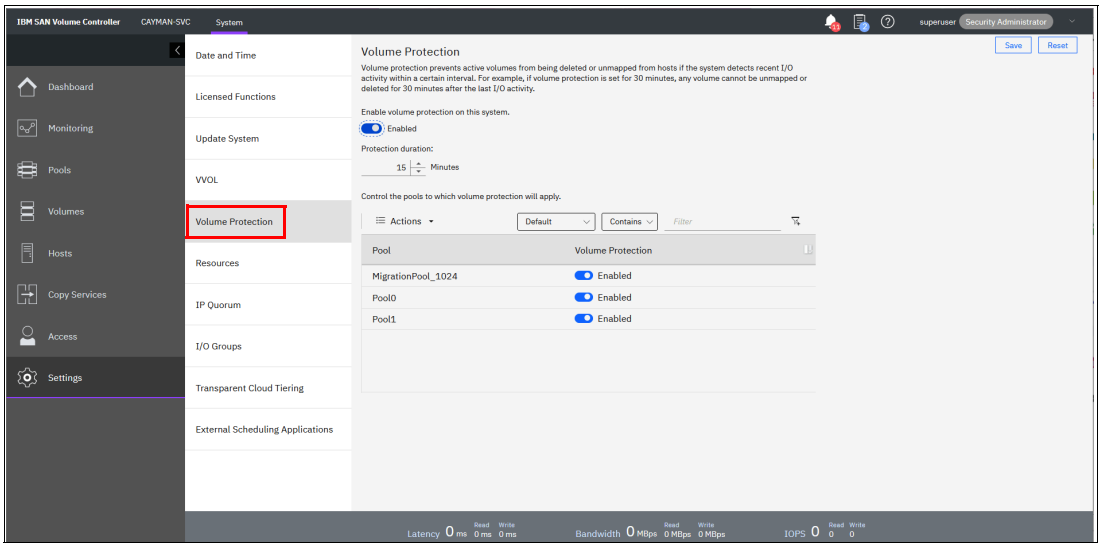


Figure 4-95 Volume protection

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 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.

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**, as shown in Figure 4-96.

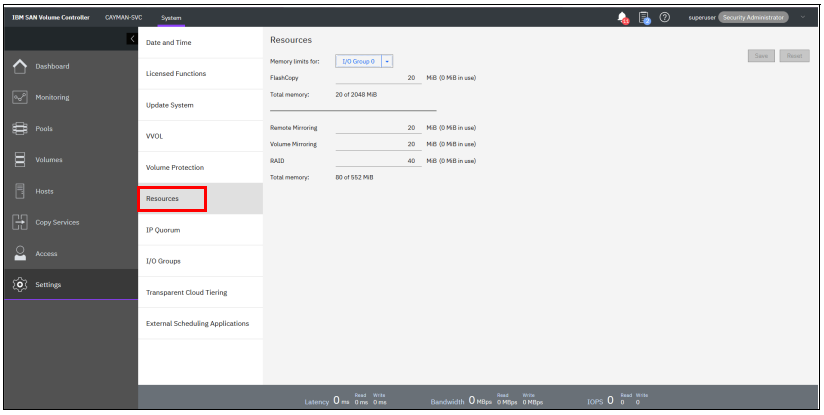


Figure 4-96 Resources allocation

Table 4-2 provides an example of 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	1 MiB of memory provides the following provisioned capacity for the specified I/O group
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.
- When a FlashCopy mapping is created, 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.
- For volume mirroring, the full 512 MiB of memory space enables 1 PiB of total provisioned capacity.
- When creating new FlashCopy relationships or mirrored volumes, more bitmap space is allocated automatically by the system if required.

IP Quorum

IBM Spectrum Virtualize also supports an IP quorum application. By using an IP-based quorum application as the quorum device for the third site, a Fibre Channel connection 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 SAN Volume Controller 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 4-97.

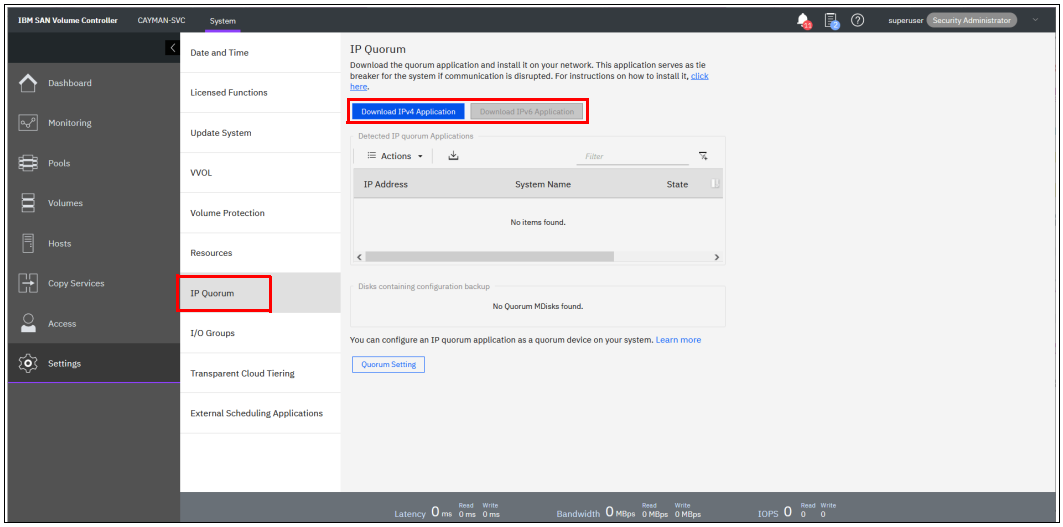


Figure 4-97 IP Quorum settings

2. When you select **Download IPv4 Application**, you are prompted if you want to download the IP quorum application with or without recovery metadata (see Figure 4-98 on page 233). 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.

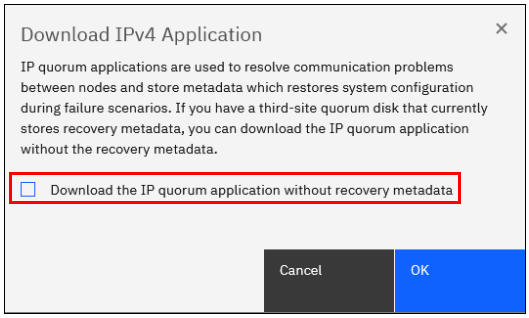


Figure 4-98 IP quorum application metadata

3. After you select your correct IP configuration, IBM Spectrum Virtualize generates an IP Quorum Java application, as shown in Figure 4-99. The application can be saved and installed in a host that is to run the IP quorum application.

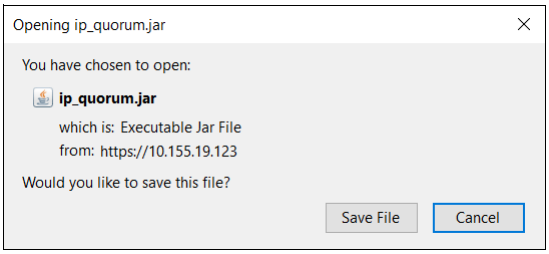


Figure 4-99 IP Quorum Java Application

4. After you download the IP quorum application, you must save the application on a separate host or server.
5. If you change the configuration by adding a node, changing a service IP address, or changing Secure Sockets Layer (SSL) certificates, you must download and install the IP quorum application again.
6. On the host, you must 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.
7. 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.
8. Change to the folder where the application is, and run the following command:

   ```
   java -jar ip_quorum.jar
   ```

   Note: The IP quorum application always must be running.

9. 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** → **Quorum Setting** and then, set the mode to **Standard**, **Preferred** or **Winner** or run the **`chsystem`** command. This configuration gives 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 4-100.

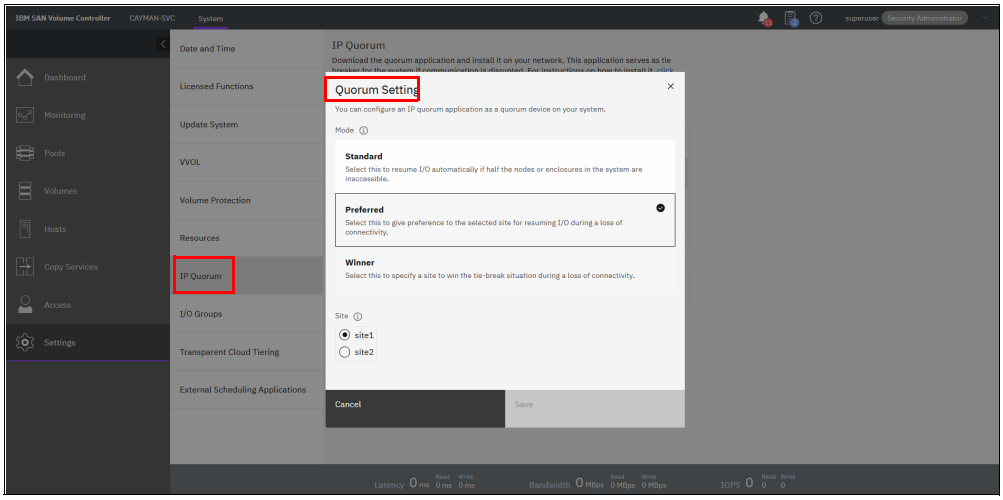


Figure 4-100 Quorum settings

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 the 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 application 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 to IP quorum applications only. It does *not* apply to Fibre Channel-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 Fibre Channel 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: The maximum number of IP quorum applications that can be deployed on a single system is *five*. 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 you have multiple San Volume Controller systems in your environment, more than one IP quorum application is allowed per host, but 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 following network requirements 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, as 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, so 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), which means 40 ms each direction.
- If you are configuring the IP quorum application without a quorum disk for metadata, a minimum bandwidth of 2 megabytes per second must be guaranteed for traffic between the system and the quorum application. However, a minimum bandwidth of 64 megabytes per second 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 megabytes 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 virtual port. When port virtualization is enabled, ports do not come up 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 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 a system, verify this setting.
- Disabled: The I/O group does not contain any virtualized ports.

- Transitional: The I/O group contains physical Fibre Channel and virtual ports that are being used. You cannot change NPIV directly from enabled to disabled states, or vice versa.

  NPIV must be in 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 NPIV is enabled.

The port virtualization settings of I/O groups are available by selecting **Settings** → **System** → **I/O Groups**, as shown in Figure 4-101.

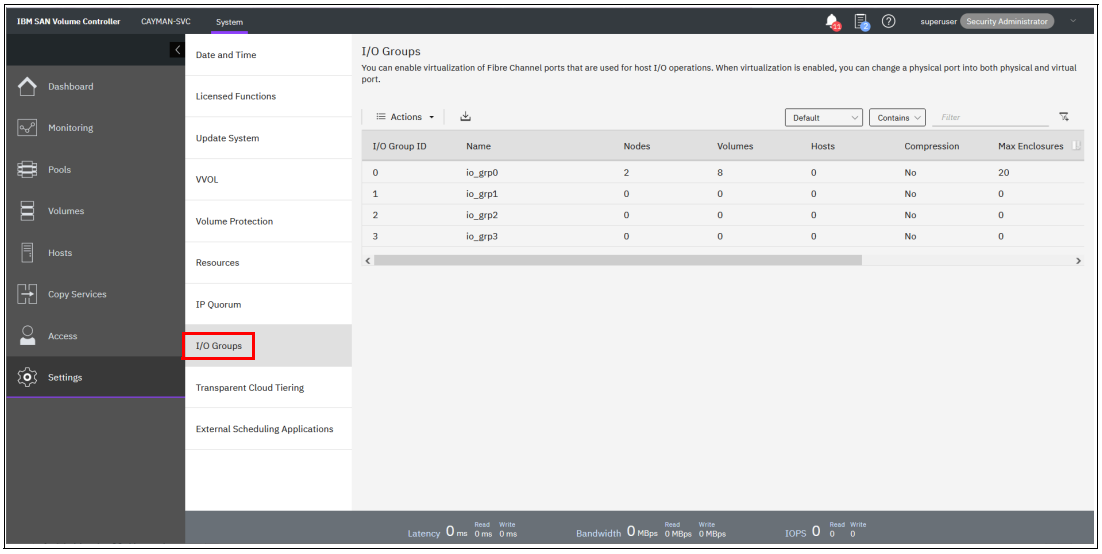


Figure 4-101 I/O groups port virtualization

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. View the **Configuring** → **Configuration details** → **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 on zoning and NPIV, view the **Configuring** → **Configuration details** → **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 more NPIV ports.

After 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 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 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, right-click the I/O group, and then, selecting **Change NPIV Settings**, as shown in Figure 4-102.

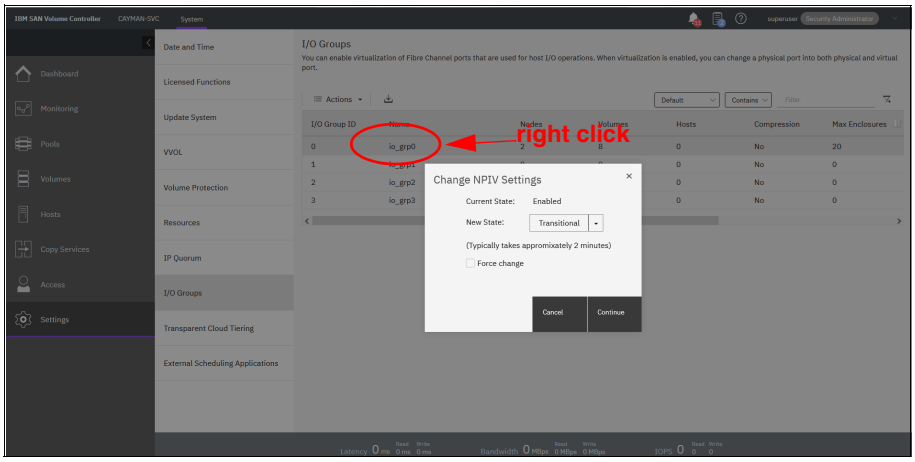


Figure 4-102 Changing the port mode

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 particular set of credentials. These credentials differ depending on the type of CSP that is being specified. Most CSPs require the host name of the CSP and an associated password, and 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 then 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 server 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 properly installed and that the 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, the following prerequisites must be met:

- 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 host names. 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.

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 also can 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. IBM SAN Volume Controller supports the following CSPs:

- IBM Cloud (formerly known as IBM SoftLayer® Object Storage)

  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 application programming interfaces (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.

To view your IBM Spectrum Virtualize cloud provider settings, from the IBM SAN Volume Controller Settings pane, click **Settings** and then, select **System**. Then, select **Transparent Cloud Tiering**, as shown in Figure 4-103.

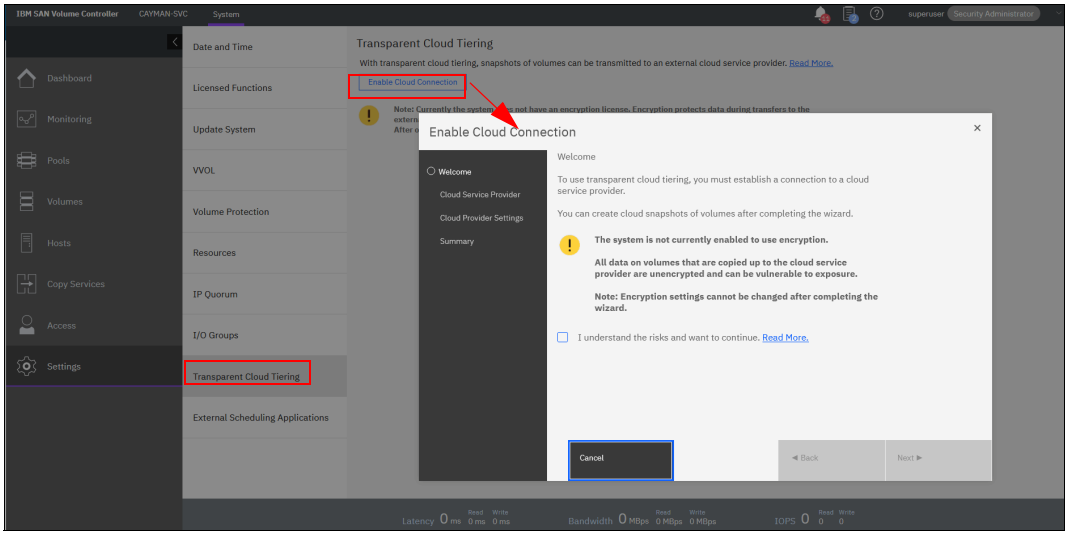


Figure 4-103 Transparent Cloud Tiering settings

By using this view, you can enable and disable features of your TCT and update the system information concerning your CSP. Use this pane to set the following options:

- CSP
- Cloud Object Storage URL
- The tenant or the container information that is associated to your Cloud Object Storage
- User name of the cloud object account

- Application programming interface (API) key
- The container prefix or location of your object
- Encryption
- Bandwidth

For more information about how to configure and enable TCT, see 10.5, “Transparent Cloud Tiering” on page 678.

Automatic configuration only for storage systems behind the IBM SAN Volume Controller

With automatic configuration, you can configure IBM FlashSystems 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 based on the technology type of the drives.
- Creates a pool for each array.
- Provisions all usable capacity in each pool to volumes 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 using this system as external storage.

Figure 4-104 shows how to enable the Automatic Configuration for Virtualization on a FlashSystem Storage System option.

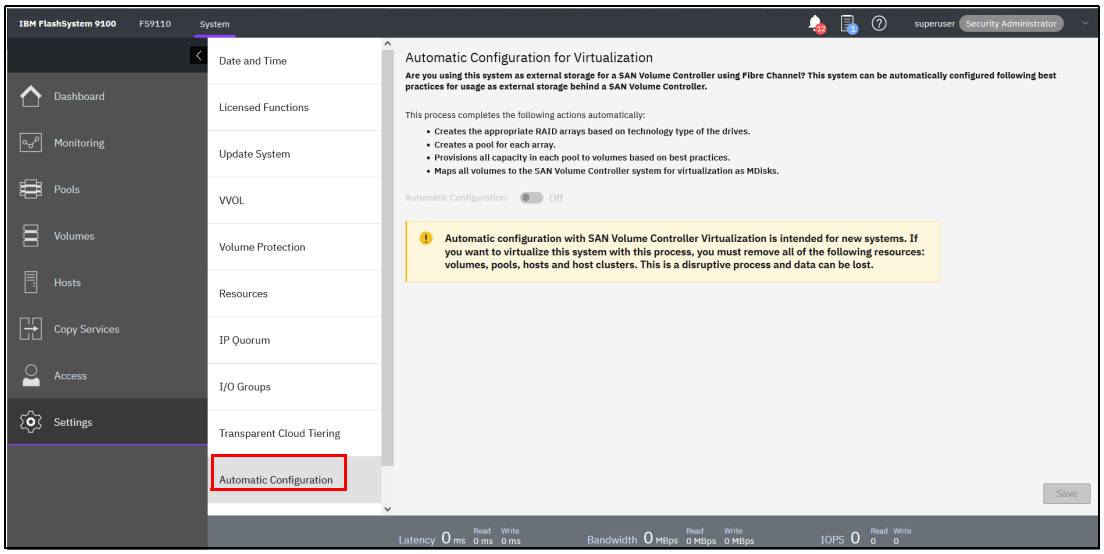


Figure 4-104 Automatic configuration for virtualization of a FlashSystem

External Scheduling Applications

The system supports IBM Copy Services Manager as an external scheduling application. IBM Copy Services Manager coordinates and automates Safeguarded Copy function across multiple systems.

IBM Copy Services Manager uses a Safeguarded backup policy to configure FlashCopy mapping and consistency groups automatically to create backup copies. When Safeguarded backup copies are created, IBM Copy Services Manager sets the retention time for the Safeguarded backup copies that are based on the settings in the Safeguarded backup policy. After copies expire, the Spectrum Virtualize software deletes these copies from the Safeguarded backup location.

As part of configuring Safeguarded Copy function on the system, you can add the web address for IBM Copy Services Manager to the system. The system uses this web address automatically when it interacts with objects that are managed with IBM Copy Services Manager.

For more information about the entire configuration process, see *Implementation Guide for SpecV/FlashSystem Safeguarded Copy*, REDP-5654.

Figure 4-105 shows the External Scheduling Applications.

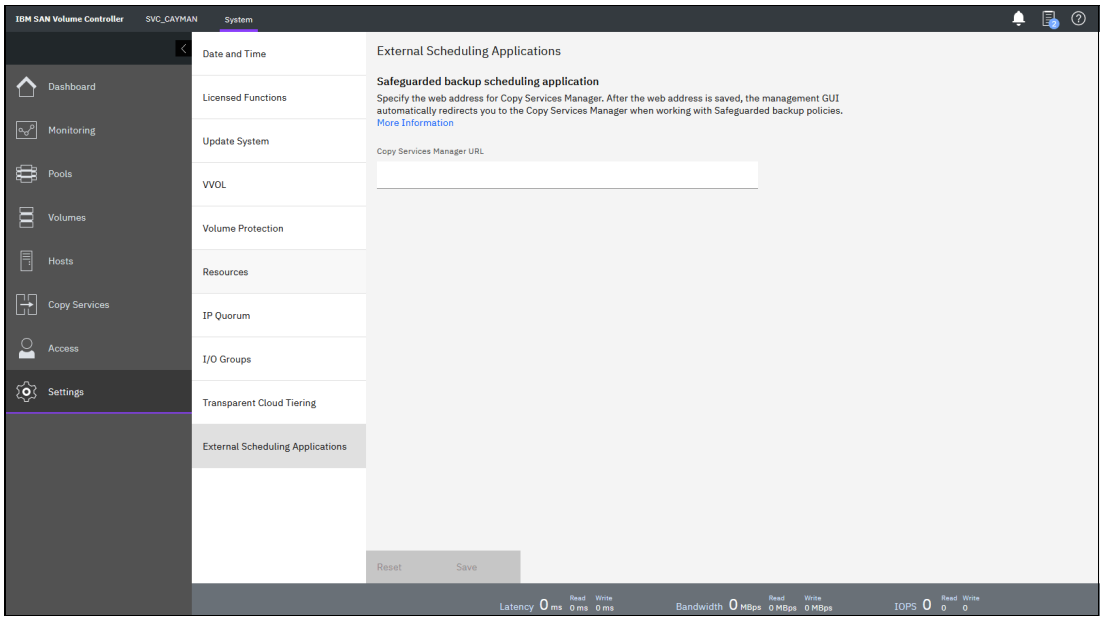


Figure 4-105 External Scheduling Applications

To add IBM Copy Services Manager as your external scheduling application, complete the following steps:

1. In the management GUI, select **Settings** → **System** → **External Scheduling Applications**.
2. Enter the web address for IBM Copy Services Manager instance that you configured with the system. Usually this address has the following format: `https://<IP address or domain name>:9559/CSM` where `<IP address or domain name>` is the IP address or domain name of IBM Copy Services Manager instance in your network.
3. Click **Save**.
4. After you specify the web address for IBM Copy Services Manager, create child pools for the Safeguarded backup locations. For more information, see *Implementation Guide for SpecV/FlashSystem Safeguarded Copy*, REDP-5654.

4.10.6 Support menu

Use the Support pane to view and change Call Home settings, configure and manage connections, and upload support packages to the IBM Support Center.

The following options are available from the menu:

- 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 4-106:

  – 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.

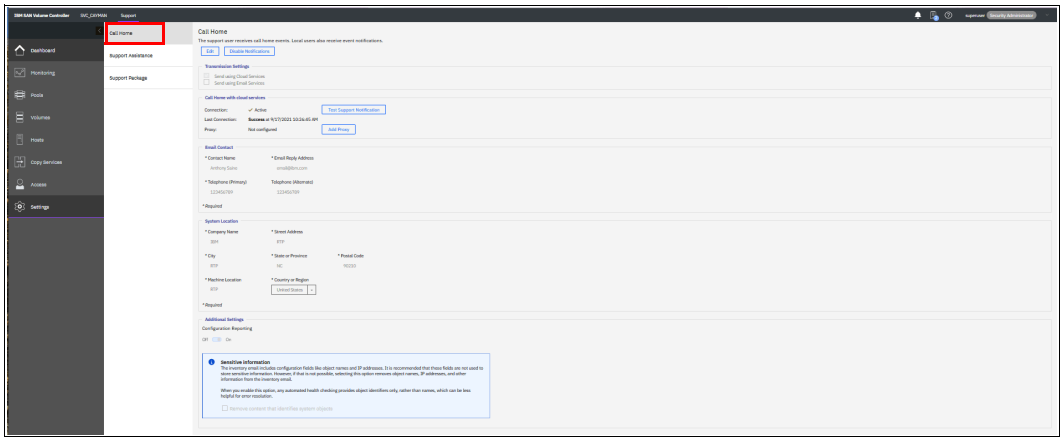


Figure 4-106 Call Home settings

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.

To update the 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 contacts 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 also is 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 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.

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, as 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 met:

- All nodes on the system have internet access.
- A valid service IP address is configured on each node on the system.

Call Home with cloud services supports the configurations that are listed in Table 4-3 that feature extra network requirements.

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 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 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 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 also can 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 also must update your network firewall settings to allow outbound traffic to `esupport.ibm.com` on port `443`. To configure a DNS server, select **Settings** → **System** → **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 option 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. If the connection continually fails for four 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 IBM Documentation.
   - `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 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 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. By using 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 require 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 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 also cam change or update current settings on the Call Home page.

Note: Call Home with email notifications also can be configured or updated by selecting **Settings** → **Notifications** → **Email**.

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 **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 also is 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 are using 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 also can use to automatically generate recommendations that are based on your actual configuration. Sensitive data can be redacted from these reports, if necessary.

8. Click **Save**.

If your system also supports both of these transmission methods, you can configure them both 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 use 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 on-site 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 by only 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 ensure that Call Home is configured and a valid email server is specified. Call Home automatically contacts IBM 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, go to **Settings** → **Network** → **Service IPs** to configure a service IP for each node on the system. Optionally, you can configure a remote proxy server if you use a firewall to protect your internal network. For more information about proxy configuration requirements, see IBM Documentation.

Figure 4-107 shows the panel for enabling or reconfigure the Support assistant.

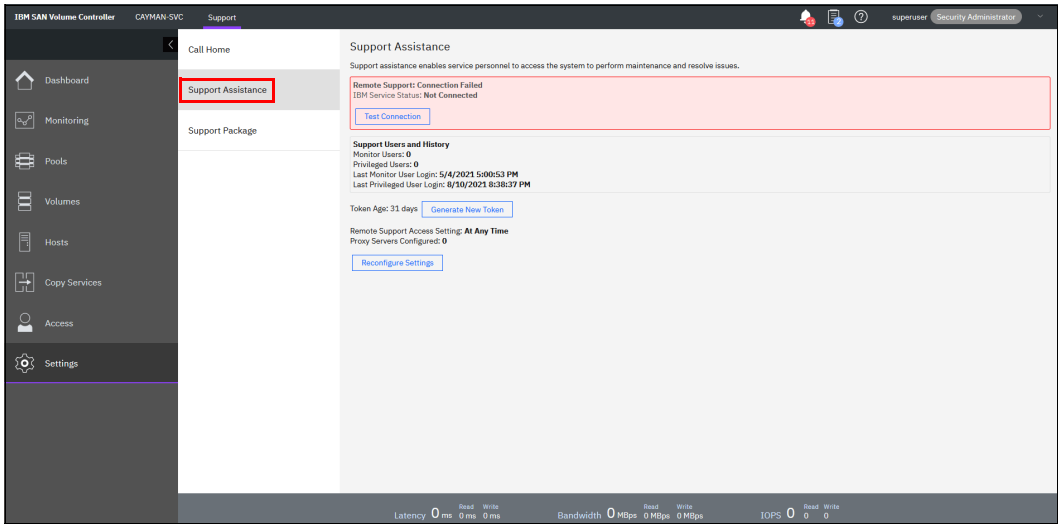


Figure 4-107 Support Assistance

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 is generated by the system and sent to IBM Support center. If the system needs support services, IBM Support personnel can be authenticated onto the system with a challenge-response mechanism.

After IBM Support personnel obtain the response code, it is entered to gain access to the system. Service personnel must enter the correct response code within three attempts. After three failed attempts, the system generates a new random challenge and IBM 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 also can 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 IBM Support to help analyze and resolve errors on the system. You can generate new support packages and upload them to the support center or select specific logs from a specific time.

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 4-108.

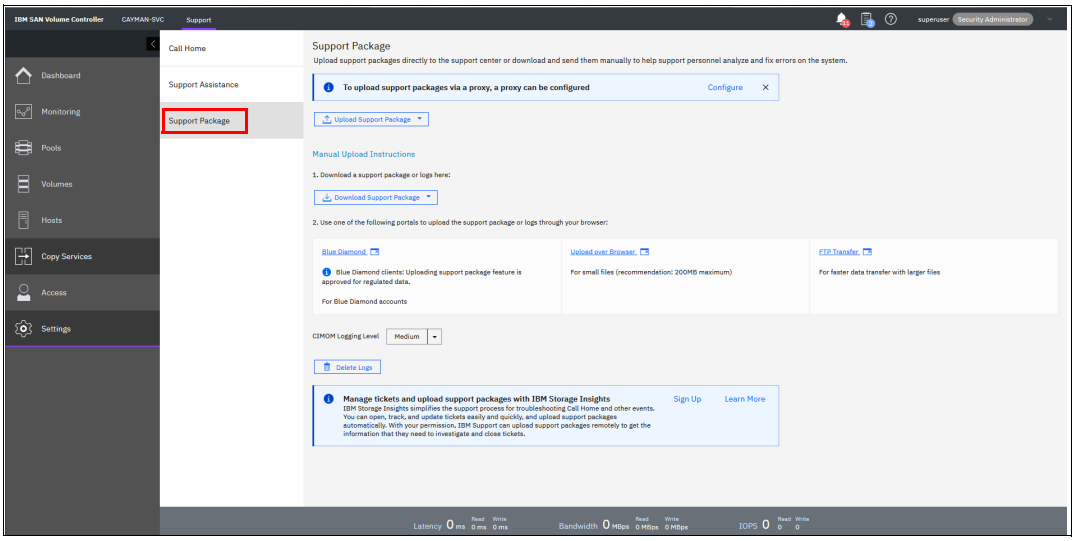


Figure 4-108 Support package

For more information about how the Support menu helps with troubleshooting your system or how to back up your systems, see Chapter 13, “Reliability, availability, serviceability, monitoring and troubleshooting” on page 853.

4.10.7 GUI Preferences menu

The GUI Preferences menu consists of the following options:

- Login
- General
- Notification Behavior
- Language
- Sidebar Accent Color

Figure 4-109 shows the GUI Preferences selection window.

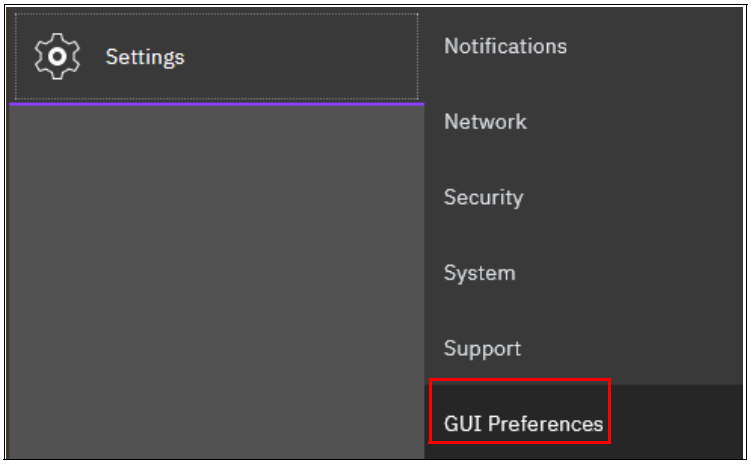


Figure 4-109 GUI Preferences selection window

Login message

IBM Spectrum 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 need to 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 4-110 on page 251).

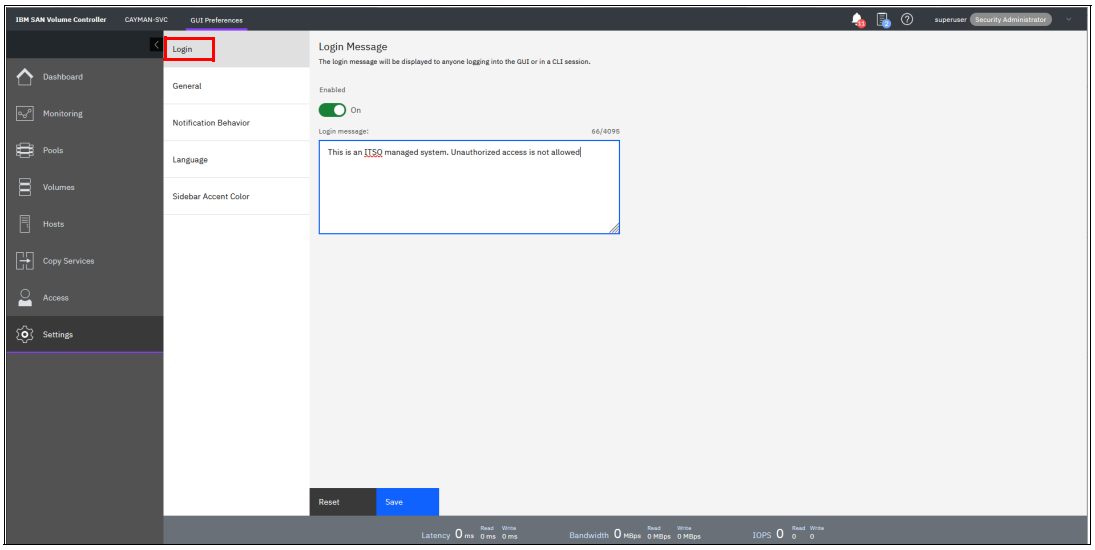


Figure 4-110 Enabling the login message

The resulting login dialog box is shown in Figure 4-111. The system shows the welcome message in the GUI before login.

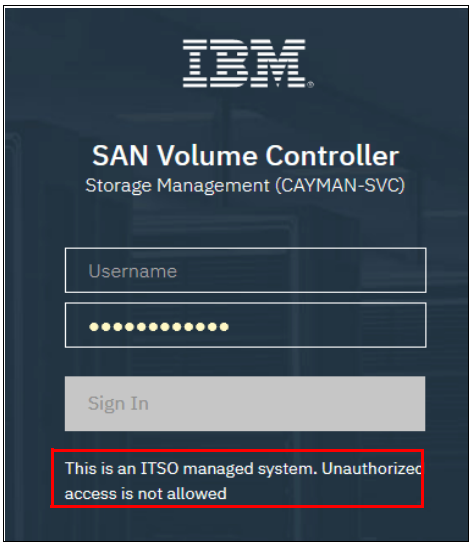


Figure 4-111 Welcome message in the GUI

Figure 4-112 shows the welcome message as it appears in the CLI.

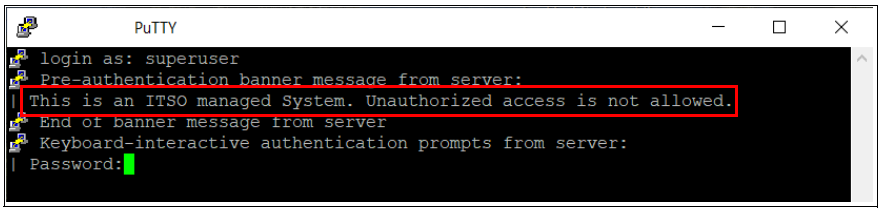


Figure 4-112 Welcome message in CLI

General Settings

By using 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 IBM SAN Volume Controller Settings window, click **Settings** and then, select **GUI Preferences** → **General** (see Figure 4-113).

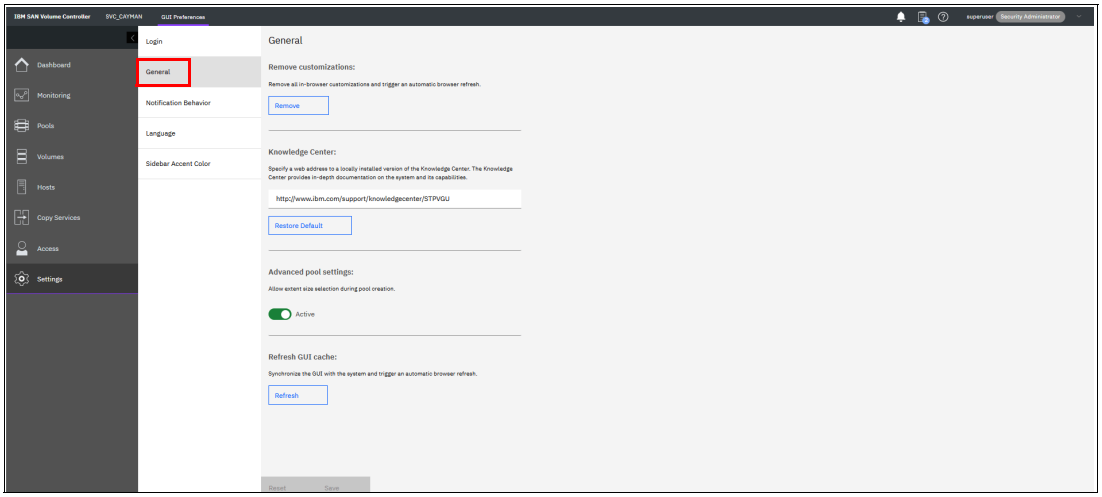


Figure 4-113 General GUI Preferences window

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 Spectrum 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 of 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.

Notification Behavior setting

Figure 4-114 shows that you can allow certain notifications to remain displayed until they are manually dismissed.

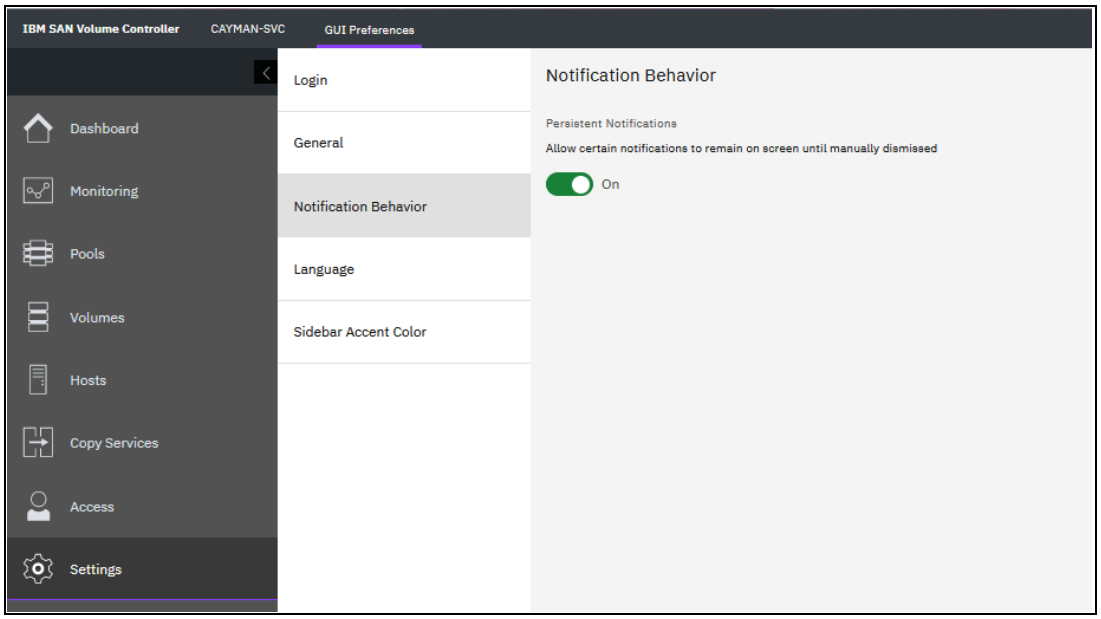


Figure 4-114 Notification Behavior window

Language

You can change the language of the GUI independently of the browser settings (see Figure 4-115).

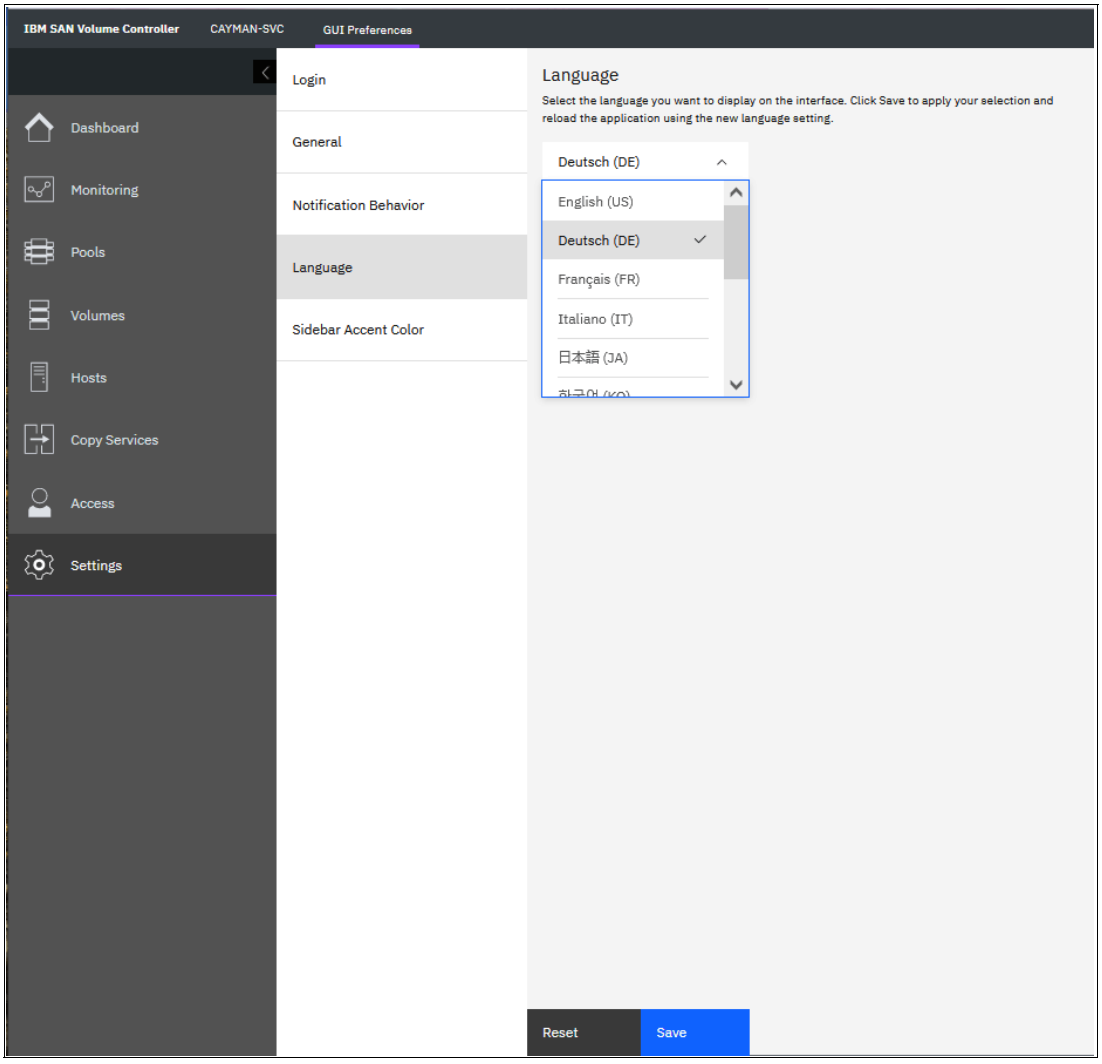


Figure 4-115 Language selection

Setting sidebar accent color

New with Version 8.4.2.0, you can change the color of the sidebars. Use the slider on the top to enable the function. This feature helps you to differentiate between session on Spectrum Virtualize Products (see Figure 4-116).

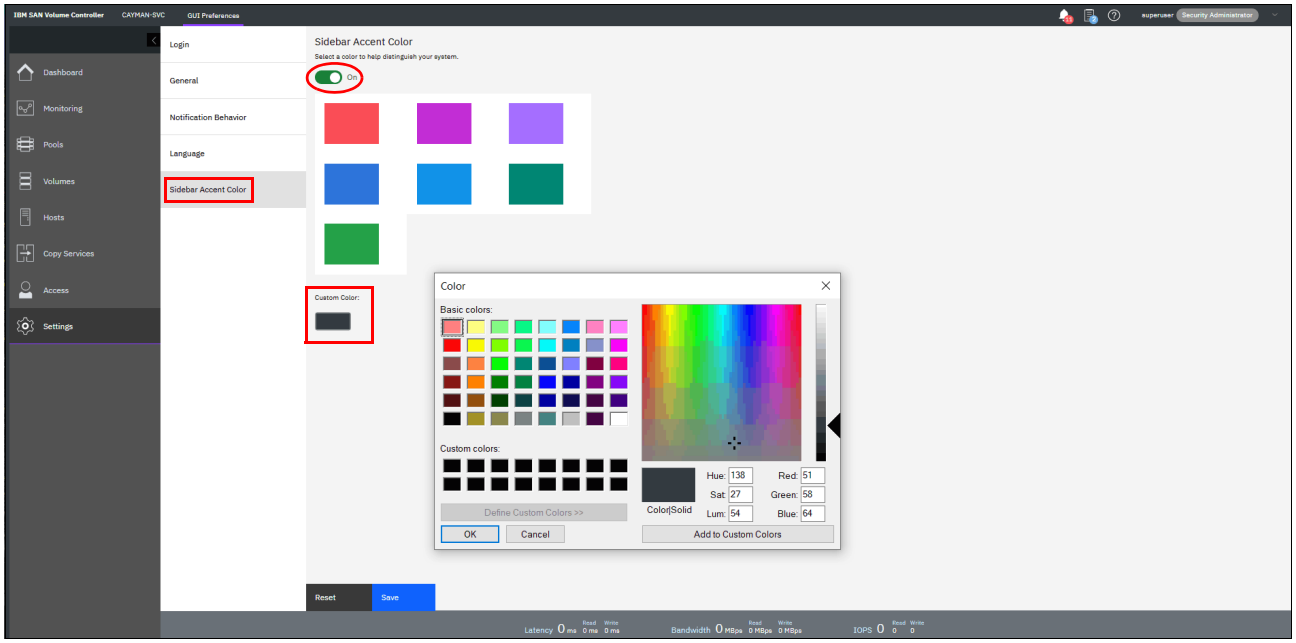


Figure 4-116 Sidebar Accent Color window

Under custom color you can choose your own color. Figure 4-117 shows a modified sidebar.

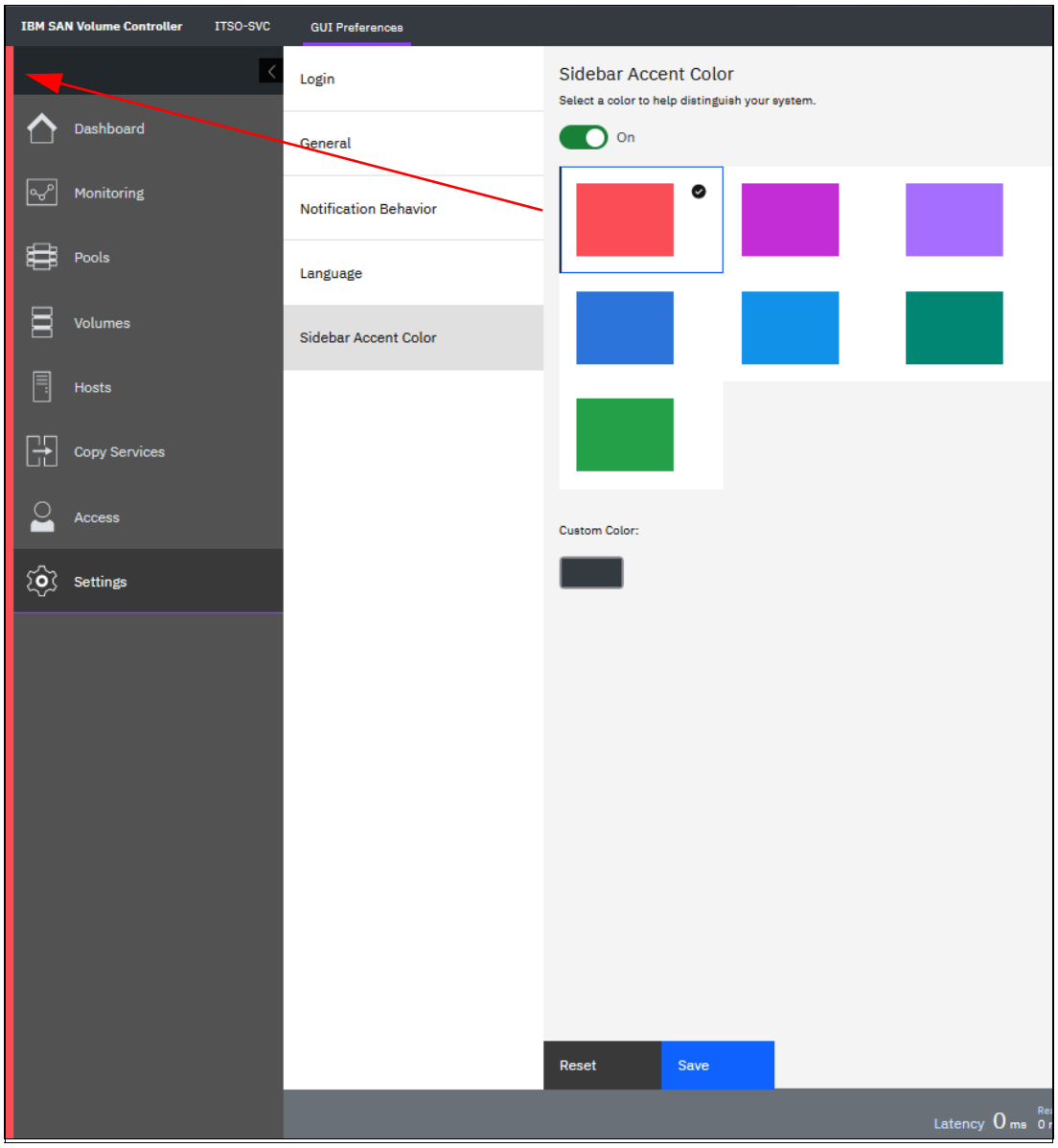


Figure 4-117 Customized sidebar

4.11 Other frequent tasks in the GUI

This section describes other options and tasks that are available in the GUI of your IBM SAN Volume Controller that are frequently used by administrators.

4.11.1 Renaming components

These sections provide guidance about how to rename your system and single nodes.

Renaming your IBM SAN Volume Controller system

All objects in the IBM SAN Volume Controller system have 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 both 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. Using 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 there are exceptions. 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 pane, complete the following steps:

1. Select **Monitoring** → **System Hardware**, and then, click **SystemActions** in the upper right corner of the pane, as shown in Figure 4-118.

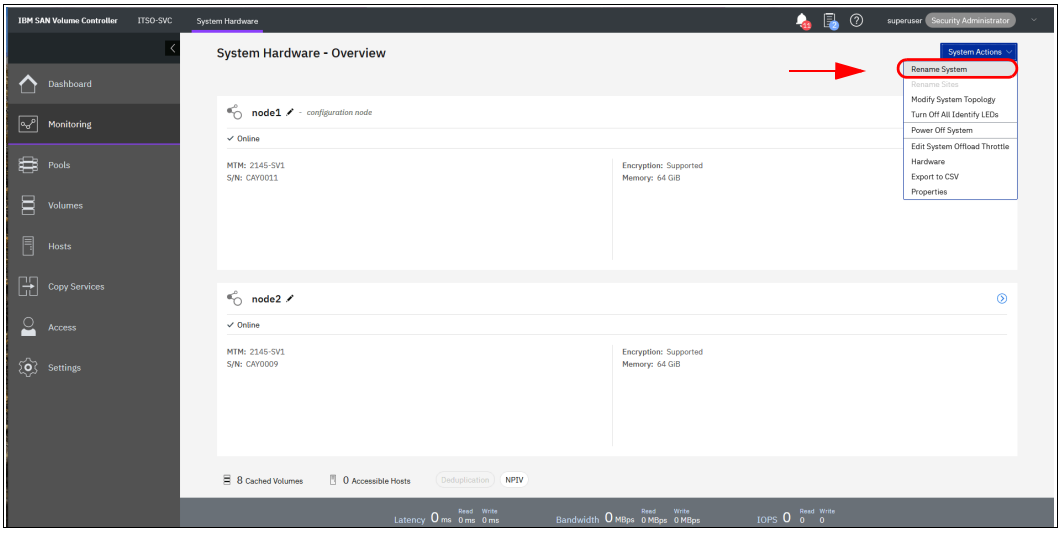


Figure 4-118 Actions on the System pane

2. The Rename System pane opens (see Figure 4-119). Specify a new name for the system and click **Rename**.

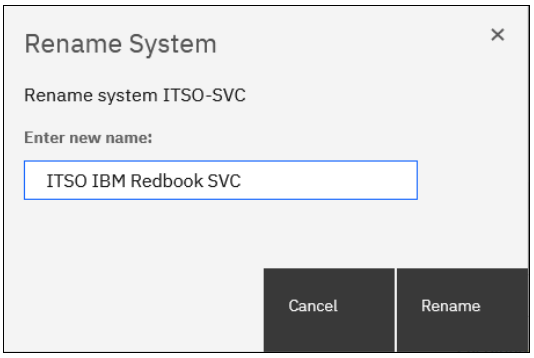


Figure 4-119 Renaming the system

Important: Consider the following points:

- For system names, the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character can be used. 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

To rename a node, complete the following steps:

1. Go to the System View window and select one of the nodes. The Node Details pane for this node opens, as shown in Figure 4-120. Select Node Actions, click **Rename**.

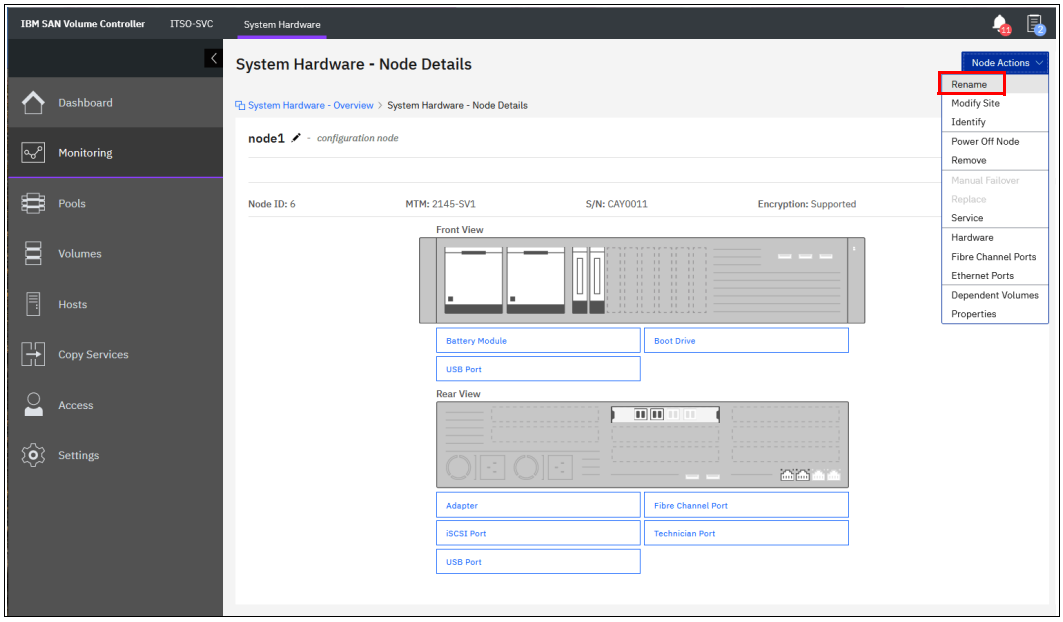


Figure 4-120 Renaming a node on the Node Details pane

2. Enter the new name of the node and click **Rename** (see Figure 4-121).

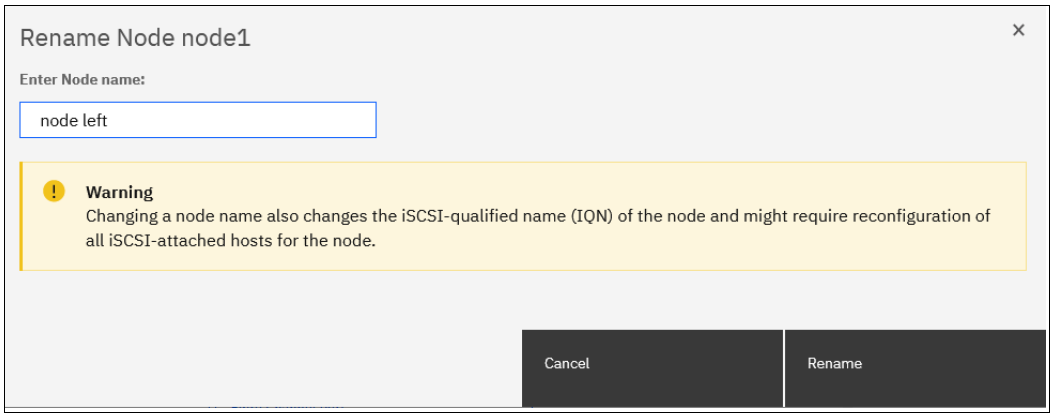


Figure 4-121 Entering the new node name

Warning: Changing the IBM SAN Volume Controller node name causes an automatic IQN update and requires the reconfiguration of all iSCSI-attached hosts.

Renaming sites

IBM SAN Volume Controller supports configuring site settings that describe the location of the nodes and storage systems that are deployed in a stretched system configuration. This site information configuration is part of the configuration process for enhanced systems. The site information makes it possible for IBM SAN Volume Controller to manage and reduce the amount of data that is transferred between the two sides of the system, which reduces the costs of maintaining the system.

Note: The renaming sites option is available only in a Stretched or HyperSwap topology.

Three site objects are automatically defined by IBM SAN Volume Controller and numbered 1, 2, and 3. IBM SAN Volume Controller creates the corresponding default names, `site1`, `site2`, and `site3` for each of the site objects. The `site1` and `site2` names are the two sites that make up the two halves of the enhanced system, and `site3` is the quorum disk. You can rename the sites to describe your data center locations.

To rename the sites, in the **System Hardware pane**, select **System Actions** in the upper-right corner, and then, select **Rename Sites**, as shown in Figure 4-122.

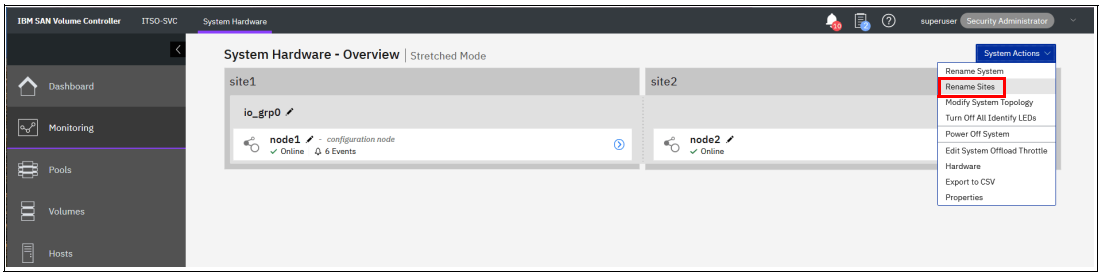


Figure 4-122 Rename Sites action

The Rename Sites window with the site information opens, as shown in Figure 4-123.

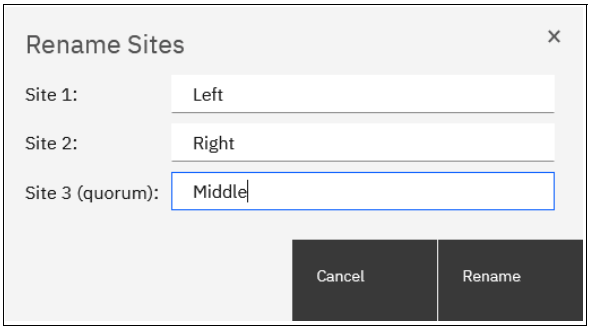


Figure 4-123 Rename Sites window

You can rename all or selected sites. Save your settings by clicking **Rename**.

4.11.2 Changing the system topology

You can create an enhanced resilient system configuration where each node on the system is physically on a different site. When used with mirroring technologies, such as volume mirroring or Copy Services, these configurations can be used to maintain access to data on the system in the event of power failures or site-wide outages.

The following options are available for an enhanced resiliency configuration:

- A Stretched topology is ideal for a high availability (HA) solution.
- HyperSwap fulfills also high availability requirements.

If you prepared you infrastructure to support enhanced topology, you can proceed with the topology change by completing the following steps:

1. From the System pane, select **SystemActions** → **Modify System Topology**, as shown in Figure 4-124.

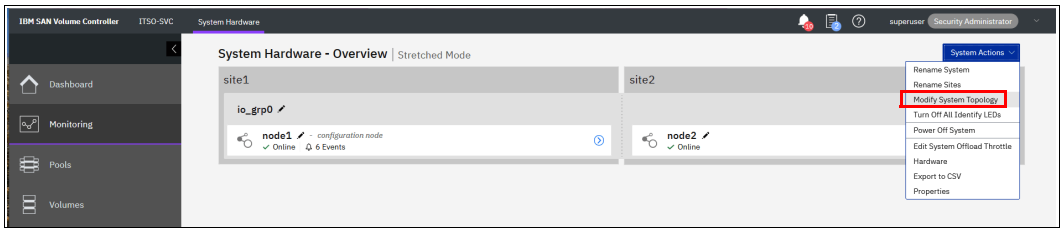


Figure 4-124 Modifying system topology

2. The wizard opens and informs you about options to change the topology to a Stretched Cluster or HyperSwap (see Figure 4-125).

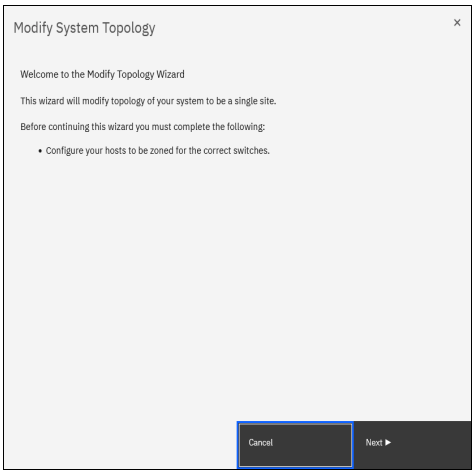


Figure 4-125 Topology wizard

3. The system requires you to define three sites: Primary, Secondary, and Quorum sites. Assign reasonable names to the sites for easy identification, as shown in Figure 4-126.

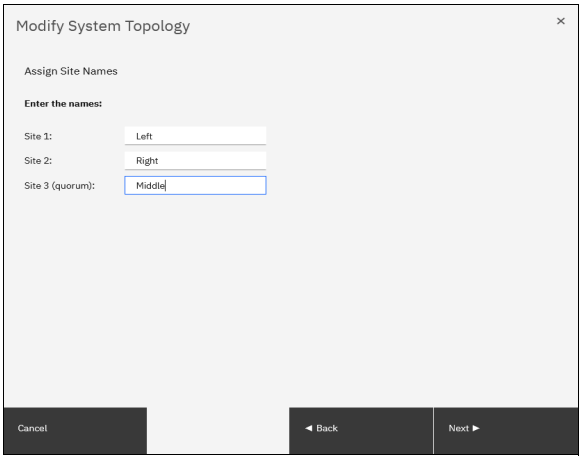


Figure 4-126 Assigning site names

4. Choose the topology. Although Stretched Cluster is optimal for DR solutions with asynchronous replication of primary volumes, HyperSwap is ideal for HA solutions with near-real-time replication. In our example, we decide to use a Stretched Cluster configuration (see Figure 4-127).

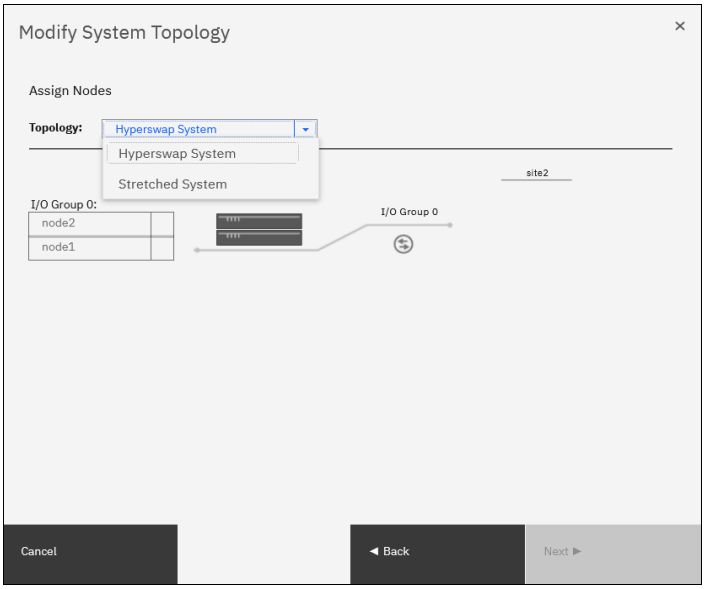


Figure 4-127 Changing the topology

5. Assign External Storage Systems to Sites.

   All external storage must have a site that is assigned before you configure a Stretched system. At least one storage system must be assigned to the **quorum site** (see Figure 4-128).

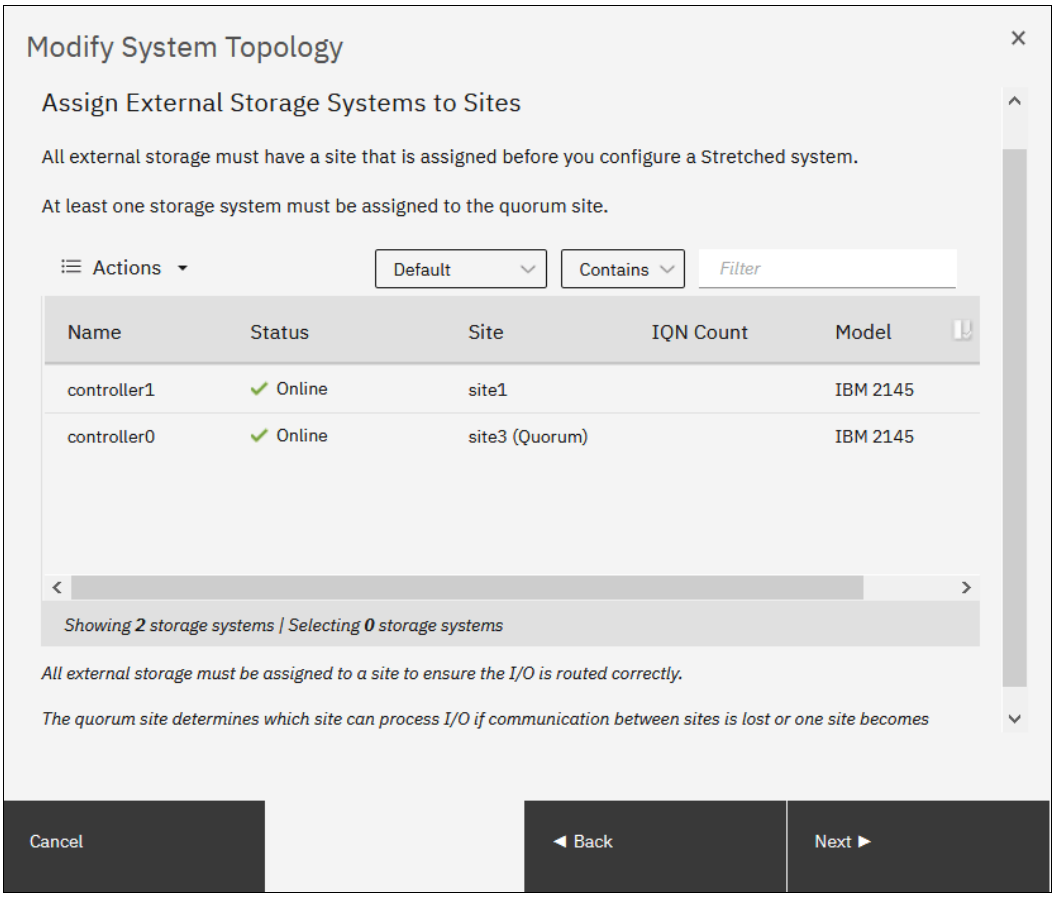


Figure 4-128 Assign external Storage Systems to sites

6. Assign hosts to primary sites. Right-click each host and modify sites for them individually (see Figure 4-129). Also, assign offline hosts to primary sites because they might be down for maintenance or any other reason.

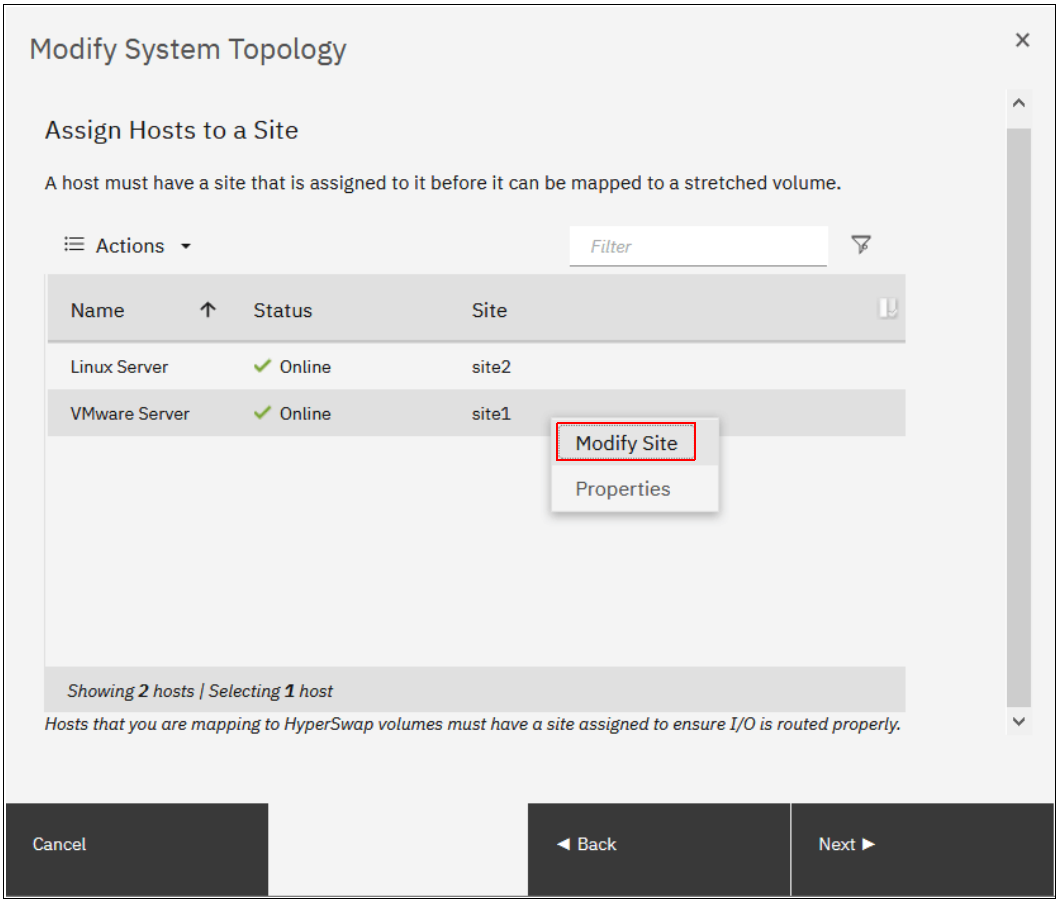


Figure 4-129 Assigning hosts to sites

7. Similarly, assign back-end storage to sites where the primary volumes are provisioned (that is, where the hosts are primarily located), as shown in Figure 4-130. At least one storage device must be assigned to the site that is planned for Quorum volumes.

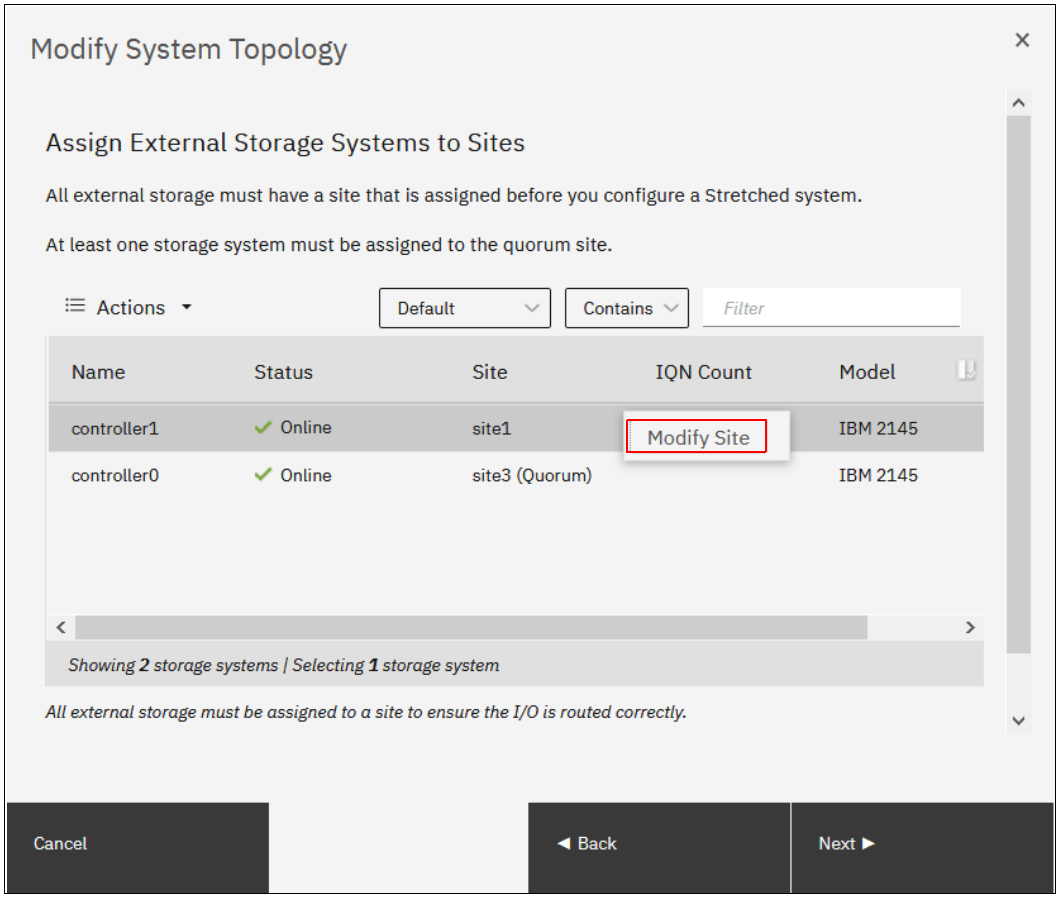


Figure 4-130 Assigning storage to sites

8. After this process is complete, the Summary window opens (see Figure 4-131). The system is ready to commit the changes. Click **Finish** to complete.

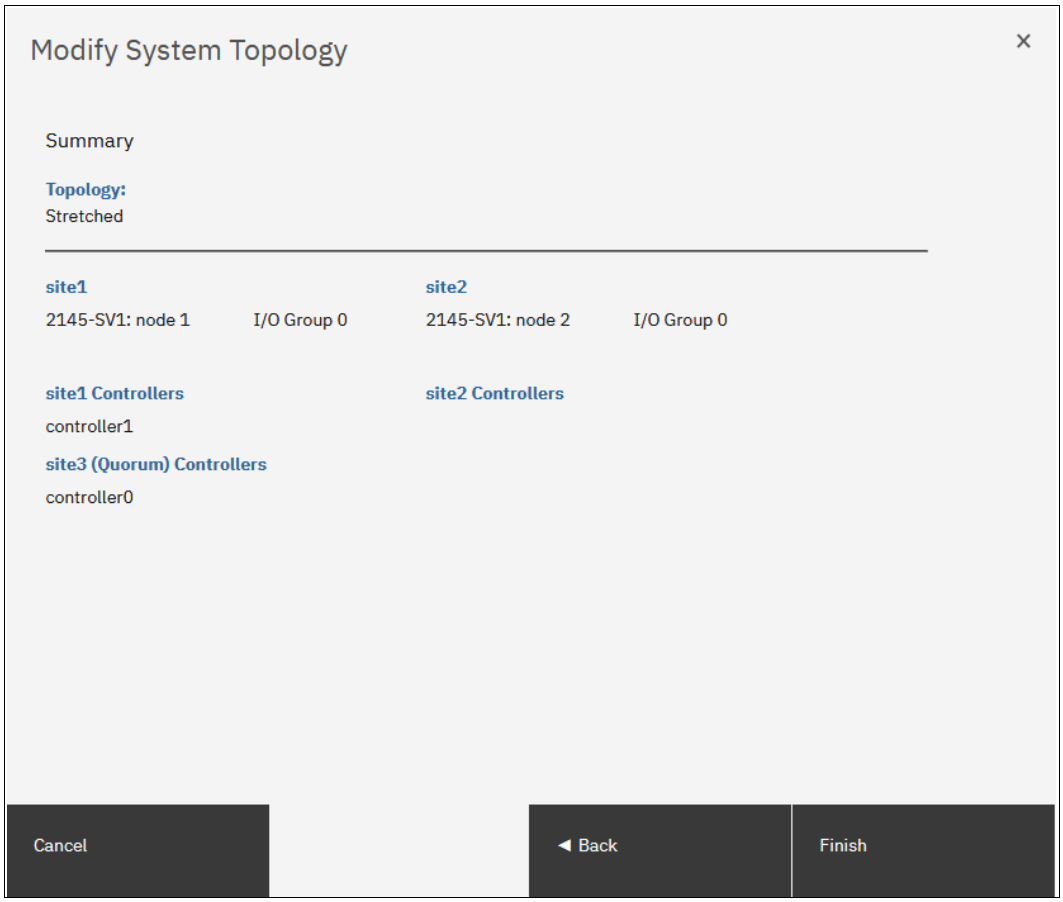


Figure 4-131 Summary of system topology changes

9. After the operation completes, select **System actions** → **Properties** to ensure that the topology now shows as Stretched, as shown in Figure 4-132.

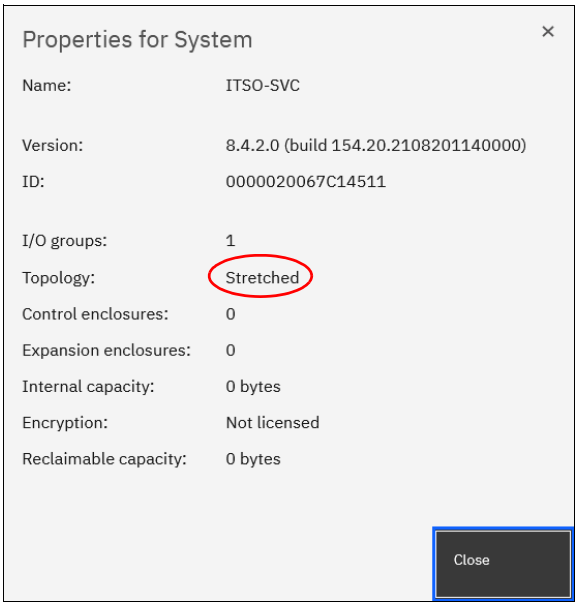


Figure 4-132 Checking system properties

As a validation step, verify that all hosts feature the correctly mapped and active online volumes and that no error appears in the event log.

Note: If you want to change your Stretched Cluster configuration to a HyperSwap configuration or vice versa, you must first configure the topology as a single-site topology. Use the wizard, which provides only the option to modify the topology as a single site before you can select a HyperSwap or Stretched Cluster configuration.

For more information about resilient solutions for your IBM SAN Volume Controller environment, see the following publications:

- *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

4.11.3 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 expectation (or you cannot connect).

Complete the following steps:

1. Log in to the Service Assistant, as shown in Figure 4-133.

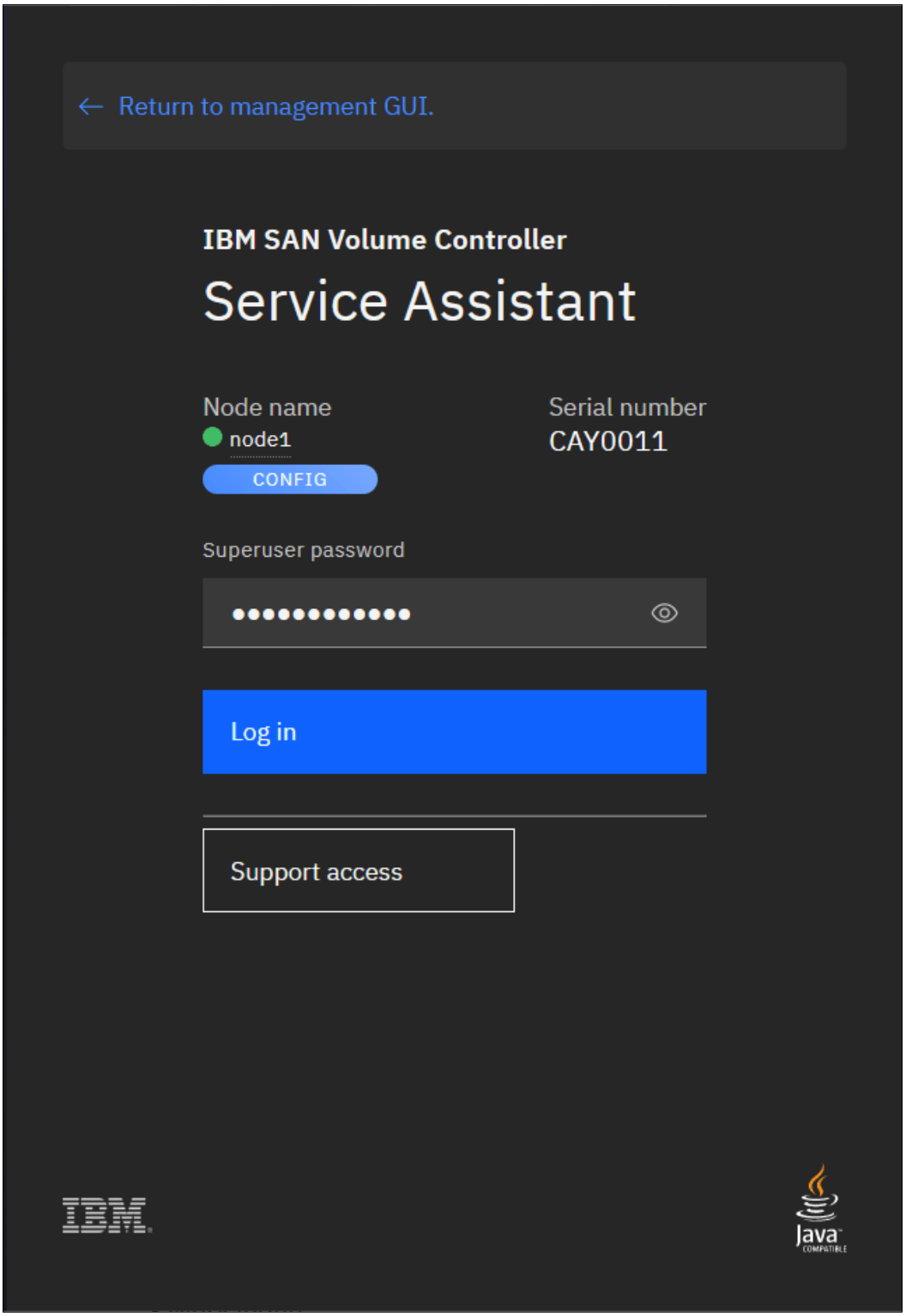


Figure 4-133 Logging into the Service Assistant

2. Identify the configuration node, as shown in Figure 4-134.

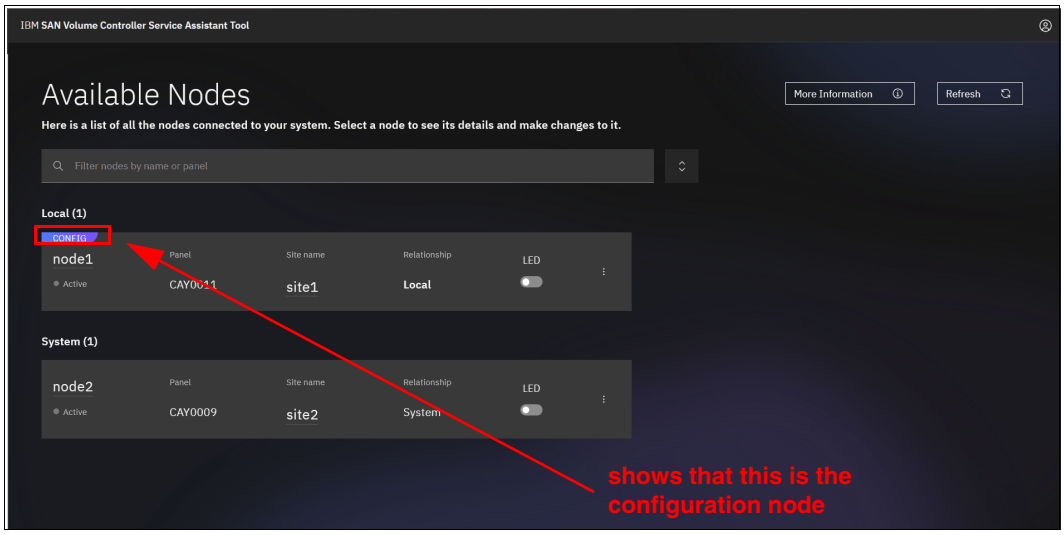


Figure 4-134 Identifying the configuration node on the service assistant

3. Select the configuration node. Node details are shown (see Figure 4-135). Select **Restart Service**.

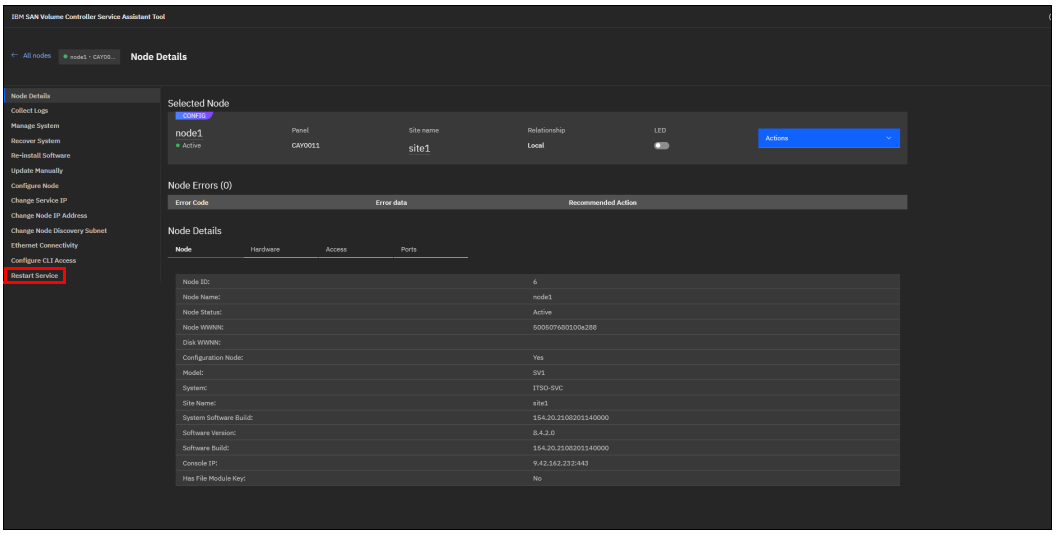


Figure 4-135 Node details

After the process completes, the Restart options appear (see Figure 4-136).

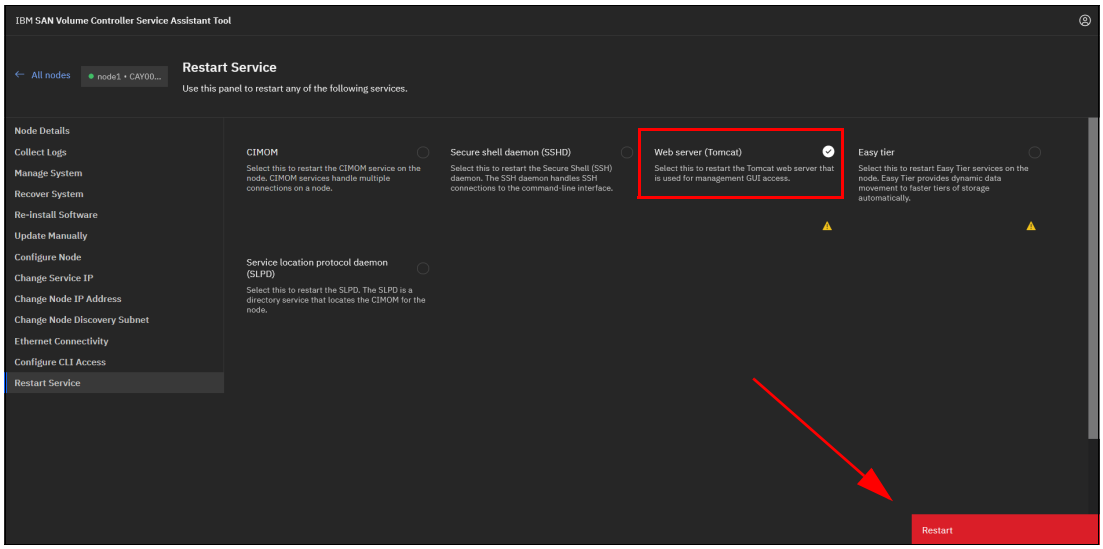


Figure 4-136 Restarting the Tomcat web server

4. Select **Web Server (Tomcat)** and then, click **Restart**. The web server that runs the GUI restarts. This task is a concurrent action, but the cluster GUI is unavailable while the server is restarting (the Service Assistant and CLI are not affected). After 5 minutes, check to see if GUI access was restored.

5

Using storage pools

This chapter describes how IBM SAN Volume Controller manages physical storage resources. All storage resources that are under system control are managed by using storage pools or managed disk (MDisk) groups (MDGs).

Storage pools aggregate internal and external capacity and provide the 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, either 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 272
- 5.2, “Working with external controllers and MDisks” on page 292
- 5.3, “Working with internal drives and arrays” on page 306

© Copyright IBM Corp. 2021. 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 either 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.

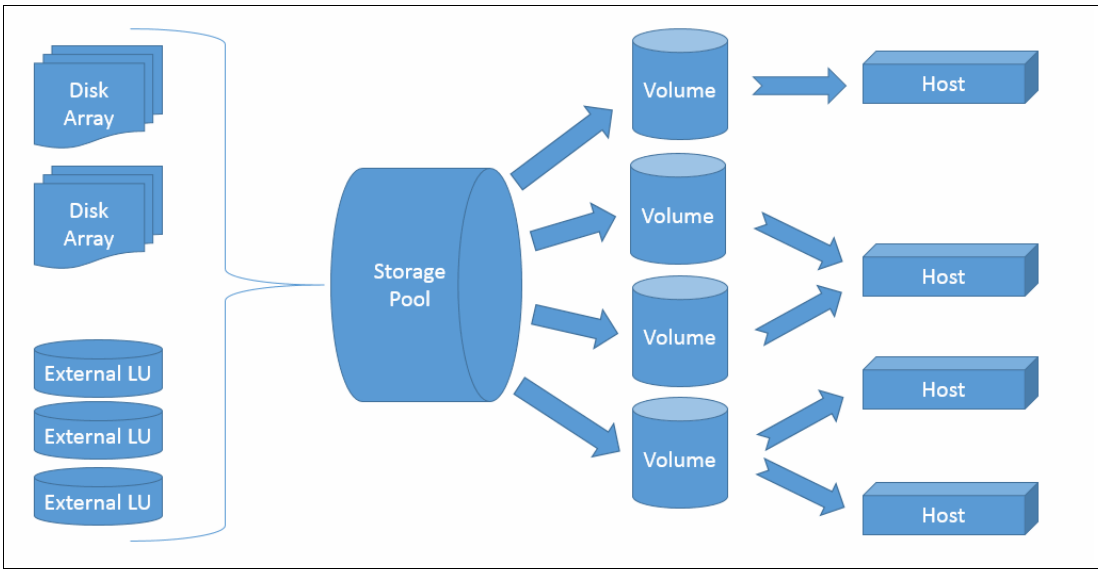


Figure 5-1 Relationship between MDisks, storage pools, and volumes

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 an existing 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 extents that are in use by volumes to other MDisks in the same pool to make sure that the data on the extents is preserved.

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 existing capacity that is assigned to a parent pool instead of being created directly from MDisks. When the 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.

DRPs use a set of techniques that can be used to reduce the amount of usable capacity that is required to store data, such as compression and deduplication. Data reduction can increase storage efficiency and performance, and reduce storage costs, especially for flash storage. 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. Child pools created from DRPs are quotaless and can use the entire parent pool capacity.

For more information about DRP planning and implementation, see Chapter 9, “Advanced features for storage efficiency” on page 559, 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. Create volumes in these pools and map them to hosts or host clusters.

You manage storage pools either in the Pools pane of the GUI or by using the CLI. To access the Pools pane, select **Pools** → **Pools**, as shown in Figure 5-2.

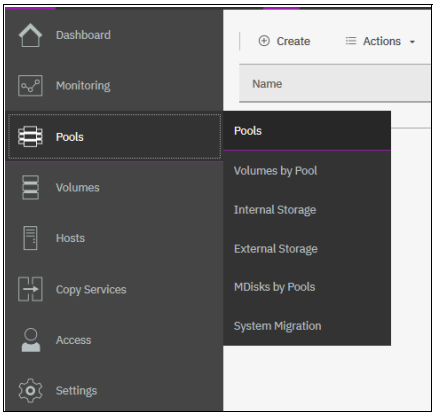


Figure 5-2 Accessing the Pools pane

The pane 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 free_capacity
0  Pool0 online 1           5           821.00GB 1024        771.00GB
1  Pool1 online 1           5           7.21TB   4096        6.14TB
```

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**, as shown in Figure 5-3.

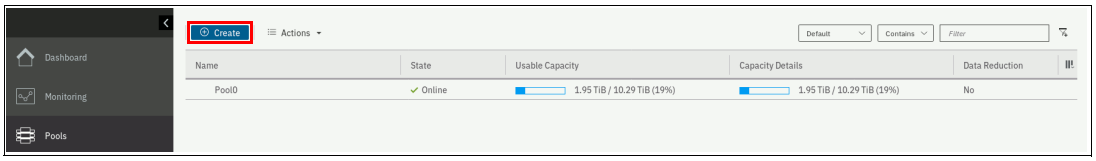


Figure 5-3 Option to create a storage pool in the Pools pane

Both alternatives open the dialog box that is shown in Figure 5-4.

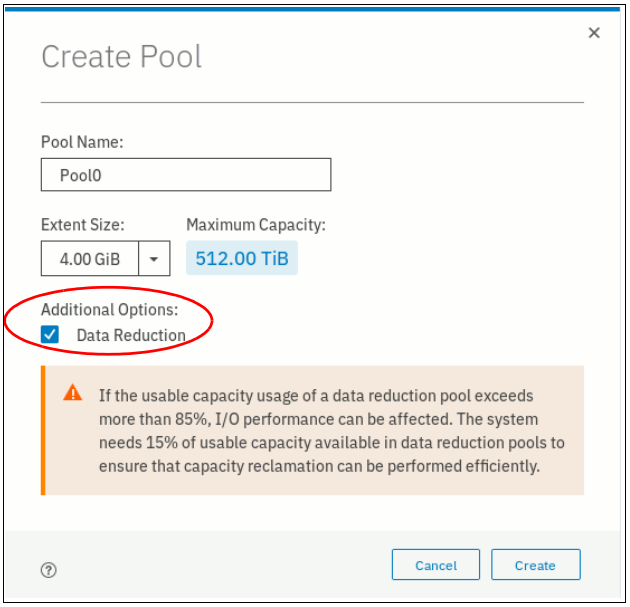


Figure 5-4 Create Pool dialog box

2. Select the **Data reduction** check box to create a DRP. Leaving it clear creates a standard storage pool.

Note: DRPs require careful planning and sizing. Limitations and performance characteristics of DRPs are different from standard pools.

A standard storage pool that is created by using the GUI has a default extent size of 1 GB. DRPs have a default extent size of 4 GB. 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 capacity after reduction in the pool itself.

For more information about the differences between standard pools and DRPs and for extent size planning, see Chapter 2, “Planning” on page 59.

Note: Do not create DRPs with small extent sizes. For more information, see this IBM Support alert. It is still valid for the actual Software Version.

For more information about the suitable system and version levels, see this IBM Support web page.

When creating a standard pool, you cannot change the extent size by using the GUI by default. If you want to specify a different extent size, enable this option by selecting **Settings** → **GUI Preferences** → **General** and checking **Advanced pool settings**, as shown in Figure 5-5. Click **Save**.

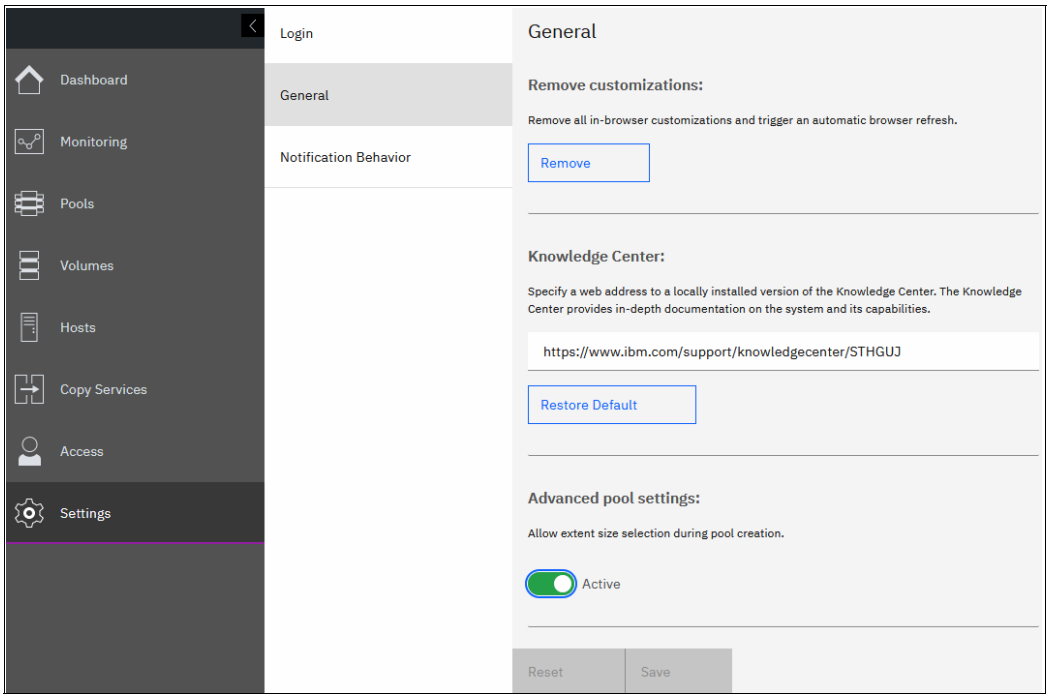


Figure 5-5 Advanced pool settings

When the **Advanced pool settings** option is enabled, you can also select an extent size for standard pools at creation time, as shown in Figure 5-6.

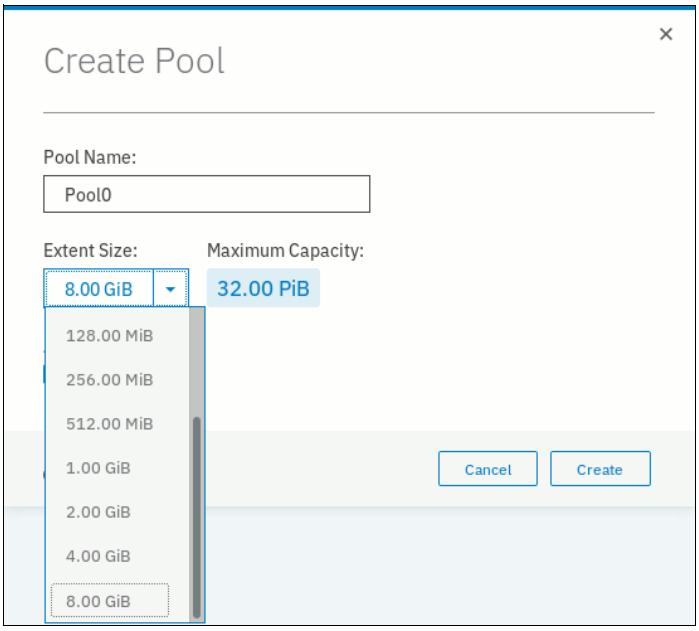


Figure 5-6 Creating a pool with Advanced settings selected

If an encryption license is installed and enabled, you can select whether the storage pool is encrypted, as shown in Figure 5-7. The encryption setting of a storage pool is selected at creation time and cannot be changed later. By default, if encryption is enabled, encryption is selected. For more information about encryption and encrypted storage pools, see Chapter 12, “Encryption” on page 793.

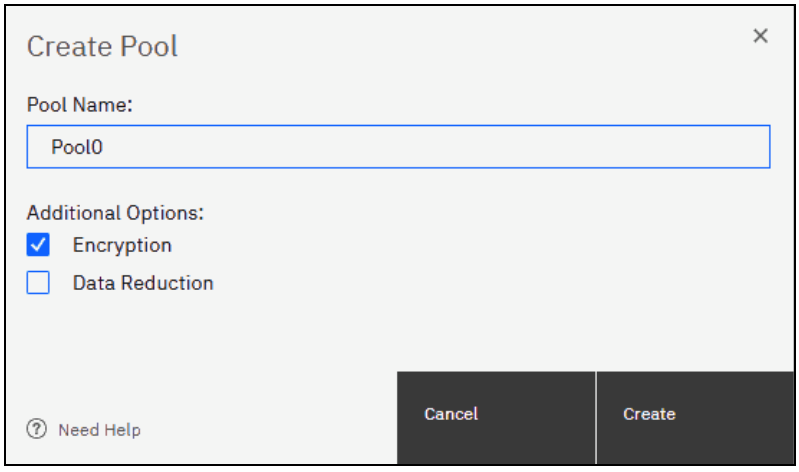


Figure 5-7 Creating a pool with encryption enabled

3. 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 with zero bytes, as shown in Figure 5-8.

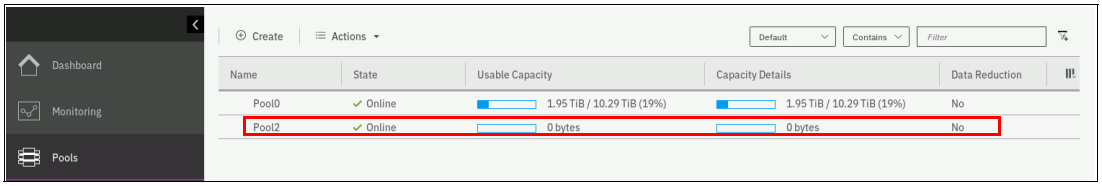


Figure 5-8 Newly created empty pool

To perform this task by using the CLI, run 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 Pool2 -datareduction yes -ext 8192
MDisk Group, id [2], successfully created
```

5.1.2 Managed disks in a storage pool

A storage pool is created as an empty container with no storage that is assigned to it. Storage is then added in the form of *MDisks*. An MDisk can be either 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 only possible to destroy an array by removing it from a pool and re-creating it within a new pool.

External MDisks can exist within a pool 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 or not.

MDisks are managed by using the MDisks by Pools pane. To access the MDisks by Pools pane, select **Pools** → **MDisks by Pools**, as shown in Figure 5-9.

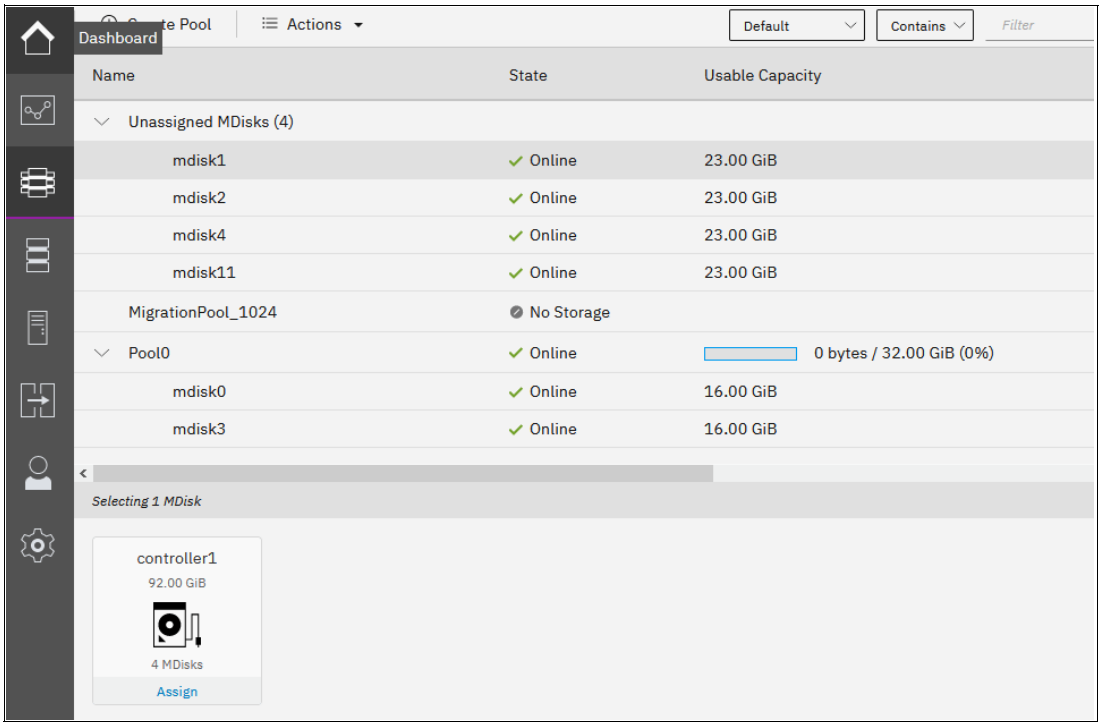


Figure 5-9 MDisks by Pools

The pane 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 5.3, “Working with internal drives and arrays” on page 306.

For more information about implementing a solution with external MDisks, see 5.2, “Working with external controllers and MDisks” on page 292.

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.

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 and click **Actions**. Alternatively, right-click the storage 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 285.

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 volumes that auto-expand (automatically use available extents from the pool), monitor the capacity usage and get 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 that 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 shown in Figure 5-10.

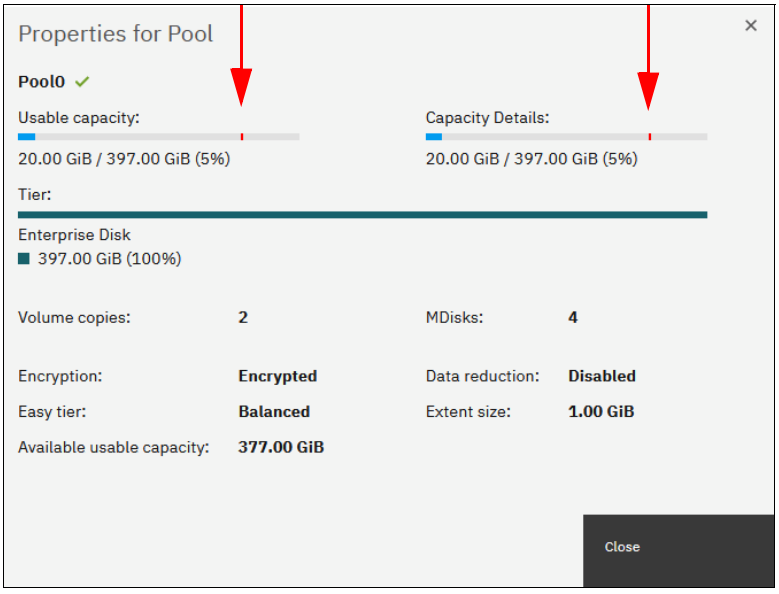


Figure 5-10 Pool properties with warning threshold

To do the task 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>
```

Add Storage to Pool window

This action starts the configuration wizard, which assigns storage to the pool, as shown in Figure 5-11.

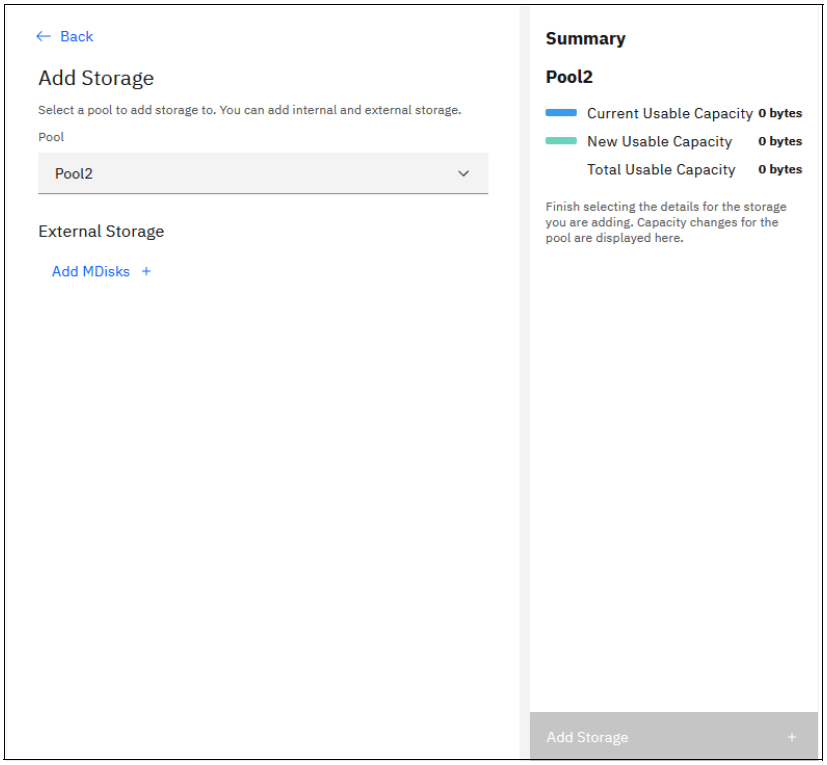


Figure 5-11 Add Storage to Pool dialog

If **Internal** Storage is chosen, the system guides you through array MDisk creation by using internal drives. If **External** Storage is selected, the system guides you through the selection of external storage MDisks. If no internal storage is attached, the **Internal** option is not shown.

Quorum Assignment

With this option, you can reset the quorum configuration to its default, which is to return quorum assignments to a set of MDisks that is chosen by system.

Edit Throttle for Pool window

Click this option 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 either 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 overwhelming the back-end storage. Only parent pools support throttles because only parent pools contain MDisks from internal or external back-end storage. For volumes in child pools, the throttle of the parent pool is applied.

You can define a throttle for input/output operations per second (IOPS), bandwidth, or both, as shown in Figure 5-12:

- `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).

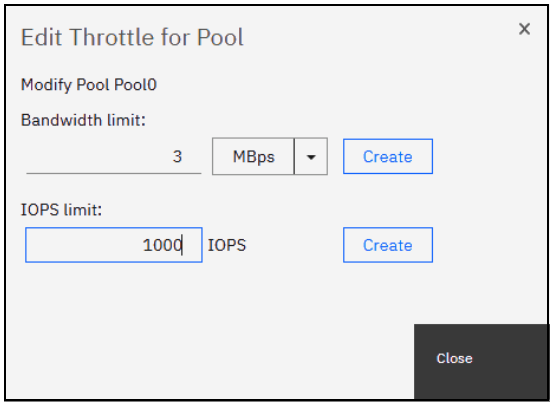


Figure 5-12 Edit throttle for Pool window

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 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 a throttle exists for the storage pool, the dialog box that is shown in Figure 5-12 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 pane. Right-click a pool and select **View all Throttles** to display the list of the pool's throttles. If you want to view the throttle of other elements (like **Volumes** or **Hosts,** for example), you can select **All Throttles** in the drop-down list, as shown in Figure 5-13.

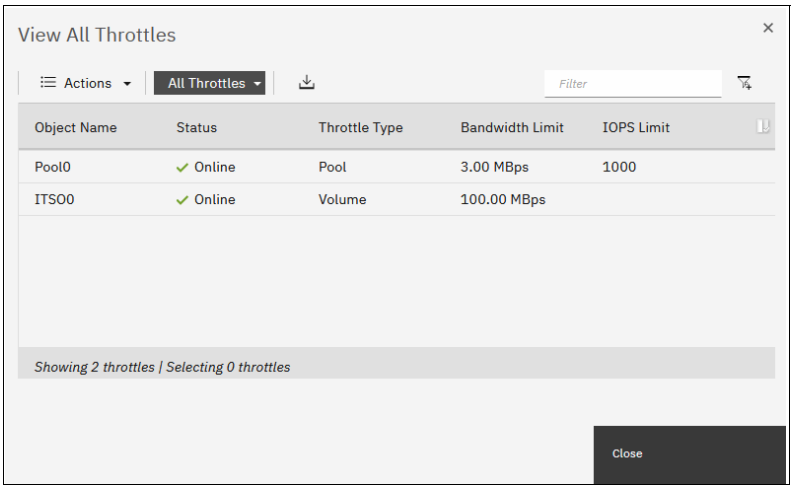


Figure 5-13 Viewing all throttles

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

To browse a list of MDisks that are part of the storage pool, click **View Resources**, which opens the window that is shown in Figure 5-14.

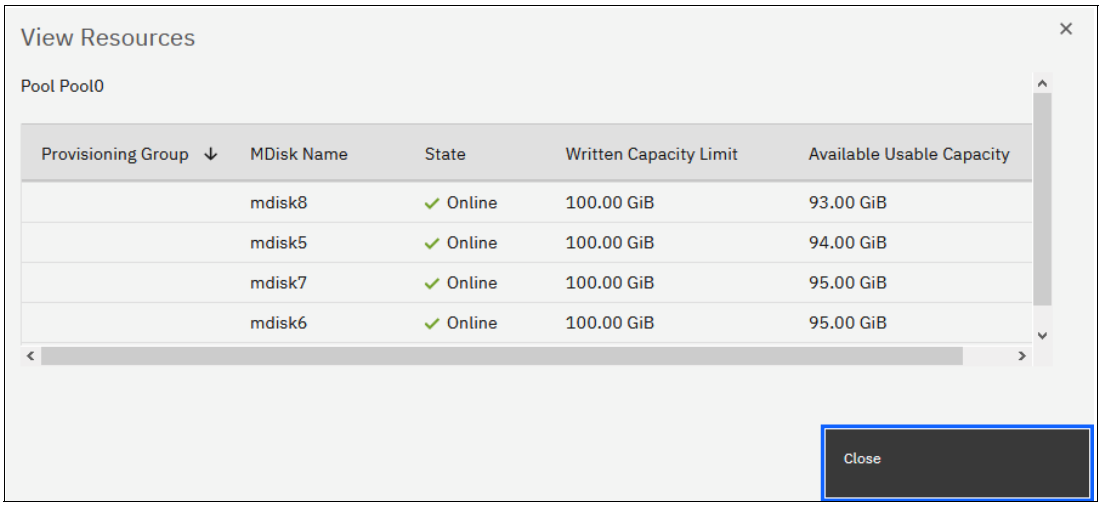


Figure 5-14 List of resources in the storage pool

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 Chapter 9.1.3 Page 458 “Monitoring Easy Tier activity”.

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 additional 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, “Creating and provisioning volumes” on page 331.

After you delete a pool, the following actions occur:

- All the external MDisks in the pool return to a mode of `Unmanaged`.
- All the array mode MDisks in the pool are deleted and all member drives return to a status of `Candidate`.

To delete a storage pool by using the CLI, run the **rmmdiskgrp** command.

Important: Be careful when you run the **rmmdiskgrp** command with the **-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. By hovering your cursor over the elements of the window and clicking **[?]**, you see a short description of each property, as shown in Figure 5-15.

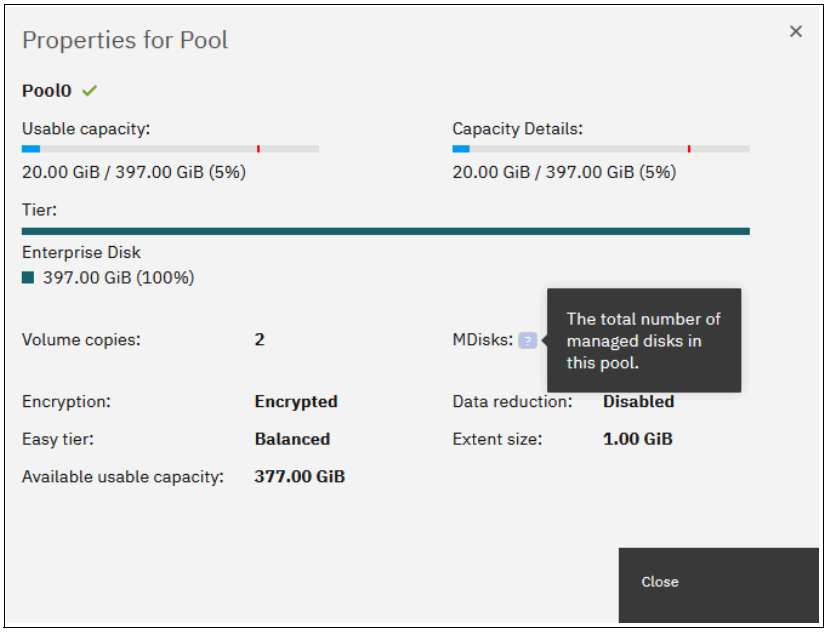


Figure 5-15 Pool properties and details

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*.

A storage pool type is parent, child_thick or child_quotaless.

Unlike a parent pool, a child pool does not contain MDisks. Its capacity is provided by the parent pool. A child pool from standard parent pool is type child_thick.

The capacity of a child pool from a standard 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. Capacity that is assigned to a child pool type child_thick is taken away from the capacity of the parent pool.

Creating a child pool within another child pool is not possible either.

Child pools type child_thick are useful when the capacity that is allocated to a specific set of volumes must be controlled. For example, child pools type child_sick can be used with VMware vSphere Virtual Volumes (VVOLs). Storage administrators can restrict access of VMware administrators to only a part of the storage pool and prevent volumes creation from affecting the rest of the parent storage pool.

Ownership groups can be used to restrict access to storage resources to a specific set of users, as described in Chapter 11, “Ownership groups” on page 781.

Child pools type child_thick 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 ever runs out of free extents, you can use the ones from the child pool.

On systems with encryption enabled, child pools type child_thick can be created to migrate volumes in a non-encrypted pool to encrypted child pools. When you create a child pool 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 pools type quotaless can be created only if the parent pool is encrypted. The data reduction child pool inherits an encryption key from the parent pool.

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 then, select **Create Child Pool**, as shown in Figure 5-16.

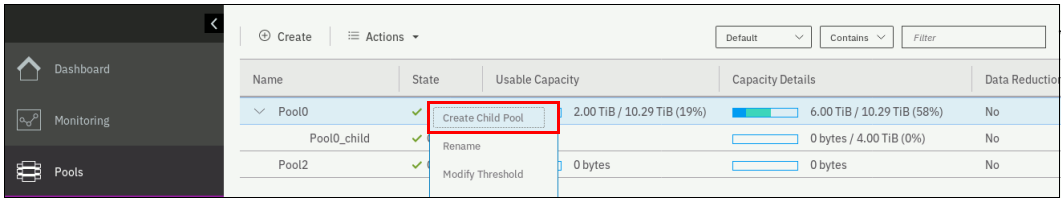


Figure 5-16 Creating a child pool

2. When the dialog box opens, enter the name of the child pool and click **Create.** Figure 5-17 shows dialog for pool type `child_quotaless`.

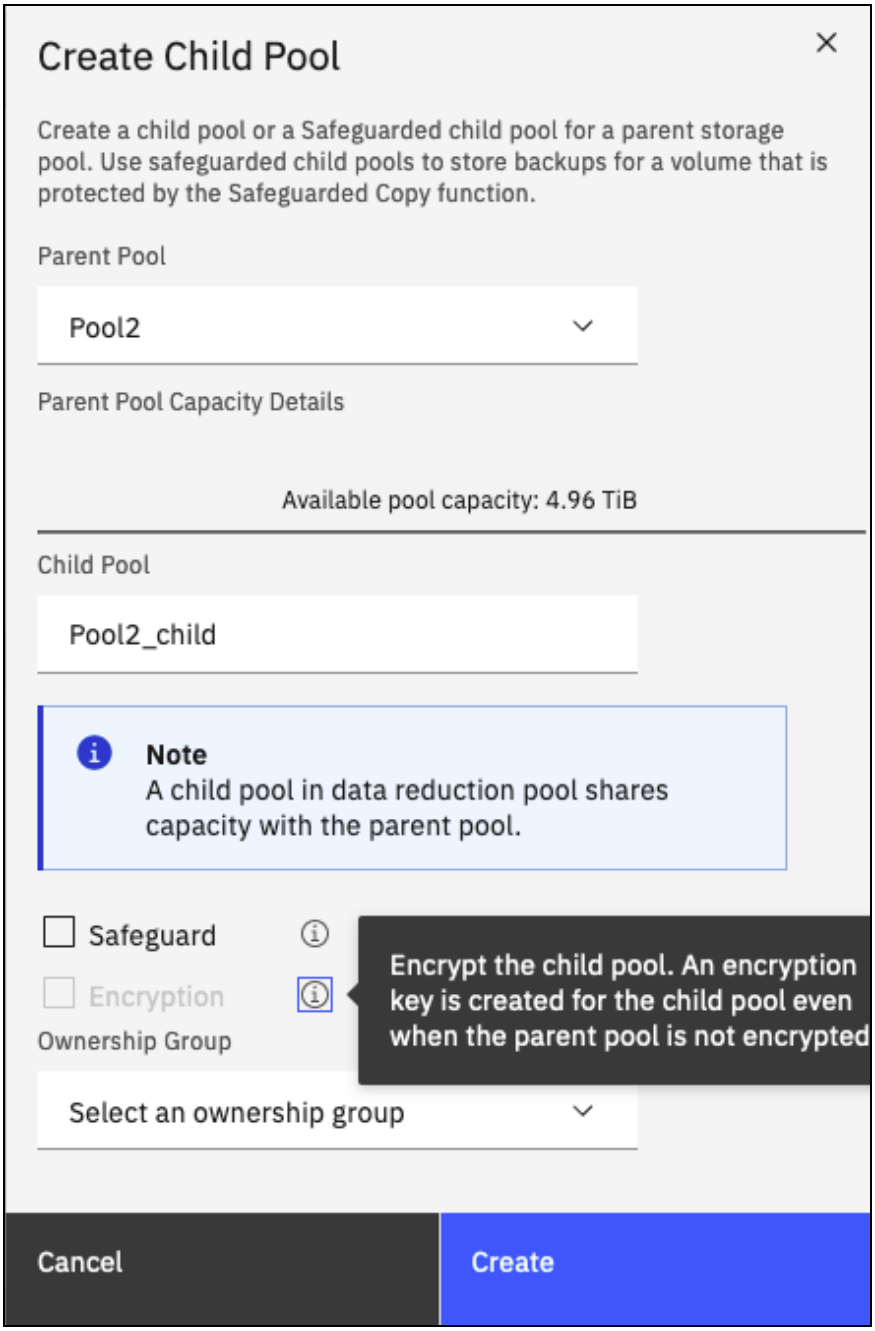


Figure 5-17 Creating a child pool type child_quotaless

3. Figure 5-18 shows dialog box for child pool from standard pool. Enter the pool capacity for pool type `Child_thick`.

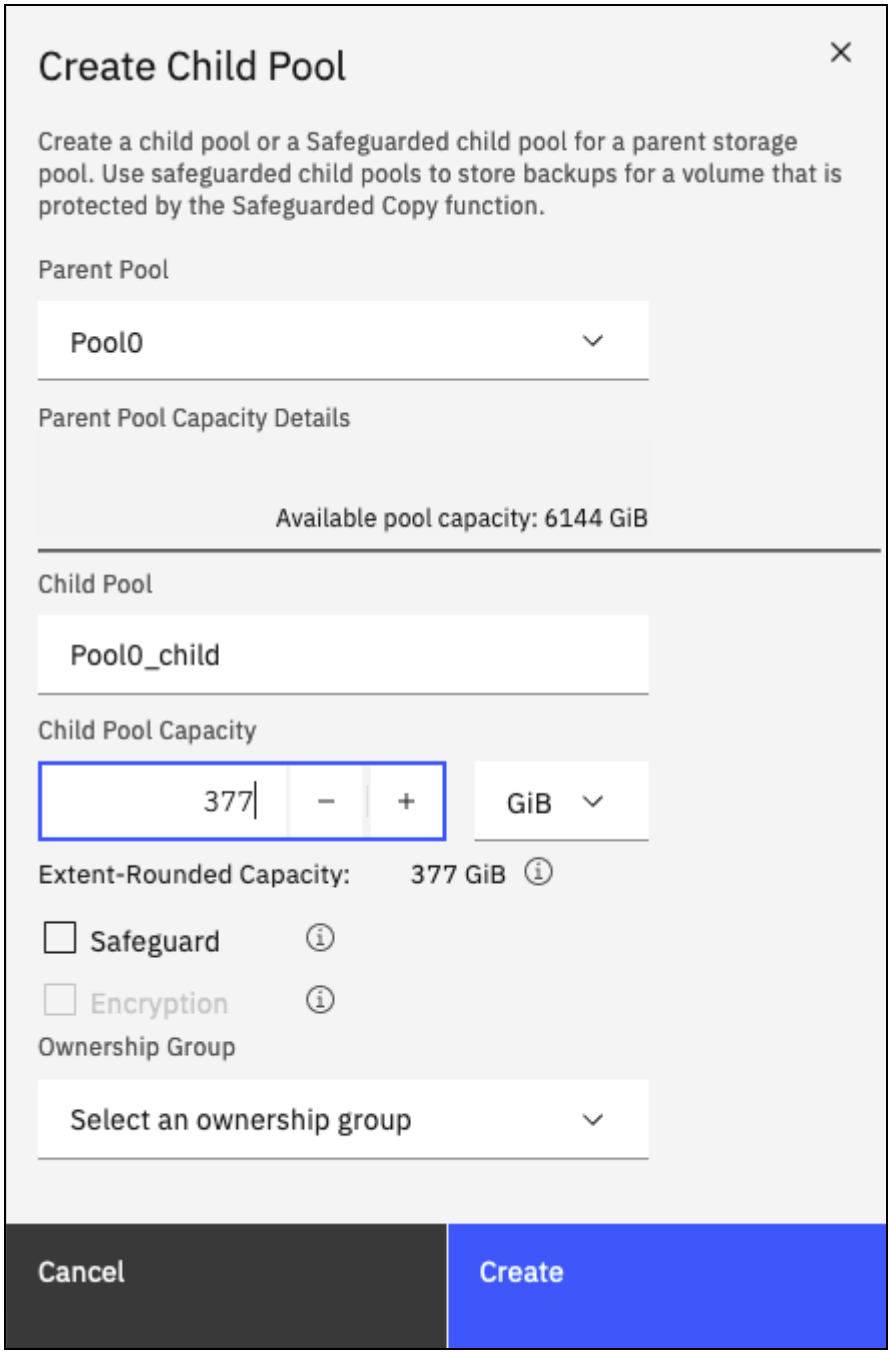


Figure 5-18 Creating a child pool type Child_thick

As you can see in Figure 5-18, you can now have an encrypted child pool, even if the parent pool is not encrypted.

You can enable the child pool to store backups of volumes protected by the Safeguarded Copy function by selecting the **Safeguard** option. For more information about Safeguarded copy function, see *Implementation Guide for SpecV/FlashSystem Safeguarded Copy*, REDP-5654.

After the child pool is created, it is listed in the Pools pane under its parent pool. Toggle the sign to the left of the storage pool icon to either show or hide the child pools, as shown in Figure 5-19. The capacity that is assigned to the child pools is not usable in the parent pool, as shown by the gray area on the capacity details bar of the parent pool.

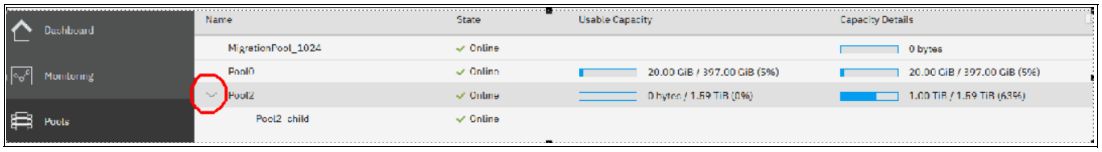


Figure 5-19 Listing parent and child pools

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
```

Actions for child storage pools

You can rename, resize (only child pools type child_thick), and delete a child pool. Also, it is possible to modify its warning threshold and assign it to an ownership group. To select an action, for example, **Resize**, complete the following steps:

1. Right-click the child storage pool, as shown in Figure 5-20. Alternatively, select the storage pool and then, click **Actions**.

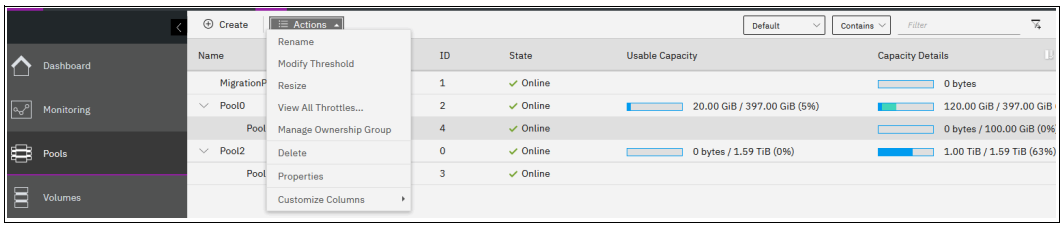


Figure 5-20 Actions for child storage pools

2. Select **Resize** to increase or decrease the capacity of the child storage pool type child_thick, as shown in Figure 5-21. Enter the new pool capacity and click **Resize**.

> **Note:** You cannot shrink a child pool below its real capacity. Thus, the new size of a child pool must be larger than the capacity that is used by its volumes.

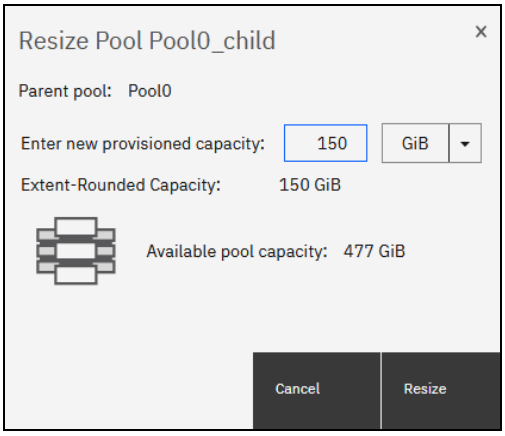


Figure 5-21 Resizing a child pool

When the child pool is shrunk, 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 `Pool0_child0` to `Pool0_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 Pool0_child_new -size 61440
Pool0_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 5-22.

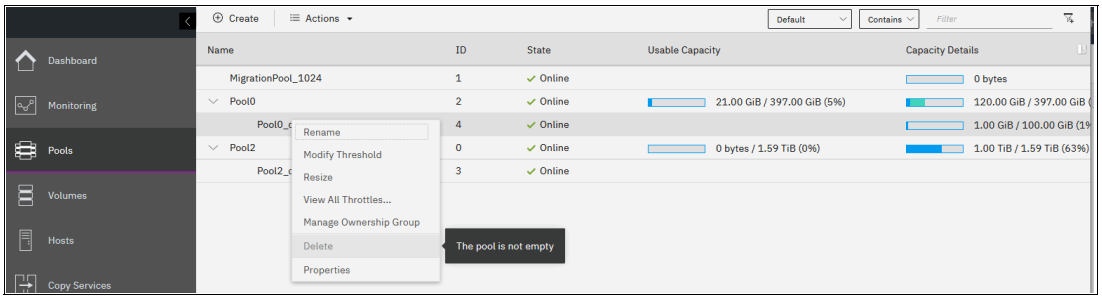


Figure 5-22 Deleting a child pool

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 5-23. All volumes that are created in the child pool inherit the ownership group of the child pool. For more information, see Chapter 11, “Ownership groups” on page 781.

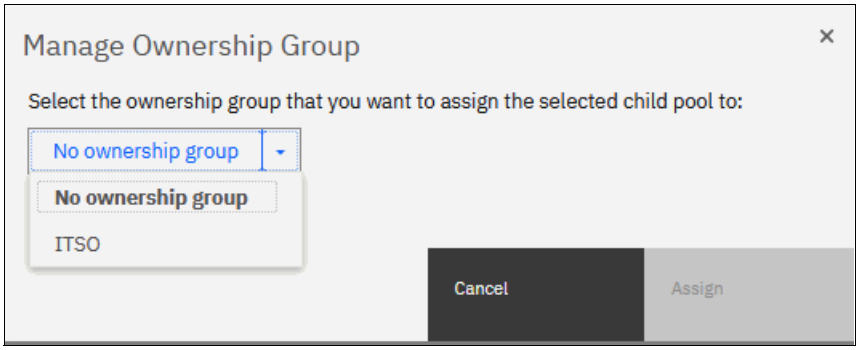


Figure 5-23 Managing the ownership group of a child pool

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, “Creating and provisioning volumes” on page 331.

The system supports migration of 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 *Implementation Guide for SpecV/FlashSystem Safeguarded Copy*, REDP-5654.

During a volume migration within a parent pool (between a child and its parent or between children with the same parent), there is no data movement, but there are extent reassignments.

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 5-24, the volume `vdisk0` was created in child pool `Pool0_child_new`. The child pools appear exactly like the parent pools in the Volumes by Pool pane.

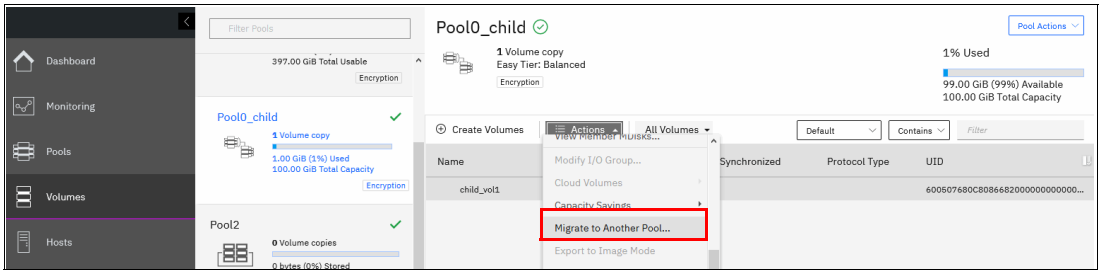


Figure 5-24 Actions menu in Volumes by Pool

For more information about CLI commands for migrating volumes to and from child pools, see Chapter 6, “Creating and provisioning volumes” on page 331.

5.1.5 Encrypted storage pools

The system supports two types of encryption: hardware encryption and software encryption.

Hardware encryption is implemented at an array level, and software encryption is implemented at a storage pool level. For more information about encryption and encrypted storage pools, see Chapter 12, “Encryption” on page 793.

5.2 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 IBM SAN Volume Controller 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 IBM SAN Volume Controller.

This concept is called *External Virtualization*, which makes your storage environment more flexible, cost-effective, and easy to manage. External Virtualization follows 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, “Planning” on page 59.

5.2.1 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 either the *storage* layer (default) or the *replication* layer.

With these default settings, IBM SAN Volume Controller can virtualize other Storwize or FlashSystem family systems. However, if the Storwize or FlashSystem family system was moved to the *replication layer*, it must be configured back to *storage layer* for the IBM SAN Volume Controller to be able to use it as external storage.

The IBM SAN Volume Controller system layer cannot be changed. The changes must be made on the external Storwize or 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 a Storwize or IBM FlashSystem family system.
- No system partnerships can be defined.
- No Storwize or IBM FlashSystem family system can be visible on the storage area network (SAN) fabric.

For more information about layers and how to change them, see IBM Documentation, Alternatively, you can expand **Product overview** → **Technical overview** → **System layers**.

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 2.6, “Fibre Channel SAN configuration planning” on page 65.

Note: Find the supported, permanently directly attached storage systems at this IBM Support web page.

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 Storwize or IBM FlashSystem family.

After the problem is corrected, rescan the FC network immediately by selecting **Pools** → **External Storage**, and then selecting **Actions** → **Discover Storage**, as shown in Figure 5-25.



Figure 5-25 Discover Storage menu

This action runs the **`detectmdisk`** command. It returns no output. Although it might appear that the command completed, some extra time might be required for it to run. The command is asynchronous and returns a prompt while the command continues to run in the background.

Attachment by using iSCSI

You must manually configure iSCSI connections between the IBM SAN Volume Controller and the external storage controller. Until you do this task, the controller is not listed in the External Storage pane. For more information about how to attach back-end storage controllers to the system, see Chapter 2, “Planning” on page 59.

To start virtualizing an iSCSI back-end controller, you must follow the documentation in IBM Documentation to perform configuration steps that are specific to your back-end storage controller. You can see find the steps by selecting **Configuring** → **Configuring and servicing storage systems** → **External storage system configuration with iSCSI connections**.

For more information about configuring IBM SAN Volume Controller 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 both FC and iSCSI storage controllers through the External Storage pane. To access the External Storage pane, select **Pools** → **External Storage**, as shown in Figure 5-26.

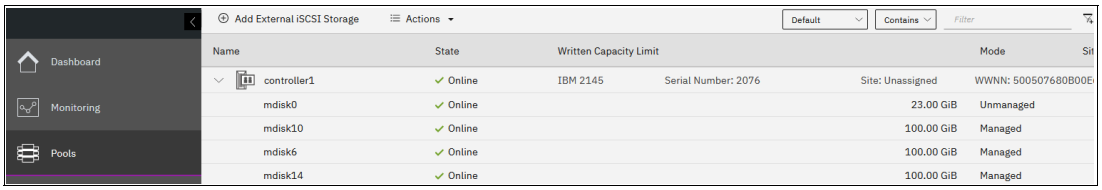


Figure 5-26 External Storage pane

Note: A controller that is connected through FC is detected automatically by the system. The cabling, the 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 pane 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, it is not possible for the system to 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 5-11.

Example 5-11 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
```

5.2.2 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 5-27. Alternatively, select the controller and click **Actions**.

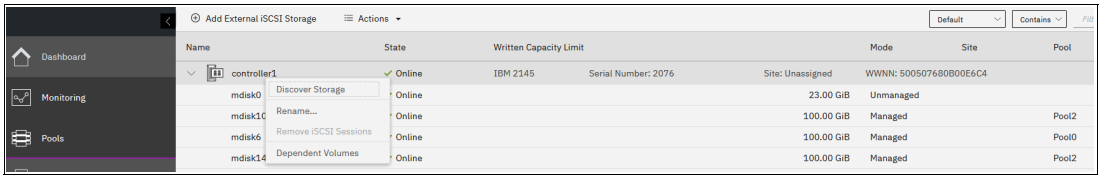


Figure 5-27 Actions for external storage

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 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.

This action runs the **`detectmdisk`** command.

Rename

To modify the name of an external controller to simplify administration tasks, click **Rename**. The naming rules are the same as for storage pools. For more information, see in 5.1.1, “Creating storage pools” on page 274.

To rename a storage controller by using the CLI, run the **`chcontroller`** command. A use case is shown in Figure 5-12 on page 296.

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 5-28.

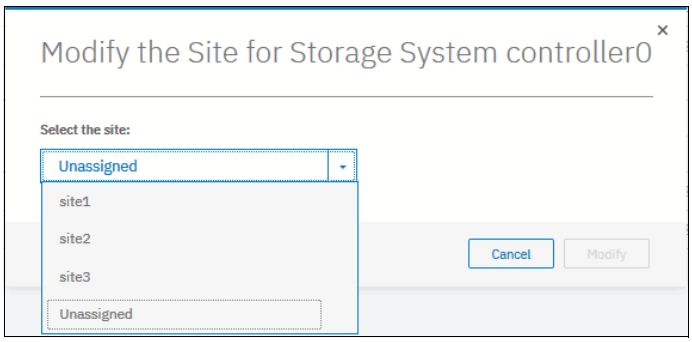


Figure 5-28 Modifying the site of an external controller

To change the controller site assignment by using the CLI, run the **`chcontroller`** command. Example 5-12 shows that `controller0` was renamed and reassigned to a different site.

Example 5-12 Changing a controller's name and site

```
IBM_2145:ITSO-SV1:superuser>chcontroller -name site3_controller -site site3
controller0
IBM_2145:ITSO-SV1:superuser>
```

5.2.3 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 existing data on an MDisk through the system to an attached host
  – Importing existing data on an MDisk into the system
  – Exporting data 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 pane, which is accessed by selecting **Pools** → **External Storage**, as shown in Figure 5-29.

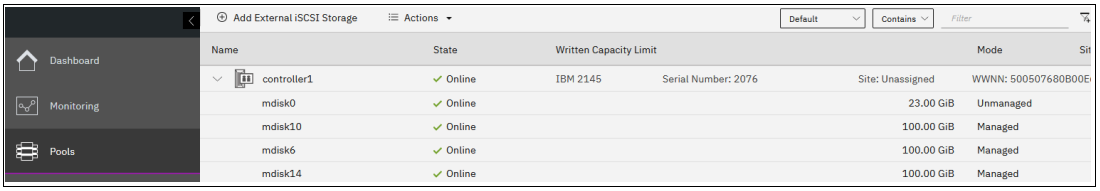


Figure 5-29 External Storage pane

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.

Assigning MDisks to pools

You can add *Unmanaged* MDisks to an existing 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 274.

Figure 5-30 shows how to add selected MDisk to an existing storage pool. Click **Assign** under the **Actions** menu or right-click the MDisk and select **Assign**.

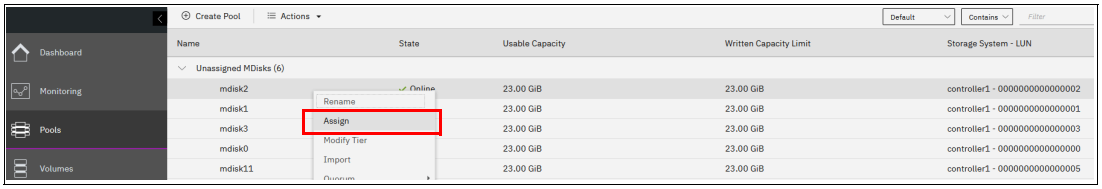


Figure 5-30 Assigning an unmanaged MDisk

After you click **Assign**, a dialog box opens, as shown in Figure 5-31. Select the target pool, MDisk storage tier, and external encryption setting.

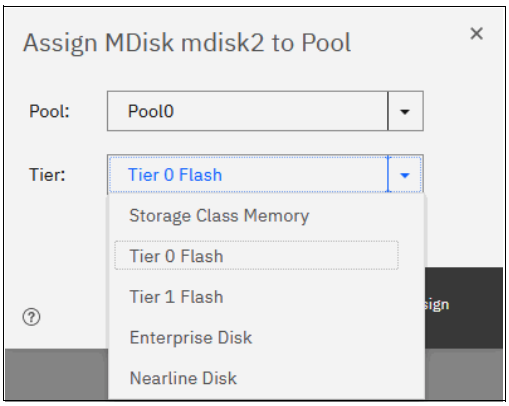


Figure 5-31 Assign MDisk dialog box

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 559.

The storage tier setting also can 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 IBM SAN Volume Controller encryption, see Chapter 5, “Using storage pools” on page 271.

After the task completes, click **Close.**

Note: If the external storage LUs that are presented to the IBM SAN Volume Controller 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 527.

To see the external MDisks that are assigned to a pool within the system, select **Pools** → **MDisks by Pools**.

When a new MDisk is added to a pool that already 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. No migration of extents occurs until there is sufficient I/O activity 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 559.

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 5-13. The command returns no feedback.

Example 5-13 The addmdisk command

```
IBM_2145:ITSO-SV1:superuser>addmdisk -mdisk mdisk2 -tier enterprise Pool0
IBM_2145:ITSO-SV1:superuser>
```

5.2.4 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 5-32. Alternatively, right-click the external MDisk.

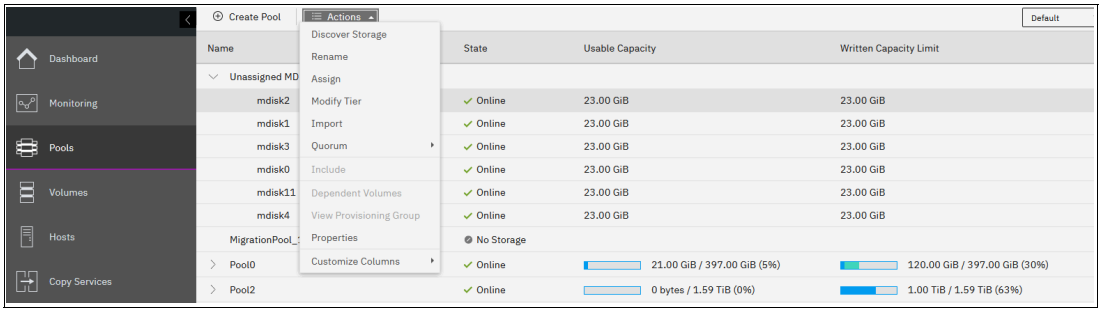


Figure 5-32 Actions for MDisks

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 these 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 297.

Modify Tier

To modify the tier to which the external MDisk is assigned, select **Modify Tier,** as shown in Figure 5-33. This setting is adjustable because the system cannot always detect the tiers that are associated with external storage automatically, unlike with internal arrays.

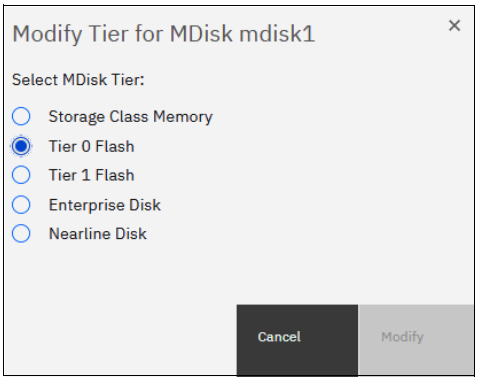


Figure 5-33 Modifying an external MDisk tier

For more information about storage tiers and their importance, see 9.1, “IBM Easy Tier” on page 560.

To change the external MDisk storage tier, run the **chmdisk** command. Example 5-14 shows setting the new tier to `mdisk2`. No feedback is returned.

Example 5-14 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 already 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 Chapter 12, “Encryption” on page 793.

To perform this task by using the CLI, run the **chmdisk** command, as shown in Example 5-15.

Example 5-15 Using chmdisk to modify encryption

```
IBM_2145:ITSO-SV1:superuser>chmdisk -encrypt yes mdisk5
IBM_2145:ITSO-SV1:superuser>
```

Import

This action is available only for unmanaged MDisks. Importing an unmanaged MDisk enables you to preserve the existing data on the MDisk. You can migrate the data to a new volume or keep the data on the external system.

MDisks are imported for storage migration. The IBM SAN Volume Controller provides a migration wizard to help with this process, which is described in Chapter 8, “Storage migration” on page 527.

Note: The wizard is the preferred method to migrate data from older storage to the IBM SAN Volume Controller. When an MDisk is imported, the data on the original LU is not modified. The system acts as a pass-through, and the extents of the imported MDisk do not contribute to storage pools.

To choose one of the following migration methods, select **Import**:

- Import to temporary pool as image mode volume

  This method does not migrate data from the source MDisk. It creates an *image mode volume* that has a direct block-for-block conversion of the MDisk. The data is preserved on the external storage controller, but it is also accessible from the IBM SAN Volume Controller system.

In this method, the image mode volume is created in a temporary migration pool and is presented through the IBM SAN Volume Controller. Choose the extent size of the temporary pool and click **Import**, as shown in Figure 5-34.

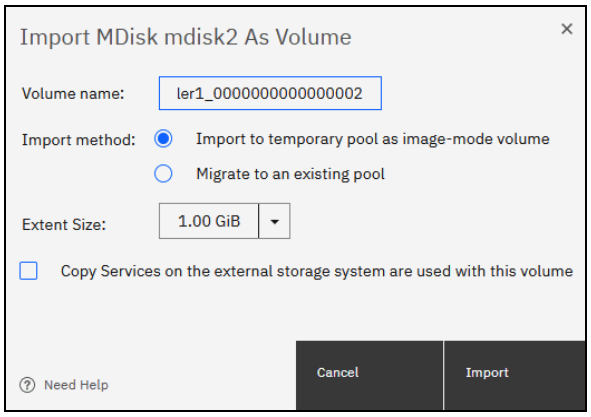


Figure 5-34 Importing an unmanaged MDisk

The MDisk is imported and listed as an `image` mode MDisk in the temporary migration pool, as shown in Figure 5-35.

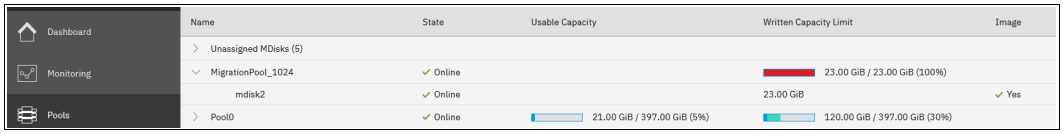


Figure 5-35 Image mode imported MDisk

A corresponding image mode volume is now available in the same migration pool, as shown in Figure 5-36.

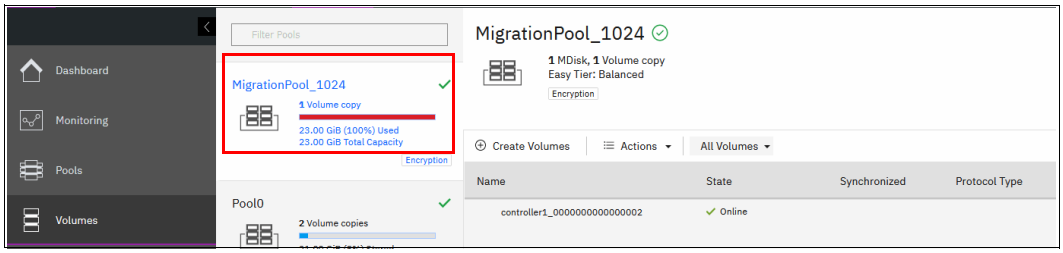


Figure 5-36 Image mode volume

The image mode volume can then be mapped to the original host. The data is still physically present on the MDisk of the original external storage controller and no automatic migration process is running. The original host sees no difference and its applications can continue to run. The image mode volume is now under the control of the system and it can optionally be migrated to another storage pool or be converted from image mode to a striped virtualized volume. You can use the **Volume Migration** wizard or perform the tasks manually.

- Migrate to an existing pool

  This method starts by creating an image mode volume like the first method. However, it then automatically migrates the image mode volume to a virtualized volume in the selected storage pool. Free extents must be available in the selected target pool so that the data can be copied there.

  If this method is selected, choose the storage pool to hold the new volume and click **Import**, as shown in Figure 5-37.

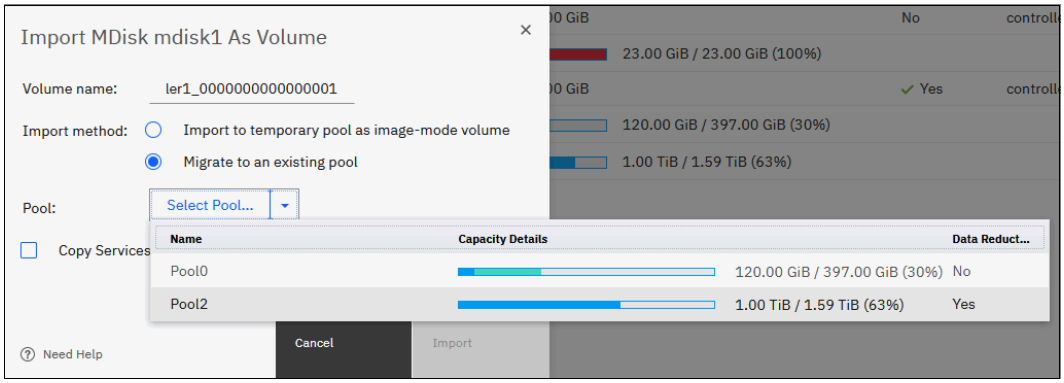


Figure 5-37 Migrating an MDisk to an existing pool

The data migration begins automatically after the MDisk is imported successfully as an image mode volume. You can check the migration progress by clicking the task under **Running Tasks**, as shown in Figure 5-38.

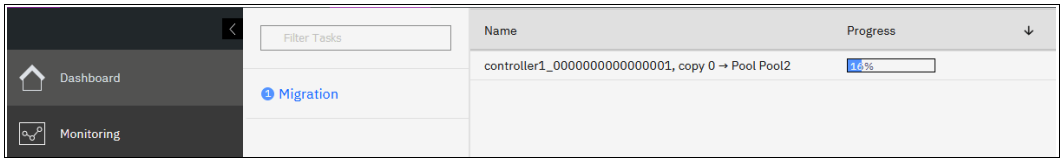


Figure 5-38 MDisk migration in the Running Tasks pane

After the migration completes, the volume is available in the chosen destination pool. This volume is no longer an image mode volume. It is now virtualized by the system.

All data is migrated off the source MDisk, and the MDisk switched its mode, as shown in Figure 5-39.

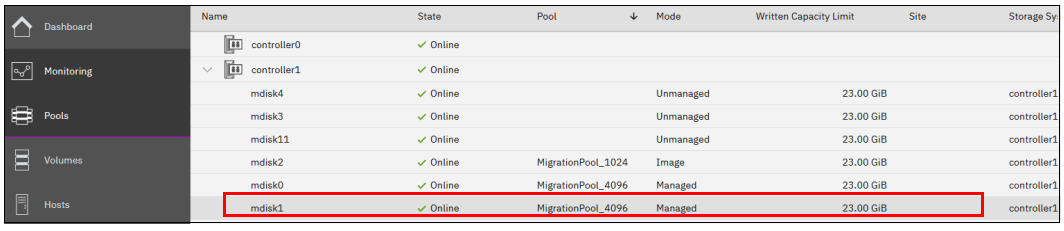


Figure 5-39 Imported MDisks appear as “Managed”

The MDisk can be removed from the migration pool. It returns to the list of external MDisks as `Unmanaged`. The MDisk can now be used as a regular managed MDisk in a storage pool, or it can be decommissioned.

Alternatively, 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 527.

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 5-40. For HyperSwap and ESC configurations, MDisks must belong to three different sites.

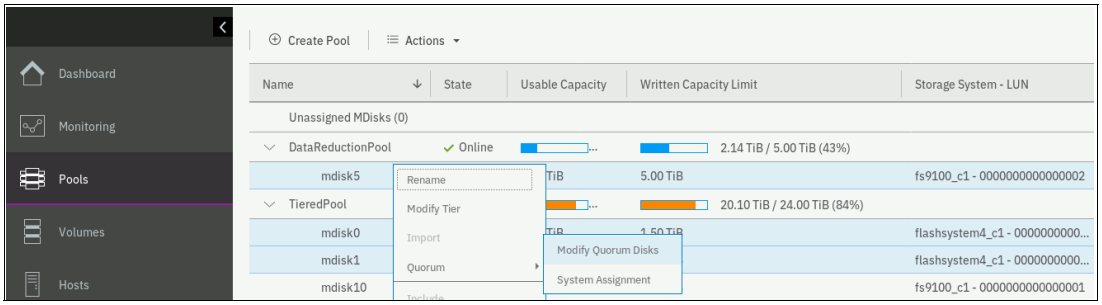


Figure 5-40 Selecting new quorum disks

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 there is no excessive error recovery that might impact 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 5-16.

Example 5-16 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 goes back to the `Unmanaged` state. If there are no volumes 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 5-41.

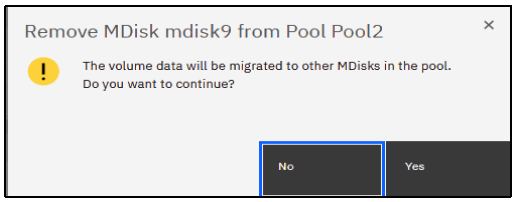


Figure 5-41 Removing an MDisk from a pool

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 returns 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 a loss of data for that volume. Use this option before you do 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 5-42.

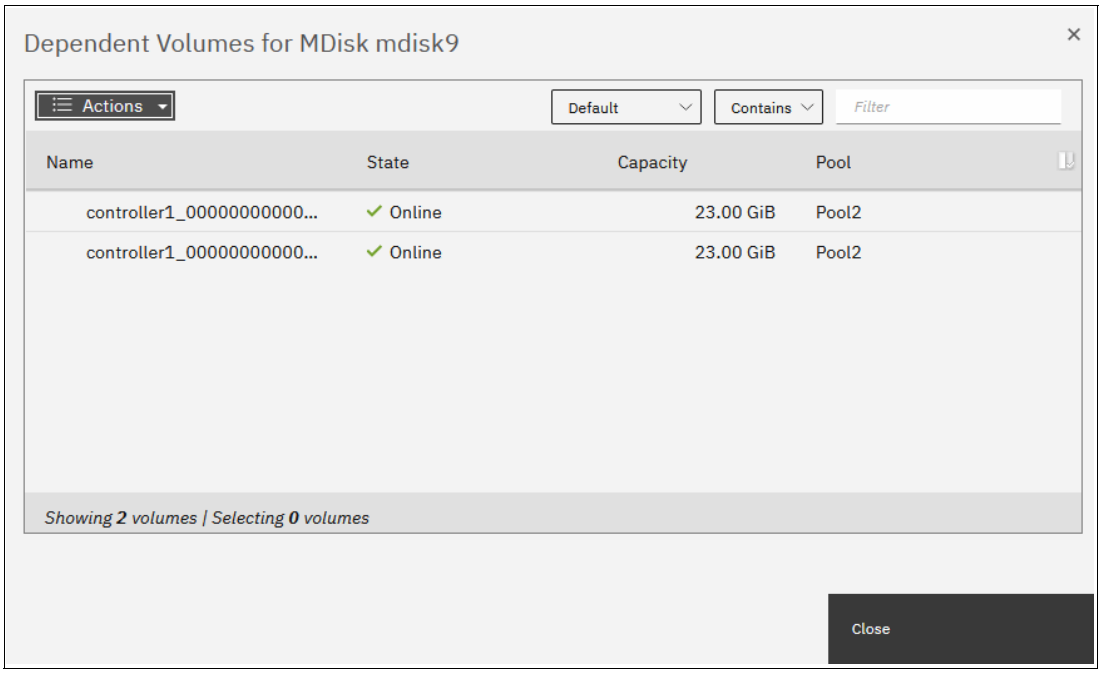


Figure 5-42 Dependent volumes for an MDisk

You can get the same information by running the **lsdependentvdisks** command (see Example 5-17).

Example 5-17 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 overprovisioned external storage controllers. 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 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 5-43.

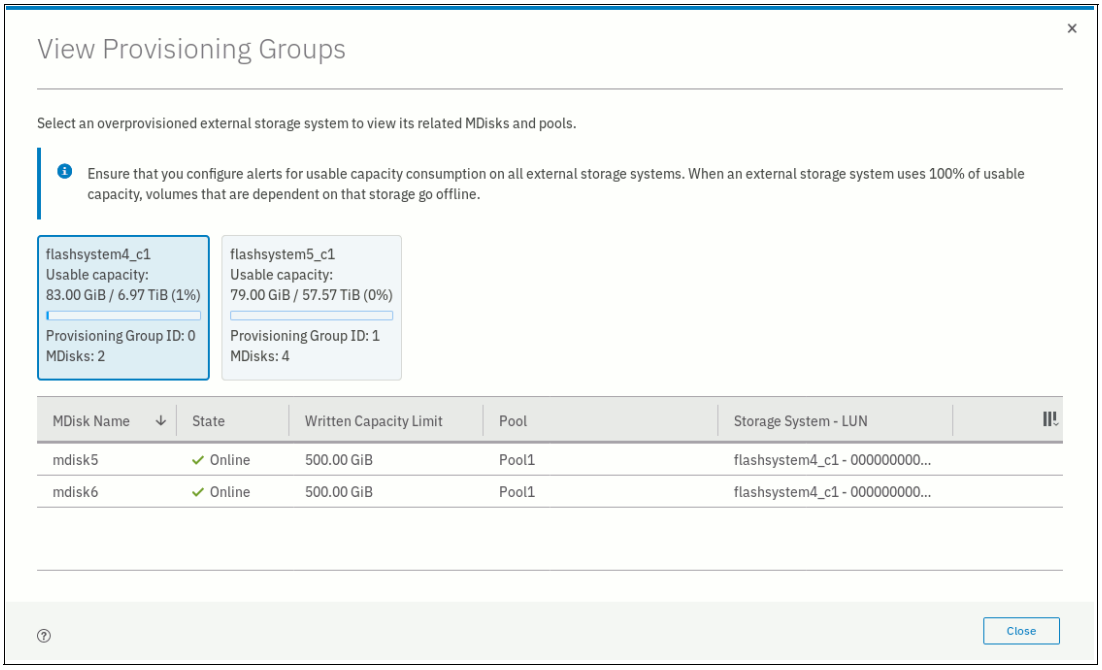


Figure 5-43 View Provisioning Groups

5.3 Working with internal drives and arrays

An *array* is a type of MDisk that is made up of disk drives (or flash drive modules). These drives are members of the array. A RAID is a method of configuring member drives to create high availability (HA) and high-performance groupings of drives. The system supports nondistributed (traditional) and distributed array configurations.

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 pane for managing all internal drives. To access the Internal Storage pane, click **Pools** → **Internal Storage**, as shown in Figure 5-44.

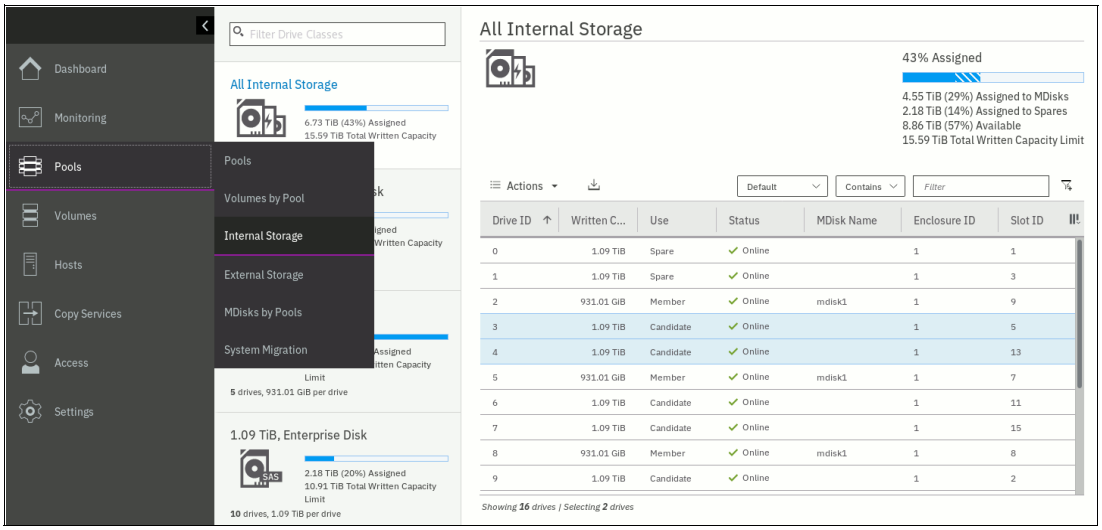


Figure 5-44 Internal storage pane

The pane 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 pane 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, as shown in Figure 5-44:

- 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 either 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 output to list only drives that belong to particular enclosure, only that have specific capacity, or by another attributes (see Example 5-18).

Example 5-18 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 status of each drive. A drive can be Online, which means that 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:

- 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 currently part of the RAID configuration. It is available to be used in an array MDisk.
- Spare: The drive is a hot spare protecting nondistributed (traditional) RAID arrays. If any member of such an 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 is 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 5-45.

		To				
		unused	**candidate**	**failed**	**member**	**spare**
From	**unused**	yes	yes	no	no	no
	candidate	yes	yes	yes	no	yes
	failed	yes	yes	yes	no	no
	member	no	no	yes	no	no
	spare	no	yes	yes	no	yes

Figure 5-45 Drive use transitions

The system automatically sets the drive use to Member when creating a RAID array. Changing the drive use from Member to Failed is allowed only if the array is not dependent 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. Initial setup or Assign storage GUI wizards do that 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 5-46. Alternatively, select the drives and click **Actions**.

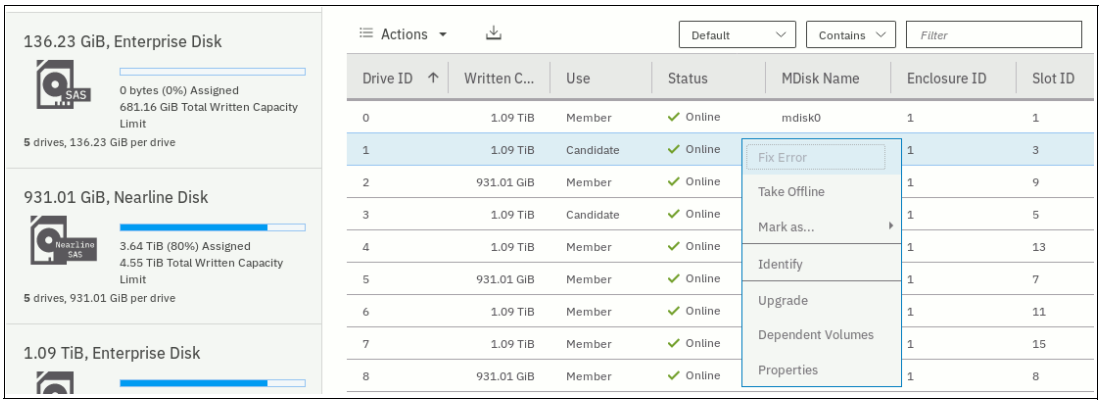


Figure 5-46 Actions on internal storage

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 associated with it. Select **Fix Error** to start the DMP for the selected drive. For more information about DMPs, see Chapter 13, “Reliability, availability, serviceability, monitoring and troubleshooting” on page 853.

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, as shown in Figure 5-47.

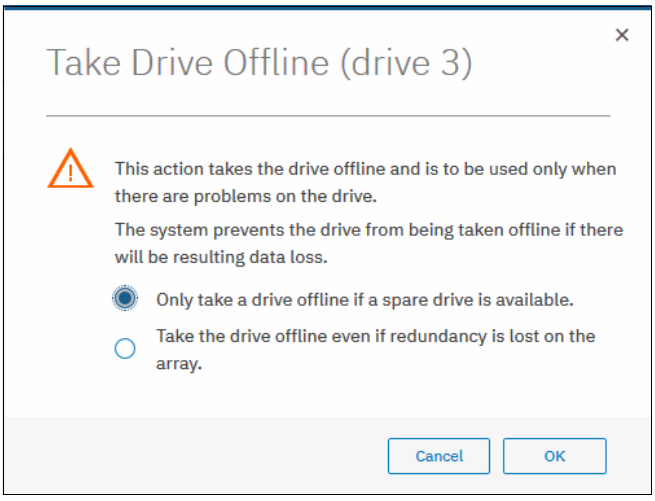


Figure 5-47 Taking a drive offline

The system prevents you from taking the drive offline if one of the following conditions is true:

- Taking the drive offline results in a loss of access to data.
- The first option was selected and no suitable spares are available to start a rebuild.

If a spare is available and the drive is taken offline, the associated MDisk remains Online and the RAID array starts a rebuild by using a suitable spare. If no spare is available and the drive is taken offline by using the second option, 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 5-48.

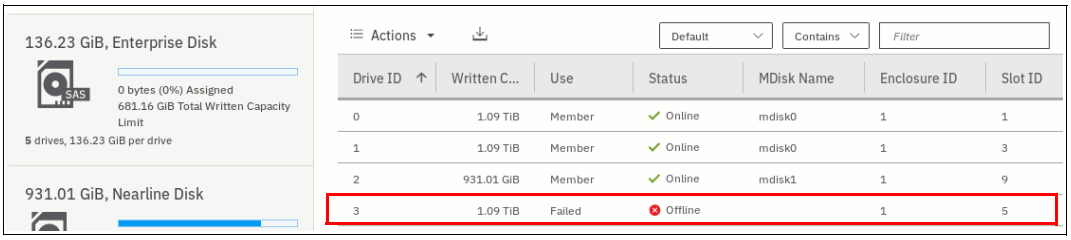


Figure 5-48 An offline drive is marked as failed

To take a drive offline with the CLI, run the **chdrive** command (see Example 5-19). This command returns no feedback. Use the **-allowdegraded** parameter to set a member drive offline, even if no suitable spare is available.

Example 5-19 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 5-49.

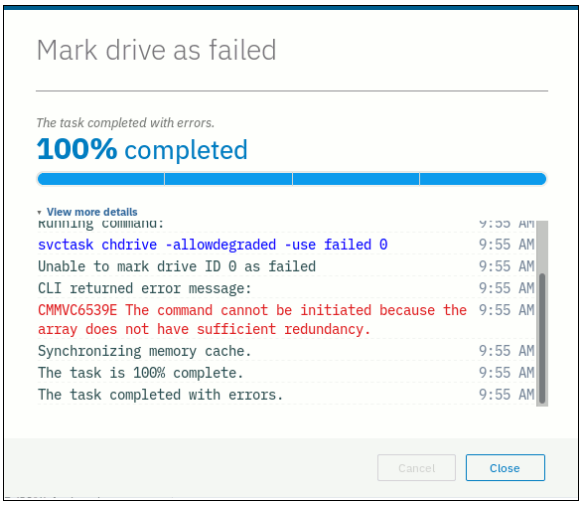


Figure 5-49 Taking a drive offline fails if it might result in a loss of access to data

Action: Mark as

Select **Mark as** to change the use that is assigned to the drive, as shown in Figure 5-50. The list of available options depends on the current drive use and state (see the allowed state transitions that are shown in Figure 5-45 on page 308).

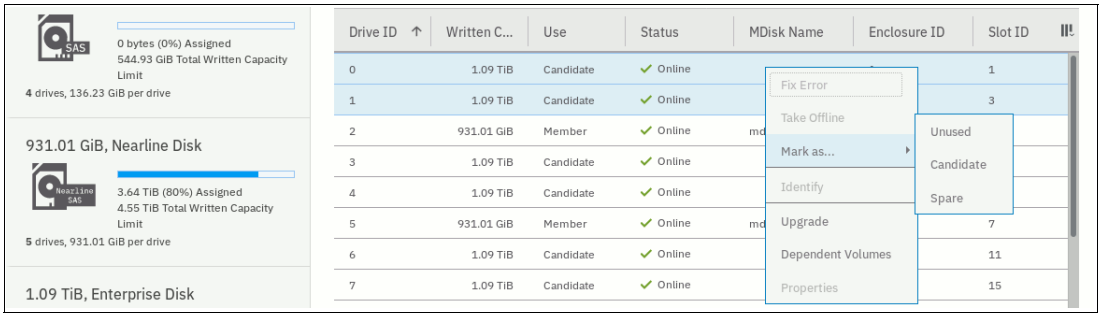


Figure 5-50 A drive can be marked as Unused, Candidate, or Spare

To change the drive role with the CLI, run the **chdrive** command (see Example 5-20). 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 to a Candidate role before. Distributed Arrays (DRAIDs) do not use spares. Marking drive as spare, fails for drive types supporting DRAIDs only.

Example 5-20 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: Identify

Select **Identify** to turn on the light-emitting diode (LED) light of the enclosure slot of the selected drive. This allows 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 5-51.

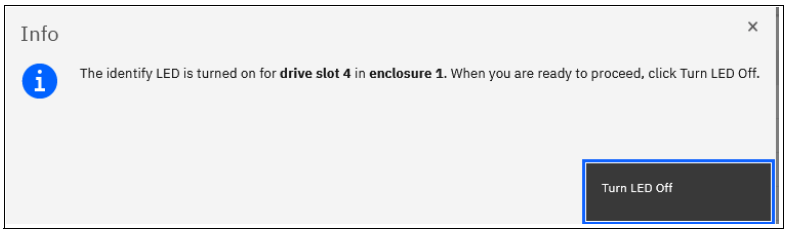


Figure 5-51 Identifying an internal drive

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 5-21 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 5-21 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: Upgrade

Select **Upgrade** to allow the user to update the drive firmware, as shown in Figure 5-52. You can choose to update an individual drive, selected drives, or all the drives in the system.

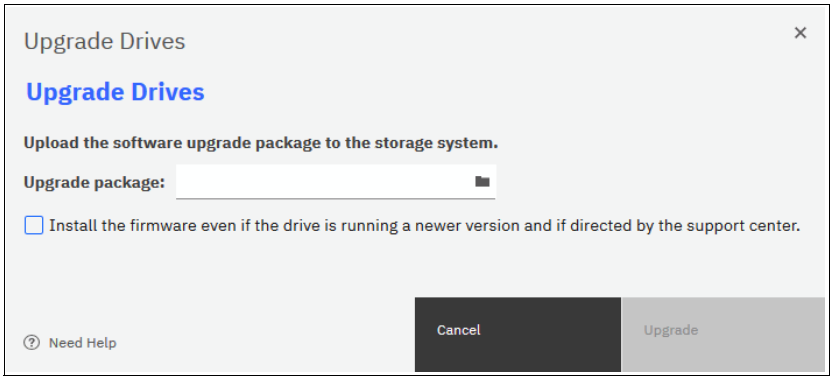


Figure 5-52 Upgrading a drive or a set of drives

For more information about updating drive firmware, see Chapter 13, “Reliability, availability, serviceability, monitoring and troubleshooting” on page 853.

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 5-53 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.

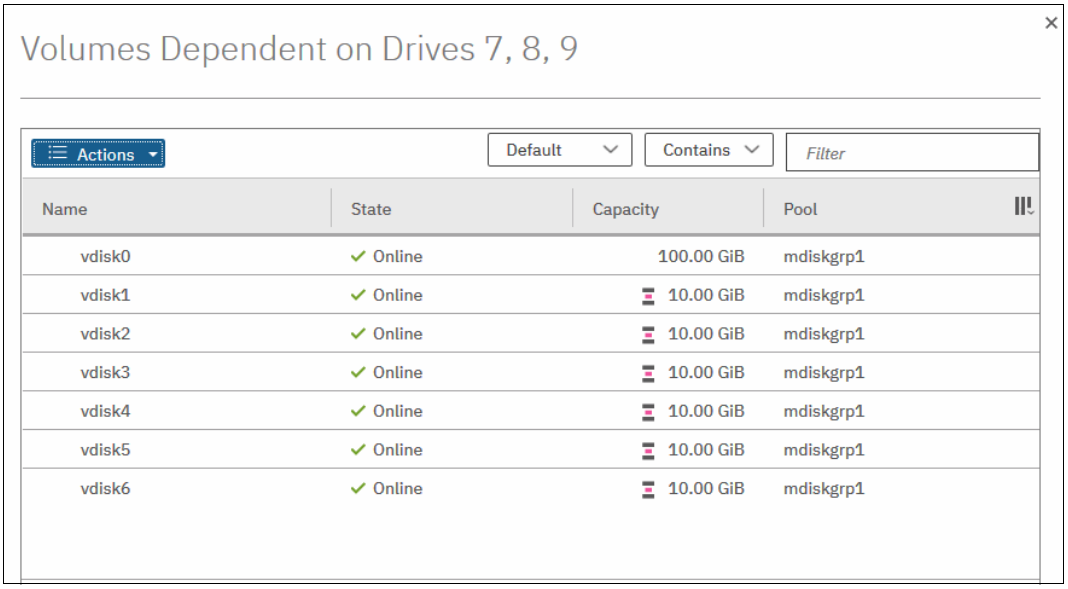


Figure 5-53 List of volumes dependent on disks 7, 8, 9

Whether there are dependent volumes depends on the redundancy of the RAID array at a 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 healthy RAID 6 array to have dependent volumes.

Note: A lack of dependent volumes does not imply that there are no volumes using 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 5-22.

Example 5-22 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 5-54.

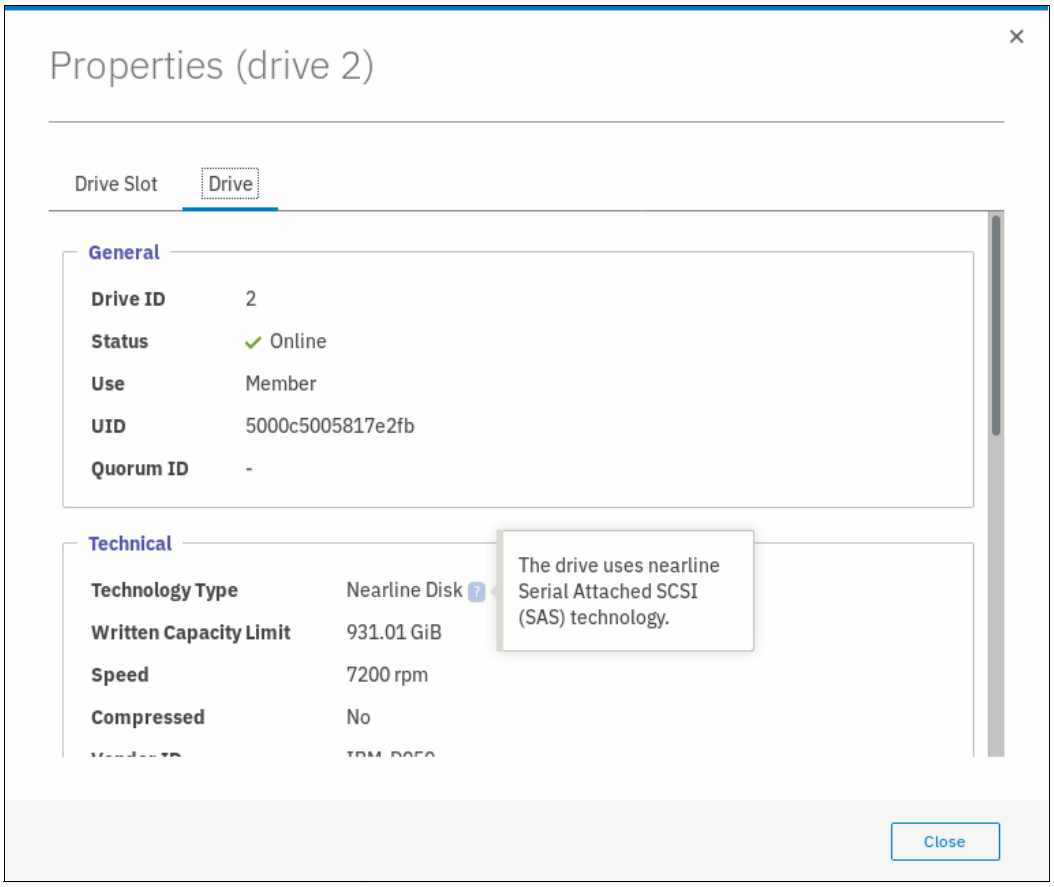


Figure 5-54 Drive properties

You can find a short description of each drive property by hovering on it and then clicking **[?]**. You can also display drive slot details by changing to the **Drive Slot** tab.

To get all available information about the particular drive, run the CLI command **`lsdrive`** with drive ID as the parameter. To get slot information, run the **`lsenclosureslot`** command.

5.3.2 RAID and distributed RAID

To use internal IBM SAN Volume Controller 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

Introduction to RAID technology

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.

When multiple physical disks are set up to use the RAID technology, they are in a RAID array. The IBM SAN Volume Controller provides multiple, traditional RAID (TRAID) levels:

- RAID 0
- RAID 1
- RAID 5
- RAID 6
- RAID 10

Note: RAID 0 does not provide any redundancy. A single drive failure in a RAID 0 array causes data loss.

In a TRAID approach, whether it is RAID 10, RAID 5, or RAID 6, data is spread among up to 16 drives in an array. There are separate spare drives 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:

- 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 TRAID 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.

Distributed RAID

In DRAID, there are no dedicated spare drives being idle most of the time. All of the 4 - 128 drives in the array process I/O requests at all times instead, which improves the overall I/O performance. Spare capacity is spread across all member drives to form one or more *rebuild areas*. During a rebuild, 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 impact 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 TRAID.

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 and to free up the rebuild area so that it can be used for another drive failure in the future.

The following DRAID types are available:

- DRAID 5
- DRAID 6

Note: IBM SAN Volume Controller is not a supported platform for DRAID 1.

Figure 5-55 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.

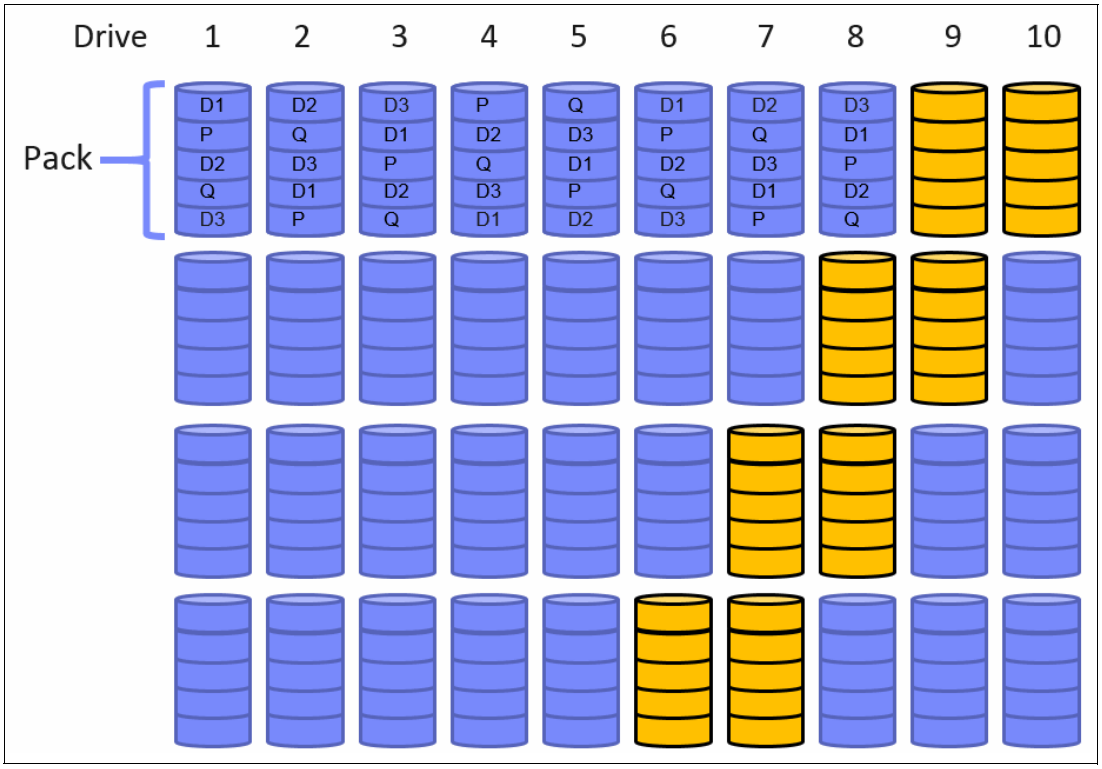


Figure 5-55 Distributed RAID 6 (for simplification, not all packs are shown)

Figure 5-56 on page 317 shows what happens after a single drive failure in this DRAID 6 array. Drive 3 failed and the array is using 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 once more after another failure.

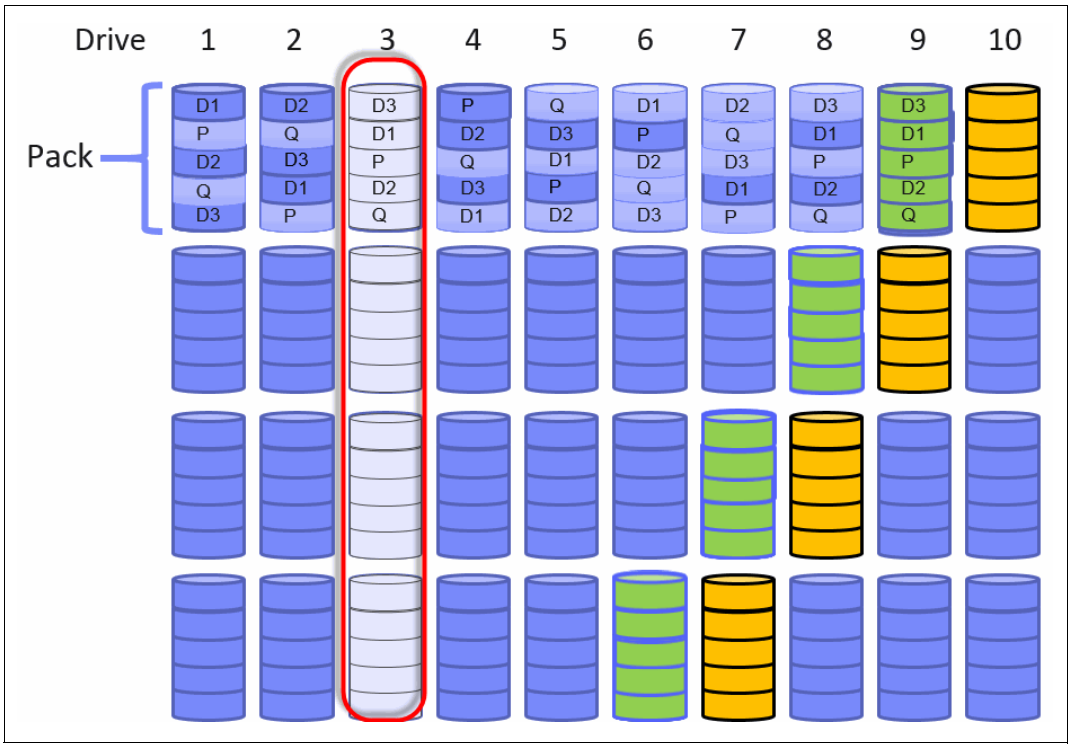


Figure 5-56 Single drive failure with DRAID 6 (for simplification, not all packs are shown)

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 case of a drive failure, data is read from many drives and written to many drives. This process minimizes the impact 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. The maximum number of drives a DRAID can contain is 128.
- Existing 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 small number of arrays.

The following minimum number of drives are needed to build a Distributed Array:

- Six drives for a DRAID 6 array
- Four drives for a DRAID 5 array

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: It is recommended to use DRAID 6 whenever possible. DRAID technology dramatically reduces rebuild times, decreases the exposure volumes have to the extra load of recovering redundancy, and improves performance.

To create a RAID array from internal storage, select **Pools** → **Pools**, **Actions** and than **Add Storage** or right-click the storage pool you want to add arrays and select **Add Storage**, as shown in Figure 5-57.

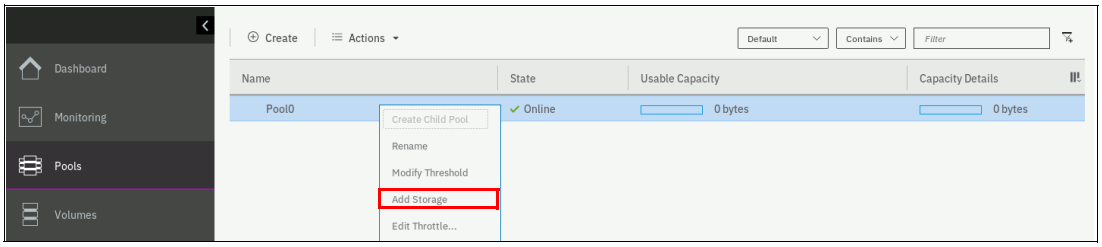


Figure 5-57 Adding storage to a pool

Alternatively, select **Pool, Mdisk by pools** and click to assign for drive class to open the configuration dialog (see Figure 5-58).

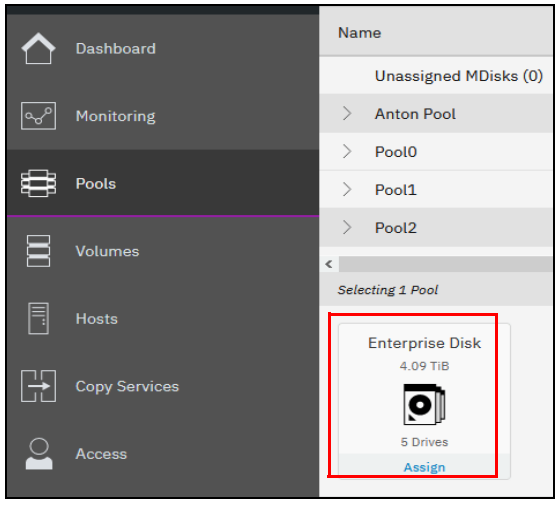


Figure 5-58 Assign drives

This action starts the configuration dialog box that is shown in Figure 5-59. If any of the drives are found with an Unused role, reconfigure them as Candidates to be included into the configuration.

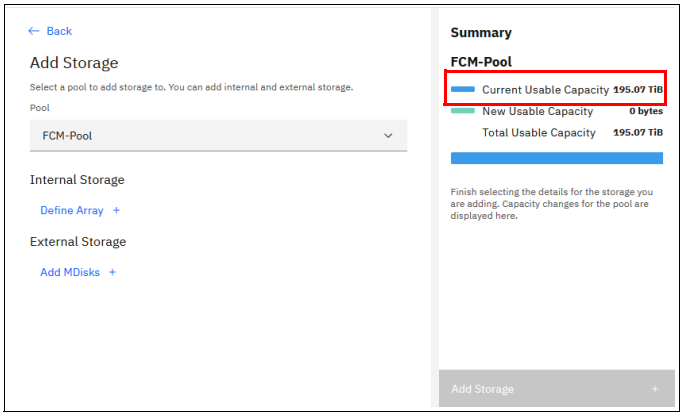


Figure 5-59 Assigning storage to a pool

If **Internal Storage** is chosen, the system guides you through the array MDisk creation process. If **External Storage** is selected, systems guide you through the selection of external storage.

Select the Pool from the drop down menu if no Pool is selected. The summary view on the right pane shows the Current Usable Capacity of the selected pool. After you defined Internal or External storage or both, click **Add storage**.

Internal storage

Select **Define Array,** and choose the drive class for the array from the drop-down menu. Only drive classes for which candidate drives exists are displayed. The system automatically recommends a raid type and level that is based on the available candidate drives.

If you are adding storage to a pool that has storage that is assigned to the pool, the existing storage configuration is considered for the recommendation. The system aims to achieve a balanced configuration, so 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 existing arrays in a pool by using the GUI.

Select **Advanced** 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 pane 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 change the drive count of a traditional array. If you need to change these properties, you must delete the array MDisk and re-create it with the required settings.

In the example that is shown in Figure 5-60, the dialog box recommends that you create one DRAID 6 with all 15 10-k enterprise drives. The summary view reflect the new usable capacity based on the selection.

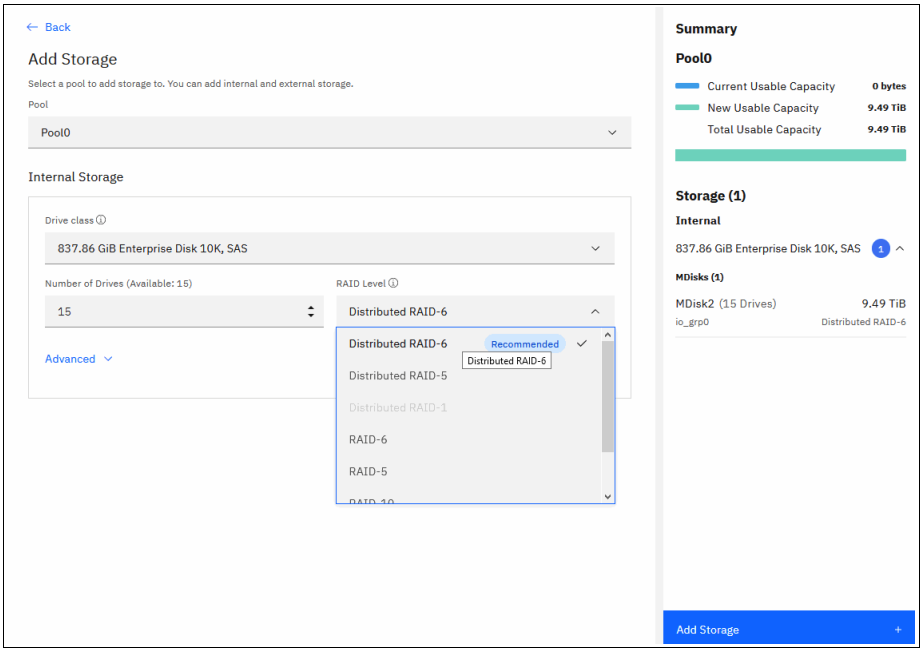


Figure 5-60 Assign Storage to Pool

Figure 5-61 shows rebuild Areas, stripe width, and array width.

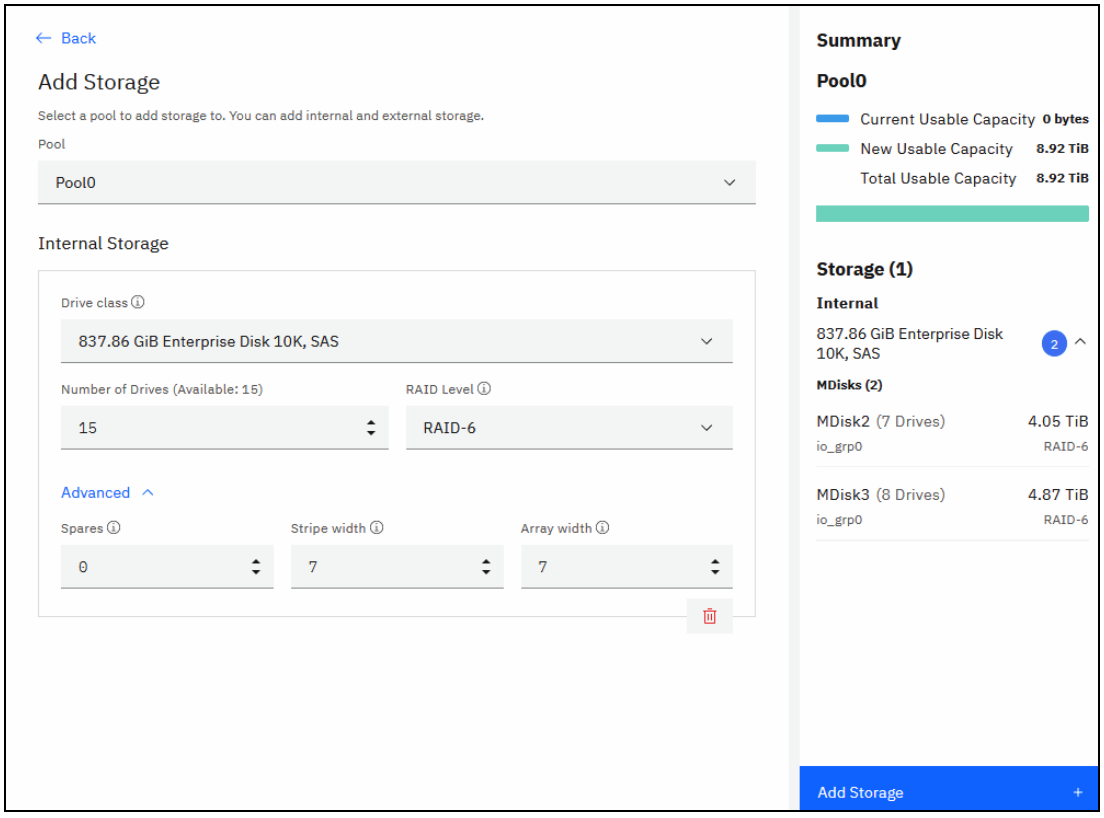


Figure 5-61 Rebuild areas, stripe width, and array width

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. By default, the system recommend DRAID 6 when possible.

If the system has multiple drive classes (for example, flash and enterprise drives), use the plus symbol to create more array from other drive classes to take advantage from 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 559.

If the pool has a DRAID 6 array 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, as shown in Figure 5-62. The array is available for I/O during this process and there is no need to wait for it to complete.

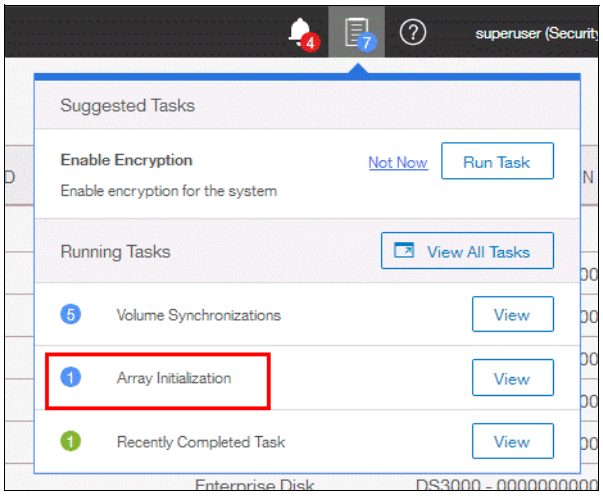


Figure 5-62 Array Initialization task

Click **View** in the Running tasks list to see the initialization progress and the time remaining, as shown in Figure 5-63. The time 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.

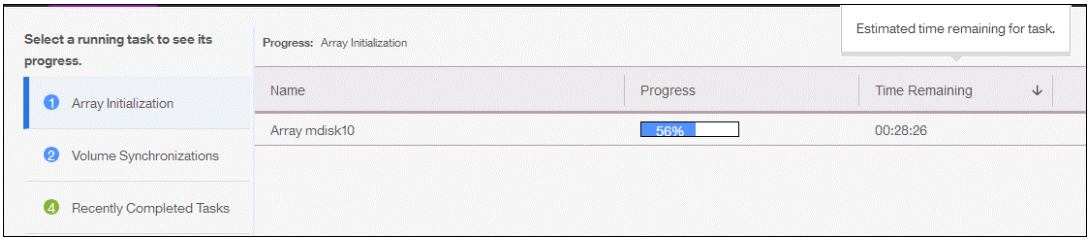


Figure 5-63 Array initialization task progress information

Configuring arrays with the CLI

When working with the CLI, run the **mkarray** command to create TRAID and 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 307.

To get the recommended array configuration on the CLI, run the **lsdriveclass** command to list the available drive classes and the **lsarrayrecommendation** commands as shown in Example 5-23. The recommendations are listed in the order of preference.

Example 5-23 Listing array recommendations using the CLI

```
IBM_2145:ITSO-SV1:superuser>lsdriveclass
id RPM   capacity tech_type       block_size candidate_count
0  10000 1.1TB    tier_enterprise 512        10
1  7200  931.0GB  tier_nearline   512        5
2  15000 136.2GB  tier_enterprise 512        1
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 5-24, 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 5-24 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
```

There are default values for the stripe width and the number of rebuild areas, depending on 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.

To create a TRAID 10 MDisk instead, you must specify a list of drives that you want to add as members, its RAID level, and the storage pool name or ID to which you want to add this array. As shown in Example 5-25, a RAID-10 array is created and Pool2 is added to it. A spare drive also is designated.

Example 5-25 Creating TRAID with mkarray

```
IBM_2145:ITSO-SV1:superuser>mkarray -level raid10 -drive 0:1:2:3:4:5:6:7 Pool2
MDisk, id [0], successfully created
IBM_Storwize:ITSOV7K:superuser>chdrive -use spare 8
```

Note: Do not forget to designate some of the drives as spares when creating traditional arrays. Spare drives are required to perform a rebuild immediately after a drive failure.

The storage pool must exist (see 5.1.1, “Creating storage pools” on page 274). 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. Some actions that are supported on TRAID arrays are not supported on DRAID arrays and vice versa.

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 5-64.

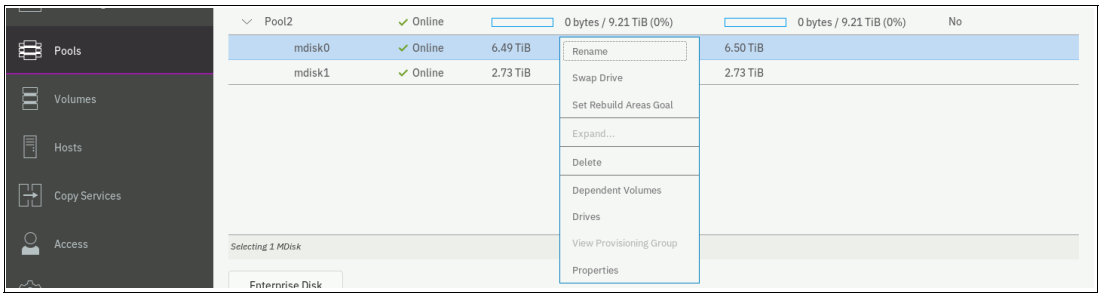


Figure 5-64 Actions on arrays

Rename

Select this option to change the name of an MDisk.

The CLI command for this operation is **`charray`** (see Example 5-26). No feedback is returned.

Example 5-26 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. 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, as indicated by an error message in the event log.

Figure 5-65 shows the dialog box that opens. Select the member drive to be replaced and the replacement drive, and click **Swap**.

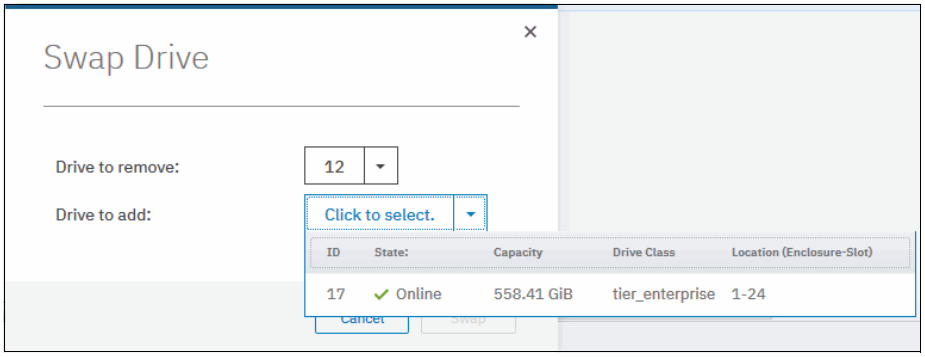


Figure 5-65 Swapping array member with another candidate or spare drive

The exchange of the drives starts running in the background. The volumes on the affected MDisk remain accessible during the process.

Swapping a drive in a traditional array performs a concurrent member exchange that does not reduce the redundancy of the array. The data of the old member is copied to the new member and after completion the old member is removed from the array.

In a distributed array, the system immediately removes the old member from the array and performs a rebuild. 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 5-27 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 5-27 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 Spare Goal or Set Rebuild Areas Goal option

Select this option to set the number of spare drives (on TRAID) or rebuild areas (on DRAID) that are expected to protect the array from drive failures.

If the number of rebuild areas that is available does not meet the configured goal, an error is logged in the event log, as shown in Figure 5-66. This error can be fixed by replacing failed drives in the DRAID array.

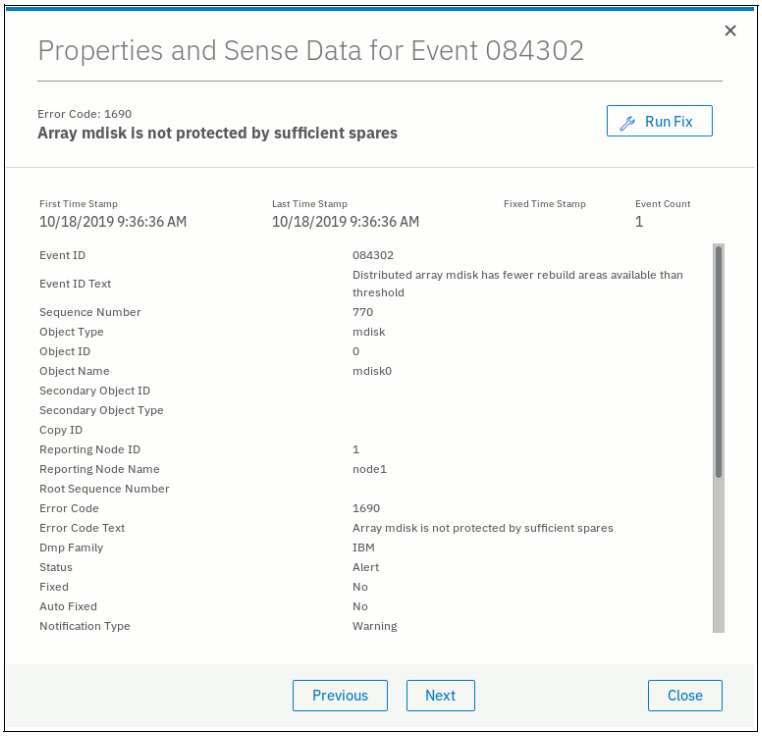


Figure 5-66 An error 1690 is logged if insufficient rebuild areas are available

Note: This option does not change the number of rebuild areas or spares that is available to the array. It specifies only at which point a warning event is generated. Setting the goal to 0 does not prevent the array from rebuilding.

In the CLI, this task is performed by using the **charray** command (see Example 5-28).

Example 5-28 Adjusting array goals with charray (some columns are not shown)

```
IBM_2145:ITSO-SV1:superuser>lsarray
mdisk_id mdisk_name        status mdisk_grp_id mdisk_grp_name distributed
0        mdisk0            online 0            mdiskgrp0 no
16       Distributed_array online 1            mdiskgrp1 yes
IBM_2145:ITSO-SV1:superuser>charray -sparegoal 2 mdisk0
IBM_2145:ITSO-SV1:superuser>charray -rebuildareasgoal 0 Distributed_array
```

Expand

Select **Expand** to expand the array by adding 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.

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 5-67 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 are updated as needed. To add rebuild areas to the array, click **Advanced Settings** and modify the number of extra spares.

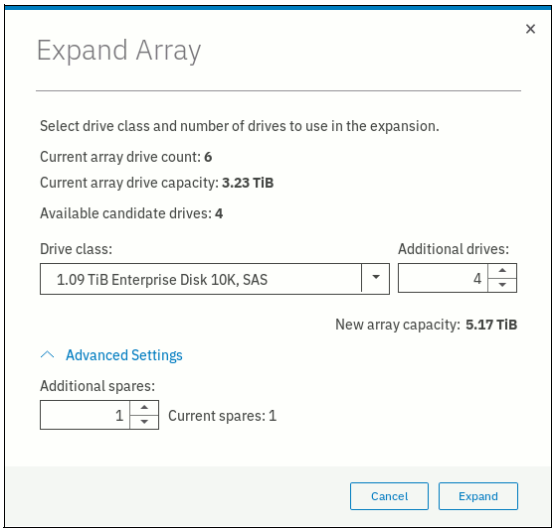


Figure 5-67 Expand a distributed array

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 5-68.



Figure 5-68 Array expansion progress

On the CLI, this task is performed by running the **`expandarray`** command. To get a list of compatible drive classes, the **`lscompatibledriveclasses`** command can be run (see Example 5-29 on page 327).

Example 5-29 Expand an array 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, including 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 into 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 5-69.

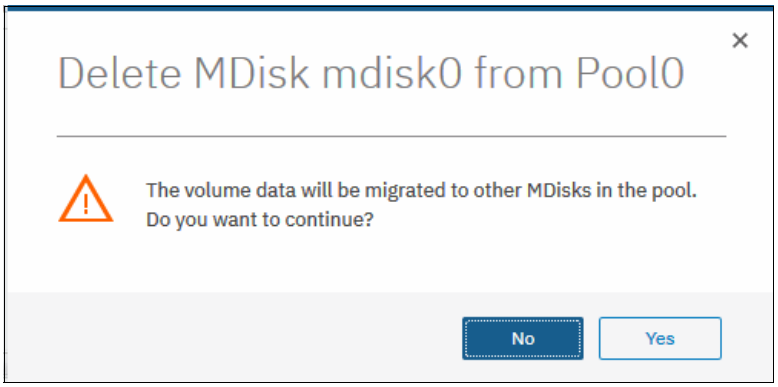


Figure 5-69 Deleting an array from a non-empty storage pool

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 is dependent 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. This means 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 5-70.

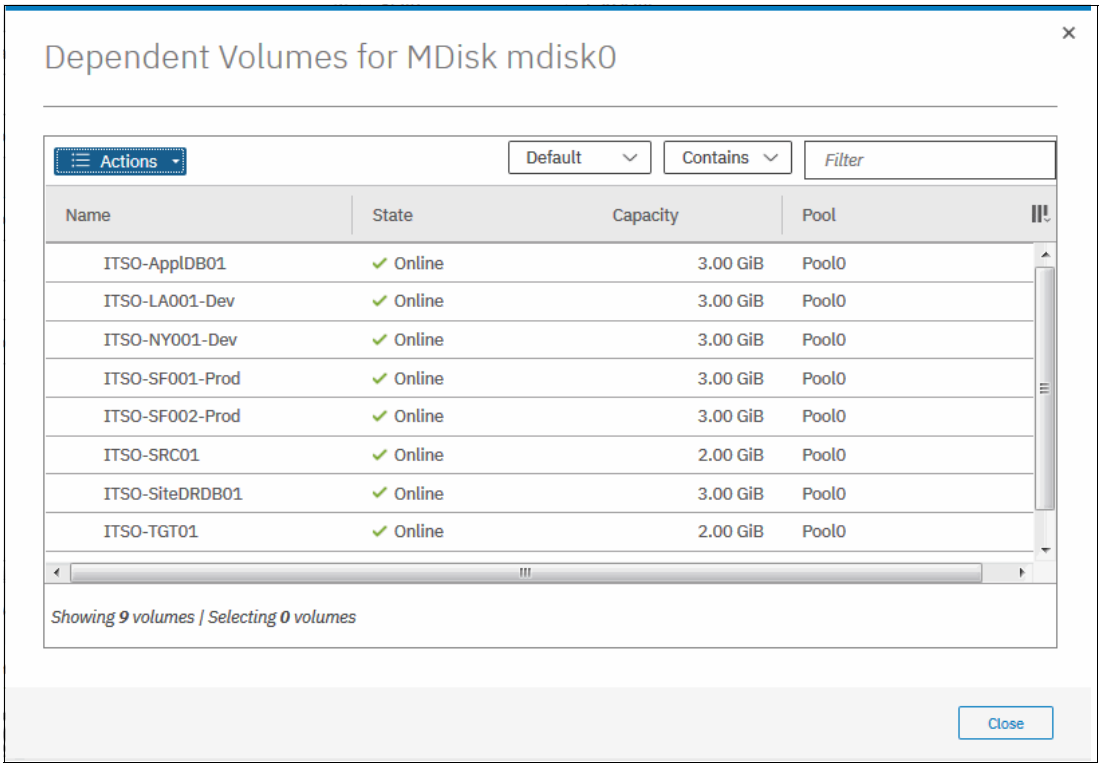


Figure 5-70 Dependent volumes for MDisk mdisk0

You can get the same information by running the CLI command **lsdependentvdisks** (see Example 5-30).

Example 5-30 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 5-71.

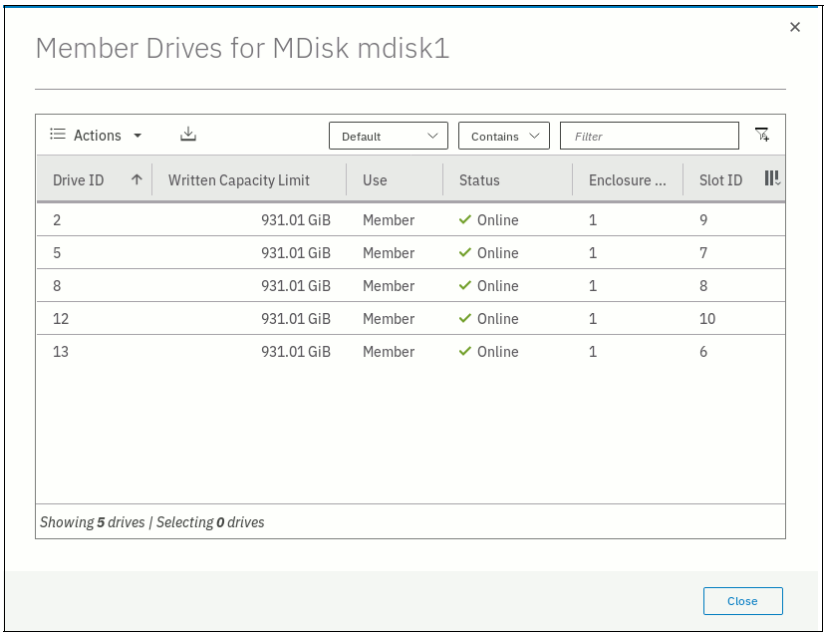


Figure 5-71 List of drives in an array

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, RAID level, and others.

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 5-31.

Example 5-31 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
<...>
```

6

Creating and provisioning volumes

In IBM Spectrum 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); that is, a logical disk.

This chapter describes how to create and provision volumes on IBM Spectrum Virtualize systems. The first part of this chapter provides a brief overview of IBM Spectrum 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 332
- 6.2, “Volume characteristics” on page 332
- 6.3, “Virtual volumes” on page 351
- 6.4, “Volumes in multi-site topologies” on page 351
- 6.4.2, “Stretched topology” on page 353
- 6.6, “Volume operations by using the CLI” on page 414

© Copyright IBM Corp. 2021. All rights reserved.

6.1 Introduction to volumes

For an IBM Spectrum 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 Spectrum 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 managed disks (MDisks), which are Redundant Arrays of Independent Disks (RAIDs) configured using internal storage, or LUs that are presented to and virtualized by an IBM Spectrum 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 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 to 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 (an application running on a host):

- Size
- Performance (input/output operations per second (IOPS), response time, and bandwidth)
- Resiliency
- Storage efficiency
- Security (data-at-rest encryption)
- Extent allocation policy
- Management mode

Also, volumes can be configured as VMware vSphere Virtual Volumes (VVOLs).

VVOLs change the approach to VMware virtual machines (VMs) disk configuration from “VM disk is a file on a VMware Virtual Machine File System (VMFS) volume” to one-to-one mapping between VM disks and storage volumes. VVOLs can be managed by the VMware infrastructure so that the storage system administrator can delegate VM disk management to VMware infrastructure specialists. This greatly simplifies storage allocation for virtual infrastructure and reduces the storage management team’s effort required to support VMware infrastructure administrators.

The downside of using VVOLs is multiplication of the number of volumes that is presented by a specific storage system, because typically multiple VM disks are configured on every VMFS volume. Because excessive proliferation of volumes presented to ESXi clusters can have negative impact on performance, it is recommended to carefully plan storage system configuration before production deployment, including in the assessment the projected system growth.

Note: If too many LUNs are presented to a sufficiently large ESXi cluster, I/O requests that are simultaneously generated by ESXi hosts can exceed the storage system command queue. Such overflow leads to I/O request retries, which reduces 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 can be *interdependent*; that is, setting one of them might affect other properties of the volume.

The volume parameters and their interdependencies are described in the following sections.

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. By default, extents are allocated from all MDisks in the storage pool using 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 if 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 and ensure that each MDisk included in the manually specified stripe set has sufficient number of free extents.
- Allow the system to automatically create the volume copy by not supplying a specific stripe set.

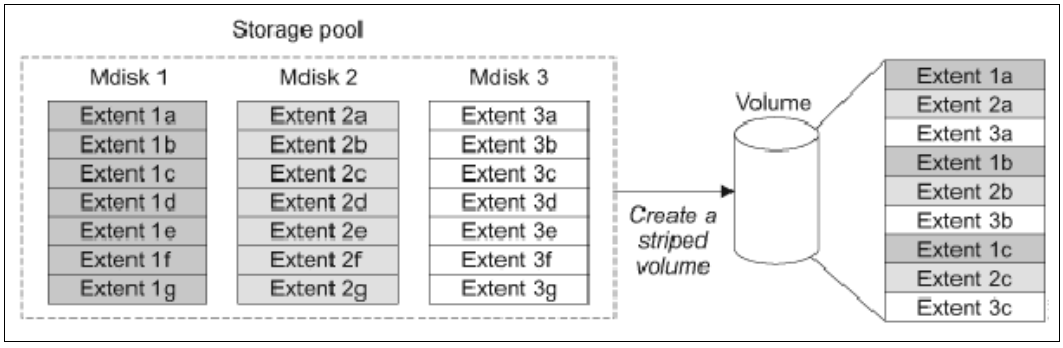


Figure 6-1 Striped extent allocation

- 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, the extents are allocated from the set of MDisks (by default, all MDisks in the storage pool):

- 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* occurs when multiple logical volumes that are defined on a set of physical storage devices (MDisks) store their metadata or file system transaction log on the same physical device (MDisk).

Because of the way the file systems work, system metadata disk regions typically are 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 successful completion of the file write.

If multiple volumes (each with its 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.

Also, some file systems allow specifying different logical disks for data and metadata storage. When taking advantage of this filesystem feature you can allocate differently configured volumes dedicated for data and metadata storage.

6.2.2 Managed mode and image mode

Volumes are configured within IBM Spectrum 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, so each MDisk extent maps to exactly one volume extent.

Note: For volumes with two copies, one volume extent maps to two MDisk extents (one for each volume copy).

Figure 6-2 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.

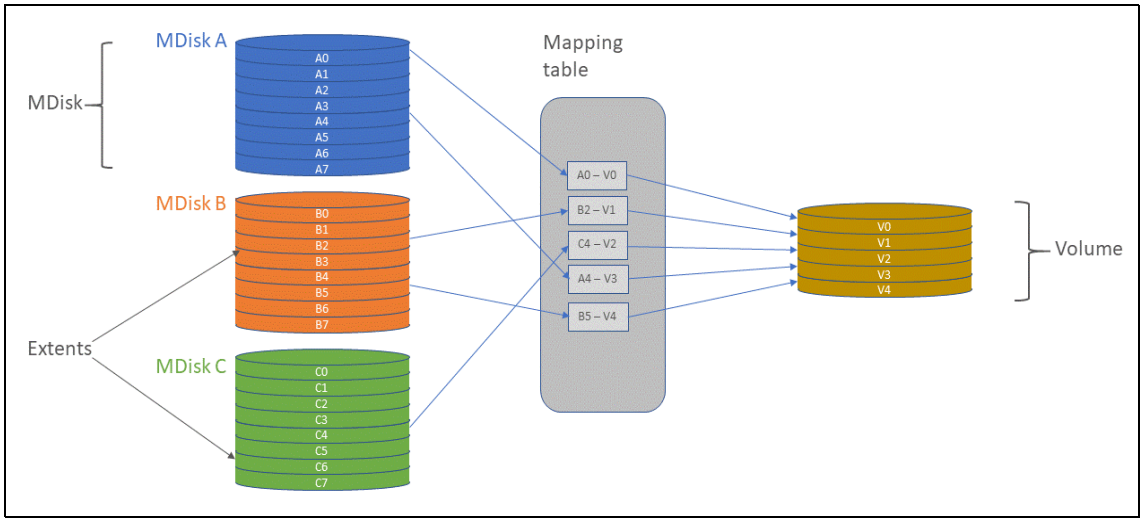


Figure 6-2 Simple view of block virtualization

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 Spectrum 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 (continuous set of logical blocks) and physical storage that is used to store these blocks. This function requires that physical storage (MDisks) is fully managed by IBM Spectrum Virtualize, which means that the LUs that are presented to IBM Spectrum 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 Spectrum Virtualize to work with LUs that were previously directly mapped to hosts, which is often required when IBM Spectrum Virtualize is introduced into a storage environment. In such scenario, image mode volumes are used to enable seamless migration of data and a smooth transition to virtualized storage.

The image mode creates one-to-one mapping of logical block addresses (LBAs) between a volume and a single MDisk (a LU that is presented by the virtualized storage). Image mode volumes have a minimum size of one block (512 bytes) and always occupy at least one extent.

An image mode MDisk cannot be used as a quorum disk and no IBM Spectrum Virtualize system metadata extents are allocated from it. All the IBM Spectrum 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 on page 336.

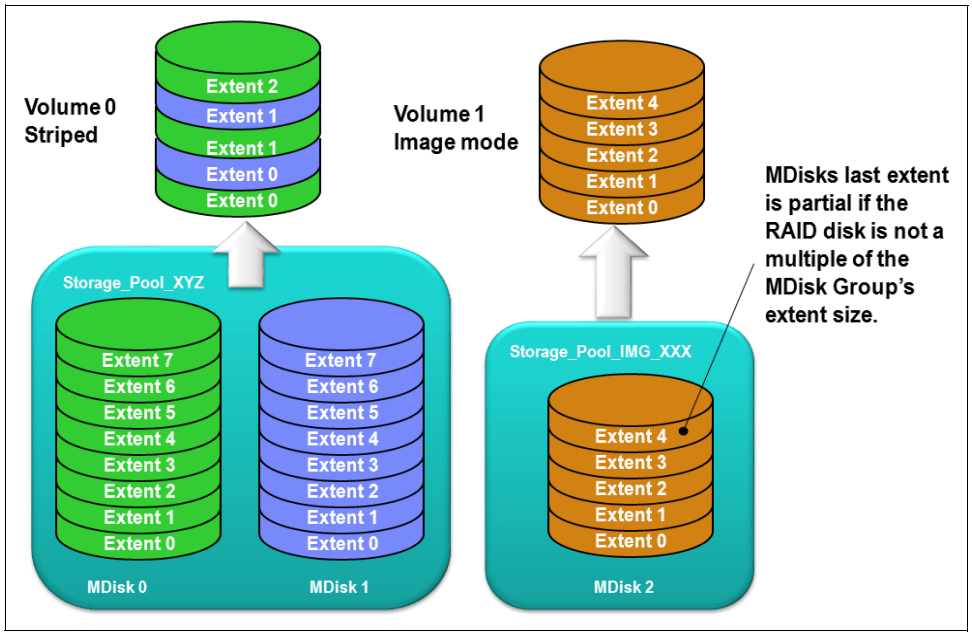


Figure 6-3 An image mode volume versus a striped volume

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 (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 managed by or based on IBM Spectrum 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 managed by the IBM Spectrum Virtualize system (see Chapter 2.9, "Back-end storage configuration" on page 79).

IBM Spectrum 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. Its mode changes to "image" only after the process completes.

6.2.3 VSize

Each volume has two associated values that describe its size: real capacity and virtual capacity:

- The real (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 any other IBM Spectrum Virtualize components or functionalities (for example, IBM FlashCopy, cache, and Remote Copy [RC]) that operate based on a volume size.

In a standard-provisioned volume, the real and virtual capacity 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.

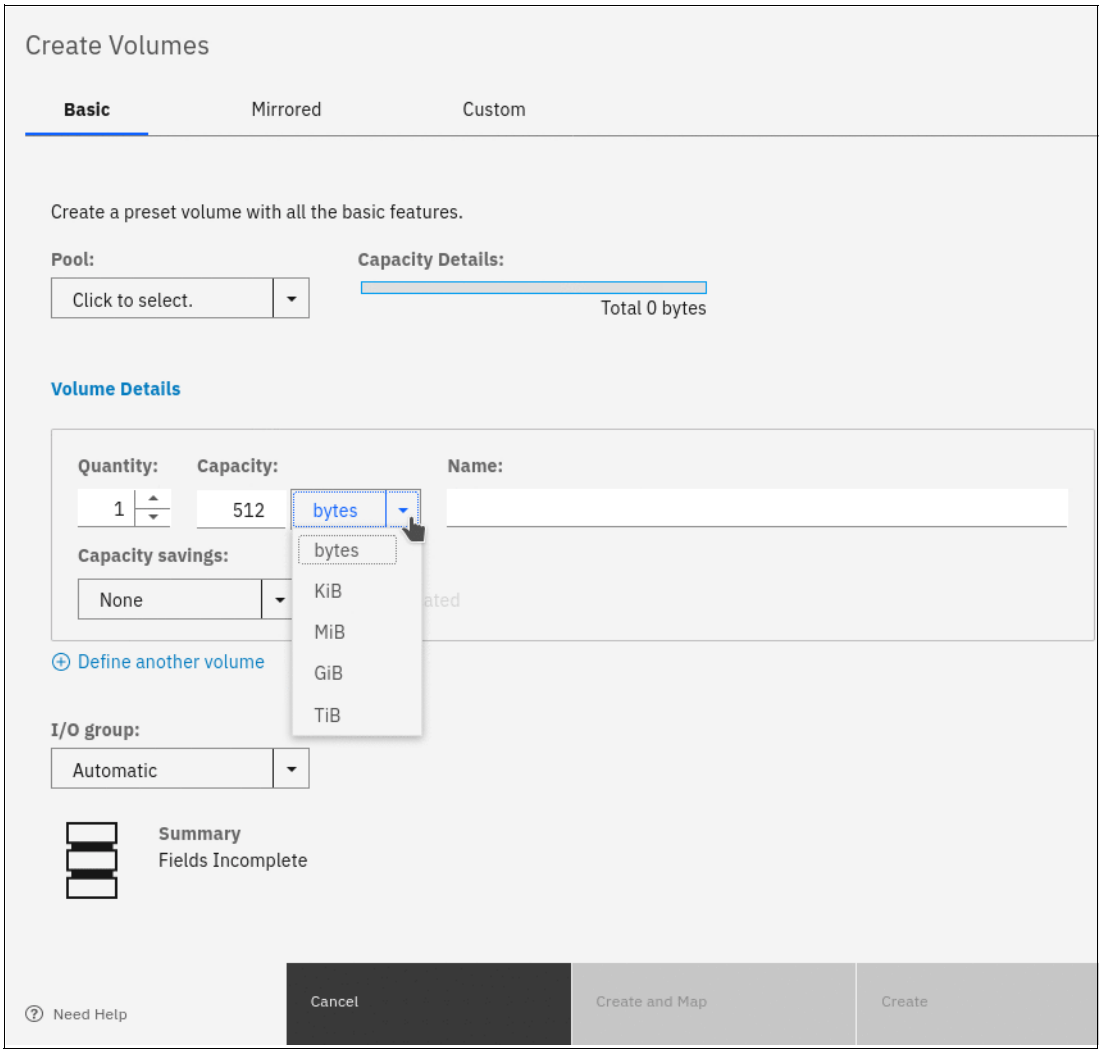


Figure 6-4 Smallest possible volume size

A volume is composed of storage pool extents, so 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 a whole 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 271.

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, 1st 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 possibility for the Easy Tier function to optimize the performance of a volume. However, volumes configured in the same storage pool or pools might still have different performance characteristics, depending on storage resiliency, efficiency, security, and allocation policy configuration settings of a specific volume.

Another factor influencing performance of volumes served by IBM SAN Volume Controller 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 also are 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:

- 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 339.

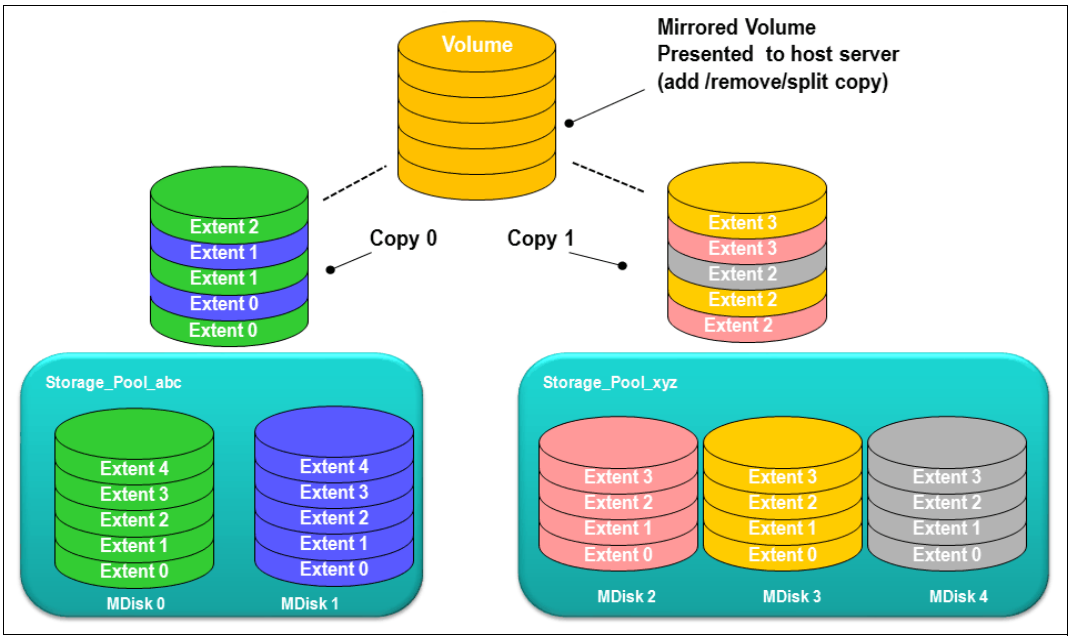


Figure 6-5 Volume mirroring overview

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 volume with more than one copy is beneficial in multiple scenarios; for example:

- 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 409.

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. While not required, using different pools 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 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.

Table 6-1 lists the bitmap space configuration options.

Table 6-1 Bitmap space default configuration

Copy service	Minimum allocated bitmap space	Default allocated bitmap space	Maximum allocated bitmap space	Minimum[a] capacity when using the default values
RC[b]	0	20 MiB	512 MiB	40 TiB of remote mirroring volume capacity
FlashCopy[c]	0	20 MiB	2 GiB	► 10 TiB of FlashCopy source volume capacity ► 5 TiB of incremental FlashCopy source volume capacity
Volume mirroring	0	20 MiB	512 MiB	40 TiB of mirrored volumes
RAID	0	40 MiB	512 MiB	► 80 TiB array capacity by using RAID 0, 1, or 10 ► 80 TiB array capacity in three-disk RAID 5 array ► Slightly less than 120 TiB array capacity in five-disk RAID 6 array

a. The actual amount of available capacity might increase based on the settings, such as grain size and strip size. RAID is subject to a 15% margin of error.
b. RC includes Metro Mirror (MM), Global Mirror (GM), and active-active relationships.
c. FlashCopy includes the FlashCopy function, Global Mirror with Change Volumes (GMVC), and active-active relationships.

The sum of all bitmap memory allocation for all functions except FlashCopy must not exceed 552 MiB.

6.2.6 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 initial configuration of a mirrored volume, but 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 non-mirrored volumes, only one volume copy exists, so 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 then responsible for destaging the data from cache to persistent storage. Figure 6-6 shows the data flow for this scenario.

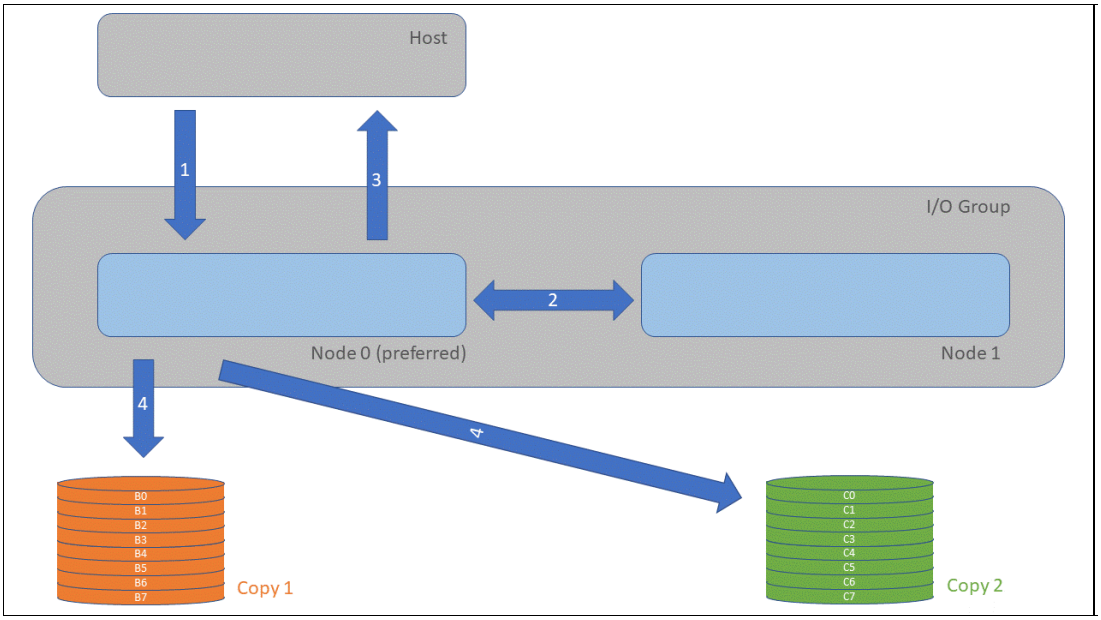


Figure 6-6 Data flow for write I/O processing in a mirrored volume

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 343 shows a case 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.

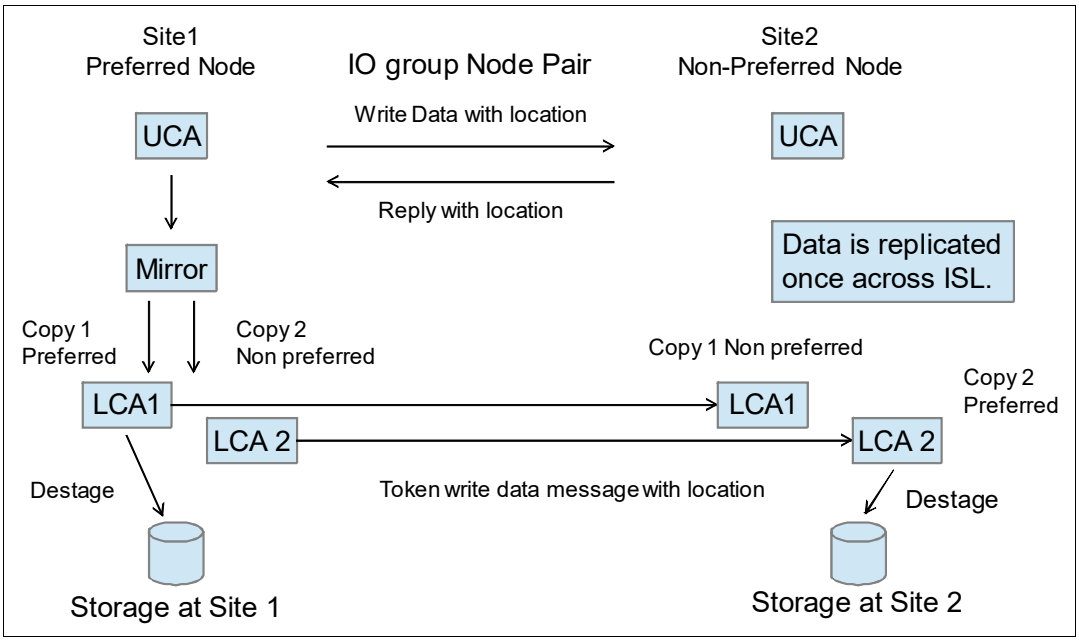


Figure 6-7 Design of an enhanced stretched cluster

6.2.7 Storage efficiency

When aiming for maximum efficiency of data storage with IBM Spectrum 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.

The 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, “Planning” on page 59, and Chapter 9, “Advanced features for storage efficiency” on page 559.

DRPs use multithreading and hardware acceleration (where available) to provide storage efficiency functions on IBM Spectrum Virtualize storage systems. When considering use of storage efficiency options, take into account that they increase the number of I/O operations the storage system must realize compared to access to 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: FCM drives include compression hardware, so it provides data set size reduction with no performance penalty.

For more information about the storage efficiency functions of IBM Spectrum Virtualize, see Chapter 5, “Using storage pools” on page 271, 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 either 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 virtual capacity larger than 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 of thin-provisioned volumes, it is recommended to monitor closely 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 sacrificial emergency space volume.

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 on the grain size:

- 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 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 shows the thin-provisioning concept.

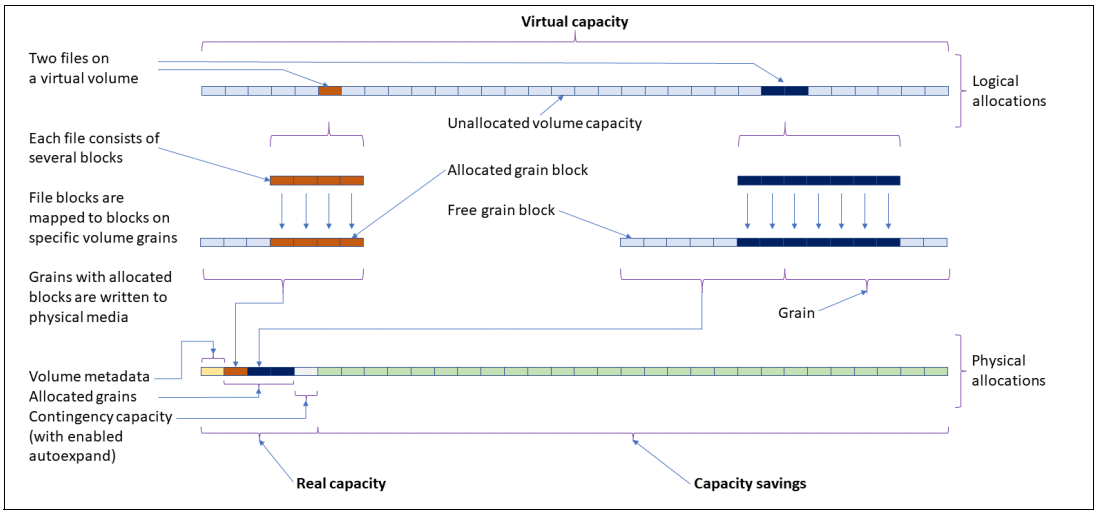


Figure 6-8 Conceptual diagram of a thin-provisioned volume

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 allocated by some 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 unallocated 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 Spectrum 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 requests that requires real capacity allocation.

To facilitate management of the auto expansion of thin-provisioned volumes, a capacity warning should 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 Spectrum Virtualize, deduplication can be enabled for thin-provisioned and compressed volumes that are created in DRPs.

IBM Spectrum 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 storage space that is required for storing the data. 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 (as are other features of IBM Spectrum 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:

- Code level V8.1.2 or higher is needed for DRPs.
- Code level V8.1.3 or higher is needed for deduplication.

Tip: Code level 8.3.1 is needed for best performance in DRP pools.

- 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 IBM Documentation, select your system, and read the "Planning for deduplicated volumes" section by expanding **Planning** → **Storage configuration planning**.

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 to enable compression on thin-provisioned volumes in most cases.

Notes: Consider the following points:

- When a volume is backed by FCM drives that compress data at line speed, the volume should be configured with compression that is switched on. IBM Spectrum 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 should 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 Spectrum Virtualize systems, which enables hosts to notify the storage controller of capacity that is no longer required and can be reused or de-allocated, which might improve capacity savings.

Note: For volumes that are outside DRPs, the complete stack from the operating system down to back-end storage controller must support UNMAP to enable the capacity reclamation. SCSI UNMAP is passed to specific back-end storage controllers only.

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 Spectrum Virtualize systems.

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 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:

- *Planning for deduplicated volumes*
- Intro*duction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430
- *IBM Spectrum Virtualize Data Reduction Best Practices* whitepaper

6.2.8 Encryption

IBM Spectrum Virtualize systems can be configured to enable data-at-rest encryption. This function is realized in hardware (self-encrypting drives or in serial-attached SCSI [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 be also realized by back-end storage that is virtualized by the IBM SAN Volume Controller. 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 Chapter 12, “Encryption” on page 793.

6.2.9 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 cache. This mode is the default cache mode for all volumes.

`readonly` Only read I/O operations that are performed on the volume are stored in cache. Writes to the volume are not cached.

`disabled` No I/O operations on the volume are stored in 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 (logical unit numbers [LUNs]) that are used as IBM Spectrum Virtualize image mode volumes. However, the use of IBM Spectrum 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 Spectrum Virtualize. Contact IBM Support before turning off the cache for volumes in your production environment to avoid performance degradation.

6.2.10 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 number of IOPS or bandwidth (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 using 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; rather, 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.

For more information about how to configure I/O throttle on a volume, see 6.5.4, “I/O throttling” on page 373.

6.2.11 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.

Two levels are available at which the volume protection must be enabled to be effective: pool level and system 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, which is new in V.8.3.1 (see 6.5.5, “Volume protection” on page 380), and CLI (see 6.6.9, “Volume protection” on page 428).

6.2.12 Secure data deletion

The system provides methods to securely erase data from a drive or from a 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.

The procedure is described in IBM Documentation.

6.3 Virtual volumes

IBM Spectrum Virtualize V7.6 introduced support for *virtual volumes*. These volumes enable support for VMware vSphere Virtual Volumes (VVOLs), which allow VMware vCenter to manage system objects, such as volumes and pools. The IBM Spectrum 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 Spectrum Virtualize, see *Configuring VMware Virtual Volumes for Systems Powered by IBM Spectrum Virtualize*, SG24-8328.

6.4 Volumes in multi-site topologies

IBM Spectrum 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 purpose of 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 spectrum), or buildings in different cities, located on different tectonic plates and powered from independent grids (the other end of the spectrum).

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 disaster recovery.
- 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 for using an enhanced stretched system rather than the HyperSwap topology might be the use of Global Mirror or Metro Mirror 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, review the Planning for high availability section of IBM Documentation (**SAN Volume Controller** → **Planning** → **Planning for high availability**) before choosing the topology.

Note: Multi-site topologies of IBM Spectrum Virtualize use two sites as storage component locations (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:

- With standard topology, the available options are: Basic, Mirrored, and Custom.
- With HyperSwap topology, the options are: Basic, HyperSwap, and Custom.
- Witch stretched topology, the options are: 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 multi-site 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.

6.4.1 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 by 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 itself to minimize data that is transmitted between sites and to minimize host read and write latency.

If the nodes or storage at either site go 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 two earlier technologies: The Non-Disruptive Volume Move (NDVM) function that was introduced in IBM Spectrum Virtualize V6.4, and the RC features that include MM, GM, and GMVC.

The HyperSwap volume configuration is possible only after the IBM Spectrum 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: It is still possible to create HyperSwap volumes in Version 7.5.

For more information, see *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317.

For more information about HyperSwap topology, see IBM Documentation.

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 6-9).

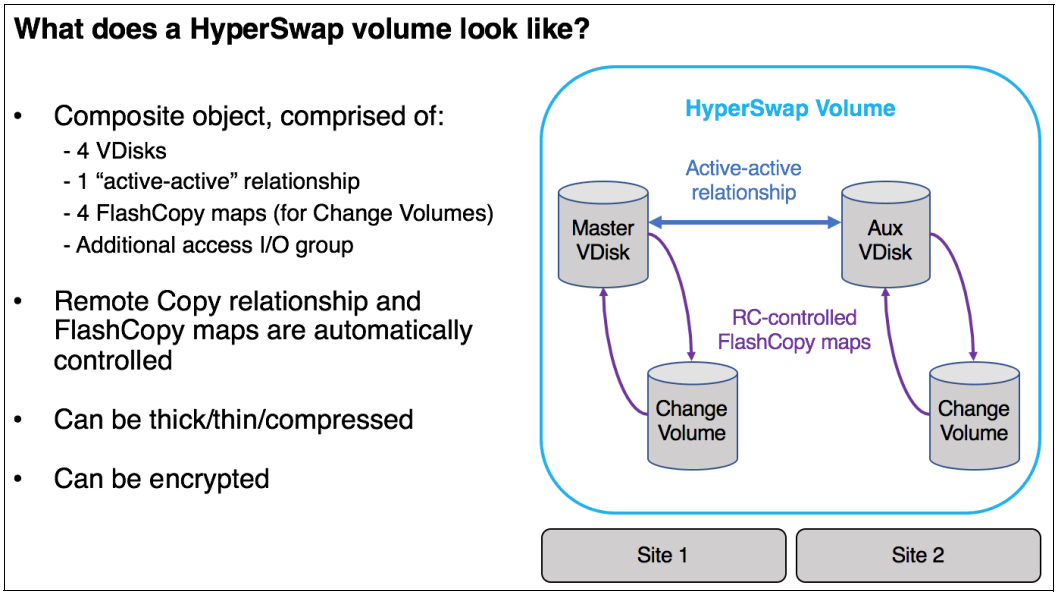


Figure 6-9 What makes up a HyperSwap volume

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 replications relationships.

An example of a HyperSwap volume configuration is shown in Figure 6-29 on page 370.

6.4.2 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 Metro Mirror or Global Mirror 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.

For more information about creating volumes in stretched topology, see “Stretched volumes” on page 372

6.5 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: With Version 7.4 and later, 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 6.6.9, “Volume protection” on page 428, and the “Enabling volume protection” topic in IBM Documentation.

6.5.1 Creating volumes

This section focuses on using the Create Volumes menu to create Basic and Mirrored volumes in a system with a standard topology. Volume creation is available with the following volume classes:

- Basic
- Mirrored
- Custom

To create a volume, complete the following steps:

1. Click the **Volumes** menu and then, click the **Volumes** option of the IBM Spectrum Virtualize GUI, as shown in Figure 6-10.

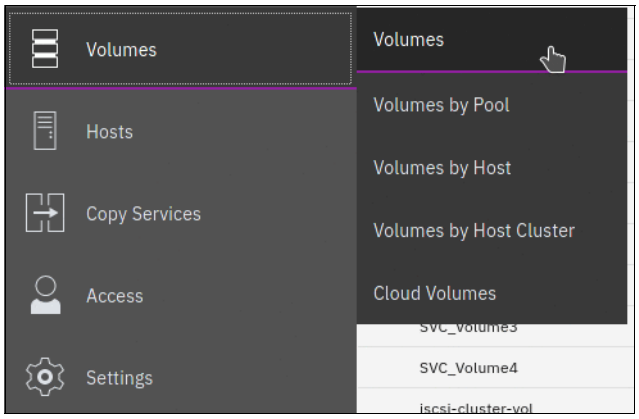


Figure 6-10 Volumes menu

A list of volumes, their state, capacity, and associated storage pools are displayed.

2. To create a volume, click **Create Volumes**, as shown in Figure 6-11.

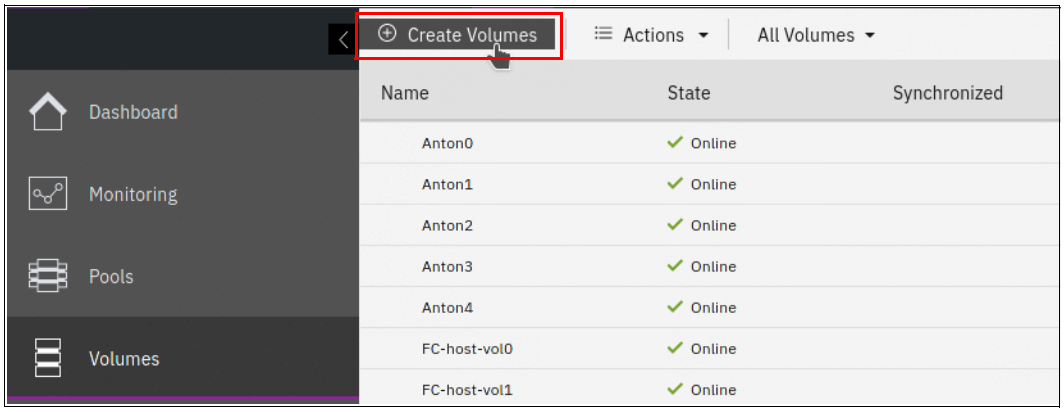


Figure 6-11 Create Volumes

The Create Volumes tab opens the Create Volumes window, which shows the available creation methods.

Note: The volume classes that are displayed in the Create Volumes window depend on the topology of the system.

The Create Volumes window for standard topology is shown in Figure 6-12.

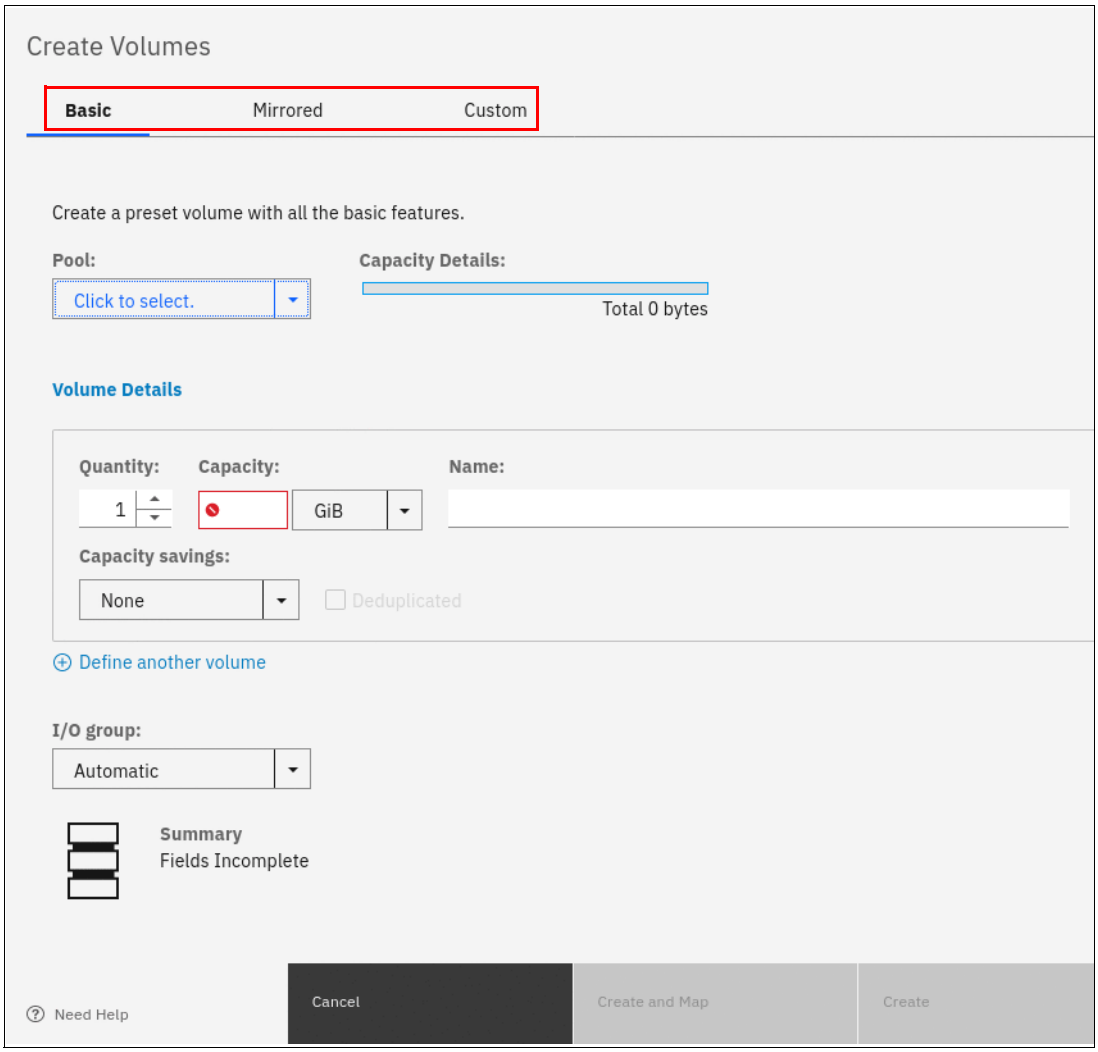


Figure 6-12 Basic, mirrored, and custom volume creation options

Note: Consider the following points:

- A *basic volume* is a volume that has only one physical copy, uses storage that is allocated from a single pool on one site, and uses read/write cache mode.
- A *mirrored volume* is a volume with two physical copies, where each volume copy can belong to a different storage pool.
- A *custom volume* (in the context of this menu) is a basic or mirrored volume with values of some of its parameters that are changed from the defaults.
- The Create Volumes window also provides (by using the Capacity Savings parameter) the ability to change the default provisioning of a basic or mirrored volume to thin-provisioned or compressed. For more information, see “Capacity savings option” on page 362.

Creating basic volumes

A *basic volume* is a volume that has only one physical copy. Basic volumes are supported in any system topology and are common to all configurations. Basic volumes can be of any type of virtualization: striped, sequential, or image. They can also use any type of capacity savings: thin-provisioning, compressed, or none.

Deduplication can be configured with thin-provisioned and compressed volumes in DRPs for added capacity savings.

To create a basic volume, click **Basic**, as shown in Figure 6-13 on page 358. This action opens the Basic volume menu, where you can define the following parameters:

- Pool: The pool in which the volume is created (drop-down menu).
- Quantity: Number of volumes to be created (numeric up or down).
- Capacity: Size of the volume in specified units (drop-down menu).
- Capacity Savings (drop-down menu):
  - None
  - Thin-provisioned
  - Compressed
- Name: Name of the volume (cannot start with a number).
- I/O group.

The Basic Volume creation window is shown in Figure 6-13.

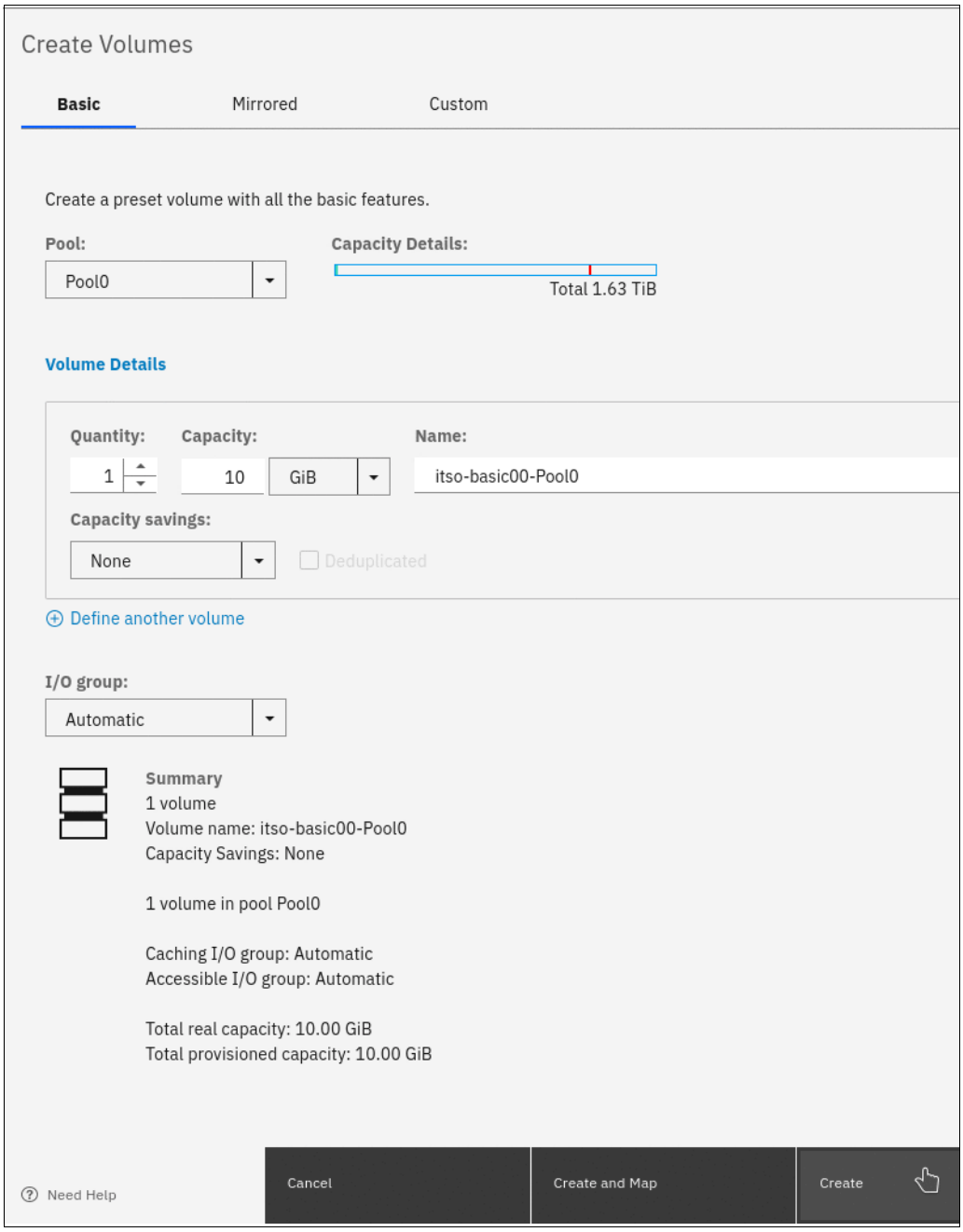


Figure 6-13 Create Volumes window

Define and consistently use a suitable volume naming convention to facilitate easy identification. For example, a volume name can contain the name of the pool or some tag that identifies the underlying storage subsystem, the host or cluster name that the volume is mapped to, and the content of this volume, such as the name of the applications that use the volume.

When all of the characteristics of the basic volume are defined, it can be created by selecting one of the following options:

- Create
- Create and Map

Note: The plus sign (**+**) icon can be used to create more volumes in the same instance of the volume creation wizard.

In the example, the Create option was selected. The volume-to-host mapping can be performed later, as described in 6.5.8, “Mapping a volume to a host” on page 395.

When the operation completes, the volume is seen in the Volumes pane in the state “Online (formatting)”, as shown in Figure 6-14.

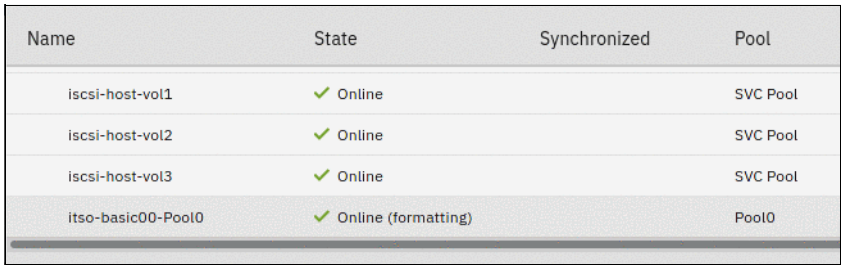


Figure 6-14 Basic volume formatting

By default, the GUI does not show any details about the commands it runs to complete a task. However, while a command runs, you can click **View more details** to see the underlying CLI commands that are run to create the volume and a report of completion of the operation, as shown in Figure 6-15.

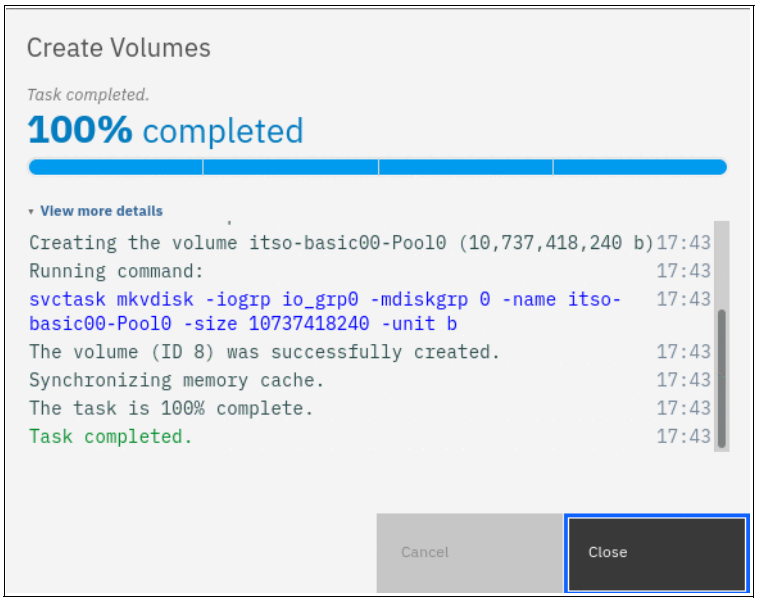


Figure 6-15 Viewing details of volume creation operation

Note: Consider the following points:

- Standard-provisioned volumes are automatically formatted through the quick initialization process after the volume is created. This process makes standard-provisioned volumes available for use immediately.
- Quick initialization requires a small amount of I/O to complete, and limits the number of volumes that can be initialized at the same time. Some volume actions, such as moving, expanding, shrinking, or adding a volume copy, are disabled when the specified volume is initializing. Those actions become available after the initialization process completes.
- The quick initialization process can be disabled in circumstances where it is not necessary. For example, if the volume is the target of a Copy Services function, the Copy Services operation formats the volume. The quick initialization process can also be disabled for performance testing so that the measurements of the raw system capabilities can take place without waiting for the process to complete.

For more information, see IBM Documentation and expand **Product overview** → **Technical overview** → **Volumes** → **Standard-provisioned volumes**.

Creating mirrored volumes

To create a mirrored volume, complete the following steps:

1. In the Create Volumes window, click **Mirrored**, and choose the **Pool** for **Copy1** and **Copy2** by using the drop-down menus. Although the mirrored volume can be created in the same pool, this setup is not typical. Generally, keep volumes copies on separate set of physical disks (pools).
2. Enter the following Volume Details:
   - Quantity
   - Capacity
   - Capacity savings
   - Name

Leave the I/O group option at its default setting of **Automatic** (see Figure 6-16).

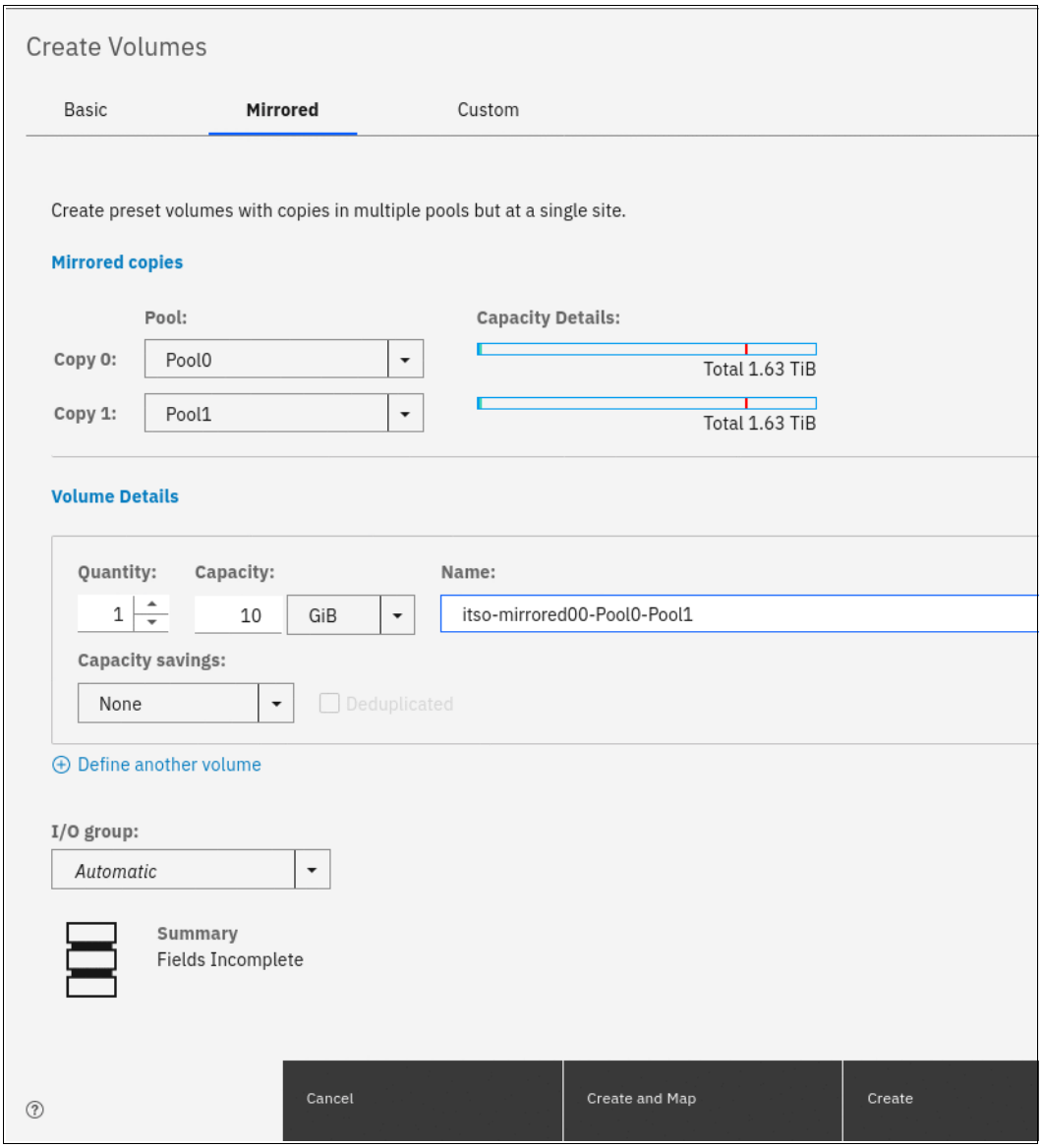


Figure 6-16 Mirrored volume creation

3. Click **Create** (or **Create and Map**).

   When the operation completes, the volume is seen in the Volumes pane in the state “Online (formatting)”, as shown in Figure 6-17.

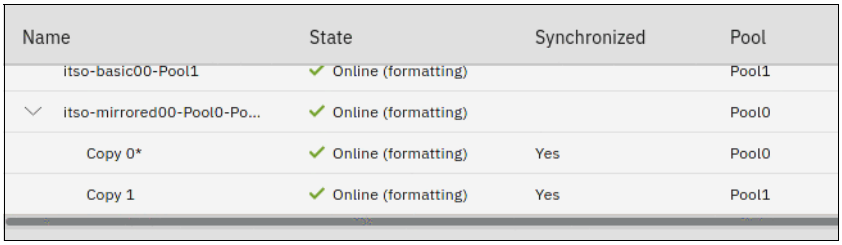


Figure 6-17 Mirrored volume formatting

A mirrored volume is displayed in the GUI as configured in the pool in which it has its primary. In this example, volume `itso-mirrored00-Pool0-Pool1` is displayed as configured in `Pool0` because it has its primary copy in `Pool0`.

Note: When creating a mirrored volume by using this menu, you are not required to specify the Mirrored Sync rate (it defaults to 2 MBps). The synchronization rate can be customized by using the Custom menu.

Capacity savings option

When the basic or mirrored method of volume creation is used, the GUI provides a Capacity Savings option, which enables altering the volume provisioning parameters without using the Custom volume provisioning method. You can select **Thin-provisioned** or **Compressed** from the drop-down menu, as shown in Figure 6-18, to create thin-provisioned or compressed volumes.

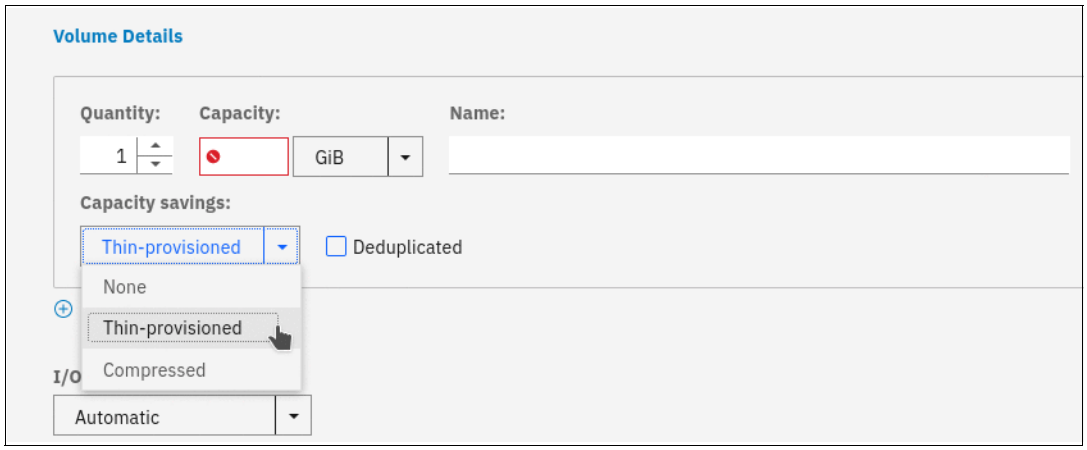


Figure 6-18 Volume creation with capacity saving option

Note: Consider the compression guidelines in Chapter 9, “Advanced features for storage efficiency” on page 559 before creating the first compressed volume copy on a system.

When a thin-provisioned or compressed volume is defined in a DRP, the Deduplicated option becomes available. Select this option to enable deduplication of the volume.

Thin-provisioned and compressed volumes have a special icon in the Capacity column of the **Volumes** menu that makes it easy to distinguish them, as shown in Figure 6-19.

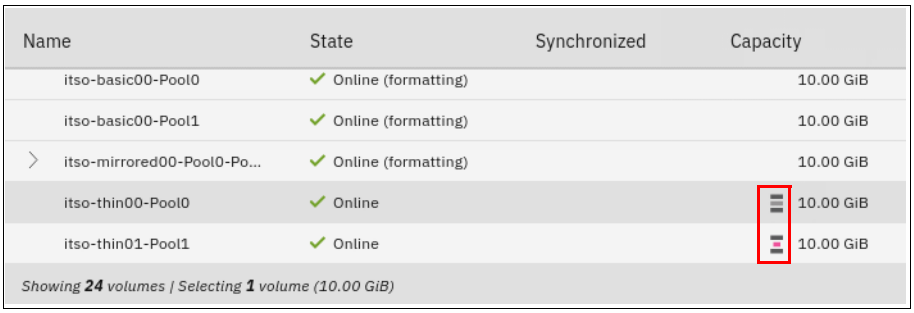


Figure 6-19 Space-efficient volumes icons

Volume `iso-thin00-Pool0` is thin-provisioned. Volume `iso-thin01-Pool1` is compressed. There is no icon indicating whether a volume is deduplicated or not.

6.5.2 Creating custom volumes

The Create Volumes window also enables Custom volume creation that expands the set of options for volume creation that are available to the administrator.

The Custom menu consists of several panes:

- Volume Location: Mandatory. It defines the number of volume copies, pools to be used, and I/O group preferences.
- Volume Details: Mandatory. It defines the Capacity savings option.
- Thin Provisioning: Enables configuration of thin-provisioning settings if this capacity saving option is selected.
- Compressed: Enables configuration of compression settings if this capacity saving option is selected.
- General: Configures cache mode and formatting.
- Summary.

Use these panes to customize your Custom volume as wanted, and then commit these changes by clicking **Create**.

You can mix and match settings on different panes to achieve the final volume configuration that meets your requirements.

Volume Location pane

The Volume Location pane is shown in Figure 6-20.

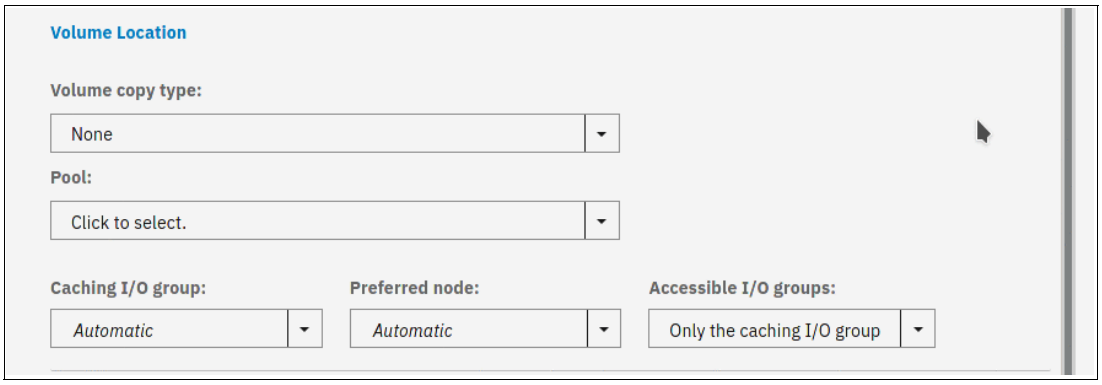


Figure 6-20 Volume Location pane

This pane has the following options:

- Volume copy type: You can choose between **None** (single volume copy) and **Mirrored** (two volume copies).
- Pool: Specifies storage pool to use for each of volume copies.
- Mirror sync rate: You can set the mirror sync rate for the volume copies. This option is displayed only for the Mirrored volume copy type, and you can set the volume copy synchronization rate to a value 128 KiBps - 64 MiBps.
- Caching I/O group: You can choose between **Automatic** (allocated by the system) and manually specifying the I/O group.
- Preferred node: You can choose between **Automatic** (allocated by the system) and manually specifying the preferred node for the volume.
- Accessible I/O groups: You can choose between **Only the caching I/O group** and **All**.

Volume Details pane

The Volume Details pane is shown in Figure 6-21.



Figure 6-21 Volume Details pane

This pane has the following options:

- Quantity: You can specify how many volumes to create.
- Capacity: Capacity of the volume.
- Name: You can define the volume name.

- Capacity savings: You can choose between **None** (standard-provisioned volume), **Thin-provisioned**, and **Compressed**.
- Deduplicated: Thin-provisioned and compressed volumes that are created in a DRP can be deduplicated.

If you click **Define another volume**, the GUI displays a subpane in which you can define the configuration of another volume, as shown in Figure 6-22.

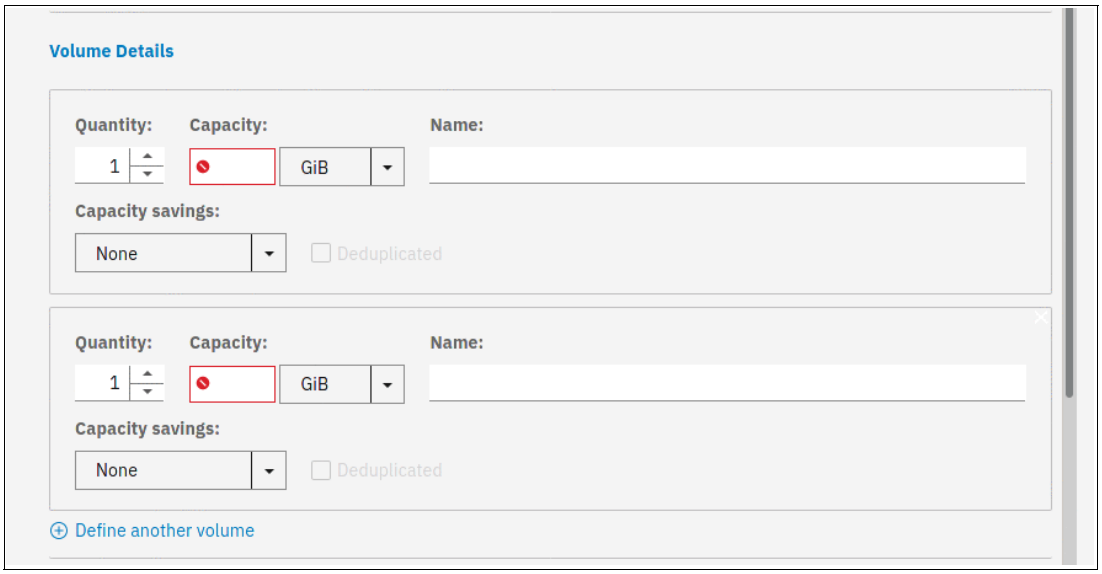


Figure 6-22 Volume Details pane with two volume subpanes

This way, you can create volumes with different characteristics in a single invocation of the volume creation wizard.

Thin Provisioning pane

If you choose to create a thin-provisioned volume, a Thin Provisioning pane is displayed, as shown in Figure 6-23.

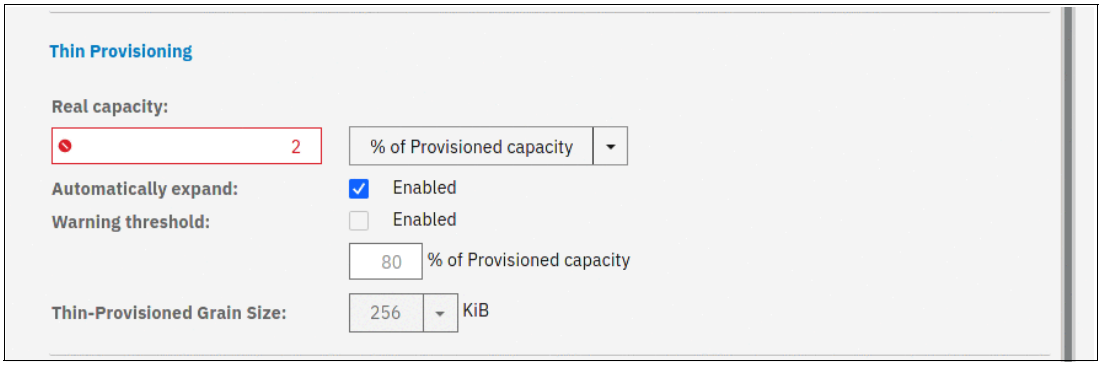


Figure 6-23 Thin Provisioning pane

This pane includes the following options:

- Real capacity: Real capacity of the volume, which is specified as a percentage of the virtual capacity or in bytes.
- Automatically expand: Whether to automatically expand the real capacity of the volume if needed. Defaults to Enabled.

- Warning threshold: Whether a warning message is sent and at what percentage of filled virtual capacity. Defaults to Enabled, with a warning threshold set at 80%.
- Thin-Provisioned Grain Size: You can define the grain size for the thin-provisioned volume. Defaults to 256 KiB.

> **Important:** If you do not use the **autoexpand** feature, the volume goes offline if it receives a write request after all real capacity is allocated.
>
> The default grain size is 256 KiB. The optimum choice of grain size depends on the volume use type. Consider the following points:
>
> - If you are *not* going to use the thin-provisioned volume as a FlashCopy source or target volume, use 256 KiB to maximize performance.
> - If you *are* going to use the thin-provisioned volume as a FlashCopy source or target volume, specify the same grain size for the volume and for the FlashCopy function.
> - If you plan to use Easy Tier with thin-provisioned volumes, see the IBM Support article *Performance Problem When Using Easy Tier With Thin Provisioned Volumes*.

Compressed pane

If you choose to create a compressed volume, a Compressed pane is displayed, as shown in Figure 6-24.

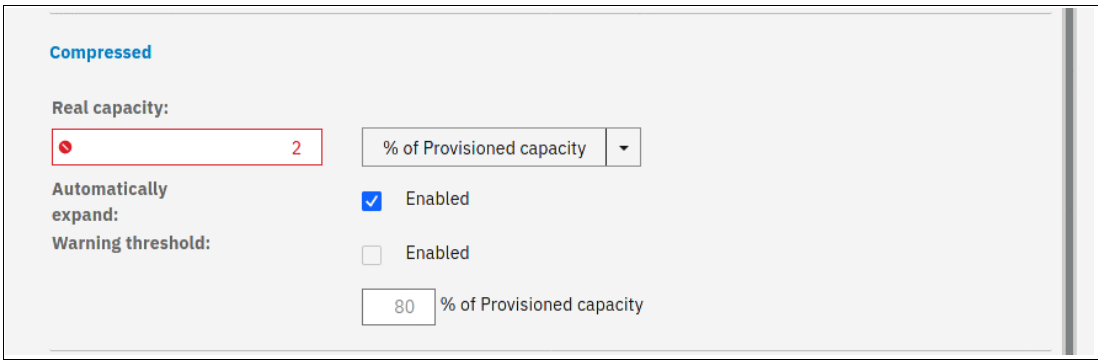


Figure 6-24 Compressed pane

This pane gives the following options:

- Real capacity: Real capacity of the volume, which is specified as percentage of the virtual capacity or in bytes.
- Automatically expand: Whether to automatically expand the real capacity of the volume if needed. Defaults to Enabled.
- Warning threshold: Whether a warning message is sent, and at what percentage of filled virtual capacity. Defaults to Enabled, with a warning threshold set at 80%.

You cannot specify the grain size for a compressed volume.

> **Note:** Consider the compression guidelines in Chapter 9, “Advanced features for storage efficiency” on page 559 before creating the first compressed volume copy on a system.

General pane

The General pane is shown in Figure 6-25.



Figure 6-25 General pane

This pane gives the following options:

- Cache mode: Controls volume caching. Defaults to Enabled. Other available options are Read-only and Disabled.
- OpenVMS unit device identifier (UDID): Each OpenVMS Fibre Channel (FC)-attached volume requires a user-defined identifier or UDID. A *UDID* is a nonnegative integer that is used in the creation of the OpenVMS device name.

6.5.3 Creating volumes in multi-site topologies

This section describes creation of volumes in multi-site topologies.

Note: For volumes in multi-site topologies the asterisk (*) does not indicate the primary copy, but the local volume copy which is used for data reads.

HyperSwap volumes

To create a HyperSwap volume, complete the following steps:

1. In the IBM Spectrum Virtualize GUI, select **Volumes** → **Volumes**, as shown in Figure 6-26.

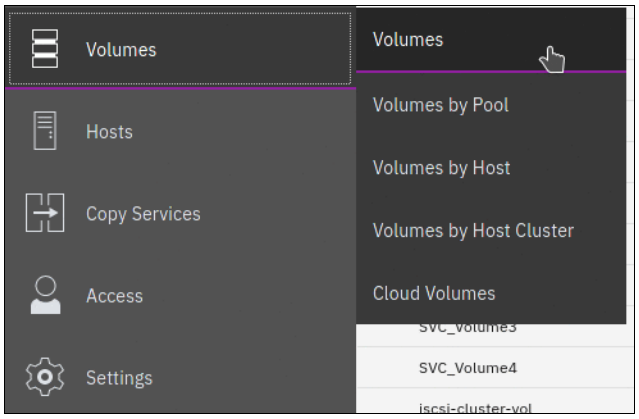


Figure 6-26 Volumes menu

A list of volumes, their state, capacity, and associated storage pools, is displayed.

2. Click **Create Volumes**, as shown in Figure 6-27 on page 368.

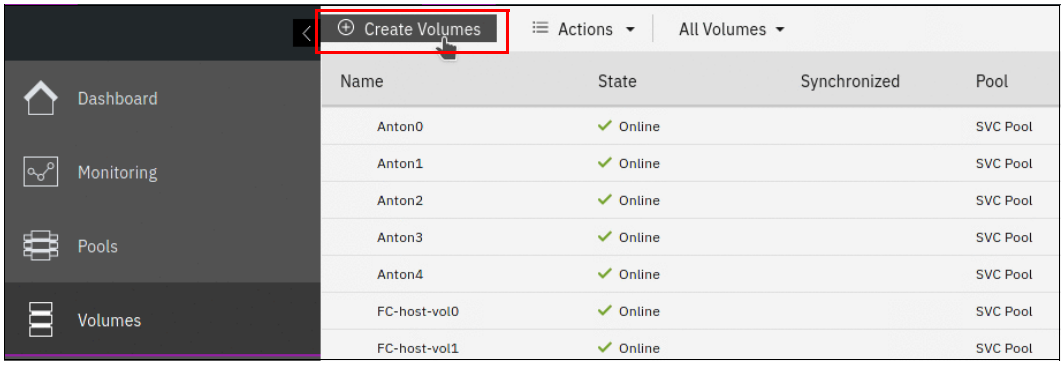


Figure 6-27 Create Volumes button

The Create Volumes tab opens the Create Volumes window, which displays available creation methods.

Note: The volume classes that are displayed in the Create Volumes window depend on the topology of the system.

The Create Volumes window for the HyperSwap topology is shown in Figure 6-28.

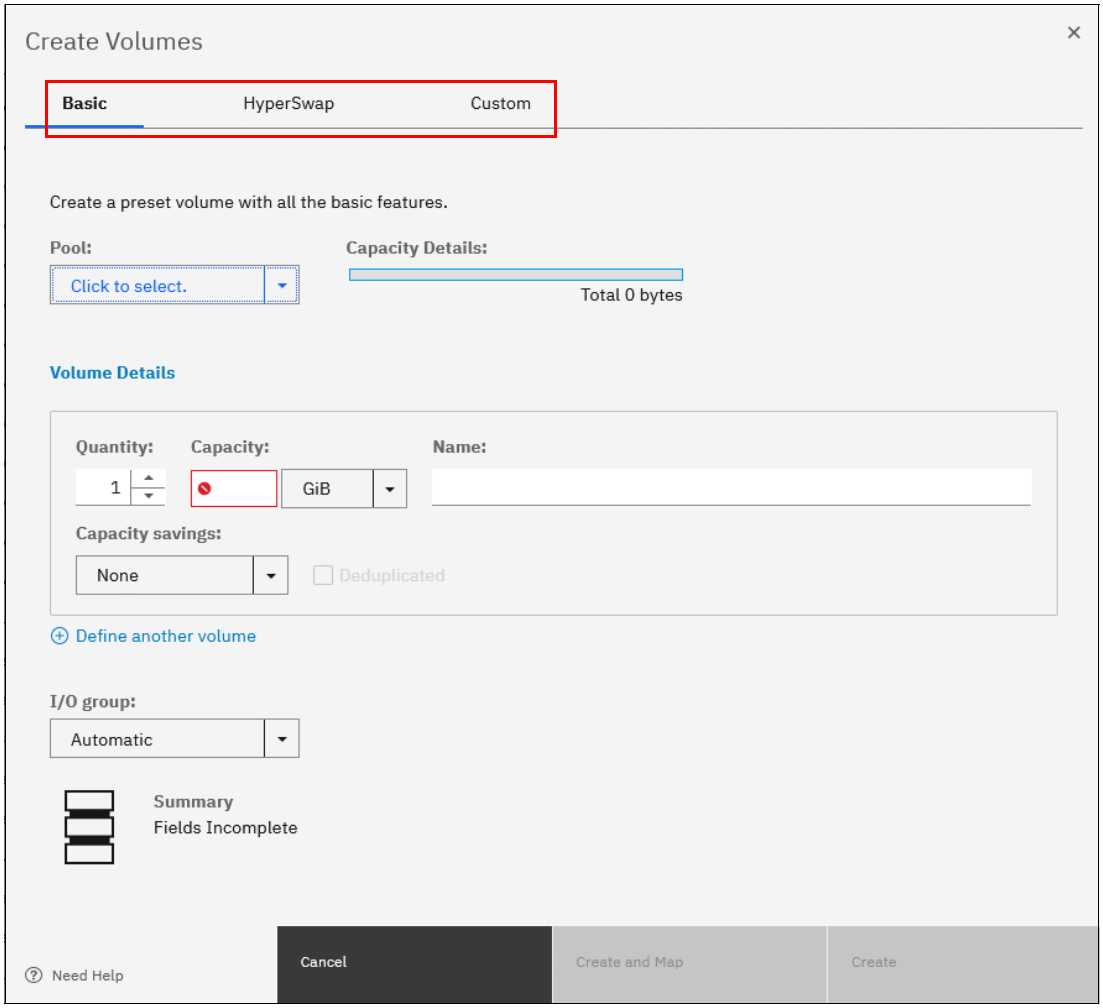


Figure 6-28 Basic, HyperSwap, and Custom volume creation options

Note: Consider the following points:

- A *basic volume* is a volume that has only one physical copy, uses storage that is allocated from a single pool on one site, and uses read/write cache mode.
- A *mirrored volume* is a volume with two physical copies, where each volume copy can belong to a different storage pool.
- A *HyperSwap volume* is a volume with two physical copies, where each volume copy is configured on a storage located in a different location. This option is available for HyperSwap system topology only.
- A Stretched volume is a volume with two physical copies, where each volume copy is configured on a storage located in a different location. This option is available for Stretched system topology only.
- A *custom volume* (in the context of this menu) is a basic or mirrored volume with values of some of its parameters that are changed from the defaults.
- The Create Volumes window also provides (by using the Capacity Savings parameter) the ability to change the default provisioning of a basic or mirrored volume to thin-provisioned or compressed. For more information, see “Capacity savings option” on page 362.

The notable difference between HyperSwap volume and basic volume creation is that HyperSwap volume creation includes specifying storage pool names at each site. The system uses its topology awareness to map storage pools to sites, which ensures that the data is correctly mirrored across locations.

The difference between HyperSwap and Stretched volume is more subtle. Enhanced stretched cluster divides two nodes from the same I/O group on two sites, while HyperSwap separates two I/O groups on the same cluster. Therefore, if one site becomes available, a system in HyperSwap topology retains a full I/O group that can cache I/O requests.

As shown in Figure 6-29, a single volume is created with volume copies in sites site1 and site2. This volume is in an active-active (MM) relationship with extra resilience that is provided by two change volumes.

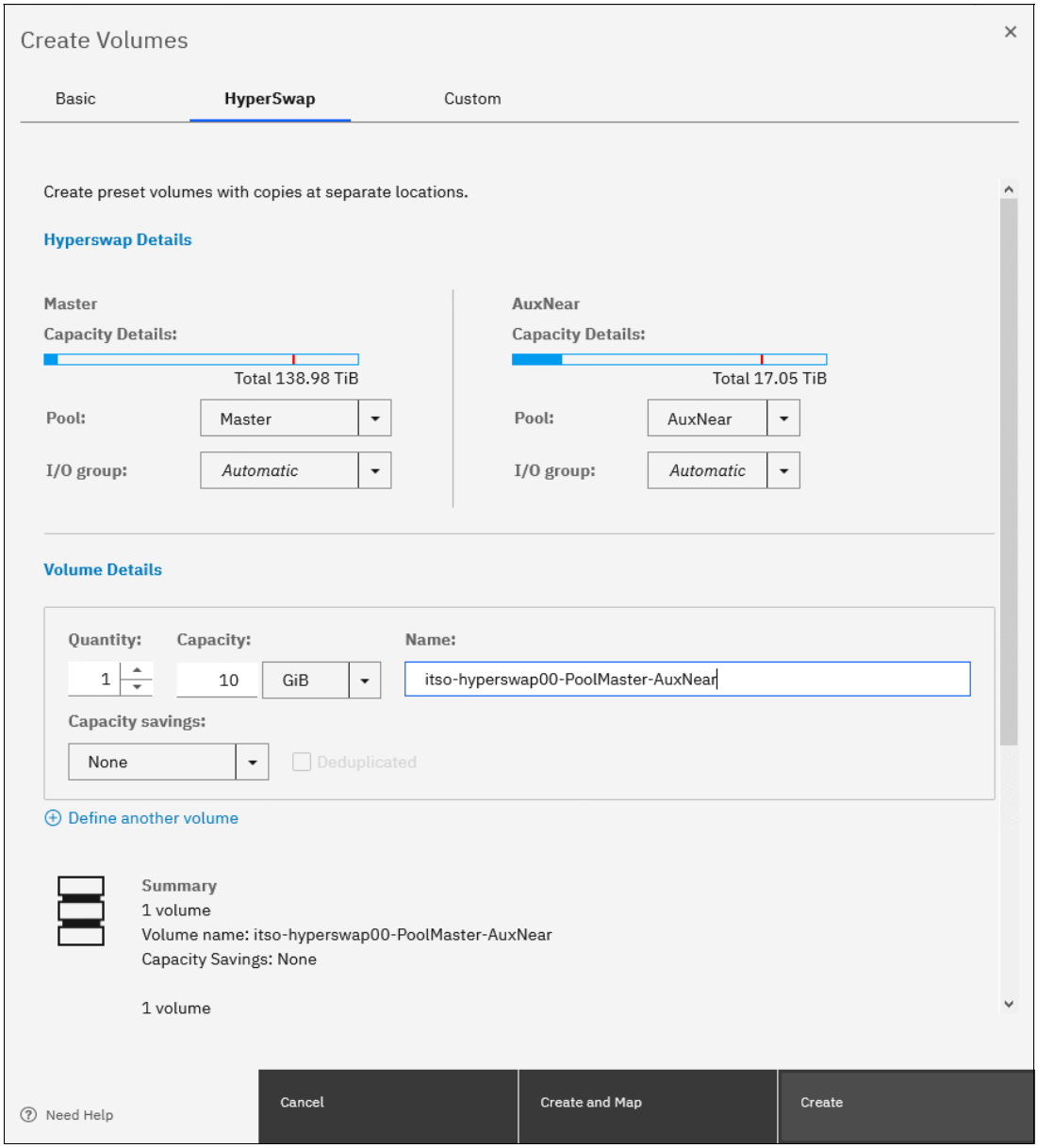


Figure 6-29 HyperSwap volume creation window

After the volume is created, it is visible in the volumes list, as shown in Figure 6-30.

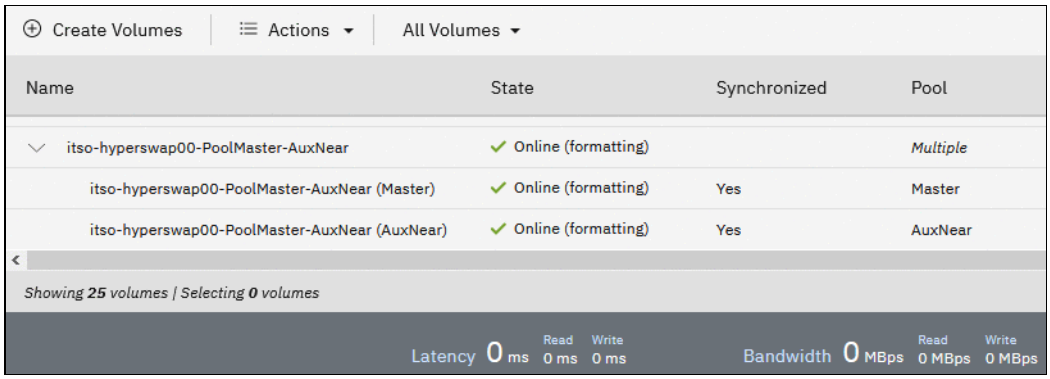


Figure 6-30 A HyperSwap volume in the list of volumes

The Pool column shows the value “Multiple”, which indicates that a volume is a HyperSwap volume. A volume copy at each site is visible, and the change volumes that are used by the technology are not displayed in this GUI view.

Note: For volumes in multi-site topologies, the asterisk (*) does not indicate the primary copy, but the local volume copy that is used for data reads.

A single **mkvolume** command can create a HyperSwap volume. Up to IBM Spectrum Virtualize V7.5, this process required careful planning and running the following sequence of commands:

```
1. mkvdisk master_vdisk
2. mkvdisk aux_vdisk
3. mkvdisk master_change_volume
4. mkvdisk aux_change_volume
5. mkrcrelationship –activeactive
6. chrcrelationship -masterchange
7. chrcrelationship -auxchange
8. addvdiskacces
```

Note: Version 8.4 extends HyperSwap support to hosts attached by way of NVMe over Fabrics (NVMe-oF), using Fibre Channel medium. Standard protocol mechanism – Asymmetric Namespace Access (ANA), which is analogous to SCSI Asymmetric Logical Unit Access (ALUA) – is used to provide this functionality to hosts attached by way of NVMe-oF.

Stretched volumes

If the IBM Spectrum Virtualize system is set up in stretched topology, the Create Volumes menu shows the Stretched option. Click **Stretched** to create a stretched volume and define its attributes, as shown in Figure 6-31.

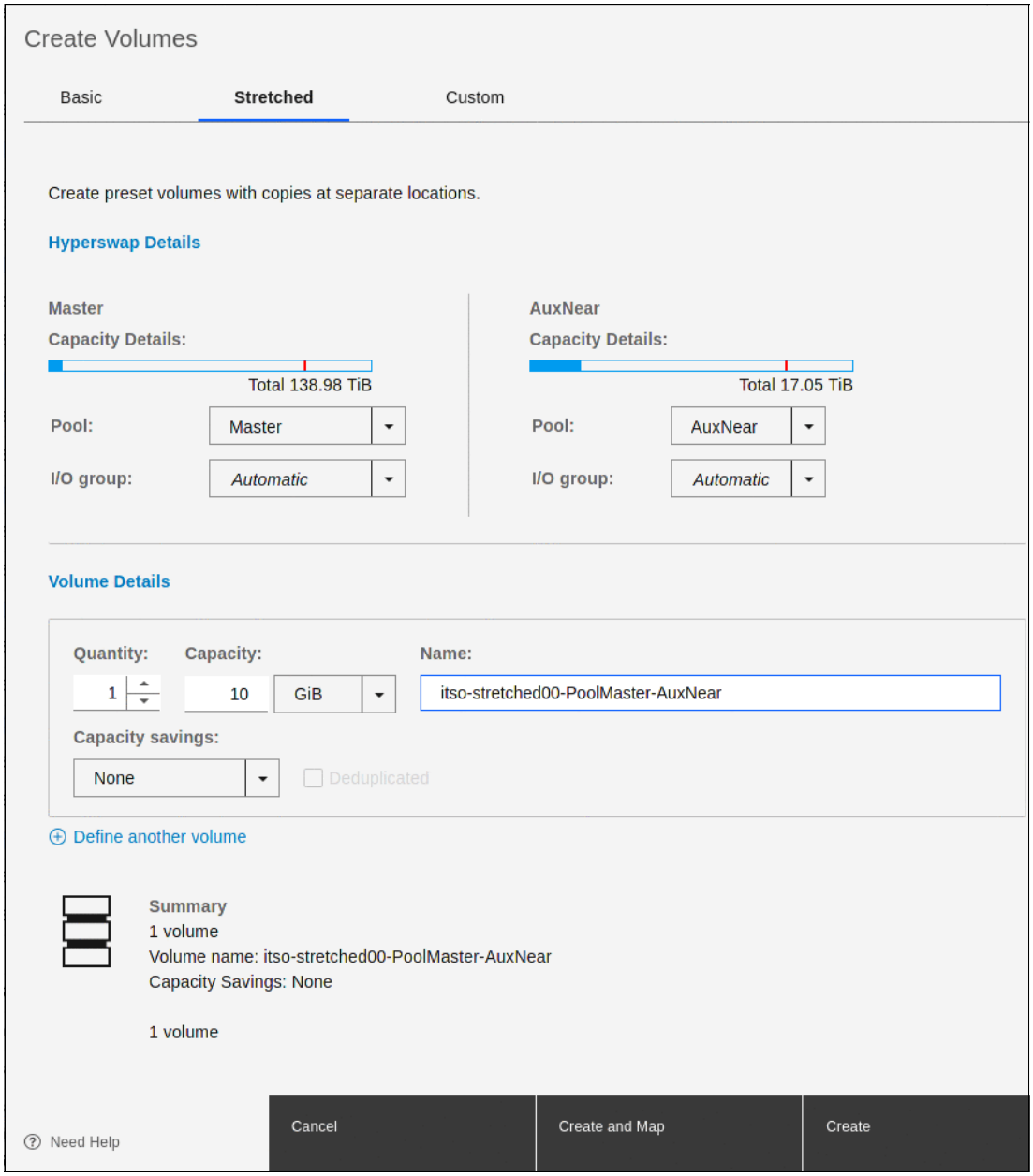


Figure 6-31 Creating a stretched volume

Note: The IBM Documentation page on Stretched topology is available at **SAN Volume Controller** → **Product overview** → **Technical Overview** → **Volumes** → **Stretched volume**s.

After you click **Create**, a volume is created and GUI shows the Volumes view. The newly created volume is visible as shown in Figure 6-32.

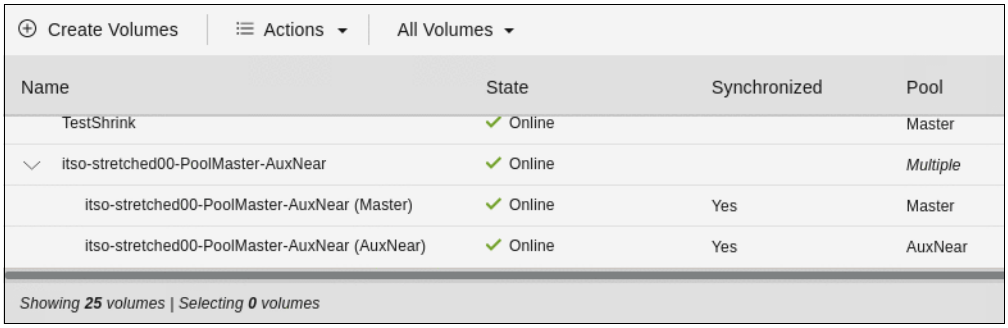


Figure 6-32 Stretched volume formatting after creation.

6.5.4 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 select the volume that you want to throttle. Select **Actions** → **Edit Throttle**, as shown in Figure 6-33.

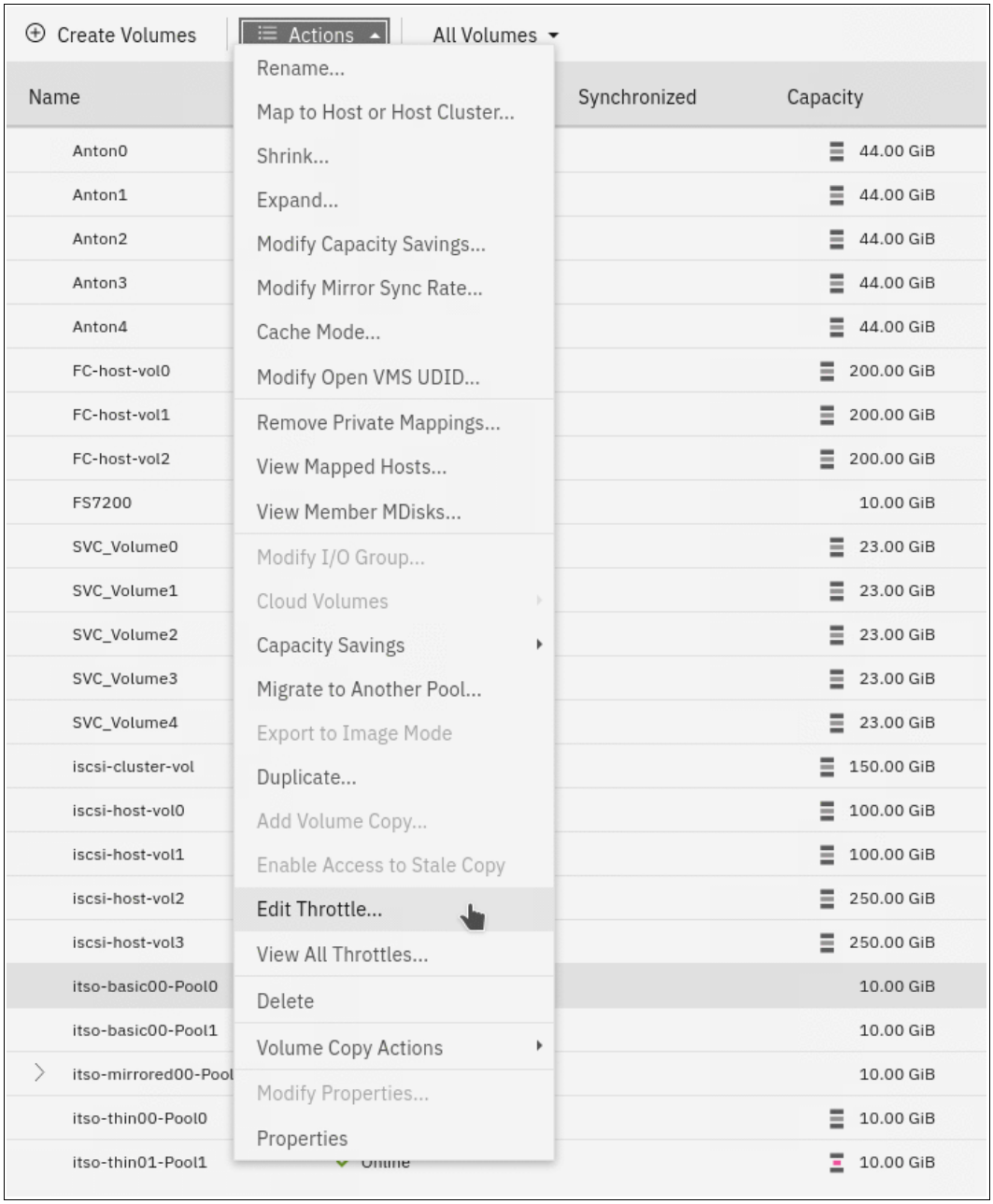


Figure 6-33 Edit Throttle menu item

2. In the Edit Throttle window, define the throttle in terms of number of IOPS or bandwidth. In our example, we set an IOPS throttle of 10,000, as shown in Figure 6-34. Click **Create**.

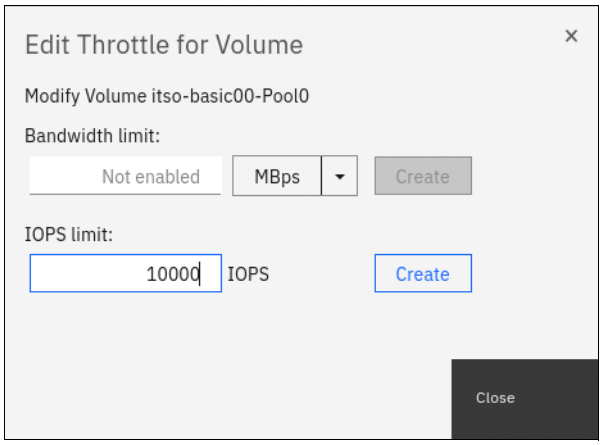


Figure 6-34 IOPS throttle on a volume

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 6-35.

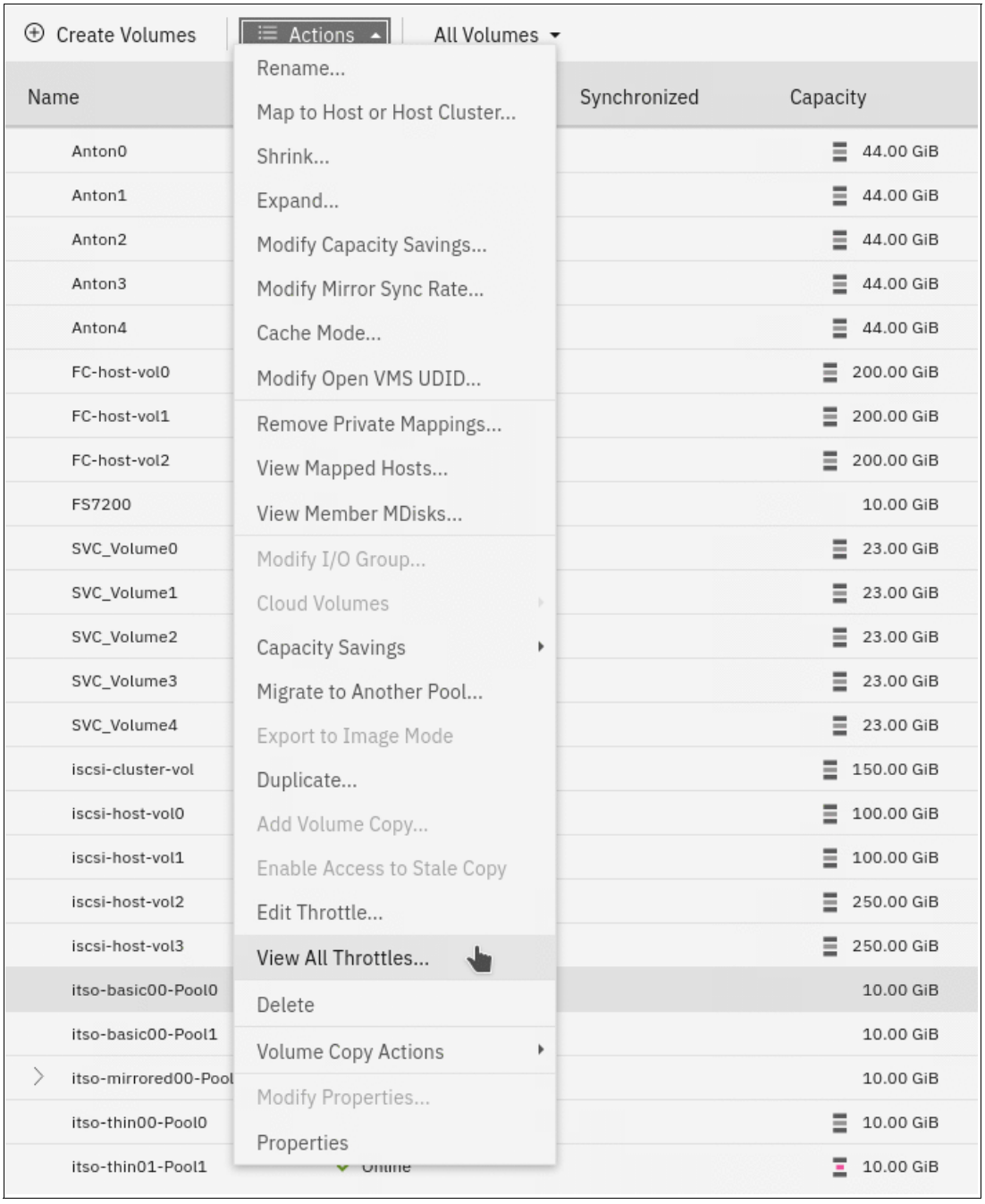


Figure 6-35 View All Throttles menu item

The **View All Throttles** menu shows all volume throttles that are defined in the system, as shown in Figure 6-36.

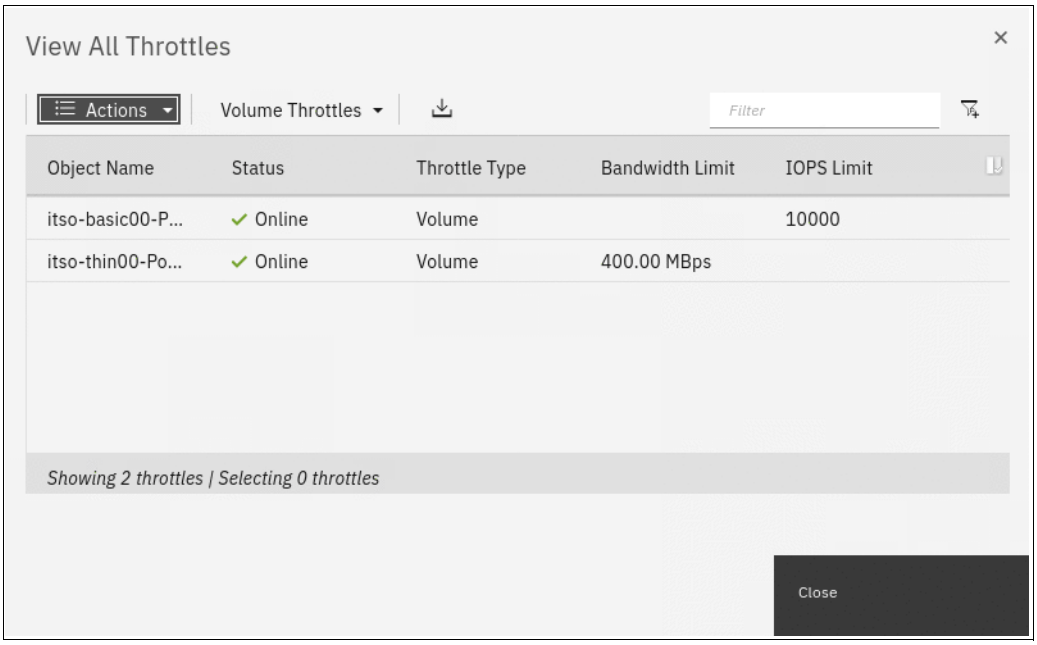


Figure 6-36 View All Throttles window

You can view other throttles by selecting a different throttle type in the drop-down menu, as shown in Figure 6-37.

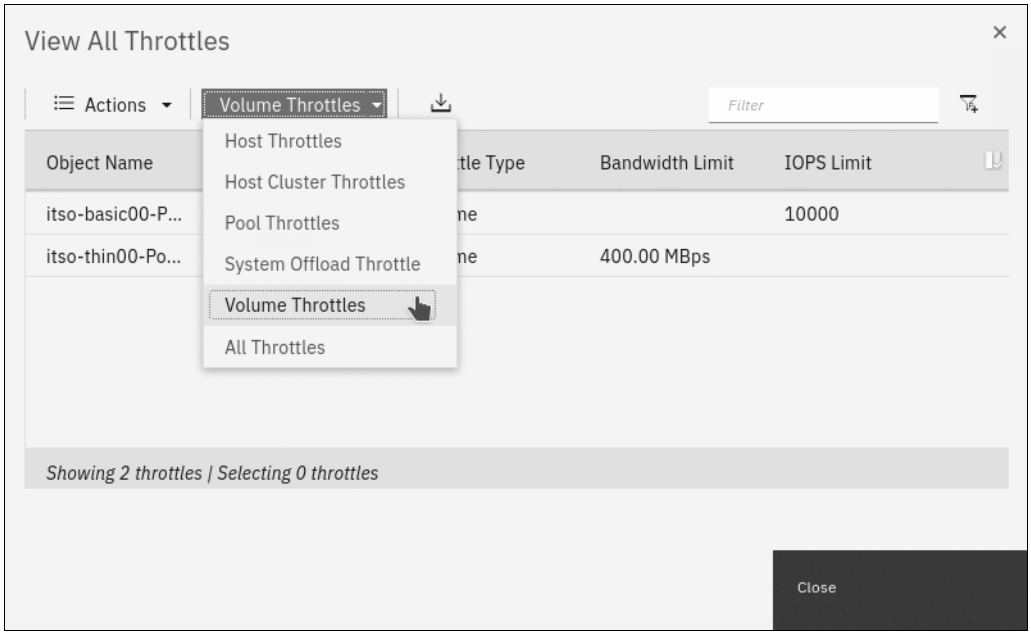


Figure 6-37 Filtering the throttle type

Modifying or removing a volume throttle

To remove a volume throttle, complete the following steps:

1. From the **Volumes** menu, select the volume that is attached the throttle that you want to remove. Select **Actions** → **Edit Throttle**, as shown in Figure 6-38.

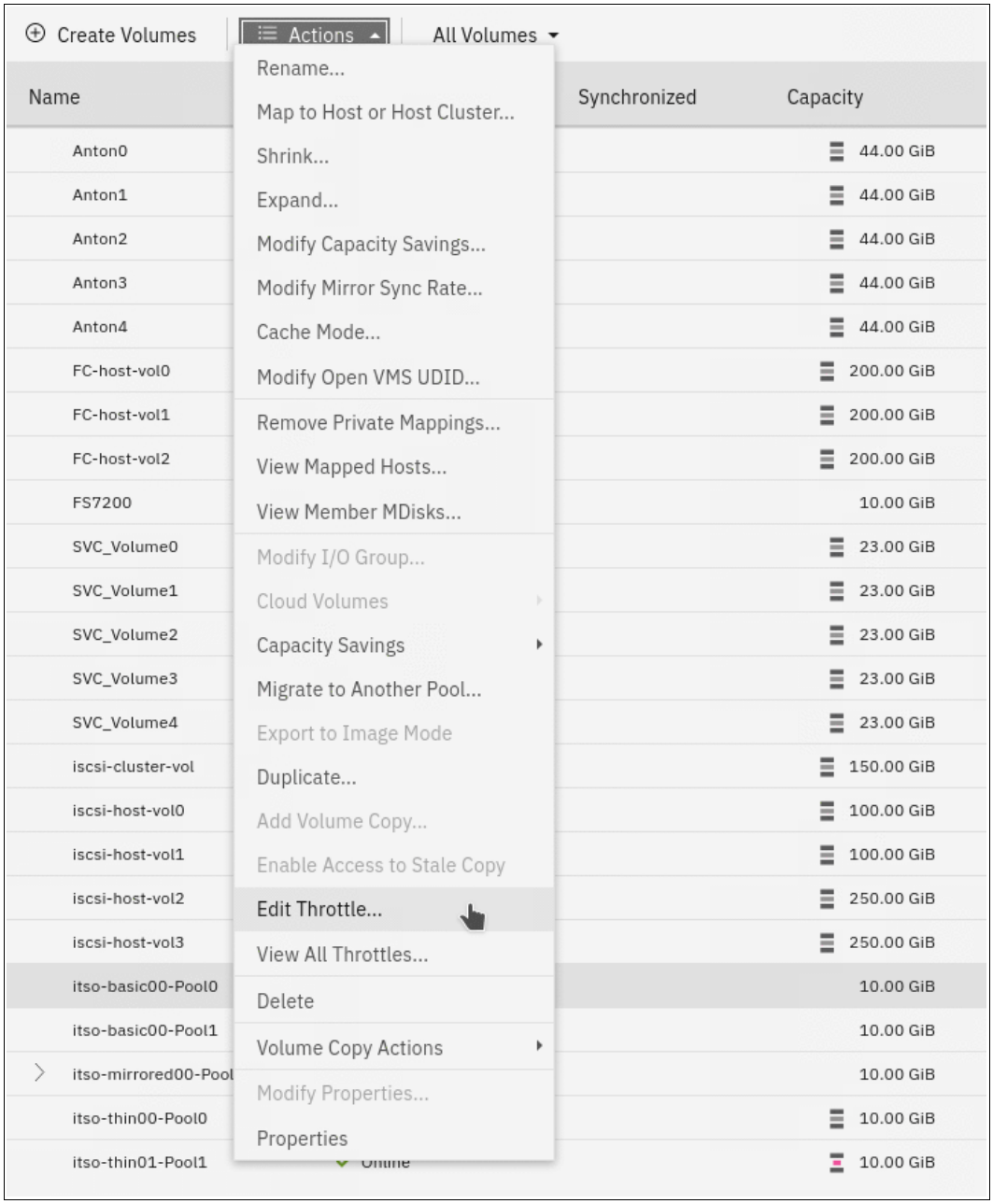


Figure 6-38 Edit Throttle menu

2. To modify the volume throttle input new throttling parameters and click **Save** as shown in Figure 6-39. In this example, the I/O throttling limit is increased to 15000 IOps.

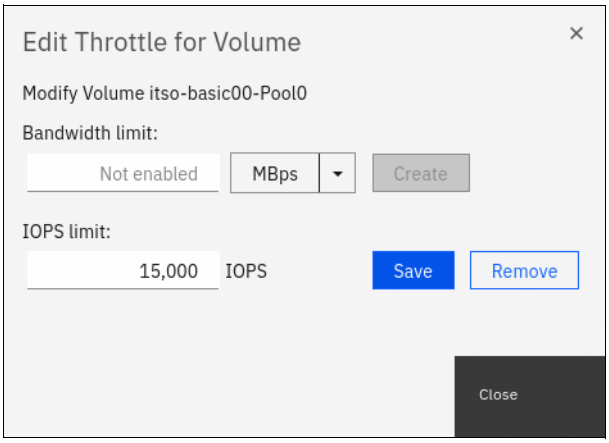


Figure 6-39 Modifying a volume throttle

To remove the throttle completely, click **Remove** for the throttle that you want to remove, as shown in Figure 6-40.

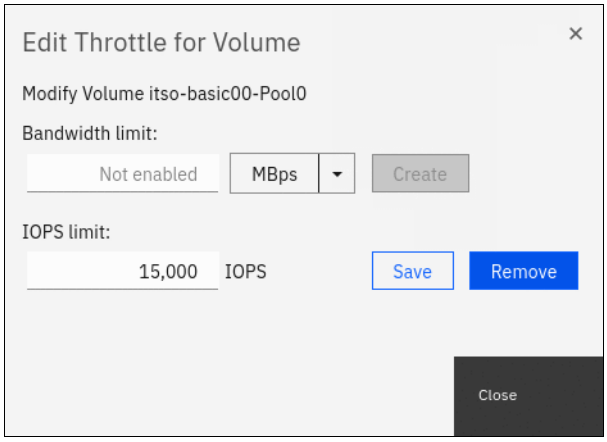


Figure 6-40 Removing a volume throttle

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 any action by clicking **Close**.

6.5.5 Volume protection

To configure volume protection, select **Settings** → **System** → **Volume Protection**, as shown in Figure 6-41.

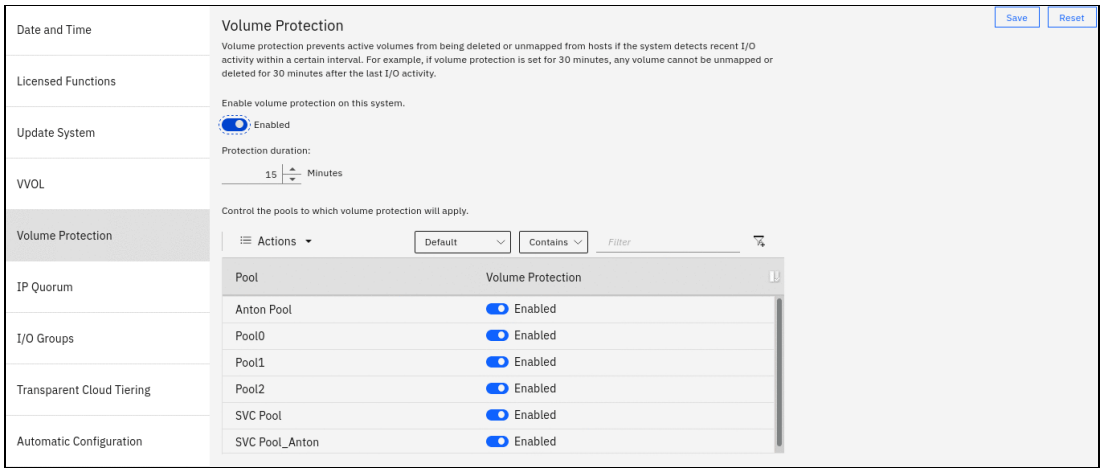


Figure 6-41 Volume Protection configuration

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.

6.5.6 Modifying a volume

After a volume is created, it is possible to modify many of its characteristics. The following sections describe 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 6-42.

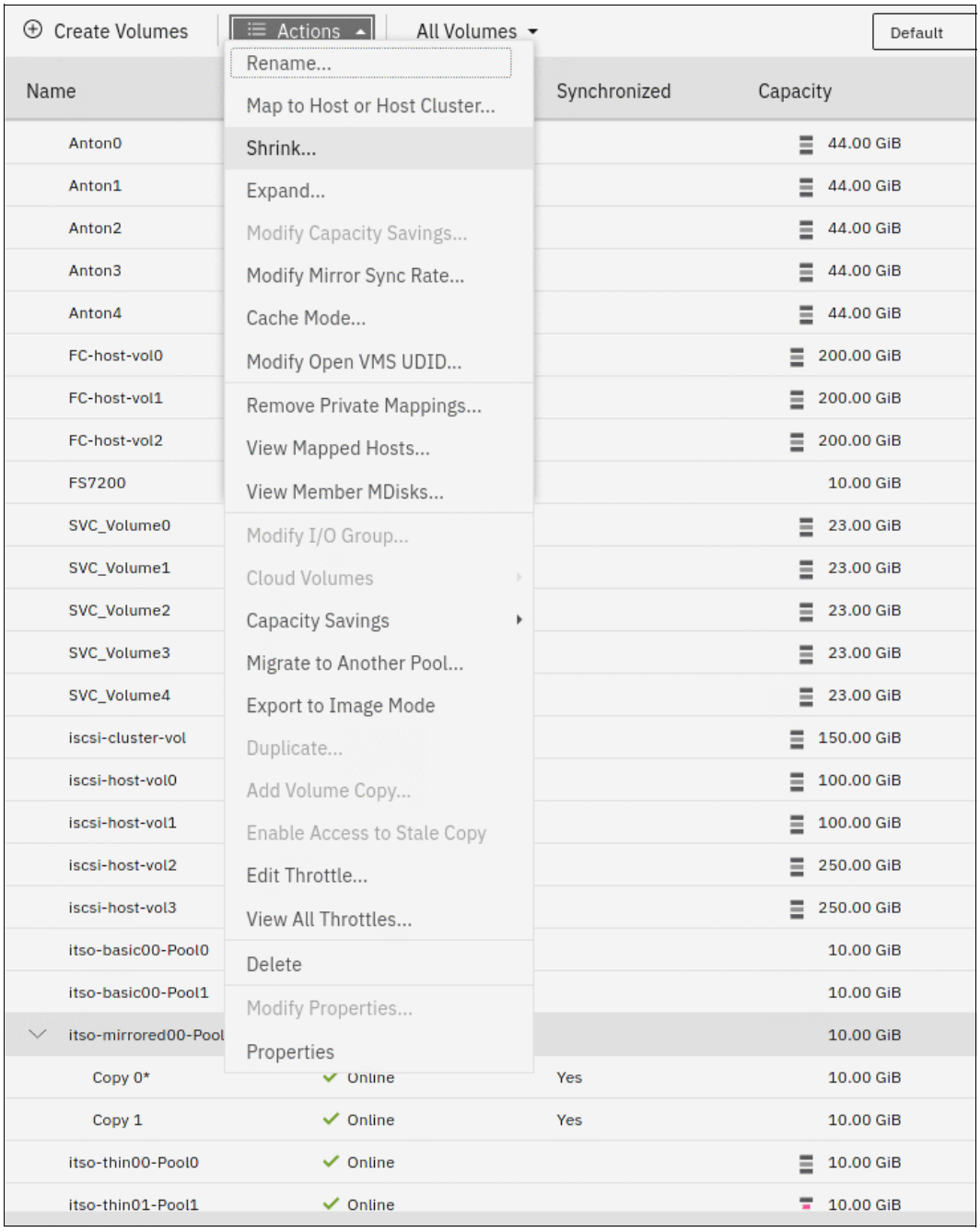


Figure 6-42 Volume Shrink menu item

3. Specify **Shrink by** or **Final size** (the other choice is calculated automatically), as shown in Figure 6-43.

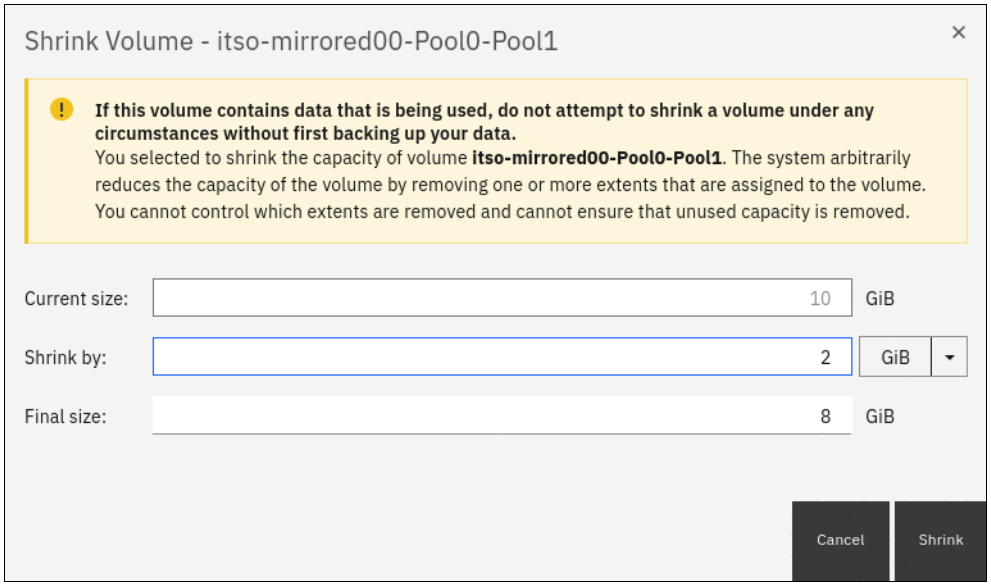


Figure 6-43 Specifying the size of the shrunk volume

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 **OK** to confirm the action, as shown in Figure 6-44.

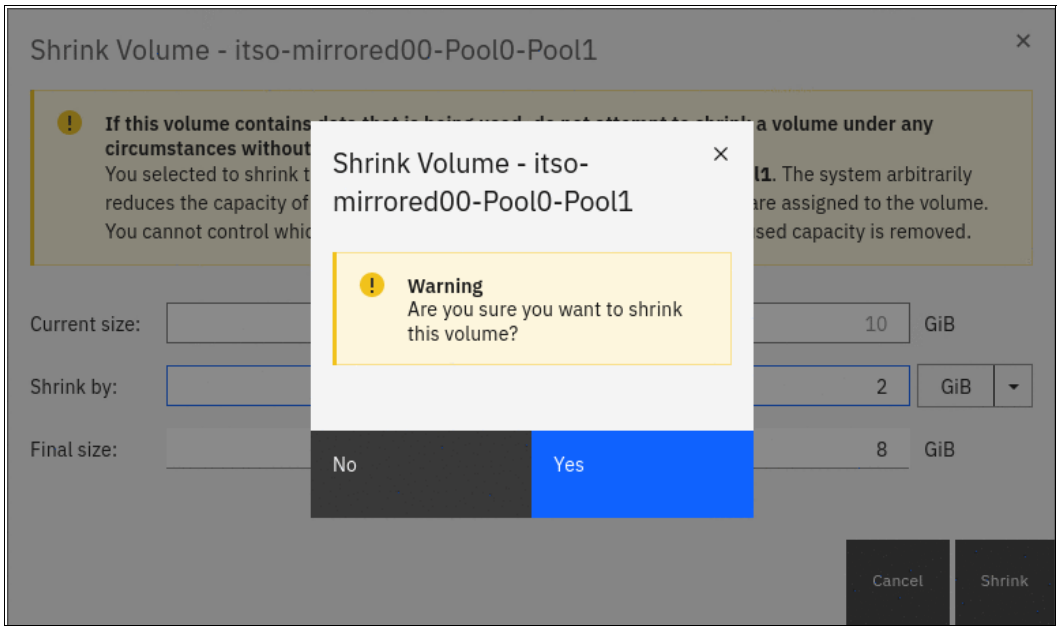


Figure 6-44 Confirming the volume shrinking operation

5. After the operation completes, you can see the volume with the new size by selecting **Volumes** → **Volumes**, as shown in Figure 6-45.

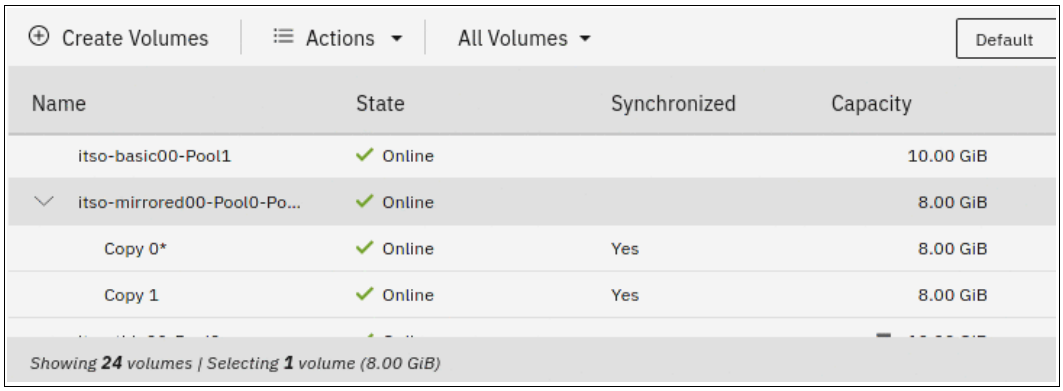


Figure 6-45 Shrunk volume size

Note: Version 8.4 introduces the ability to shrink a volume while it is formatting.

Expanding

To expand a volume, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to expand. Select **Actions** → **Expand…**, as shown in Figure 6-46 on page 384.

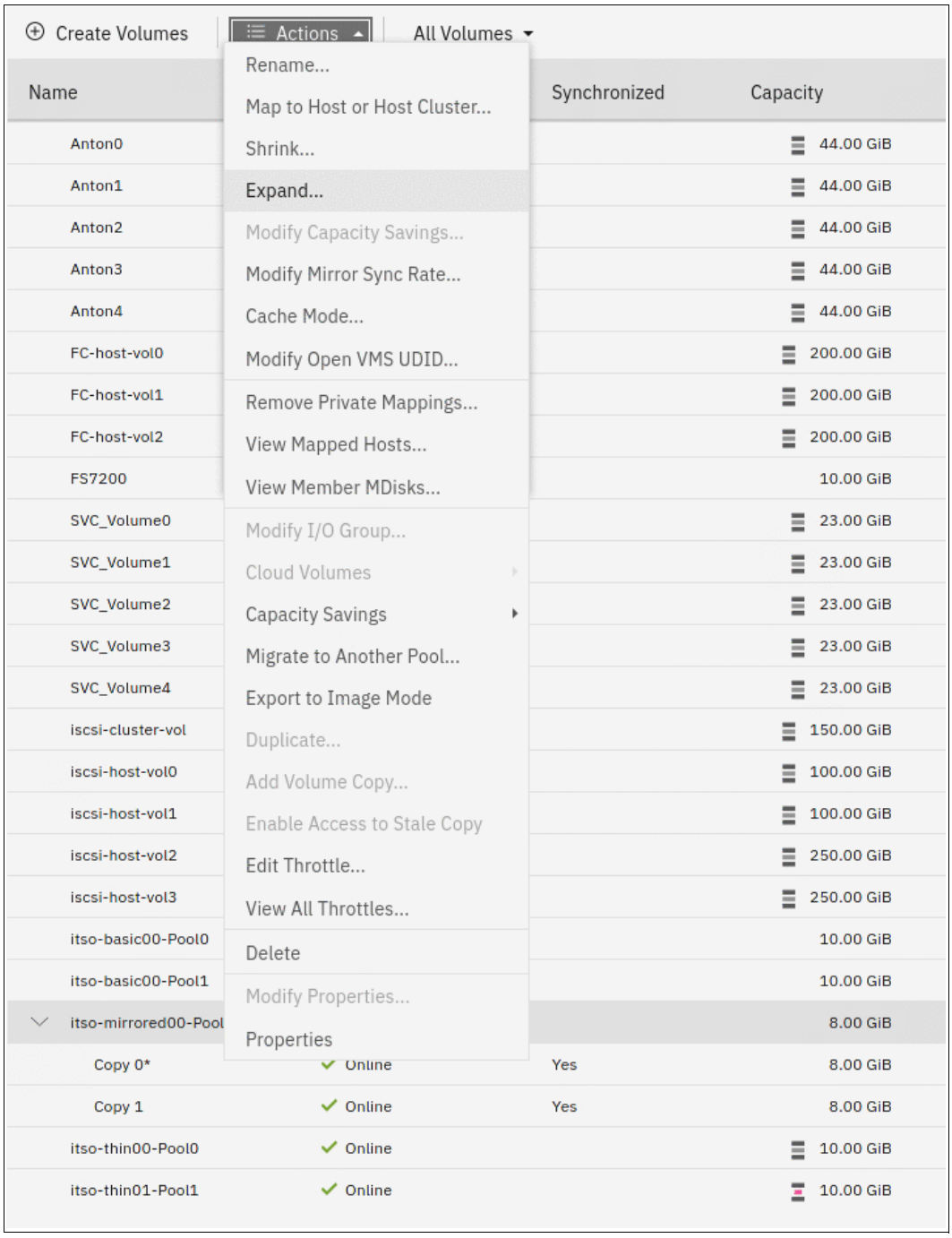


Figure 6-46 Volume Expand menu item

2. Specify either **Expand by:** or **Final size:** (the other value is calculated automatically) and click **Expand**, as shown in Figure 6-47.

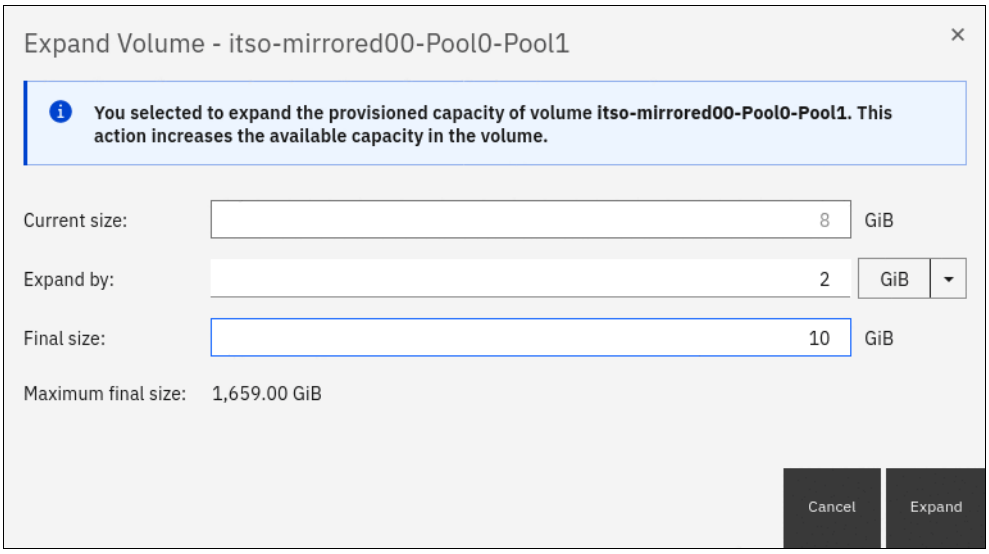


Figure 6-47 Specifying the expanded volume size

3. After the operation completes (including the formatting of the extra space), you can see the volume with the new size by selecting **Volumes** → **Volumes**, as shown in Figure 6-48.

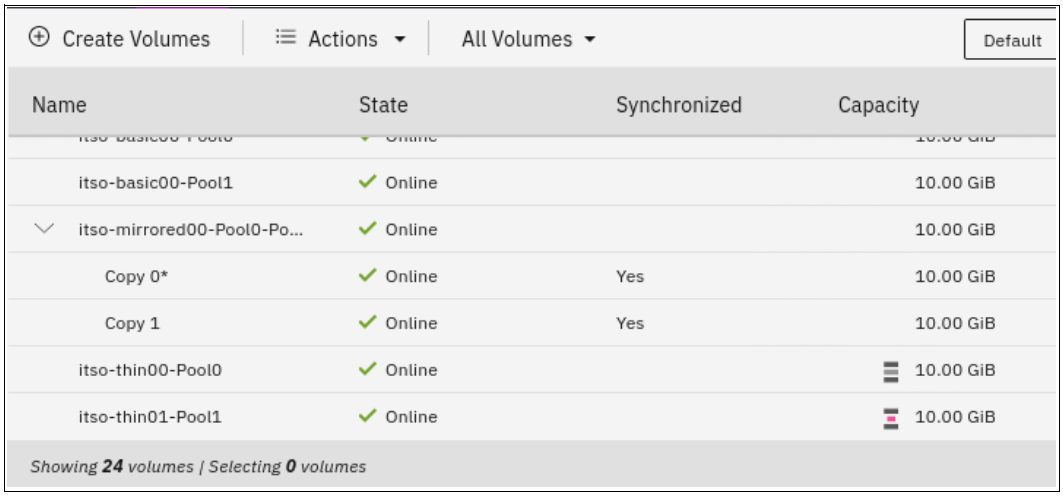


Figure 6-48 Expanded volume size

Notes: Consider the following points:

- 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; for example, through 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.
- Spectrum Virtualize Version 8.4 introduced the ability to expand a volume while it is formatting.

Modifying capacity savings

This action is available only for space-efficient volumes. 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 6-49.

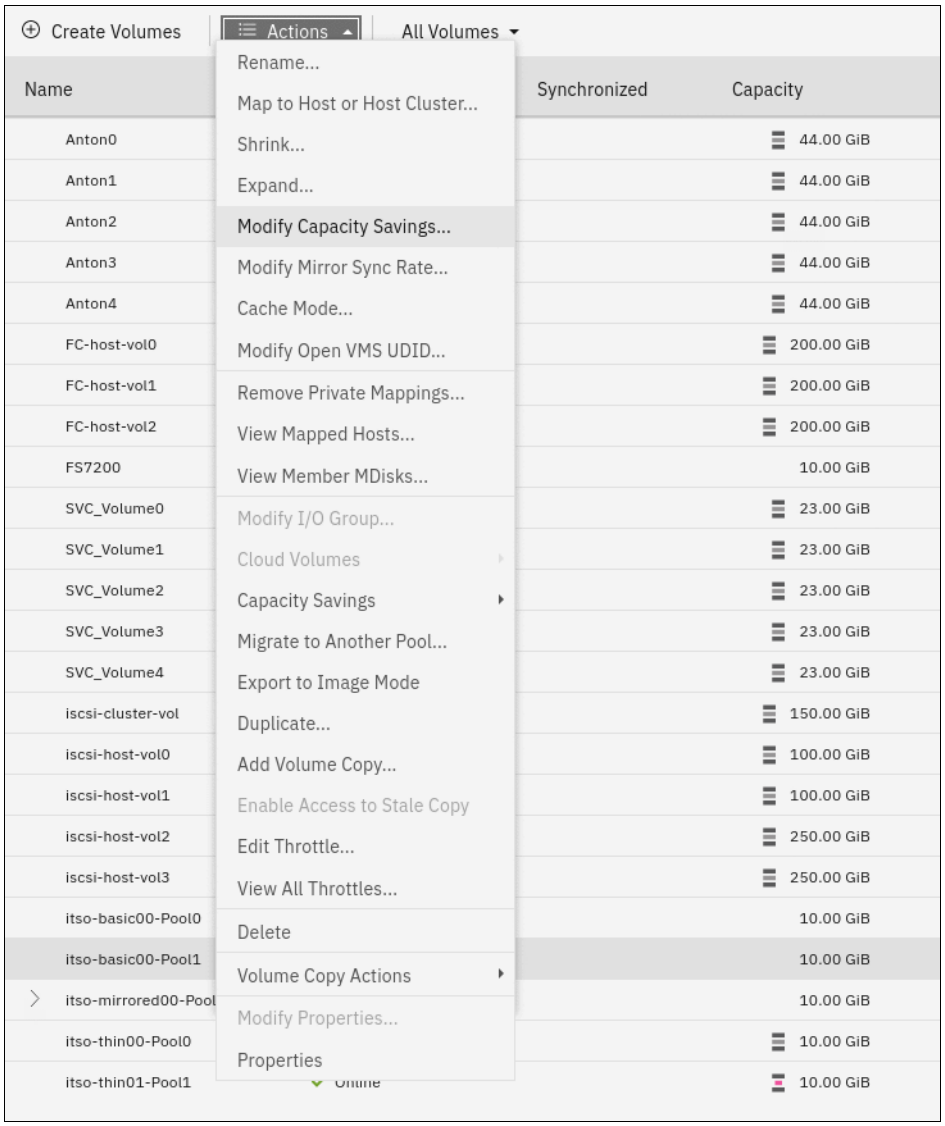


Figure 6-49 Modify Capacity Savings menu item

2. Select the capacity savings option that you want, as shown in Figure 6-50.

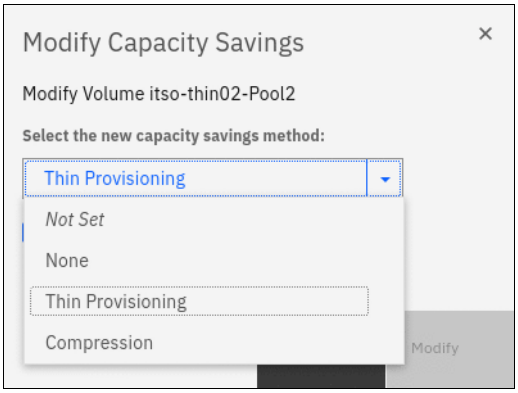


Figure 6-50 Capacity savings options for a volume

3. For volumes that are configured in a DRP, it is possible to enable deduplication, as shown in Figure 6-51.

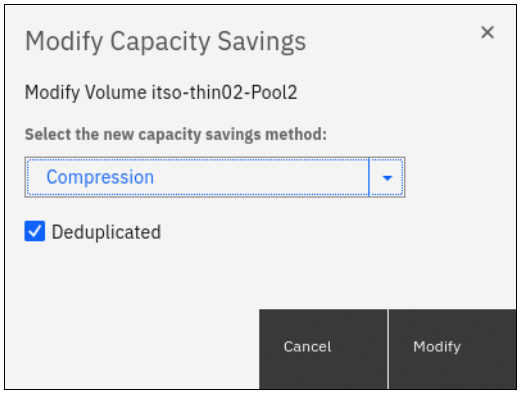


Figure 6-51 Enabling deduplication on a volume

After you configure the capacity savings options of a volume, click **Modify** to apply them. When the operation completes, you are returned to the Volumes view.

Modifying the mirror sync rate

This action is available for mirrored volumes only. 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 6-52.

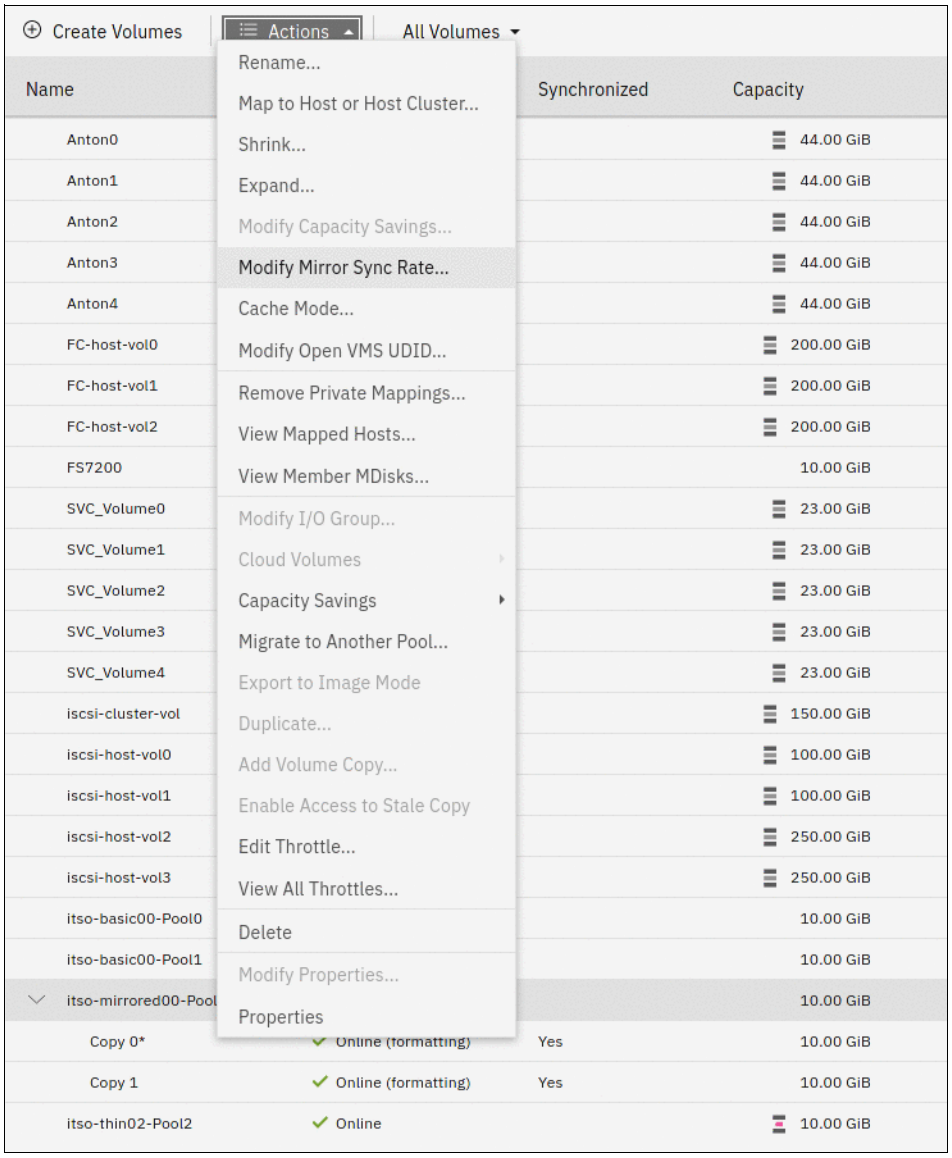


Figure 6-52 Modify Mirror Sync Rate menu item

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 6-53.

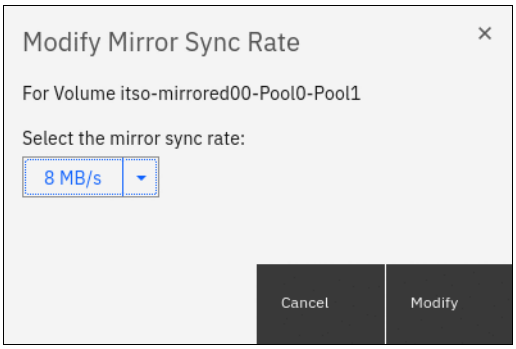


Figure 6-53 Setting the volume mirror sync rate

When the operation completes, you are returned to the Volumes view.

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 6-54.

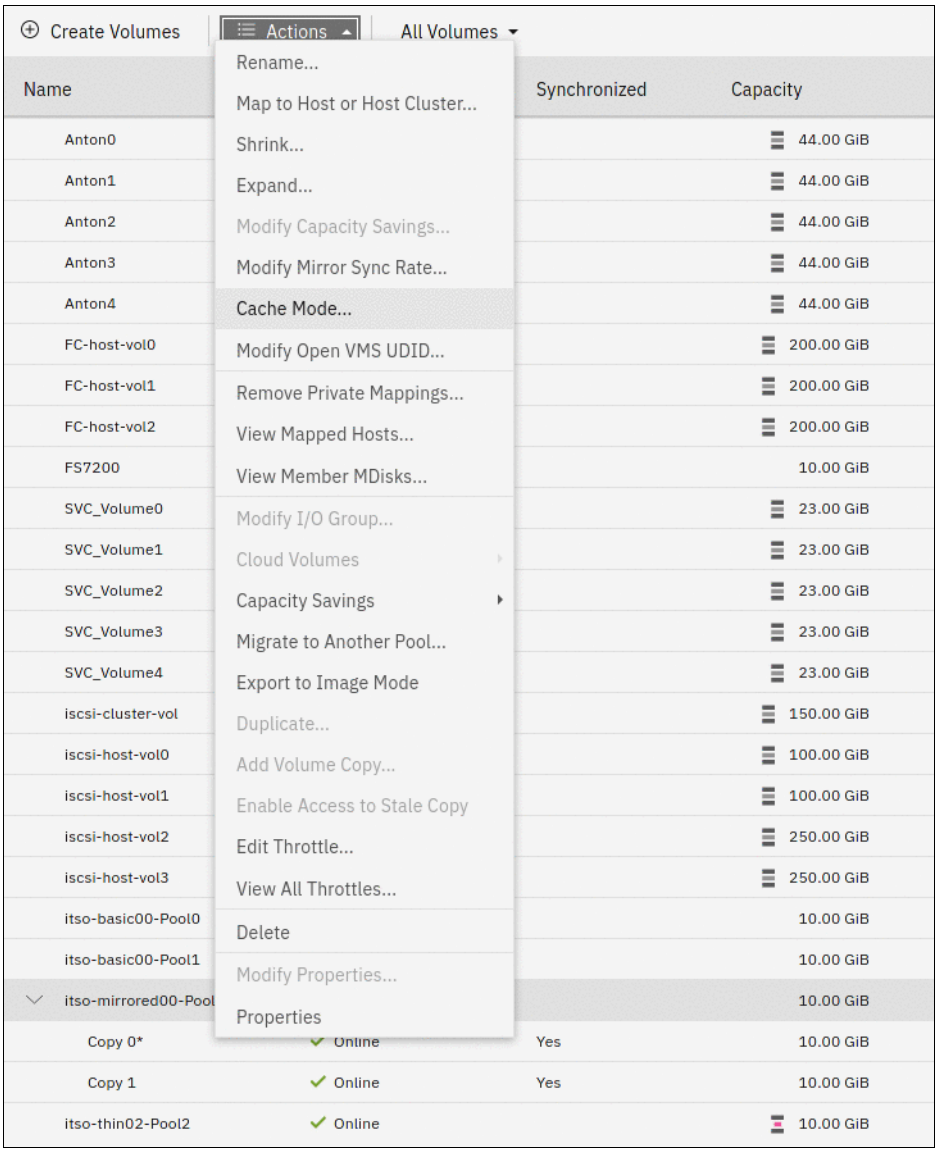


Figure 6-54 Modifying the volume cache mode

2. Select the cache mode that you want for the volume from the drop-down list and click **OK**, as shown in Figure 6-55.

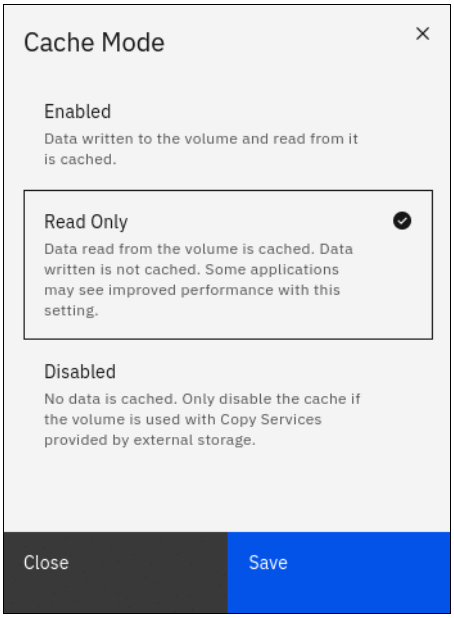


Figure 6-55 Setting the volume cache mode

When the operation completes, you are returned to the Volumes view.

Modify OpenVMS UDID menu

To change the OpenVMS UDID, use the **Modify OpenVMS UDID** menu.

A *UDID* is a nonnegative integer that is used in the creation of the OpenVMS device name. All fibre-attached volumes have an allocation class of 1, followed by the letters DGA, and then followed by the UDID.

All storage unit LUNs that you assign to an OpenVMS system need an UDID so that the operating system can detect and name the device. LUN 0 must also have a UDID, but the system displays LUN 0 as `$1$GGA<UDID>`, not as `$1$DGA<UDID>`.

For more information about Fibre-attached storage devices, see *Guidelines for OpenVMS Cluster Configurations*.

To change a volume OpenVMS UDID, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Modify Open VMS UDID…**, as shown in Figure 6-56.

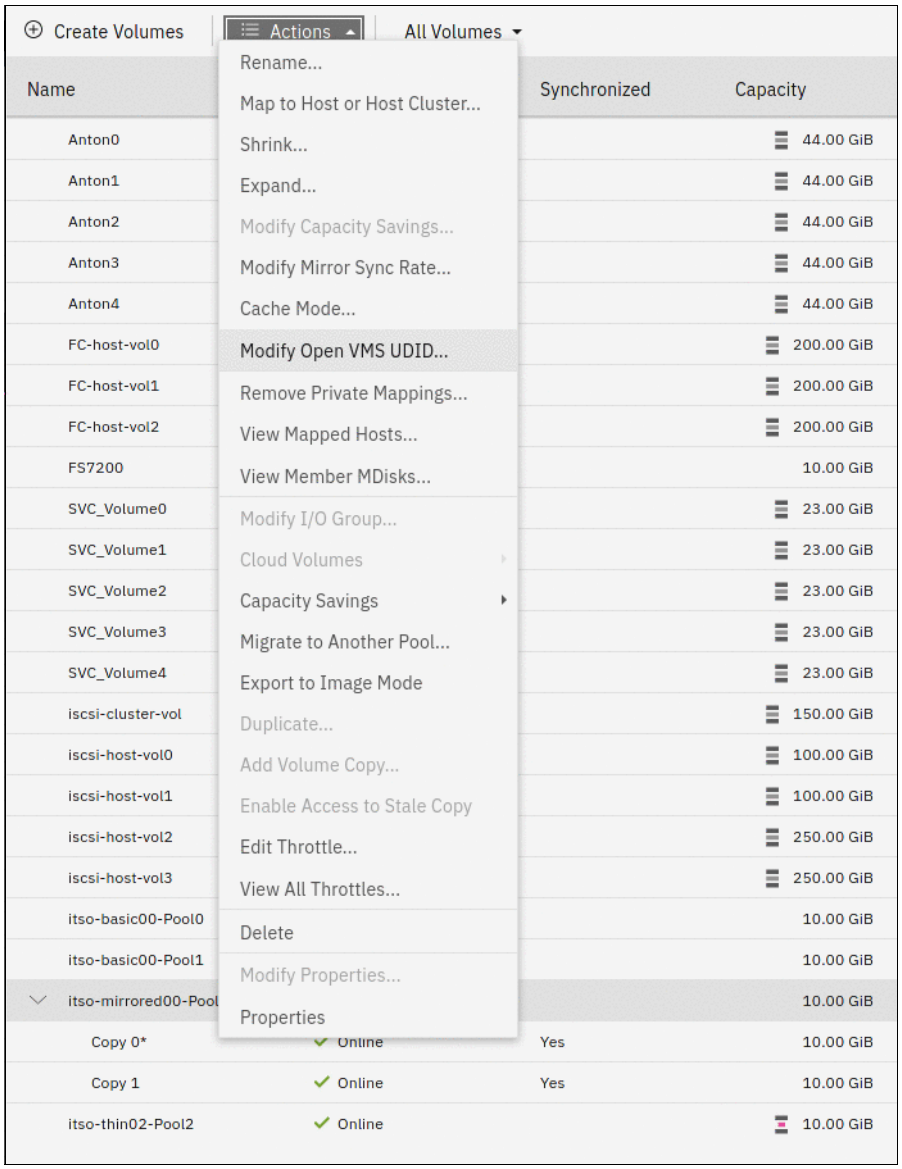


Figure 6-56 Modify Open VMS UDID menu item

2. Specify the UDID for the volume and click **Modify**, as shown in Figure 6-57.

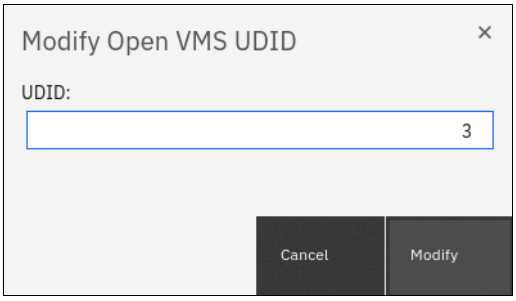


Figure 6-57 Setting the volume UDID

When the operation completes, you are returned to the Volumes view.

6.5.7 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, a delete fails (even if the **-force** parameter is specified) if the system detects any I/O activity to the volume within configured time frame. 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.

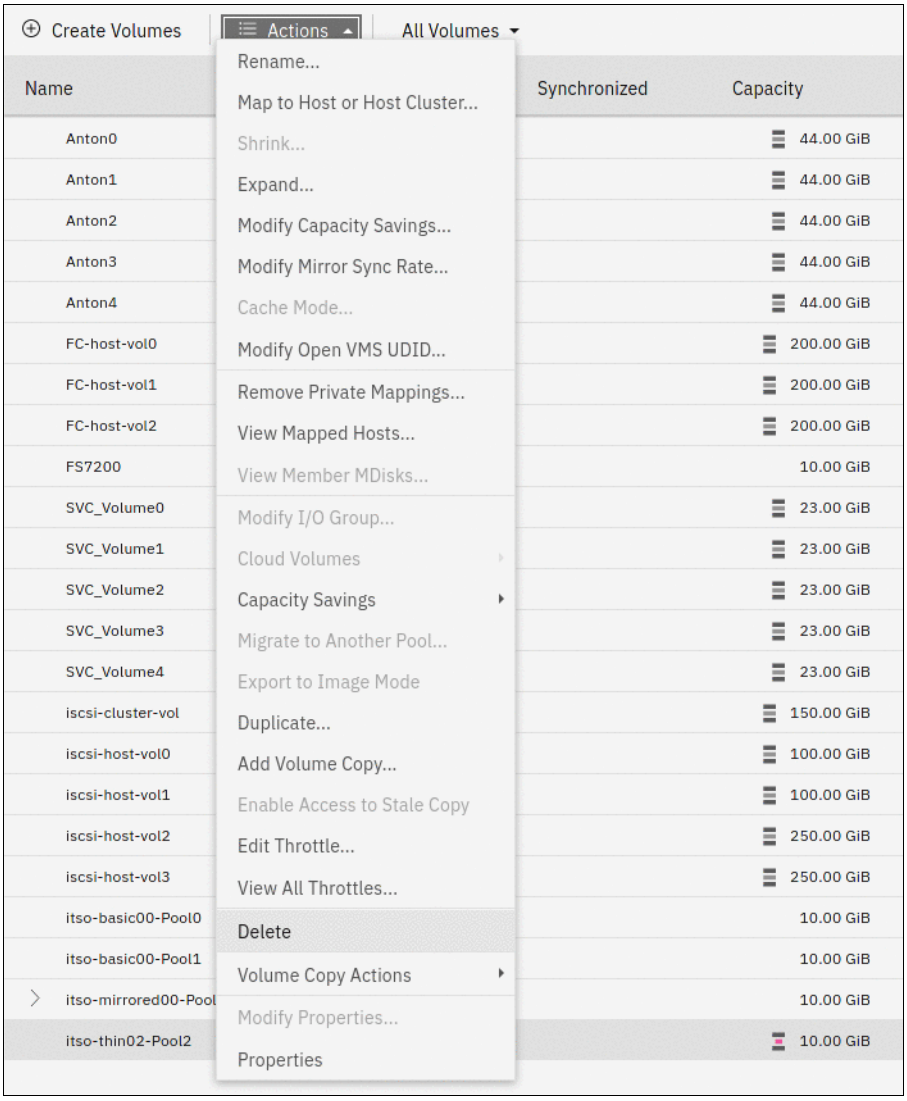


Figure 6-58 Volume Delete menu item

2. Review the list of volumes that is selected for deletion and provide the number of volumes that you intend to delete, as shown in Figure 6-59. Click **Delete** to remove the volume from the system configuration.

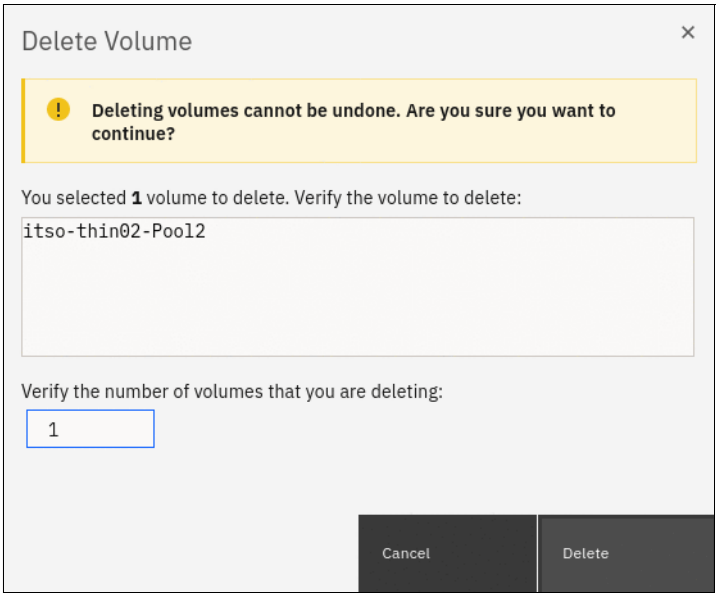


Figure 6-59 Confirming the volume deletion

When the operation completes, you are returned to the Volumes view.

6.5.8 Mapping a volume to a host

A volume must be mapped to make it available to a host or cluster of hosts. 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 on page 396.

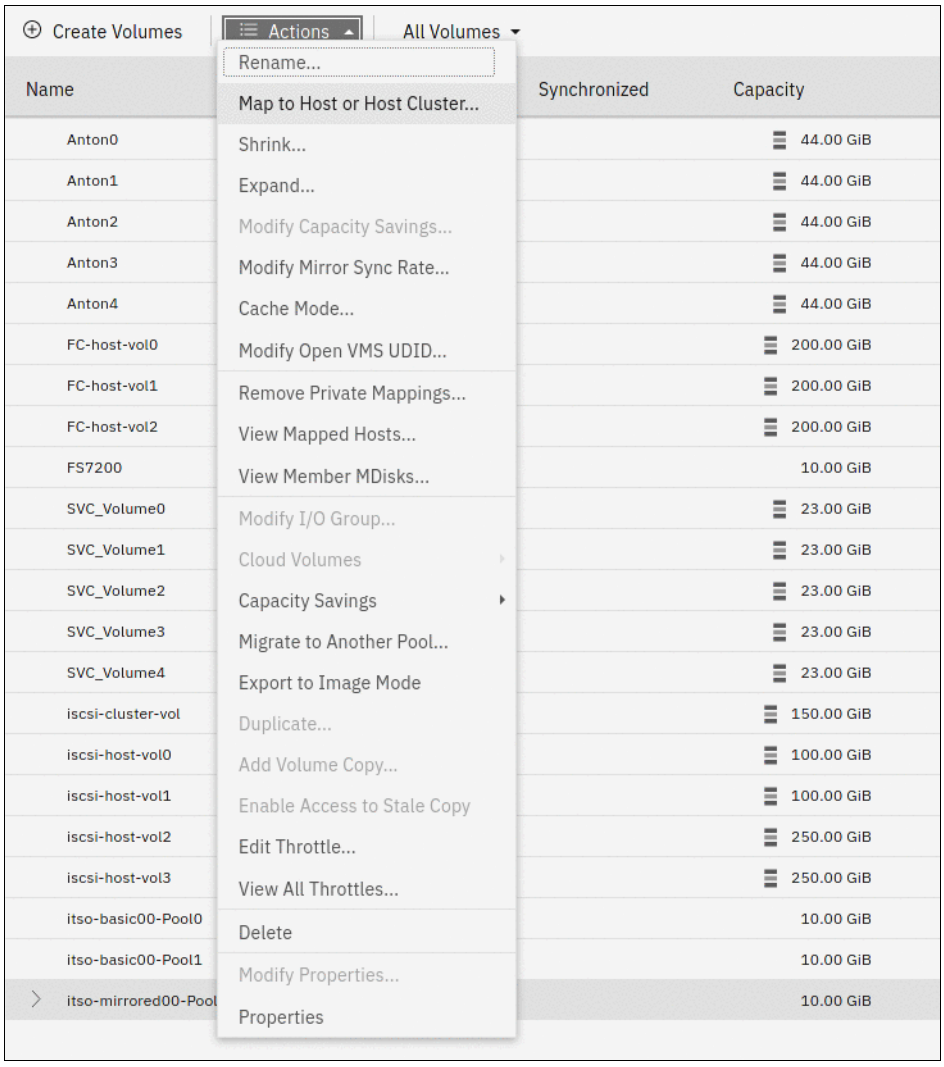


Figure 6-60 Volume mapping menu item

Tip: An alternative way of opening the Actions menu is to highlight (select) a volume and right-click.

2. The Create Mapping window opens. In this window, select whether to create a mapping to a host or 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.

As shown in Figure 6-61, a single volume is mapped to a host and the storage system assigns the SCSI LUN IDs.

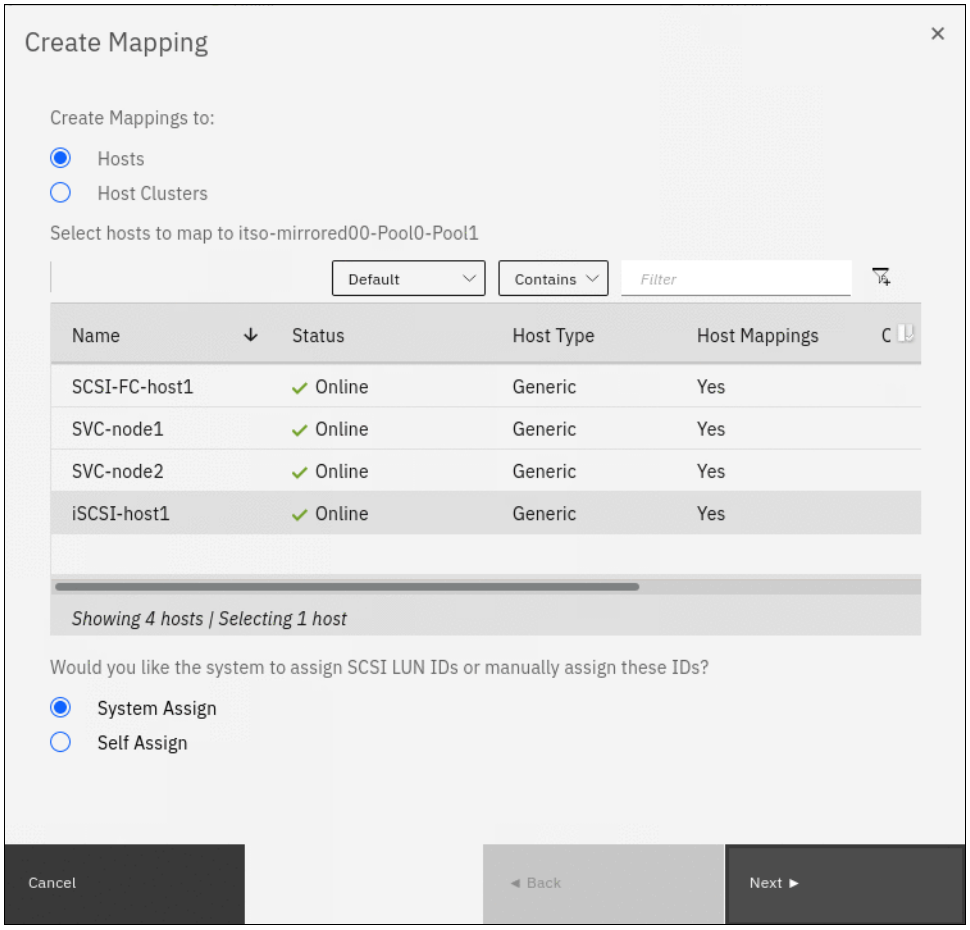


Figure 6-61 Create Mapping panel

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 on page 398. Review the future configuration state and click **Map Volumes** to map the volume.

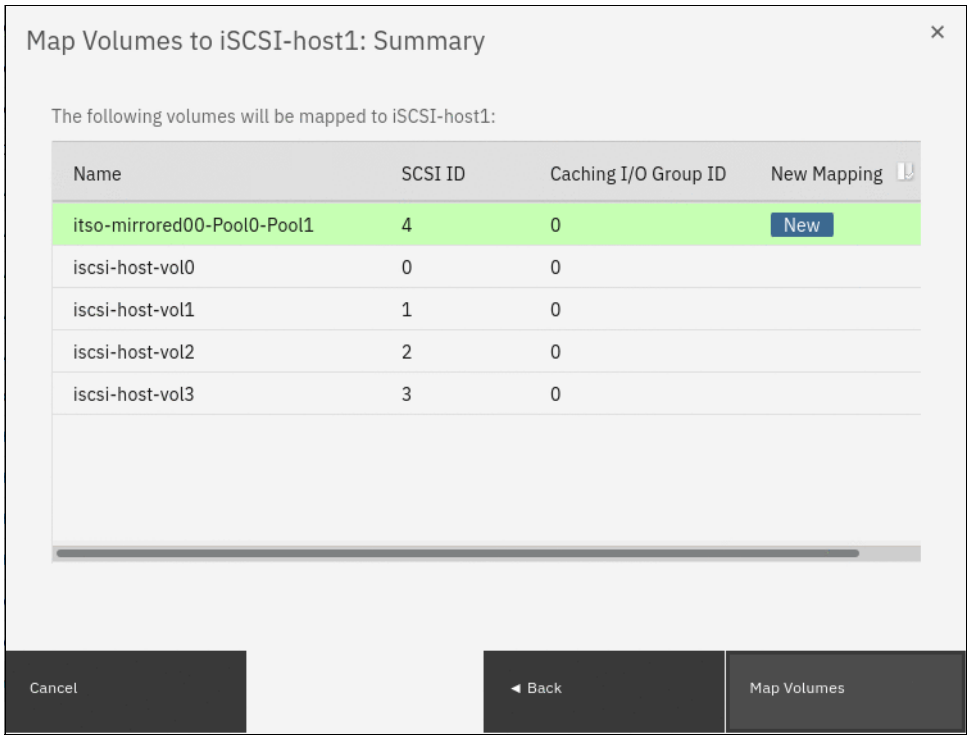


Figure 6-62 Mapping a volume to host: Summary

4. After the task completes, the wizard returns to the Volumes window. You can list the volumes that are mapped to the host by selecting **Hosts** → **Mappings**, as shown in Figure 6-63.

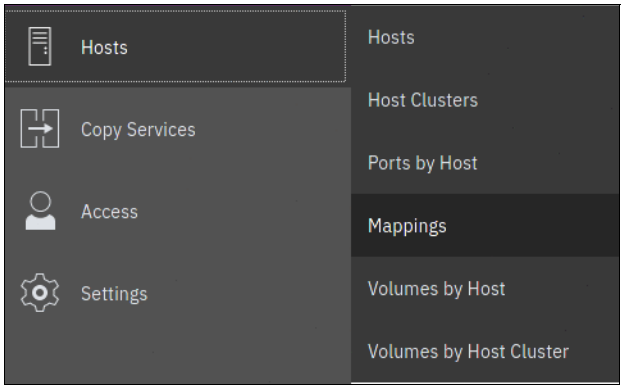


Figure 6-63 Accessing the Hosts Mapping view

A window with a list of volumes that are mapped to all hosts opens, 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

To see volumes that are mapped to clusters instead of hosts, change the value that is shown in the upper left corner (see Figure 6-64) 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 7, “Hosts” on page 443.

To remove the volume to host mapping, in the **Hosts** → **Mappings** view, select the volume or volumes and click **Unmap Volumes** from the context menu, as shown in Figure 6-65.

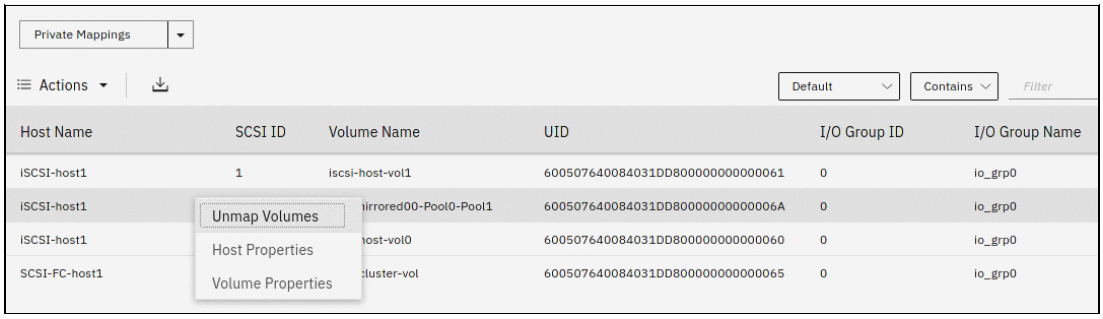


Figure 6-65 Unmap Volumes menu item

In the Delete Mapping pane, enter the number of volumes that you intend to unmap, as shown in Figure 6-66 on page 400. This task is intended as a security measure to minimize changes for an accidental unmap of 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. Improperly run volume unmap operations can result in data unavailability and data loss.

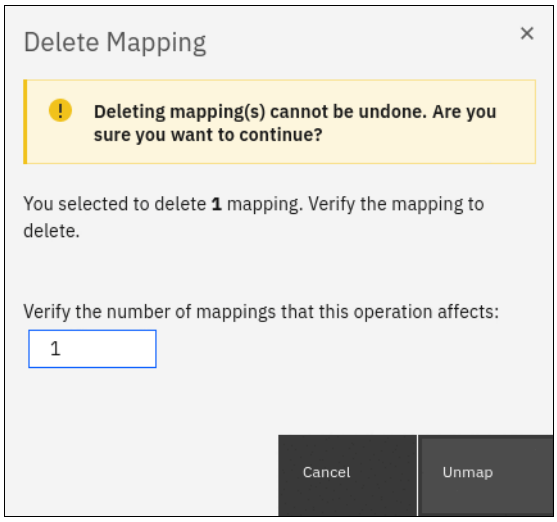
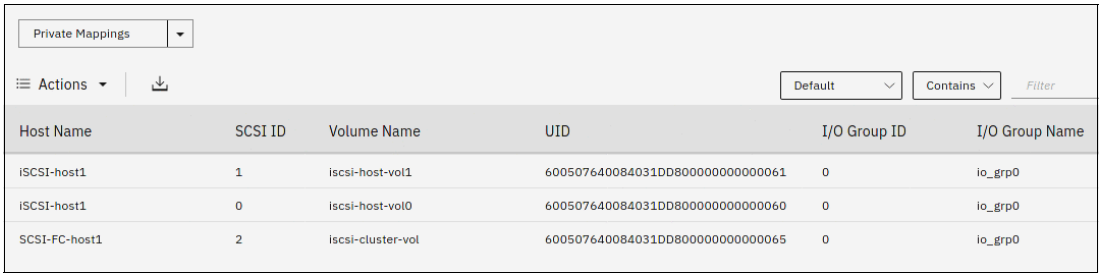


Figure 6-66 Volume unmap confirmation pane

Click **Unmap** to complete the operation. Volume mapping is removed and it is no longer displayed in the volume map view, as shown in Figure 6-67.



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	0	iscsi-host-vol0	600507640084031DD800000000000060	0	io_grp0
SCSI-FC-host1	2	iscsi-cluster-vol	600507640084031DD800000000000065	0	io_grp0

Figure 6-67 Volume mapping removed

6.5.9 Modify I/O Group or Non-disruptive Volume Move (NDVM)

The NDVM function was introduced in version 6.3 of the IBM SAN Volume Controller and Storwize code. This function is also used for changing the preferred node within the I/O group a volume or volumes.

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 backend storage traffic. In enclosure based systems, we generally want to align the I/O group with the enclosure that contains the pool in which the volume is stored.

The primary use case for this function is the upgrade of environments from Storwize V7000 to the FlashSystem 7200, 9150, or 9200 because those systems can be clustered to a Storwize V7000.

Note: Specific conditions prevent the changing of I/O group dynamically with NDVM for a volume. If the volume uses Data Reduction in a DRP, or if the volumes is a member of a FlashCopy map, or is in a remote copy relationship, the first command in the sequence, **`addvdiskaccess`**, fails.

To start the process, select the volumes and then, access the **Modify I/O Group...** menu item from the context menu or use the Actions drop-down men, as shown in Figure 6-68.

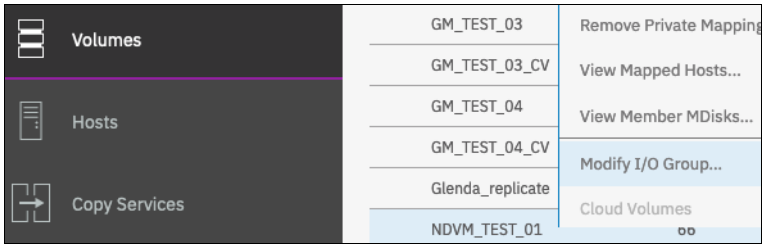


Figure 6-68 Modify I/O Group menu item

If no host mappings for the volumes are available, the operation immediately displays the target I/O group selection panel, as shown in Figure 6-69.

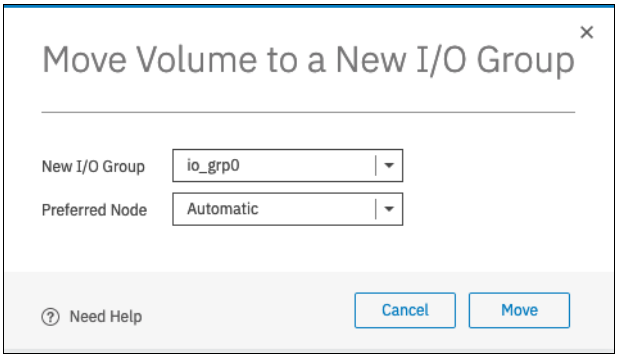


Figure 6-69 I/O Group selection panel

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}`**, which 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}`**, which moves the preferred node of the volume to the new (target) caching I/O group.
- **`rmvdiskaccess -iogrp {old i/o group} {volume}`**, which removes the old (source) I/O group from the set of I/O groups in which the volume can be made accessible to hosts.

In the likely case where 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 402, to ensure that the correct steps are performed in the correct order.

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. The steps that are required to modify I/O group of a mapped volume are described next.

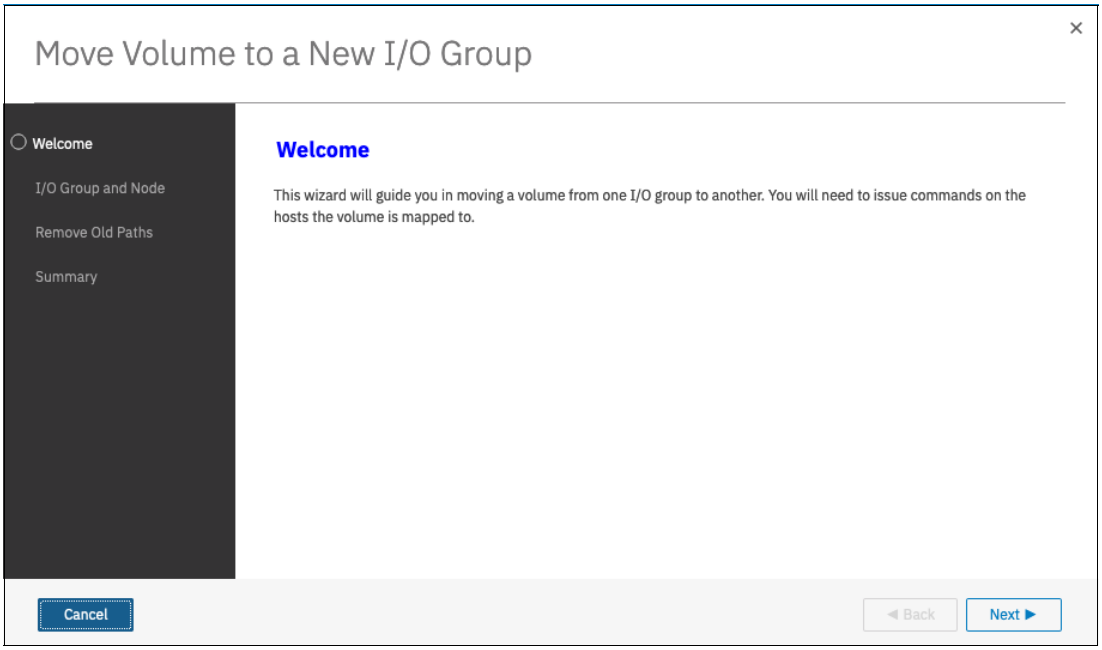


Figure 6-70 Modify I/O group for a mapped volume wizard – welcome panel

Verify that all hosts that use the volume are zoned to the target I/O group and click **Next** to proceed to the new I/O group selection panel (see Figure 6-71).

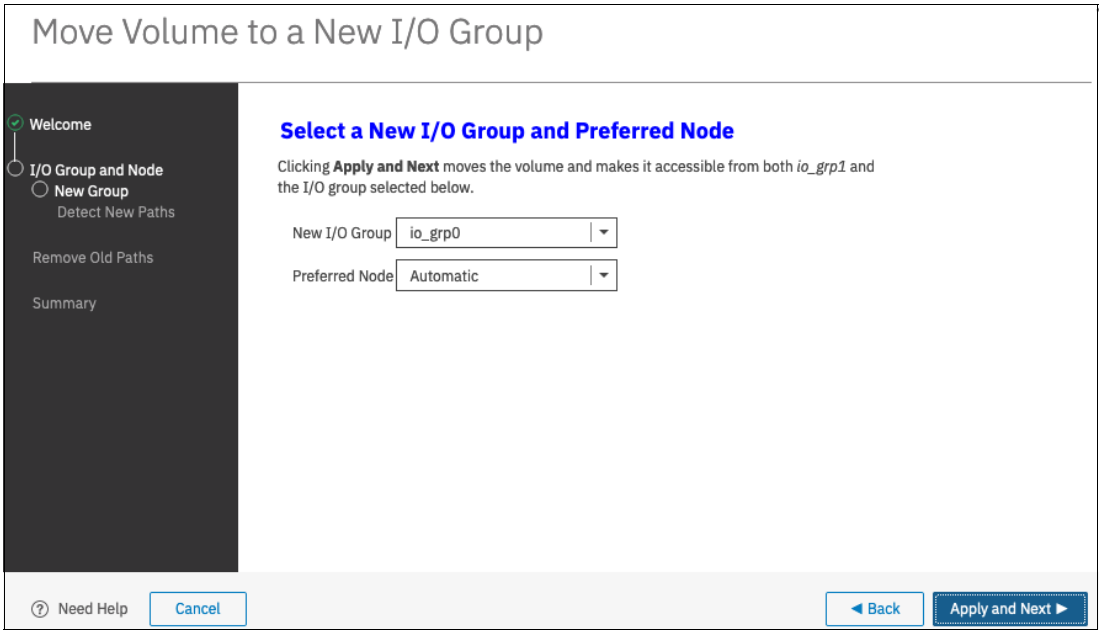


Figure 6-71 Modify I/O Group for a mapped volume wizard – I/O group selection panel

Select the new (target) I/O group and preferred node and click **Apply and Next** to proceed to the next step. The GUI runs the following commands:

- **`addvdiskaccess -iogrp {new i/o group} {volume}`**, which 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}`**, which moves the preferred node of the volume to the new (target) caching I/O group.

You are notified of the successful completion of the commands, as shown in Figure 6-72.

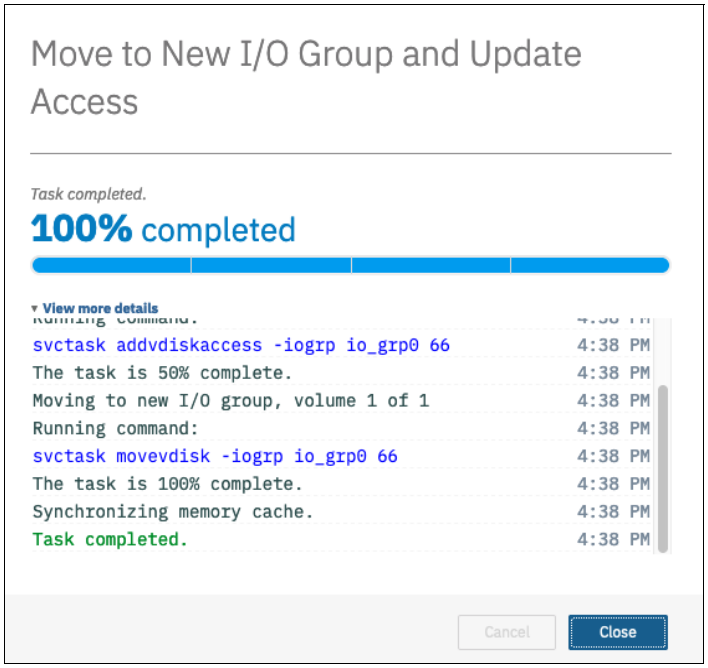


Figure 6-72 Modify I/O Group for a mapped volume wizard – first stage completed

Click **Close** to proceed to the validation window (see Figure 6-73). Click **Need Help** to see information about how to prepare the host for the volume move. After the host is ready for volume path change, select the option to confirm that path validation was performed on the host.

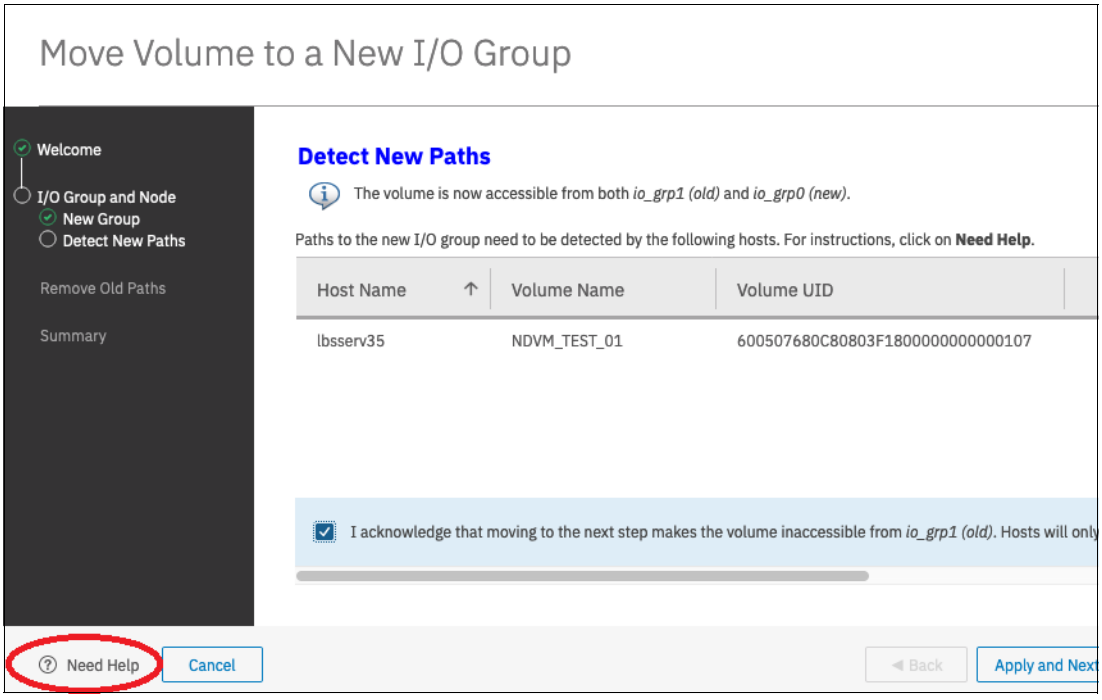


Figure 6-73 Modify I/O Group for a mapped volume wizard – validation

Note: Failure to ensure the host discovered the new paths to all the volumes can result in this process being disruptive and the host to lose access to the moved volume or volumes.

After validation is complete and the acknowledgment box is selected, click **Apply and Next** to advance the wizard to the final phase of removing access from the old I/O group (see Figure 6-74).

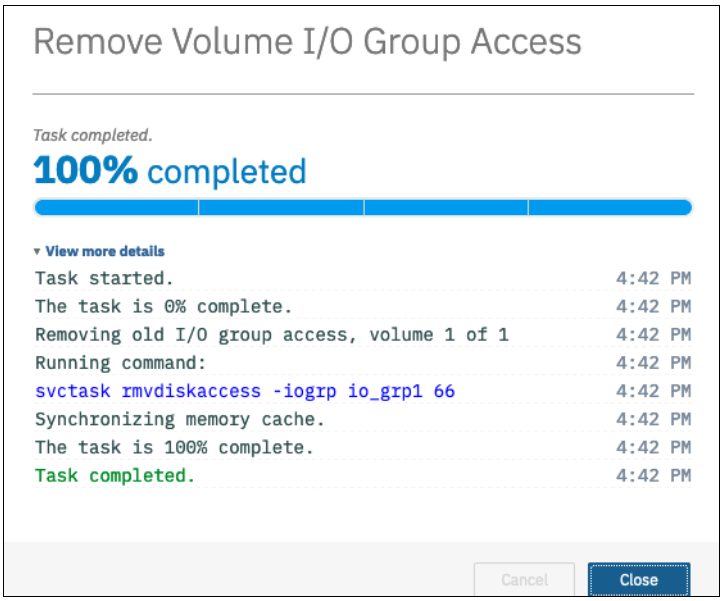


Figure 6-74 Modify I/O Group for a mapped volume wizard – second stage completed (details)

Close the detail panel to proceed to the final panel of the Wizard (see Figure 6-75). At this point, hosts cannot access the volume by way of the old I/O group and, depending on the operating system, can require a restart to remove the dead or stale paths.

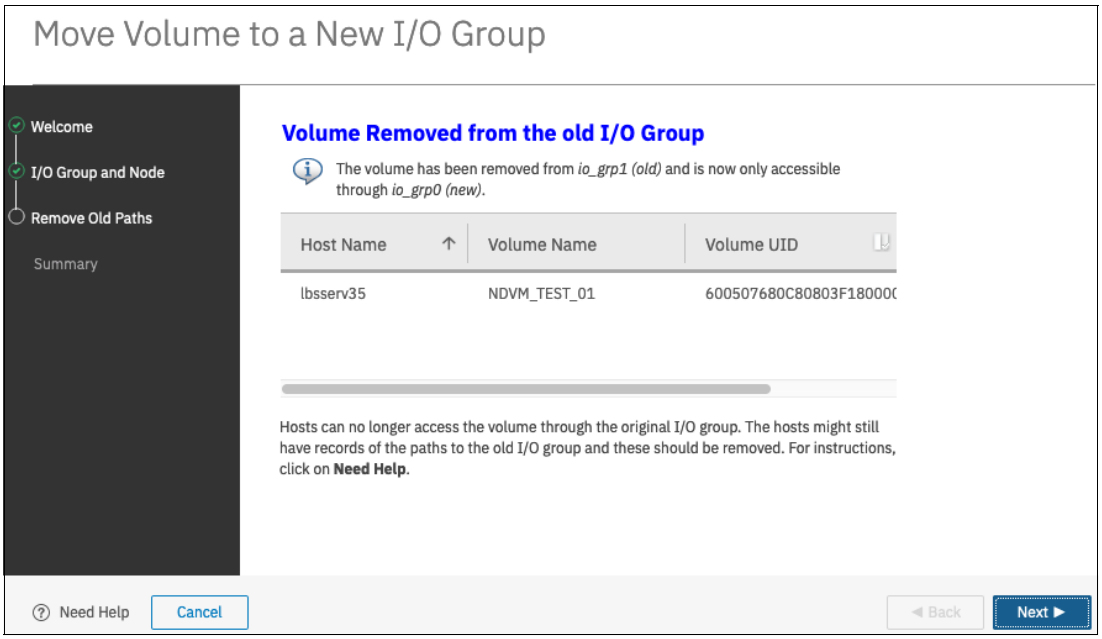


Figure 6-75 Modify I/O Group for a mapped volume wizard: Operation completed

6.5.10 Migrating a volume to another storage pool

IBM Spectrum Virtualize enables online volume migration with no applications downtime. Volumes can be moved between storage pools without affecting business workloads that are running on these volumes.

Volume migration can be performed by using one of two methods: the use of the volume migration feature and, 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 Spectrum 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.

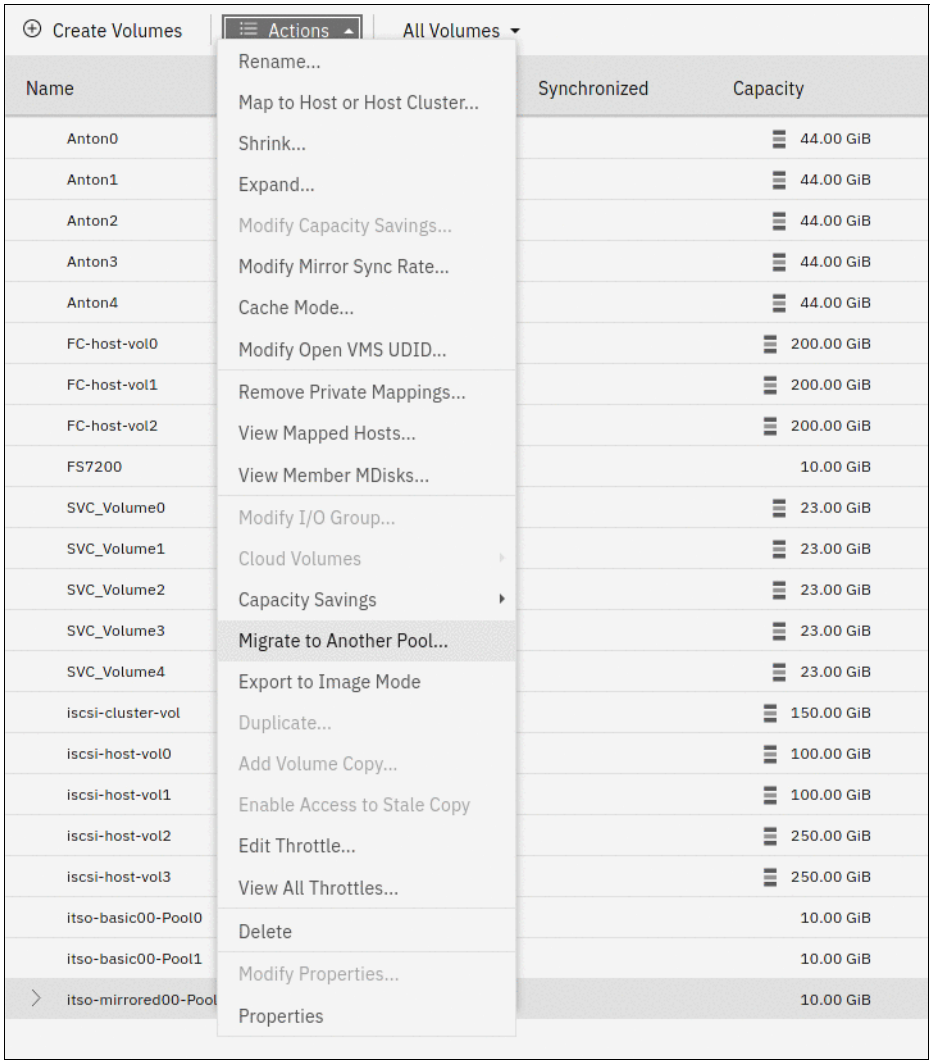


Figure 6-76 Migrate Volume Copy: Selecting menu item

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 pane is not displayed.

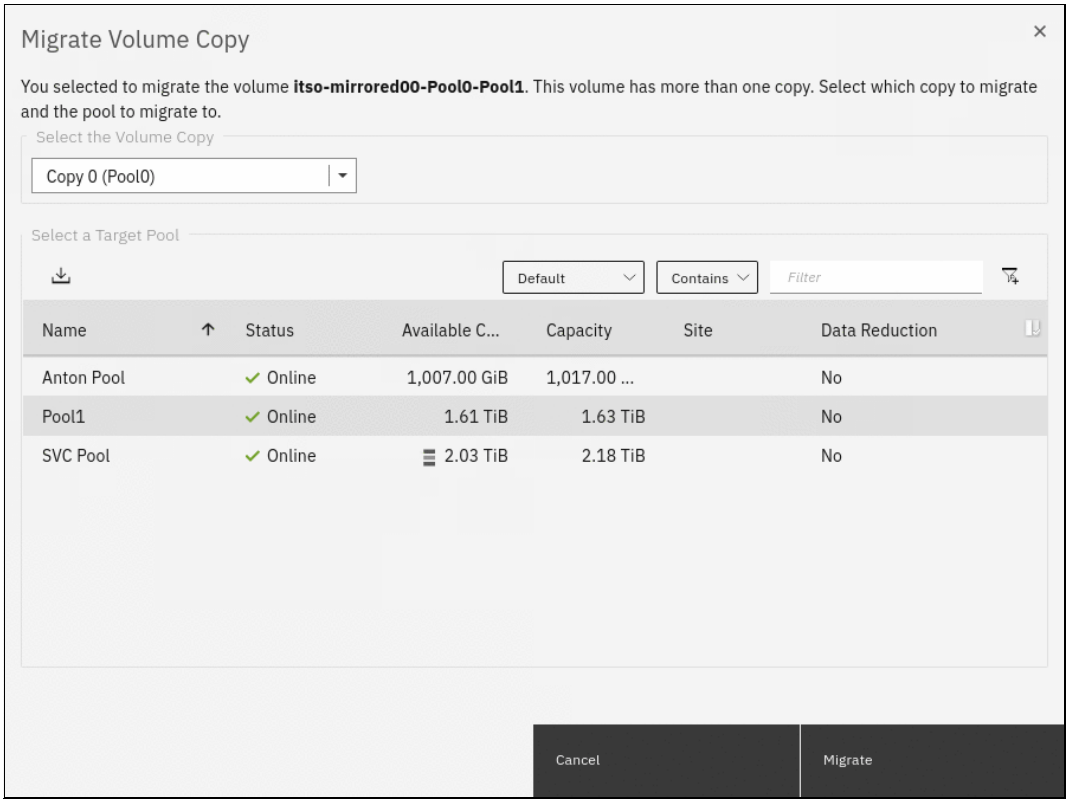


Figure 6-77 Migrate Volume Copy: Selecting the volume copy

3. Select the new target storage pool and click **Migrate**, as shown in Figure 6-77. The Select a Target Pool pane displays the list of all pools that are a valid migration copy target for the selected volume copy.
4. You are returned to the Volumes view. 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 pane of the **Background Tasks** menu, as shown in Figure 6-79.



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.

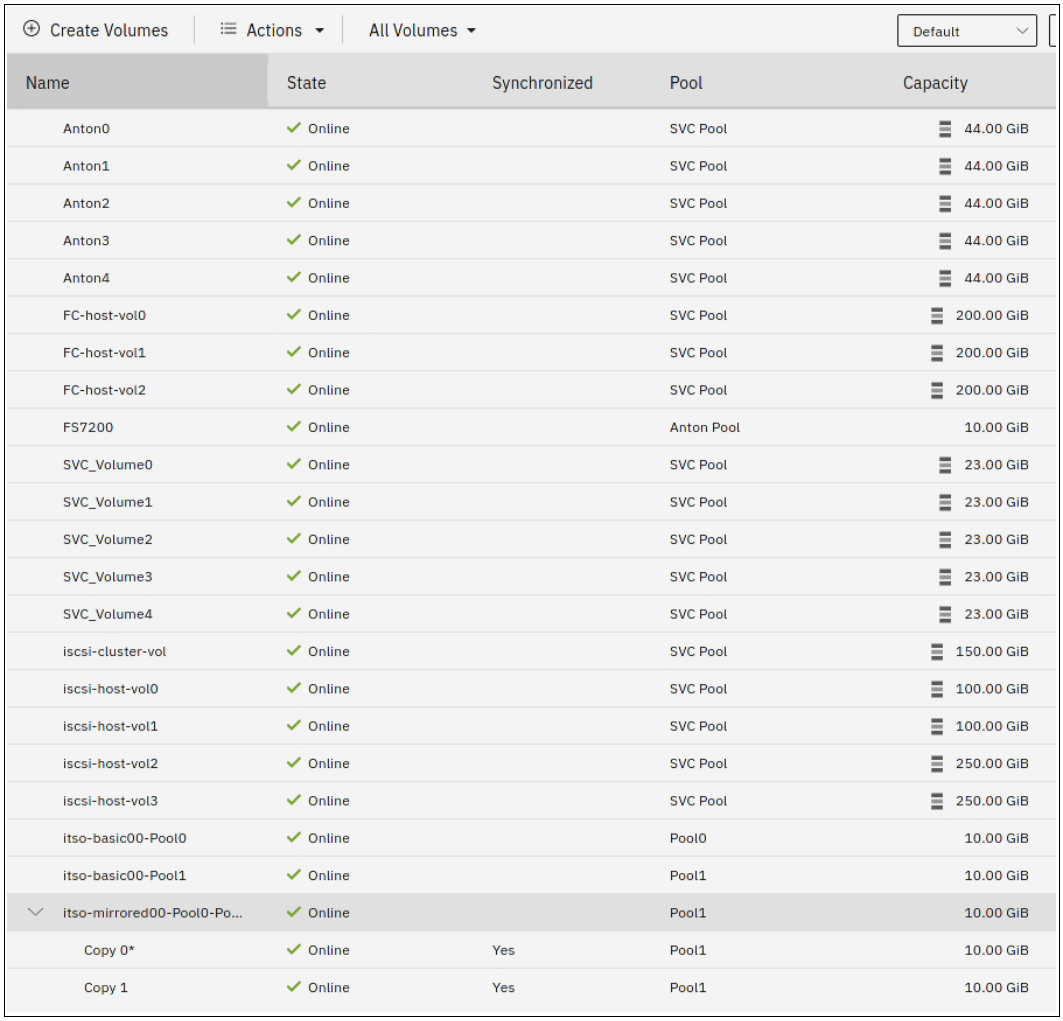


Figure 6-80 Volume copy after migration

The volume copy is now migrated without any host or application downtime to the new storage pool.

Another way to migrate single-copy volumes to another pool is to use the volume copy feature, as described in “Volume migration by adding a volume copy” on page 409.

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 Spectrum 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, as described in 6.5.10, “Migrating a volume to another storage pool” on page 405. 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 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 only for single-copy volumes. If you need to 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.

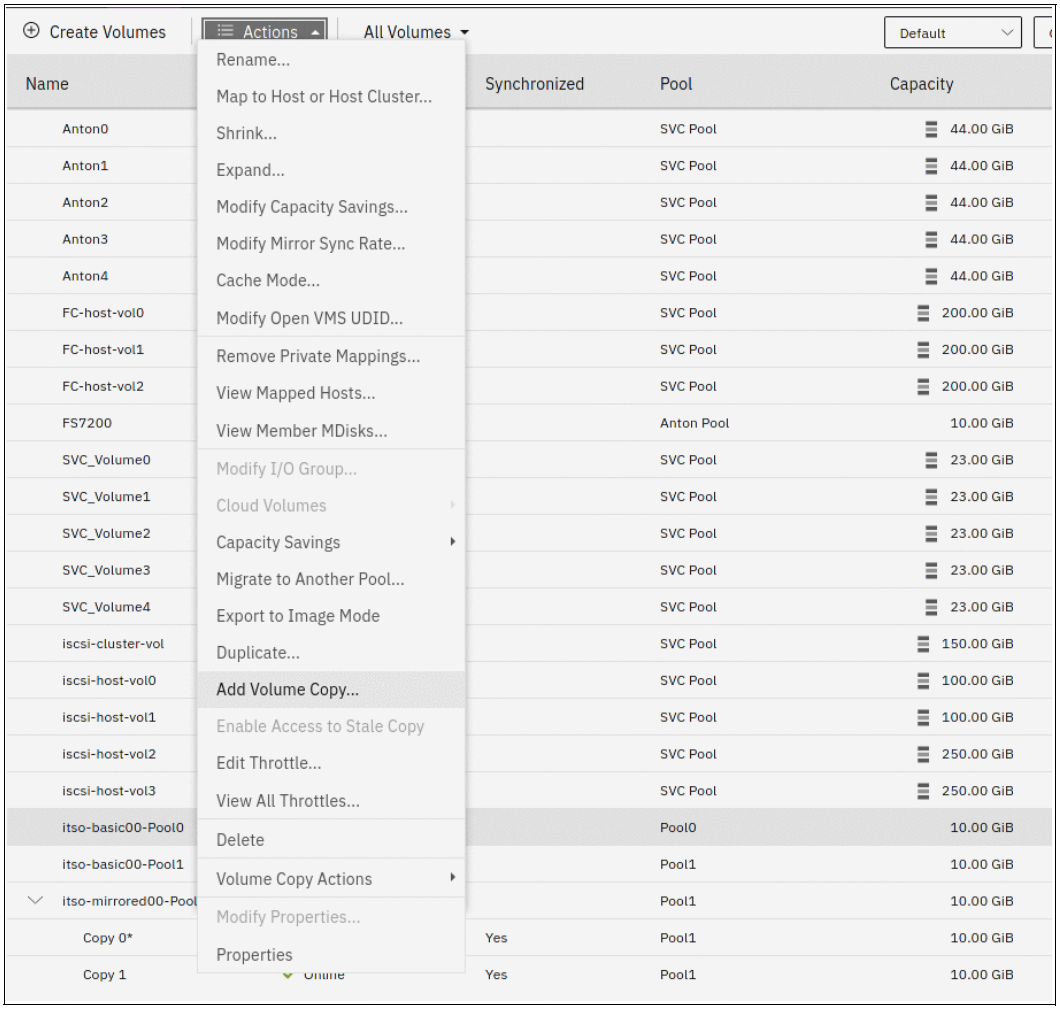


Figure 6-81 Add Volume Copy menu item

2. Create a second copy of your volume in the target storage pool, as shown in Figure 6-82. Note that you can modify capacity savings options for the new volume copy. In our example, a compressed copy of the volume is created in target pool Pool2. The Deduplication option is not available if either of the volume copies is not in a DRP. Click **Add** to proceed.

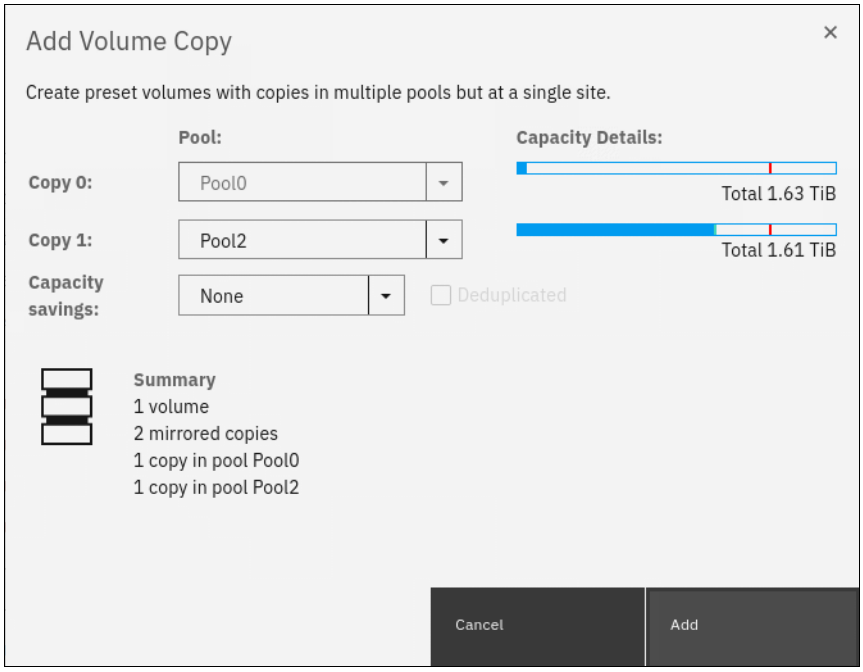


Figure 6-82 Adding a volume copy

Wait until the copies are synchronized, as shown in Figure 6-83.

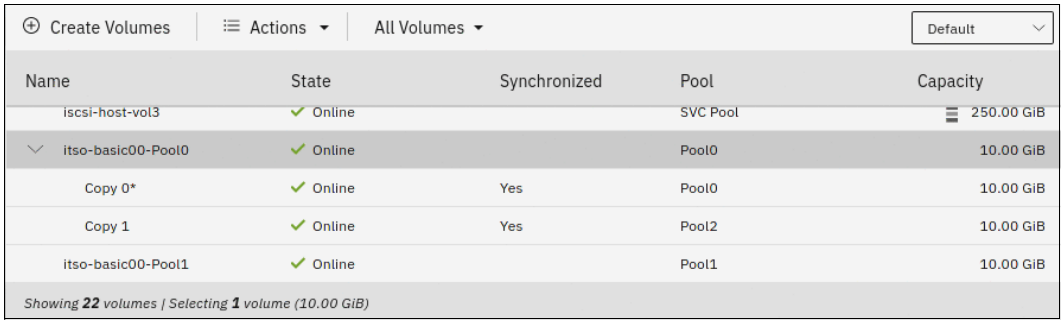


Figure 6-83 Verifying that the volume copies are synchronized

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.

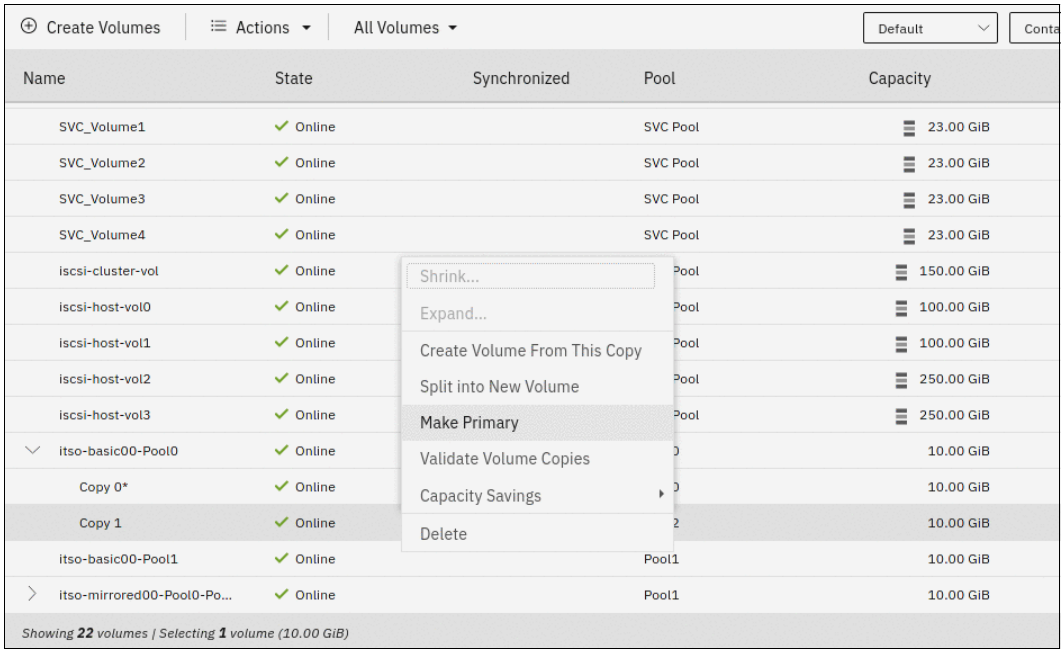


Figure 6-84 Setting the volume copy in the target storage pool as the primary copy

4. Split or delete the volume copy in the source pool, as shown in Figure 6-85.

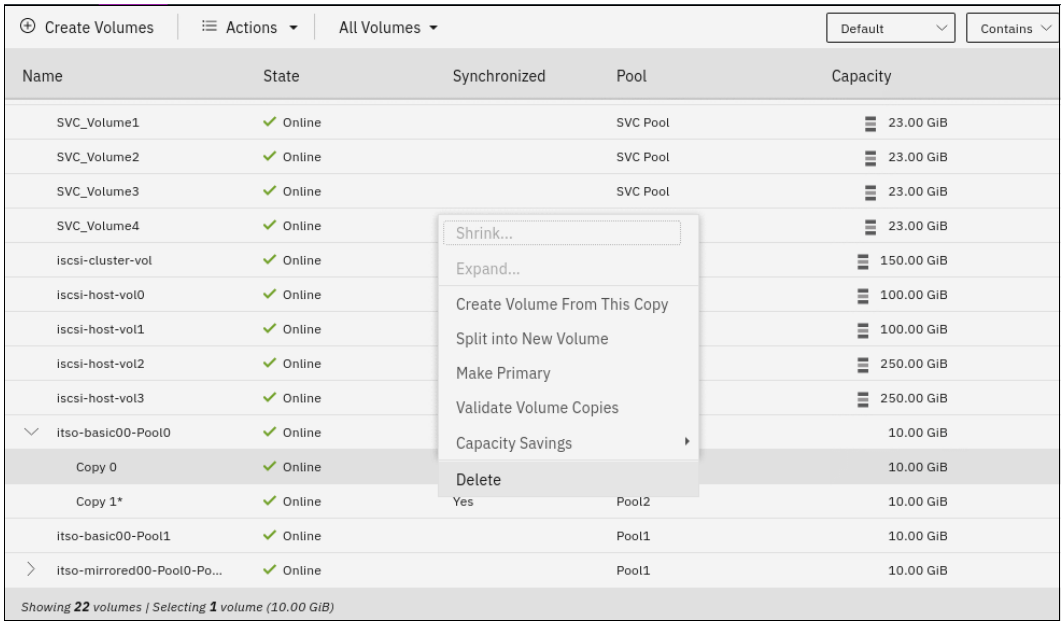


Figure 6-85 Deleting the volume copy in the source pool

5. Confirm the removal of the volume copy, as shown in Figure 6-86.

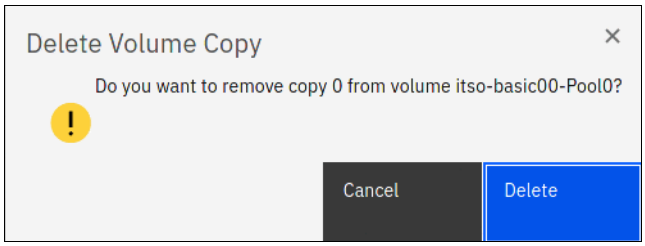


Figure 6-86 Confirming the deletion of a volume copy

6. The Volumes view now shows that the volume has a single copy in the target pool, as shown in Figure 6-87.

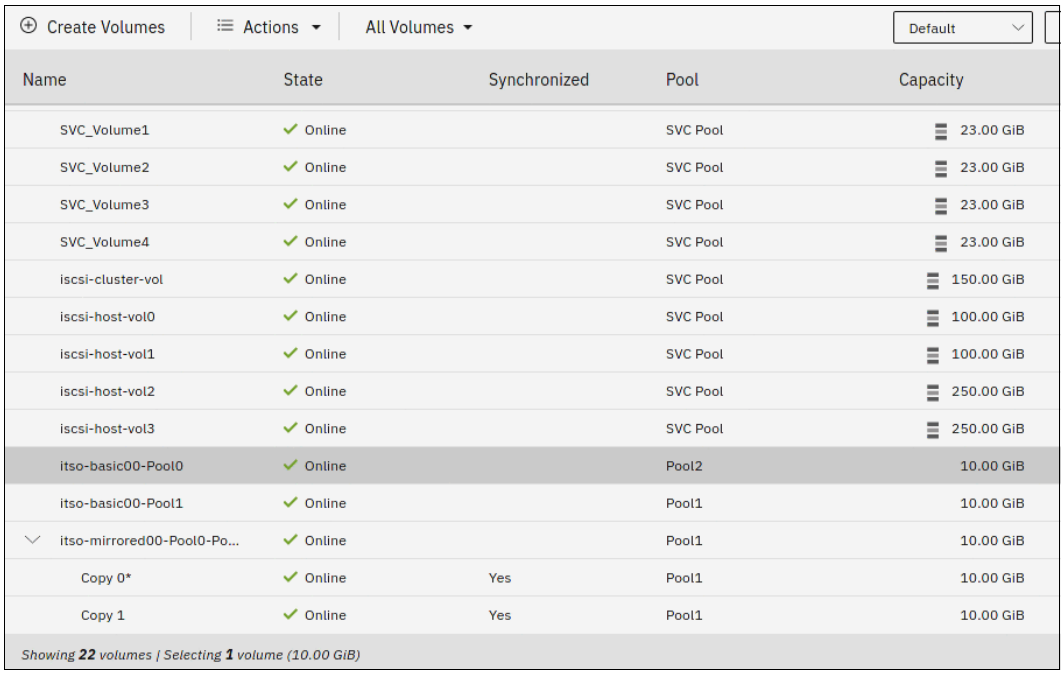


Figure 6-87 Volume copy in the target storage pool

Migrating volumes by using the volume copy feature requires more user interaction, but might be a preferred option for particular 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 only done from the primary copy.

To test the performance of the volume in the new pool, switch the roles of the volume copies making 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 backed by tier 1 storage.

This method gives you ability to migrate between storage tiers with fast and secure back-out option.

6.6 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 B, “Command-line interface setup” on page 951.

6.6.1 Displaying volume information

To display information about all volumes that are defined within the IBM Spectrum 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
6005076400F58004980000000000000002 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
6005076400F58004980000000000000003 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     6005076400F58004980000000000000004 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
6005076400F58004980000000000000005 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
6005076400F58004980000000000000006 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
6005076400F58004980000000000000007 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
6005076400F58004980000000000000008 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
6005076400F58004980000000000000009 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     6005076400F5800498000000000000010 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
6005076400F5800498000000000000000E 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
6005076400F5800498000000000000000F 0 1 empty 0 no 0 0 Pool0 no yes 13 vdisk0_restore1
```

6.6.2 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 Spectrum 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 6005076400F5800498000000000000010
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.

6.6.3 Creating a thin-provisioned volume

Example 6-4 shows an example of 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 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.

6.6.4 Creating a volume in image mode

Use an image mode volume to bring a non-virtualized disk (for example, from a pre-virtualization environment) under the control of the IBM Spectrum 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 thus far unmanaged MDisk from which it is created. Therefore, except for a thin-provisioned image mode volume, the volume's LBA *x* equals MDisk LBA *x*.

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 this example, 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
```

6.6.5 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 be also 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 an uncompressed volume with a single volume copy.

Example 6-12 An uncompressed 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 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.

6.6.6 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
```

6.6.7 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.

6.6.8 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. Using 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 perspective 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 Spectrum 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 Spectrum 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
```

6.6.9 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
```

6.6.10 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 Spectrum 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 give it a total size of 40 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.
```

6.6.11 HyperSwap volume modification with CLI

The following new CLI commands for administering volumes were released in IBM Spectrum Virtualize V7.6. However, the GUI uses the new commands only for HyperSwap volume creation (**mkvolume**) and deletion (**rmvolume**):

- mkvolume
- mkimagevolume
- addvolumecopy
- rmvolumecopy
- rmvolume

In addition, the **lsvdisk** output shows more fields: volume_id, volume_name, and function, which help 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 an 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 exactly what protection they are bypassing.

6.6.12 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: SCSI-3 standard requires LUN 0 to exist on every SCSI target (see SCSI Architecture Model – 2, section 4.7.2 *SCSI target model*). This LUN must implement a number of standard-specified 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 are using 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
```

6.6.13 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 host name is specified by the **lshostvdiskmap** command, it returns all defined host-to-volume mappings.

Specifying the flag before the host name: Although the **-delim** flag normally comes at the end of the command string, you must specify this flag before the host name in this case. Otherwise, it returns the following message:

```
CMMVC6070E An invalid or duplicated parameter, unaccompanied argument, or
incorrect argument sequence has been detected. Ensure that the input is as per
the help.
```

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
```

6.6.14 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.

6.6.15 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.

6.6.16 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 Spectrum Virtualize environment by using image mode. For more information about migration, see Chapter 8, “Storage migration” on page 527.

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 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.

6.6.17 Migrating a fully managed volume to an image mode volume

Migrating a fully managed volume to an image mode volume enables the IBM Spectrum Virtualize system to be removed from the data path. This feature might be useful when the IBM Spectrum 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
```

6.6.18 Shrinking a volume

The **shrinkvdisksize** command reduces the capacity that is allocated to the particular 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 in size. 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 from a total size of 80 GB by 44 GB to the new total size of 36 GB.

Example 6-32 The shrinkvdisksize command

```
IBM_2145:ITSO_CLUSTER:superuser>shrinkvdisksize -size 44 -unit gb volume_D
```

6.6.19 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
```

6.6.20 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.

6.6.21 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,,,,,6005076400F5800498000000000
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,
```

6.6.22 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 6005076400F58004980000000000000002
```

```
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
```

To learn more about these storage pools, run the **lsmdiskgrp** command as described in Chapter 5, “Using storage pools” on page 271.

6.6.23 Tracing a volume from a host back to its physical disks

In some cases, you might need to verify exactly 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 exactly one MDisk. This configuration usually is not used for striped volumes unless the volume size is lesser than the extent sizes. Therefore, a single striped volume uses multiple MDisks in a typical case.

For example, on a Linux host running a native multipath driver, you can use the output of 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 Spectrum 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 on page 439, 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 lists the 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 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 on page 440, 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 60080e50001b0b62000007b04e731e4d0000000000000000000000000000000000
preferred_WWPN 20580080E51B09E8
active_WWPN 20580080E51B09E8
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.

On the back-end storage, you can identify which physical disks make up the LUN that was presented to the Storage Virtualize system by using the volume ID that is displayed in the `UID` field.

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 an IBM SAN Volume Controller storage system. It also documents the available host attachment protocols, including Non-Volatile Memory Express (NVMe) over Fabric (NVMe-oF), Fibre Channel Small Computer System Interface (SCSI) (FC-SCSI), serial-attached SCSI (SAS), and internet Small Computer Systems Interface (iSCSI).

In addition, this chapter explains host clustering and N_Port ID Virtualization (NPIV) support for a host-to-storage system communication and how to configure portsets to advantage of the IP Multi-tenancy function that is new to IBM Spectrum Virtualize 8.4.2.

This chapter includes the following topics:

- 7.1, “Host attachment overview” on page 444
- 7.2, “Host objects overview” on page 445
- 7.3, “NVMe over Fibre Channel” on page 446
- 7.4, “N_Port ID Virtualization support” on page 447
- 7.5, “IP multi-tenancy” on page 457
- 7.6, “Hosts operations by using the GUI” on page 465
- 7.7, “Performing hosts operations by using CLI” on page 502
- 7.8, “Host attachment practical examples” on page 512
- 7.9, “Container Storage Interface drivers” on page 526

© Copyright IBM Corp. 2021. All rights reserved.

7.1 Host attachment overview

IBM SAN Volume Controller systems support 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)

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, “Planning” on page 59.

Note: Specific host operating systems can be connected directly to IBM SAN Volume Controller storage systems without the use of SAN switches. For more information, see the IBM System Storage Interoperation Center (SSIC).

To enable multiple access paths and correct volume presentation, a host system must have a multipathing driver that is installed, which serves the following purposes:

- Protects from:
  - Fabric path failures, including port failures on IBM SAN Volume Controller system nodes
  - A host bus adapter (HBA) failure (if two HBAs are used)
  - Fabric failures if the host is connected through two HBAs across two separate fabrics
- Enables load balancing across the host HBAs.

For more information about the native operating system multipath drivers that are supported for IBM SAN Volume Controller systems, see the SSIC.

For more information about how to attach specific supported host operating system to the storage systems, see the “Host Attachment” section of IBM Documentation.

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 can be configured by using the graphical user interface (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 Spectrum Virtualize V 8.4.2 supports configuring the following host objects:

- Host
- Host cluster (supported since Version 7.7.1 and later)

A *host cluster object* groups clustered servers and treats them as a single entity. This feature 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 manage host and host cluster objects:

- Commands to provide information about defined hosts and host clusters start with **`ls`** (list):
  - `lshostcluster`
  - `lshostclustermember`
  - `lshostclustervolumemap`
  - `lshost`
  - `lshostiogrp`
  - `lshostiplogin`
  - `lsiscsiauth`
- Commands to define or configure a host object on a storage system start with **`mk`** (make):
  - `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)
- Commands to remove or delete defined host objects from a storage system configuration start with **`rm`** (remove):
  - `rmhostclustermember`
  - `rmhostcluster`
  - `rmvolumehostclustermap`
  - `rmhost`
  - `rmhostiogrp`
  - `rmhostport`
- Commands to change defined host objects start with **`ch`** (change):
  - `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)

- Commands to add a host to a host cluster, to an I/O group, or to add a port to an existing host object begin with **`add`**:
  - `addhostclustermember`
  - `addhostiogrp`
  - `addhostport`

For more information about each command, see IBM Documentation and select **Command-line interface** → **Host commands**.

The instructions to perform basic tasks on host clusters are provided in 7.7, “Performing hosts operations by using CLI” on page 502.

7.3 NVMe over Fibre Channel

IBM SAN Volume Controller systems running IBM Spectrum Virtualize V8.2 or later can be attached to an NVMe host through NVMe-oF, which uses FCP (FC-NVMe) as its underlying transport protocol. This configuration 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, a limit of 512 host objects per I/O group and specific limits for each protocol are imposed, as shown in Table 7-1.

Table 7-1 Defined host object limits per I/O group

SCSI host objects	NVMe host objects	Total host objects
496	16	512

Note: The specifications that are listed in Table 7-1 are the maximum number of each type of host attachment allowed. However, not having the maximum number of one type of host object defined does *not* reduce the total number of host objects allowed.

For example, if 10 NVMe host objects are defined, up to 502 SCSI host objects can be defined. The only hard limit is 16 NVMe host objects per I/O group.

When planning a parallel SCSI and NVMe deployment, refer to the information that is provided in Chapter 2, “Planning” on page 59 because this parallel SCSI and NVMe deployment is resource-intensive, especially if many hosts are deployed.

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 Spectrum Virtualize 8.4.2 also supports HyperSwap for NVMe-oF attached hosts.

Note: IBM Spectrum Virtualize version 8.4.2 uses Asymmetric Namespace Access (ANA), which enables HyperSwap and Non-disruptive 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 non-optimized path, even if the primary site is available.
- The ability to fail over to the secondary site if the primary site is down.

7.4 N_Port ID Virtualization support

The IBM SAN Volume Controller systems use a pair of distinct control modules, which are known as *nodes* that share active/active access to all volumes within the same I/O group. Each node has its own FC worldwide node name (WWNN). Each node'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 the following purposes:

- Internode communication (that is, communication between storage system nodes)
- Back-end controllers communication for external storage virtualization
- Host communications

If a node fails or is removed, its path to the host goes offline. This result requires the host's native operating system multipathing software 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 to *only* 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 IBM Spectrum Virtualize system nodes, each physical WWPN reports up to four virtual WWPNs in addition to the physical one, as listed in Table 7-2.

Table 7-2 IBM Spectrum Virtualize NPIV ports

NPIV port	Port description
Primary port	The WWPN that communicates with back-end storage. It can be used for node to node traffic (local or remote)
Primary SCSI host attach port	The WWPN that communicates with hosts. It is a target port only. It is the primary port, so it is based on this local node's WWNN

NPIV port	Port description
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, so 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 an IBM SAN Volume Controller port when NPIV is enabled.

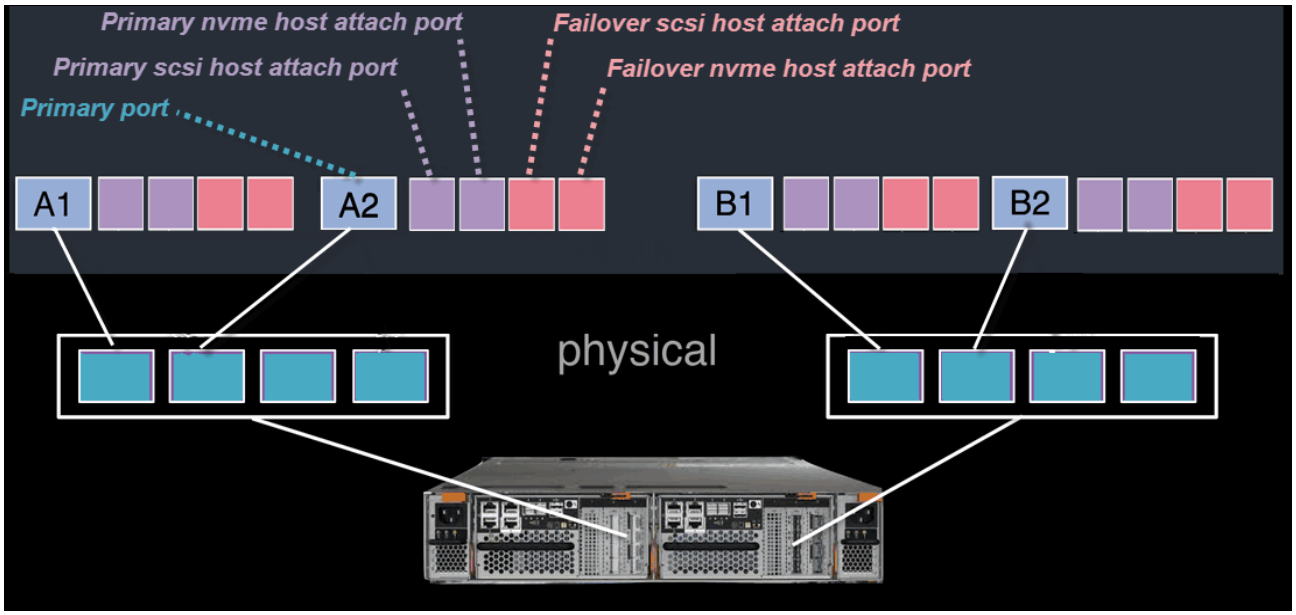


Figure 7-1 Allocation of NPIV virtual WWPN ports per physical port

Figure 7-2 shows what occurs when the partner node fails. The failover host attach ports on the remaining node become active and take on the WWPN of the failed node's *primary* host attach port.

Note: Although Figure 7-2 shows only two ports per node canister in detail, the same situation applies for all physical ports, including the NVMe ports because they use the same NPIV structure, but with the NVMe topology instead of SCSI.

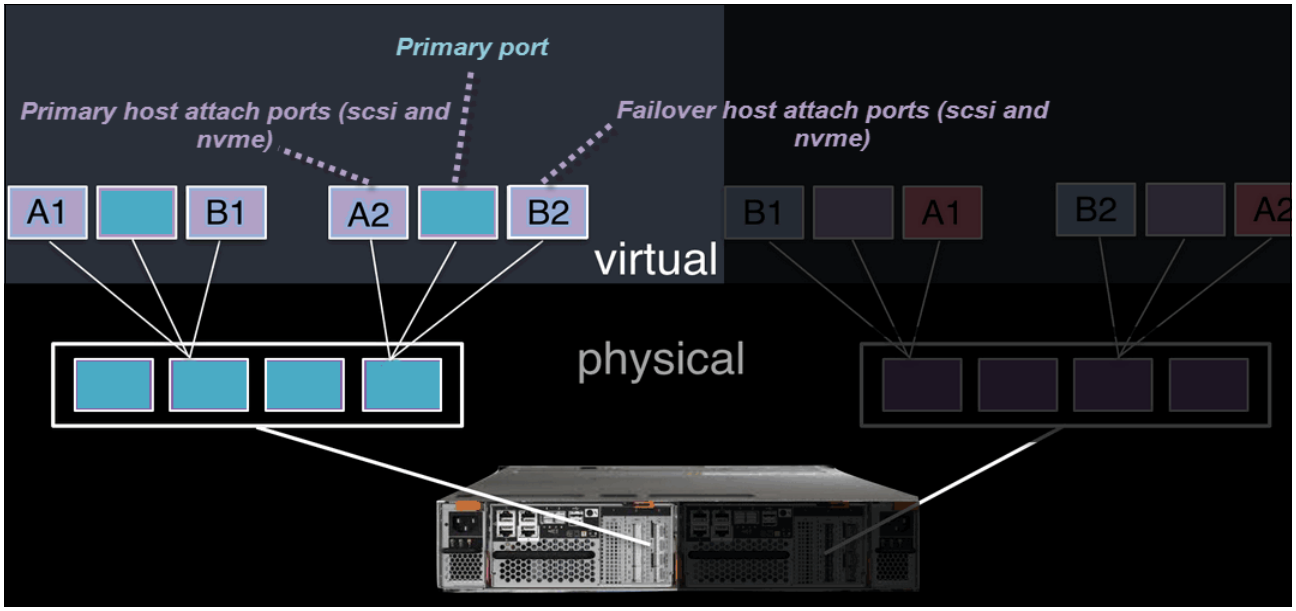


Figure 7-2 Allocation of NPIV virtual WWPN ports per physical port after a node failure

Note: When the Hot Spare Node function is enabled, the NPIV ports move to the partner node in the I/O group in case of node failure. This process makes the ports immediately available on the online node. After the hot spare node joins the I/O group and is ready to process I/O, the NPIV ports automatically move to the hot spare node.

The NPIV mode can be in one of the following states:

- 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: Both 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 through NPIV target ports only.

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 IBM Documentation.

Note: NPIV is supported for FCP and FC-NVMe only. It is *not* supported for the FCoE or iSCSI protocols.

7.4.1 NPIV prerequisites

Before enabling NPIV, the storage system must meet the following prerequisites:

- The system is running IBM Spectrum Virtualize version 7.7 or later.
- Both nodes within an I/O group include identical hardware.
- The FC switches to which the system ports are attached support NPIV and this feature is enabled.
- Node connectivity is done according to “Zoning requirements for N_Port ID virtualization” at IBM Documentation. Both nodes in one I/O group features their equivalent ports that are connected to their equivalent fabrics (switch). For example, port 1 of node1 is the same fabric as port 1 of the node2.

Note: A system running IBM Spectrum Virtualize version 7.7 or later with NPIV enabled as back-end storage for an IBM SAN Volume Controller storage system that is running a version of IBM Spectrum Virtualize that is earlier than 7.7 is *not* supported.

7.4.2 Verifying the NPIV modes state for a new system installation

NPIV is enabled by default on all current IBM SAN Volume Controller systems running IBM Spectrum Virtualize 8.4.2.

However, if you are unsure, complete the following steps to verify:

1. Run the **`lsiogrp`** command to list the I/O groups present in a system, as shown in Example 7-1.

Example 7-1 Listing the I/O groups in a system

```
IBM_2145:ITSO-SV1:superuser>lsiogrp
id name            node_count vdisk_count host_count site_id site_name
0  io_grp0         2          8           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
IBM_2145:ITSO-SV1:superuser>
```

Example 7-1 shows one full I/O group with ID `0`, two nodes in the group, and eight VDisks. 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 with the fctargetportmode field

```
IBM_2145:ITSO-SV1:superuser>lsiogrp 0 | grep fctargetportmode
fctargetportmode enabled
IBM_2145:ITSO-SV1:superuser>
```

3. Run the **lstargetportfc** command for a list of the available virtual WWPNs (see Example 7-3). The `host_io_permitted` and `virtualized` columns should display a status of `yes`, which indicates that those WWPNs belong to a primary host attach port and should must be used when zoning the hosts to the SAN Volume Controller system.

Example 7-3 Listing the virtual WWPNs

```
IBM_2145:ITSO-SV1: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  no                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  no                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  no                no          scsi
8  500507680112A288 500507680100A288 3       1             1               010501  yes               yes         scsi
9  500507680114A288 500507680100A288 3       1             1               010502  yes               yes         nvme
...
IBM_2145:ITSO-SV1:superuser>
```

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 where the virtualized ports are marked in **bold**).

> **Note:** If supported on the host, NVMe ports can be used in the zoning. However, the IBM SAN Volume Controller system supports only one topology per host. NVMe and SCSI ports cannot be used together for the same host object.

5. If the status of **fctargetportmode** is `disabled` and this installation is a new, run the **chiogrp** command to set NPIV mode to first `transitional` state and then to `enabled`, as shown in Example 7-4.

Example 7-4 Changing the NPIV mode to enabled

```
IBM_2145:IBM Redbook SVC:superuser>chiogrp -fctargetportmode transitional 0
IBM_2145:IBM Redbook SVC:superuser>chiogrp -fctargetportmode enabled 0
```

6. Verify the NPIV enablement by checking the `fctargetportmode` field, as shown in Example 7-5.

Example 7-5 NPIV enablement verification

```
IBM_2145:ITSO-SV1:superuser>lsiogrp 0
id 0
name io_grp0
node_count 2
vdisk_count 8
host_count 0
flash_copy_total_memory 20.0MB
flash_copy_free_memory 19.9MB
remote_copy_total_memory 20.0MB
remote_copy_free_memory 20.0MB
mirroring_total_memory 20.0MB
mirroring_free_memory 19.9MB
raid_total_memory 40.0MB
raid_free_memory 40.0MB
maintenance no
compression_active no
```

```
accessible_vdisk_count 8
compression_supported no
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:2145.0000020067214511.iogroup0
```

7. Configure zones for hosts by using the primary host attach ports (virtual WWPNs) of the IBM SAN Volume Controller IBM Spectrum Virtualize ports, as shown in **bold** in the output that is shown in Example 7-3 on page 451.

7.4.3 Enabling NPIV on a system

If an IBM SAN Volume Controller system running back-level code (Spectrum Virtualize Version 7.7.1 or earlier) is upgraded to the latest version, the NPIV feature must be enabled manually because this process 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 features 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 SAN Volume Controller 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 7-6. Those virtualized ports are not yet enabled for host I/O communication (see the `host_io_permitted` column).

Example 7-6 lstargetportfc command

```
IBM_2145:ITSO-SV1: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
...
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
IBM_2145:ITSO-SV1:superuser>
```

4. Enable transitional mode for NPIV on the system (see Example 7-7) to enable virtualized ports for host IO communication and still keep access for the hosts that are using hardware defined ports (not in bold in Example 7-6).

Example 7-7 NPIV in transitional mode

```
IBM_2145:ITSO-SV1:superuser>chiogrp -fctargetportmode transitional 0
IBM_2145:ITSO-SV1:superuser>lsiogrp 0 | grep fctargetportmode
fctargetportmode transitional
IBM_2145:ITSO-SV1:superuser>
```

Alternatively, to activate NPIV in transitional mode by using the GUI, go to the GUI and select **Settings** → **System** → **I/O Groups**, as shown in Figure 7-3.

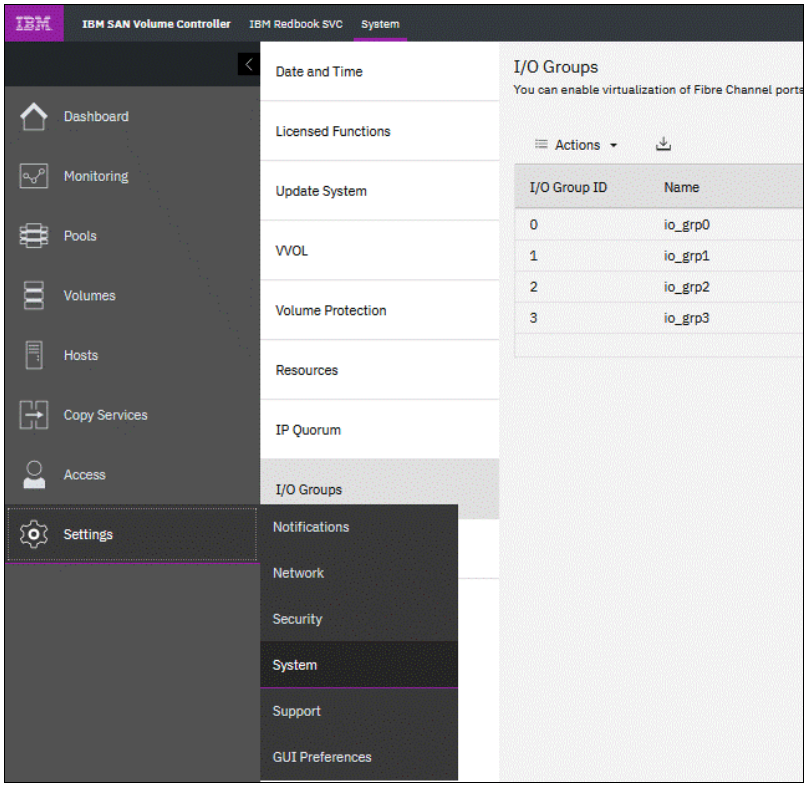


Figure 7-3 I/O Groups menu

Then, check the current NPIV setting by viewing the NPIV column, which shows `disabled` if NPIV is not enabled. Select the I/O group on which you want to enable NPIV and then, select **Actions** → **Change NPIV Settings**, as shown in Figure 7-4.

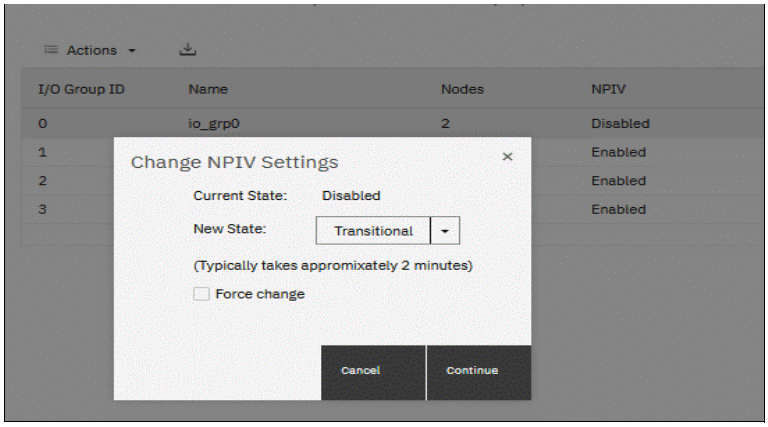


Figure 7-4 Enabling NPIV to transitional state

Select **Continue** to complete the process and enable Transitional Mode.

5. Ensure that host traffic is now allowed on the primary host attach WWPNs (virtual WWPNs), as shown in **bold** in Example 7-8.

Example 7-8 Host-attached WWPNs (virtual WWPNs) permitting host traffic

```
IBM_2145:ITSO-SV1: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
IBM_2145:ITSO-SV1:superuser>
```

6. Add the primary host attach ports (virtual WWPNs) to the host zones, but do not remove the IBM SAN Volume Controller WWPNs. Example 7-9 shows a host zone that was added to the Primary Port WWPNs of the IBM SAN Volume Controller nodes.

Example 7-9 Legacy host zone

```
zone: WINDOWS_HOST_01_IBM_ITSOSV1
                10:00:00:05:1e:0f:81:cc
                50:05:07:68:01:40:A2:88
                50:05:07:68:01:10:A2:88
                50:05:07:68:0C:11:00:09
                50:05:07:68:0C:13:00:09
```

Example 7-10 shows that the primary host attach ports (virtual WWPNs) were added to the example host zone without disrupting its availability.

Example 7-10 Transitional host zone

```
zone: WINDOWS_HOST_01_IBM_ITSOSV1
                10:00:00:05:1e:0f:81:cc
                50:05:07:68:01:40:A2:88
                50:05:07:68:01:10:A2:88
                50:05:07:68:0C:11:00:09
                50:05:07:68:0C:13:00:09
                50:05:07:68:01:42:A2:88
                50:05:07:68:01:12:A2:88
                50:05:07:68:0C:15:00:09
                50:05:07:68:0C:17:00:09
```

7. After activating transitional zoning in the fabrics, validate that the host uses the new NPIV ports for host I/O. Example 7-11 shows the pathing to the host before and after adding the new host attach ports by using the old SDDDSM multipathing driver. The select count increases on the new paths and stopped on the old path.

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 7-11 Host device pathing: Before and after

```
C:\Program Files\IBM\SDDDSM>datapath query device

Total Devices : 1


DEV#:   0  DEVICE NAME: Disk2 Part0  TYPE: 2145       POLICY: OPTIMIZED
SERIAL: 60050768OC838020E800000000000002        LUN SIZE: 20.0GB
============================================================================
Path#            Adapter/Hard Disk        State  Mode       Select     Errors
    0     Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL      10626          0
    1     Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL      10425          0
    2 *   Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL          0          0
    3 *   Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL          0          0
    4     Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL    1128804          0
    5     Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL    1129439          0
    6 *   Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL          0          0
    7 *   Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL          0          0

C:\Program Files\IBM\SDDDSM>datapath query device

Total Devices : 1


DEV#:   0  DEVICE NAME: Disk2 Part0  TYPE: 2145       POLICY: OPTIMIZED
SERIAL: 60050768OC838020E800000000000002        LUN SIZE: 20.0GB
============================================================================
Path#            Adapter/Hard Disk        State  Mode       Select     Errors
    0 *   Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL      10630          1
    1 *   Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL      10427          1
```

```
 2 *   Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL          0          0
 3 *   Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL          0          0
 4 *   Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL    1128809          2
 5 *   Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL    1129445          1
 6 *   Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL          0          0
 7 *   Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL          0          0
 8     Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL      76312          0
 9     Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL      76123          0
10 *   Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL          0          0
11 *   Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL          0          0
12     Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL        623          0
13     Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL        610          0
14 *   Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL          0          0
15 *   Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL          0          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 show `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 might 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 7-12.

Example 7-12 Enabling NPIV

```
IBM_2145:ITSO-SV1: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**. The NPIV Settings window opens, as shown in Figure 7-5.

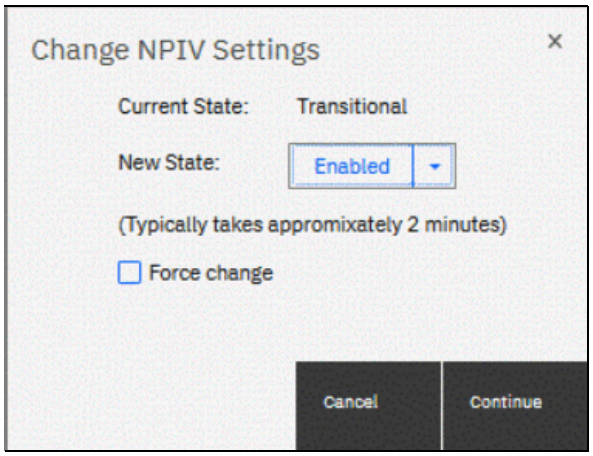


Figure 7-5 Changing NPIV settings - Enable

9. Click **Continue,** NPIV is enabled on the I/O group.

   To complete the NPIV implementation, modify the host zones to remove the old primary attach port WWPNs. Example 7-13 shows final zoning with the host HBA and the IBM SAN Volume Controller virtual WWPNs.

Example 7-13 Final host zone

```
zone:  WINDOWS_HOST_01_IBM_ITSOSV1
                10:00:00:05:1e:0f:81:cc
                50:05:07:68:01:42:A2:88
                50:05:07:68:01:12:A2:88
                50:05:07:68:0C:15:00:09
                50:05:07:68:0C:17:00:09
```

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 IO and force flag not set.
```

7.5 IP multi-tenancy

A new feature in IBM Spectrum Virtualize 8.4.2 is the ability to have multiple IP addresses defined per port for use by an Ethernet host. This definition includes all servers that use iSCSI and iSER host attachment protocols.

Use cases for this feature include the following examples:

- Allowing multiple host tenants to share a single IBM SAN Volume Controller 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 are associated with the specific traffic types. The IBM SAN Volume Controller storage systems support portsets for host attachment, backend storage connectivity, and IP replication traffic.

7.5.1 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 Limitations for configuring IP addresses

Limit	Description
Maximum number of portset objects	64 per system
Maximum number of IP address objects (includes shared IP address objects)	2048 per system
Maximum number of IP addresses objects per port	64 (IPv4 or IPv6)[a]
Maximum number of routable IP addresses objects per port	1 (IPv4 or IPv6)
Maximum number of IP address objects per node	256 per node
Maximum number of IP addresses objects per node per portset	► Host portsets - 4 per node per portset ► Replication portsets - 1 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 - 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 1 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]

Type of HBA	Limit
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 any more IP addresses or VLANs after this limit is reached.

[b] If you are using iSER for clustering (node-to-node) connections, you need to lower the number of IP addresses per port based on the number that you are currently using for clustering.

7.5.2 Prerequisites

Specific prerequisites must be met before creating a portset and assigning it to an IP address, a host object, or an IP partnership.

Requirements for portsets

In general, the portsets features 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 that are 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. It 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.

Requirements for host portsets

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 4 IP addresses per node.
- A single portset can contain IPv4, IPv6, or a mix of both 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.

Requirements for replication portsets

Replication portsets for IP replication have 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 can also 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 remote a 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.

Requirements for a storage portset

The system supports maximum of 1 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 both IPv4 and IPv6 addresses.

Requirements for ownership group portsets

Portsets can be assigned to ownership groups to restrict access to specific set of users, including the following examples:

- Restricted users can create a portset and IP addresses and assign the hosts to portset.
- Restricted users cannot view portsets that are assigned to a different ownership group.
- Restricted users cannot own replication or storage type portset.

7.5.3 Configuring the portset

You can configure the portset from the GUI or command line.

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).

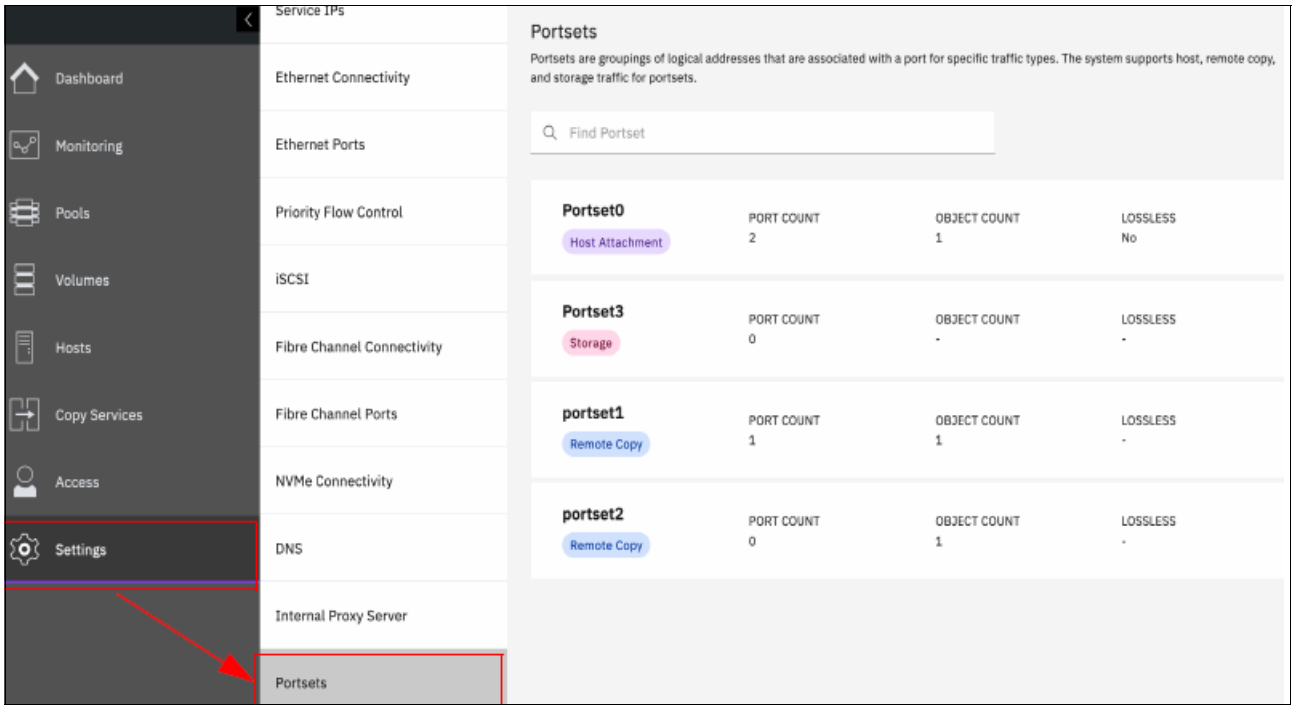


Figure 7-6 Selecting the portset option

2. Select **Create Portsets** (see Figure 7-7).

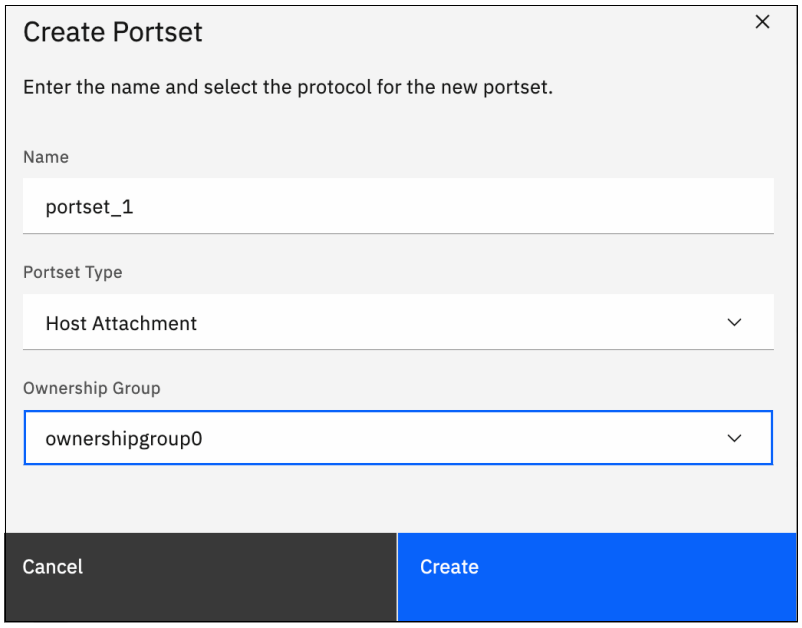


Figure 7-7 Creating a portset

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
```

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. 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 "portset_2" for node1 in ownershipgroup0

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. All are optional values and default host type portset will be created.

7.5.4 Assigning an IP address to a portset

You can assign an IP address to a Portset from the GUI or command line.

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 (see Figure 7-9):

1. Select **Settings** → **Network** → **Ethernet Ports**. See Figure 7-9.

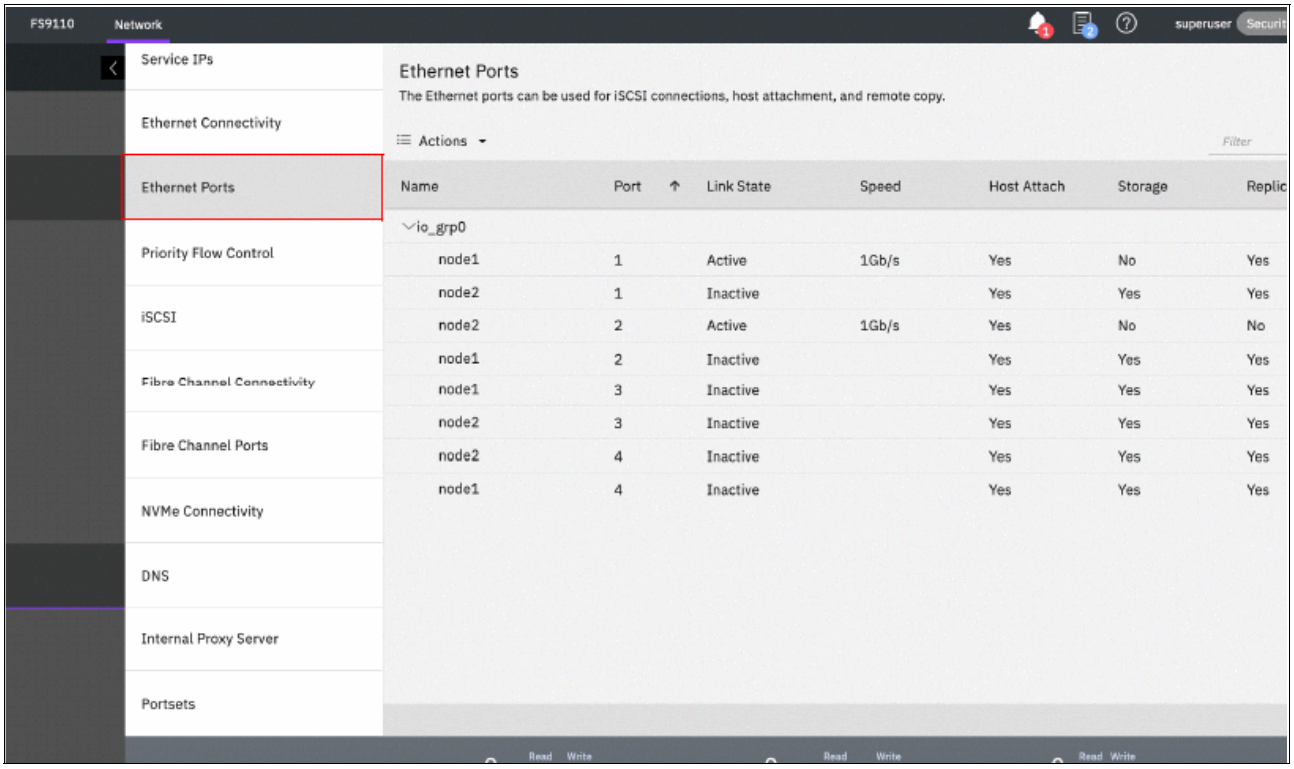


Figure 7-9 Selecting Ethernet ports

2. Right-click the port and select **Manage IP addresses** (see Figure 7-10 on page 464).
3. On Manage IP Addresses page, select **Add IP Address**. Enter the relevant information for the IP address that are being added to the selected port:
   - IP Address: Enter the IP address that is associated with the selected port.
   - Type: Select the IP protocol version of the IP address.
   - Subnet Mask or Prefix: Enter the subnet mask for the IPv4 addresses or enter the prefix for IPv6 addresses.
   - VLAN: Enter the corresponding VLAN tag that this IP address belongs to.
   - Portset: Select the name or ID of the portset and ensure that 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 corresponding portsets.

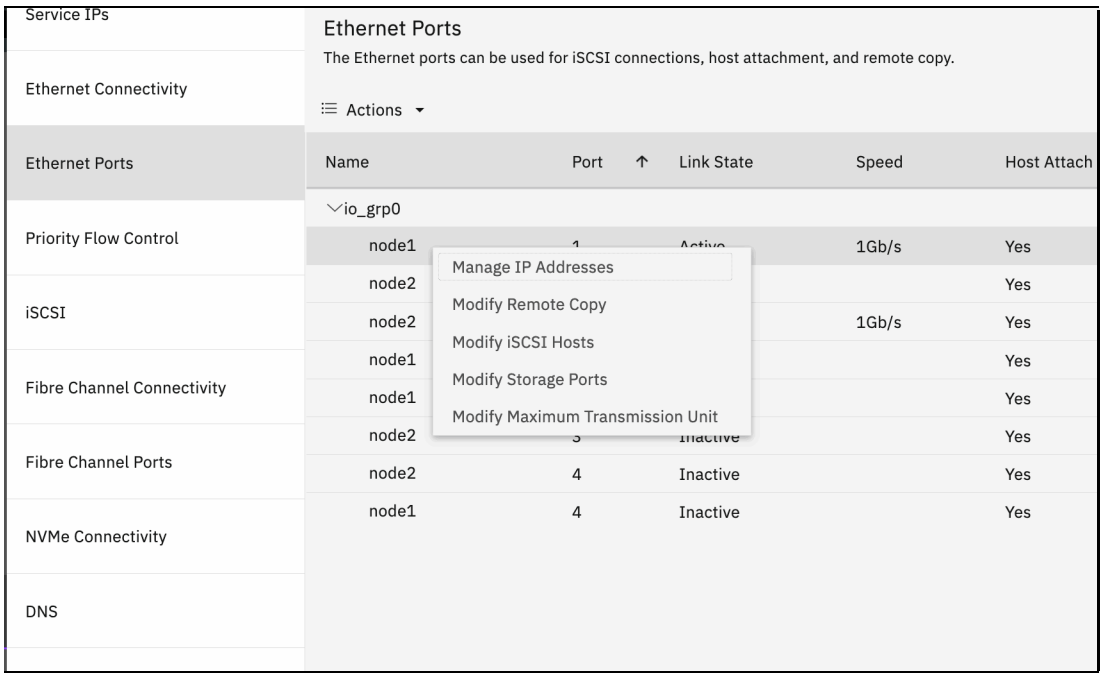


Figure 7-10 After right-clicking the suitable node and port

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 **`-ip`** parameter. This address is assigned to the portset and more can be added to the portset through 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

7.5.5 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 `portset_id` or `portset_name`.

7.6 Hosts operations by using the GUI

This section describes how to perform the following host operations by using IBM Spectrum Virtualize GUI:

- Creating hosts
- Advanced host administration
- Adding and deleting host ports
- Host mappings overview

7.6.1 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 IBM Documentation.

To create a host, complete the following steps:

1. Open the host configuration window by clicking **Hosts** (see Figure 7-12).

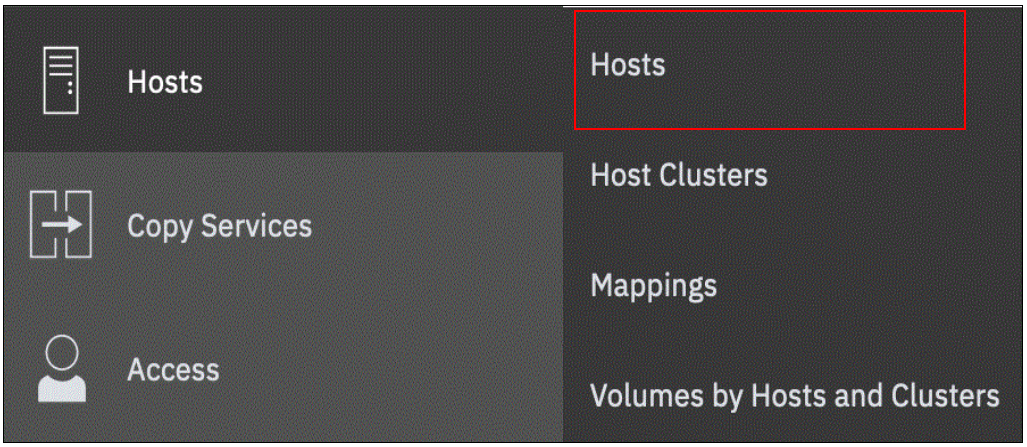


Figure 7-12 Opening the Hosts window

2. To create a host, click **Add Host**. For more information about creating an FC host, see “Creating Fibre Channel hosts objects” on page 466. For more information about creating an iSCSI host, see “Preparing for iSCSI connection” on page 469.

Creating Fibre Channel hosts objects

To create FC hosts, complete the following steps:

1. Select **Fibre Channel** in the Host Connections list and enter the suitable information in the configuration fields (see Figure 7-13).

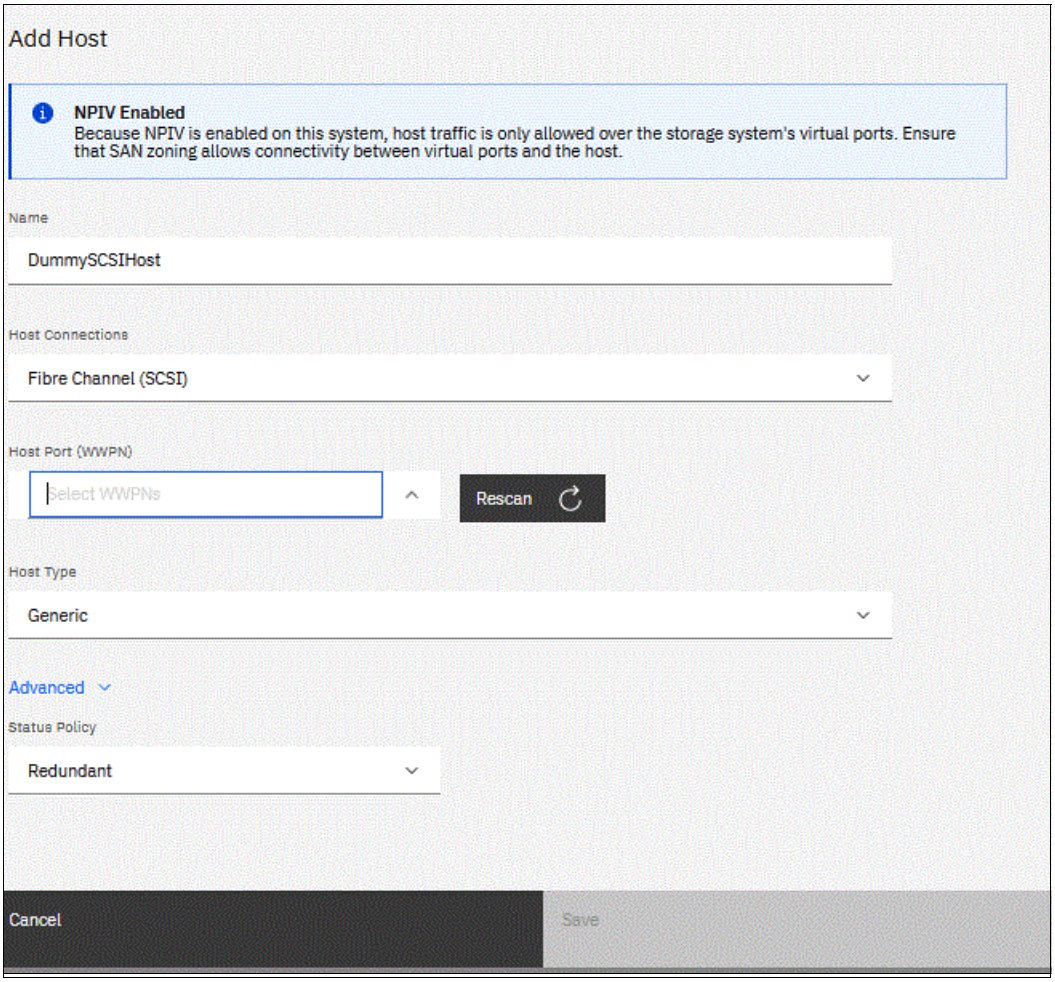


Figure 7-13 Fibre Channel host configuration

2. Enter a host name and click the **Host Port** menu to get a list of all discovered WWPNs (see Figure 7-14).

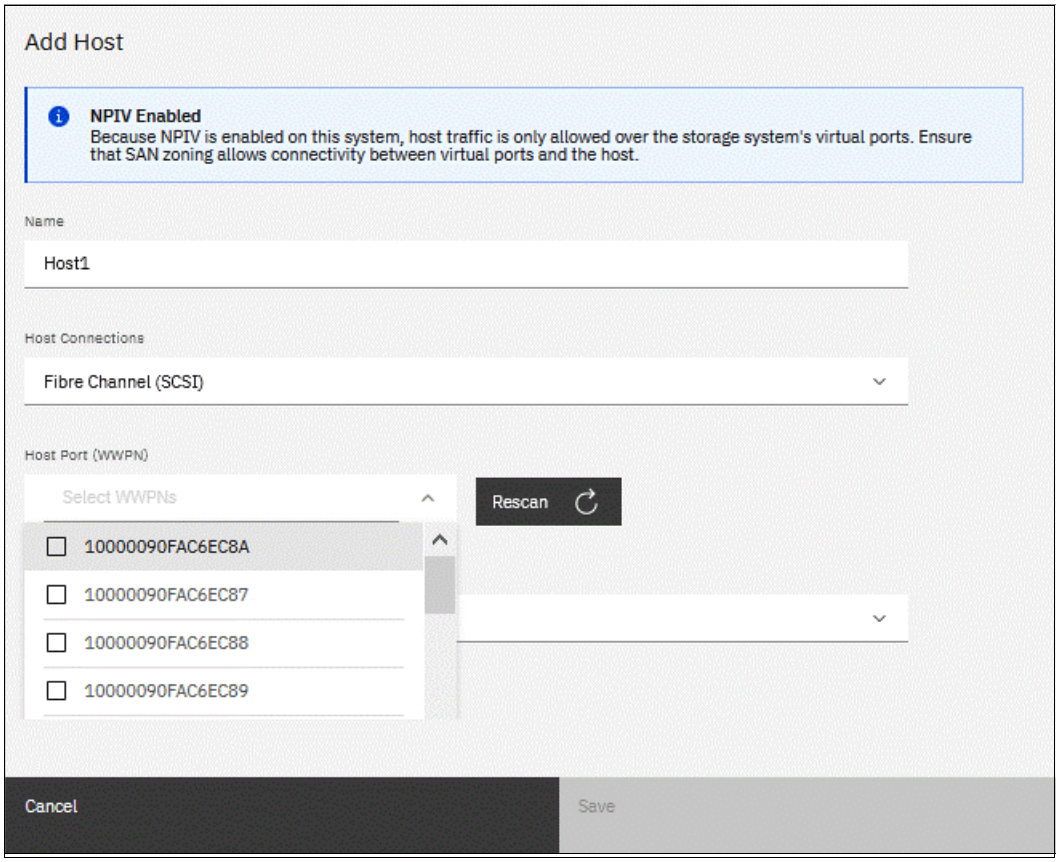


Figure 7-14 Setting up Host object WWPNs

3. Select one or more host WWPNs from the list. These WWPNs are visible on the IBM SAN Volume Controller system if the hosts are correctly zoned and presented.

   If the host information does not appear on the list, scan for new paths as required by the respective operating system and click the **Rescan** icon next to the WWPN box. If they still do not appear, check the SAN zoning and ensure that hosts are properly connected and running. Repeat the scan.

   Creating offline hosts: To create hosts that are offline or not currently connected, enter the WWPNs manually into the **Host Ports** field to add them to the list.

4. To add other ports to the Host, select the wanted WWPNs from the list and add them to the suitable host.

5. If creating a Hewlett-Packard UNIX (HP-UX) or Target Port Group Support (TPGS) host, click the **Host type** list (see Figure 7-15). Select your host type. If the specific host type is not listed, select **Generic**.

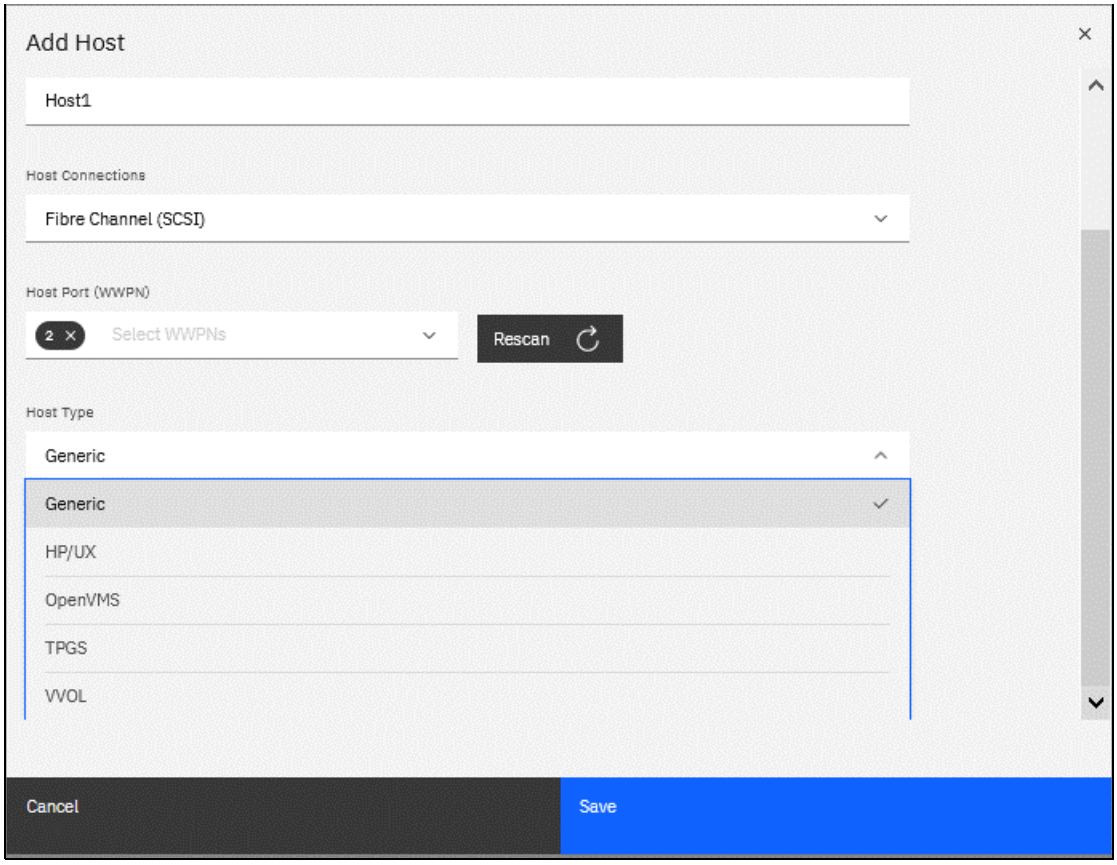


Figure 7-15 Host types selection

6. If setting up object-based access control (OBAC) as described in Chapter 11, “Ownership groups” on page 781, select the **Advanced** section and choose the suitable ownership group from the **Ownership Group** menu, as shown in Figure 7-16.

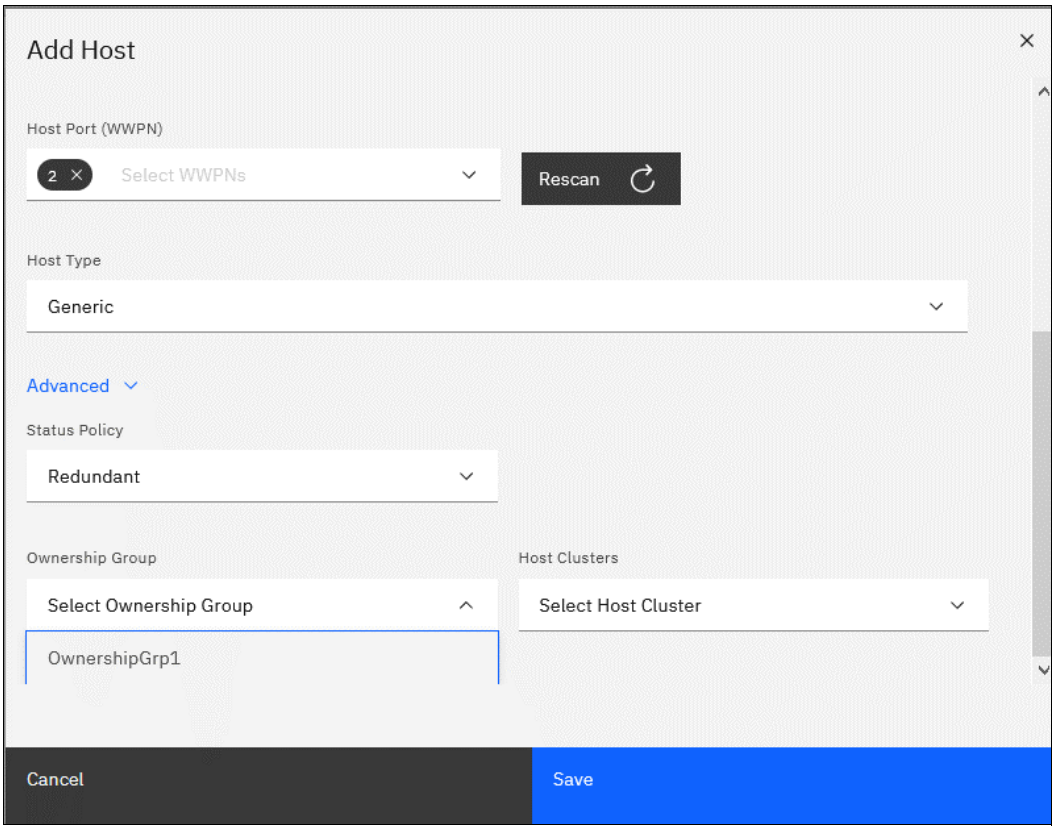


Figure 7-16 Adding the new host to an ownership group

7. Click **Save** to create the host object.
8. Repeat these steps for all of your FC hosts. Figure 7-17 shows the All Hosts view after creating a host.

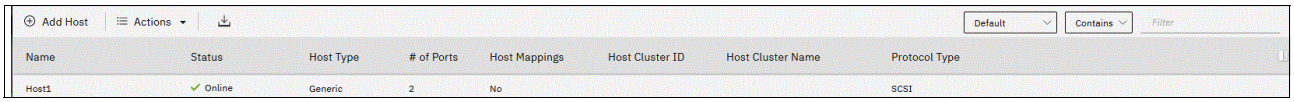


Figure 7-17 Hosts view after creating a host

Volumes can be created and mapped to the defined FC hosts as described Chapter 6 “Volumes”.

Preparing for iSCSI connection

Before creating iSCSI host objects on IBM SAN Volume Controller systems, ensure that iSCSI connectivity was correctly configured on the host initiators. Configuring the iSCSI connectivity is based on host system and operating system.

To enable iSCSI connectivity, complete the following steps:

1. Select **Settings** → **Network** and click the **iSCSI** tab.
2. In the iSCSI Configuration window, enter or modify the system name, node names, and provide an optional iSCSI Alias for each node as needed (see Figure 7-18 on page 470).

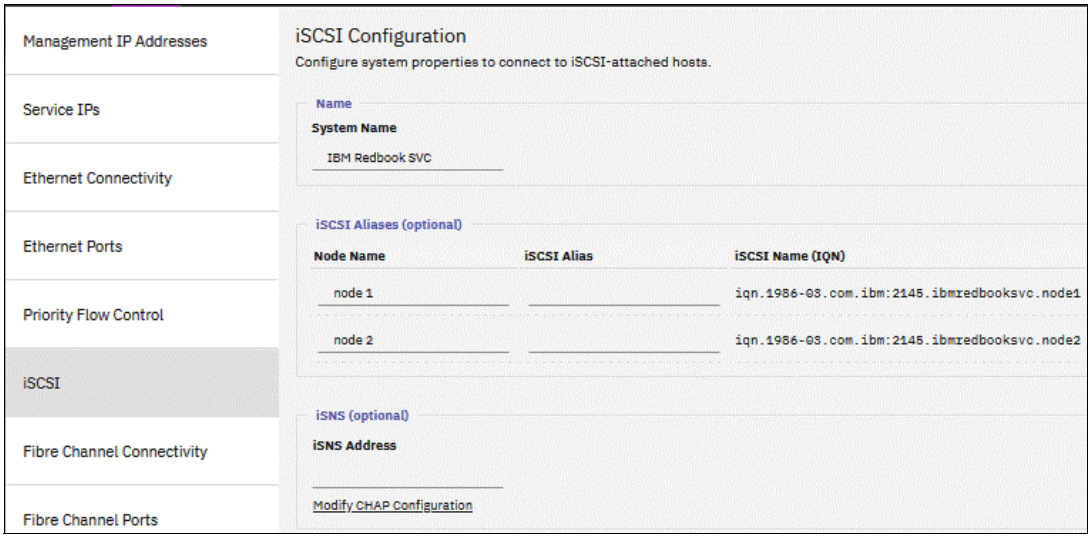


Figure 7-18 iSCSI Configuration view

A warning message about renaming the nodes is shown in Figure 7-19.

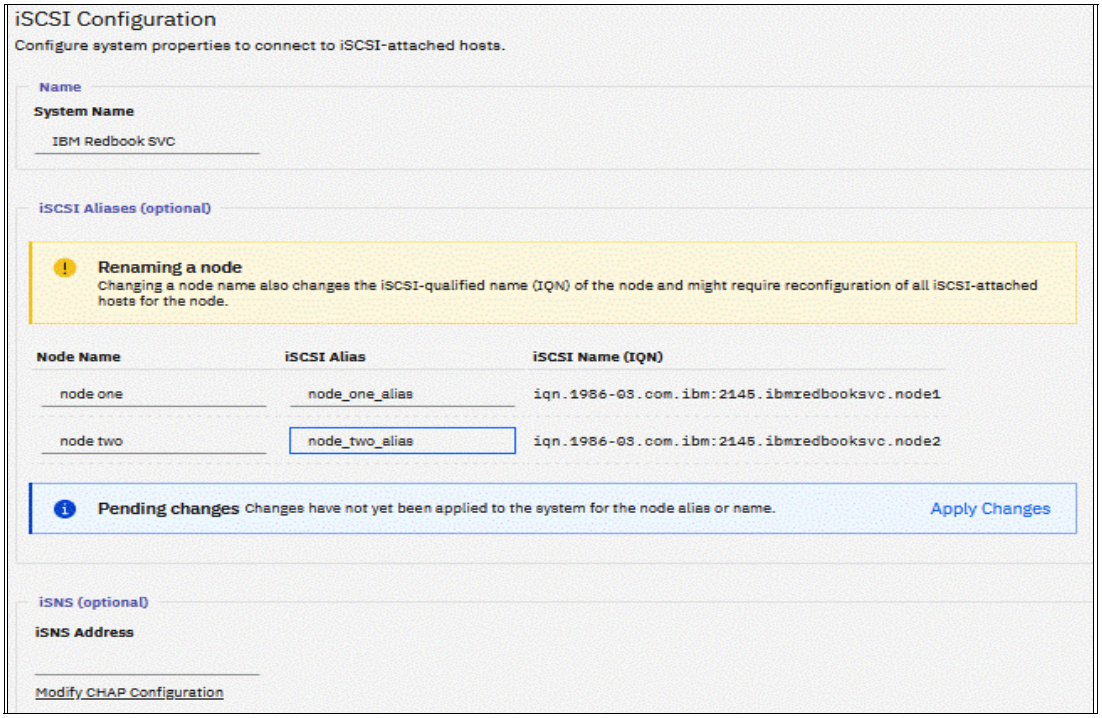


Figure 7-19 Applying modifications to iSCSI configuration

3. Before continuing, select **Apply Changes** after the prompt to accept any modifications that are made.

   If suitable the current environment, the iSNS addresses and Challenge Handshake Authentication Protocol (CHAP) can be configured in the lower left corner from the iSCSI Configuration window.

Note: 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 SAN Volume Controller system does not allow it to perform I/O to volumes.

4. Configure the Ethernet ports that are used for iSCSI communication by selecting **Settings** → **Network**. Click the **Ethernet Ports** tab to see a list of the available ports and their corresponding IP addresses (see Figure 7-20).

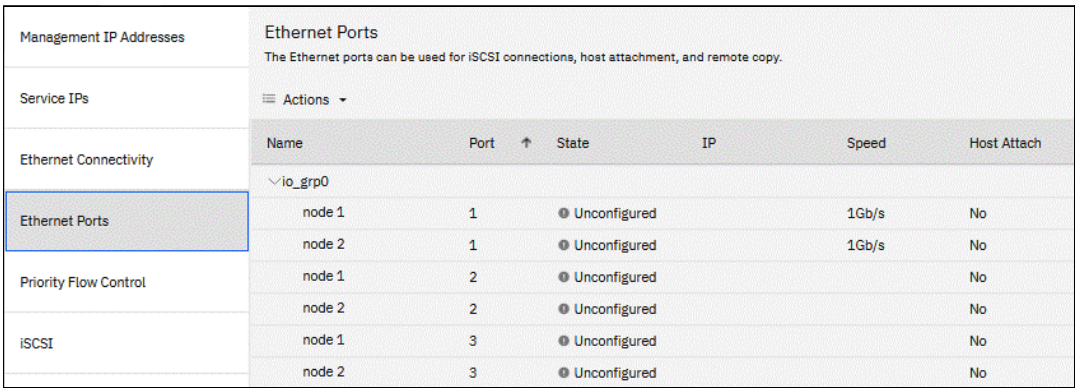


Figure 7-20 Ethernet ports configuration view

5. Select the port to set the iSCSI IP information and then, select **Actions** → **Manage IP Addresses** (see Figure 7-21).

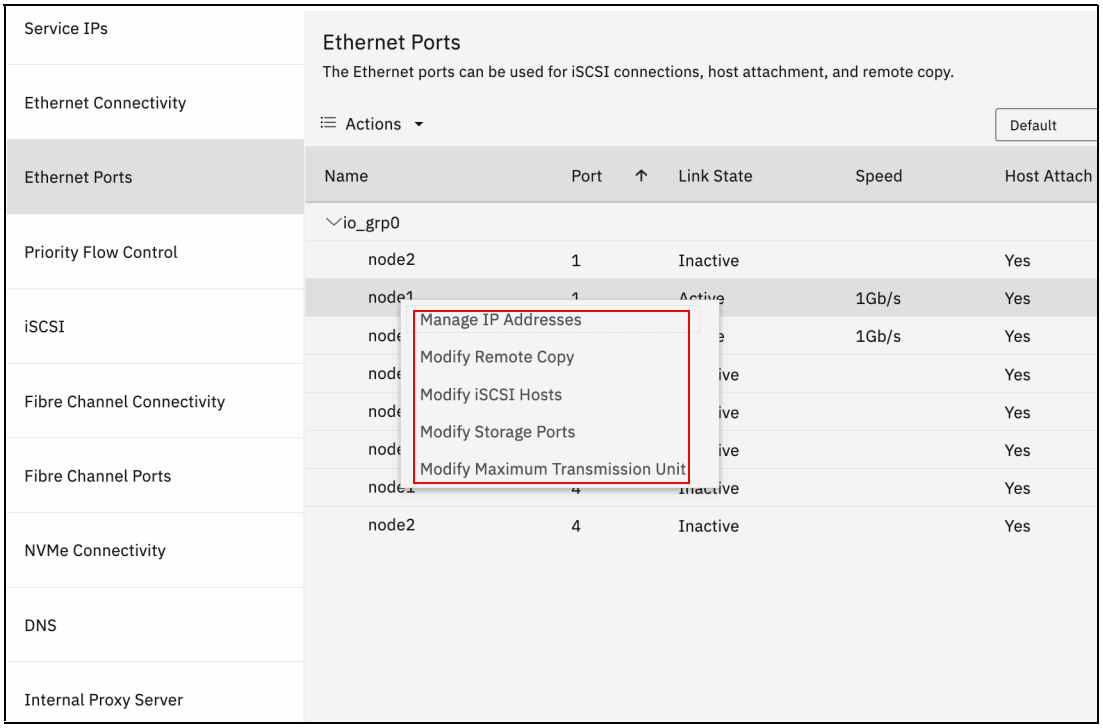


Figure 7-21 Select Manage IP Addresses from the “Actions” menu

As shown in Figure 7-22, IP addresses can be added, deleted, or modified.

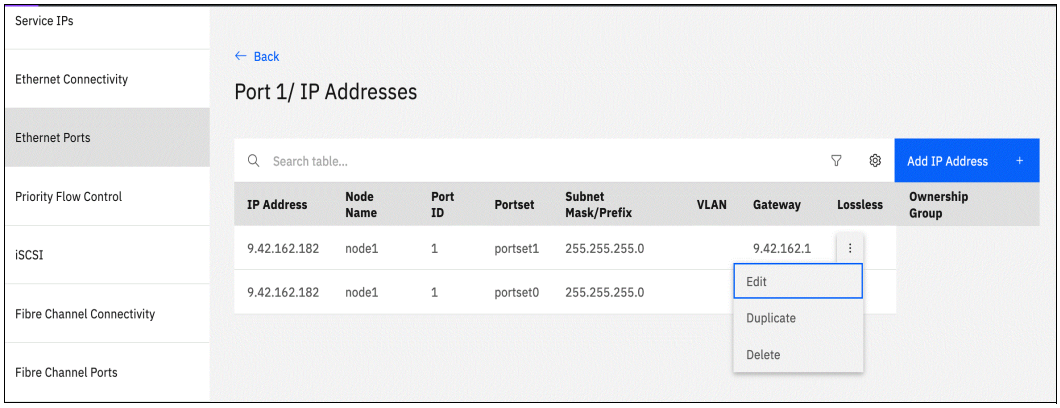


Figure 7-22 Verifying, editing or modifying the IP address

6. After the ports are configured, multiple options are available from the Actions drop-down, as shown in Figure 7-23. For example, to disable any interfaces that you do not want to be used for host connections and might be used for replication only, select the configured port, and then, select **Actions** (or click-right while hovering over the chosen port) → **Modify iSCSI Hosts** (see Figure 7-23).

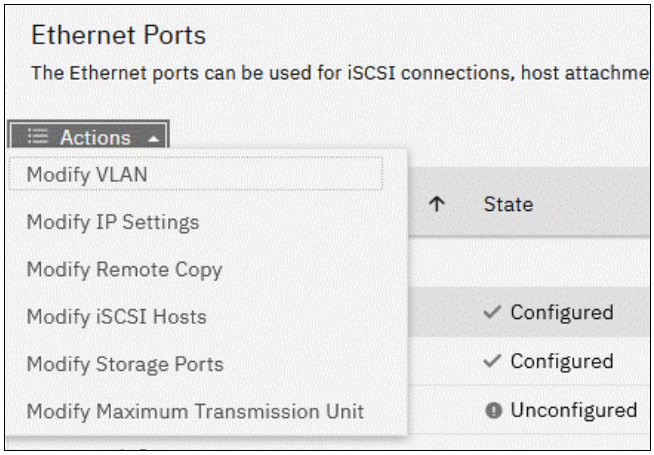


Figure 7-23 Available actions with configured ports

7. Make any necessary changes as prompted in the dialog box that is shown in Figure 7-24 and click **Modify**.

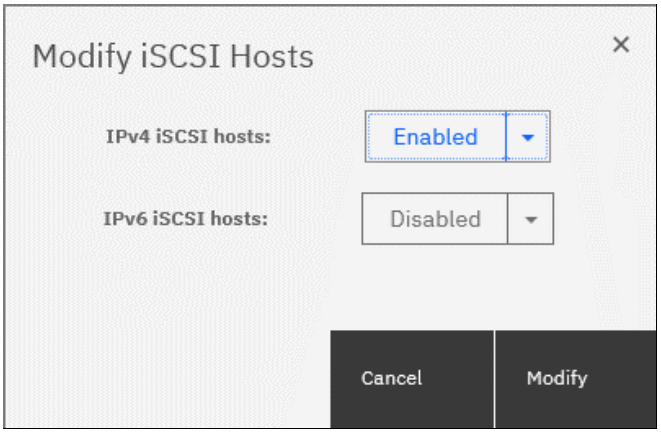


Figure 7-24 Modifying the port for host connectivity

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 7-25.

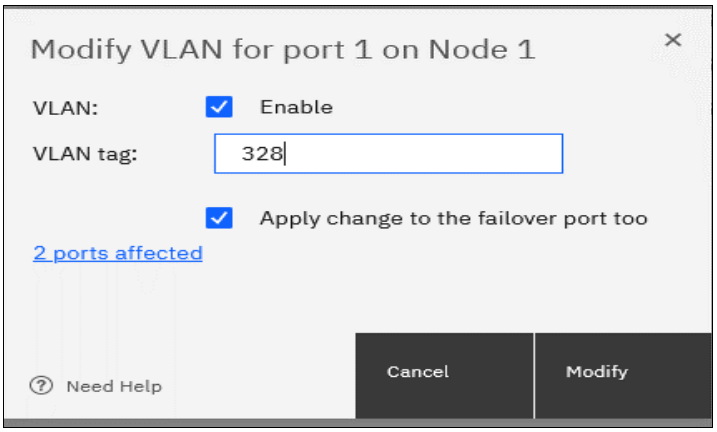


Figure 7-25 Modifying the VLAN settings

The system notification states that at least two ports are affected. To view the details, click **2 ports affected** (see Figure 7-26). Make any necessary changes and click **Modify**.

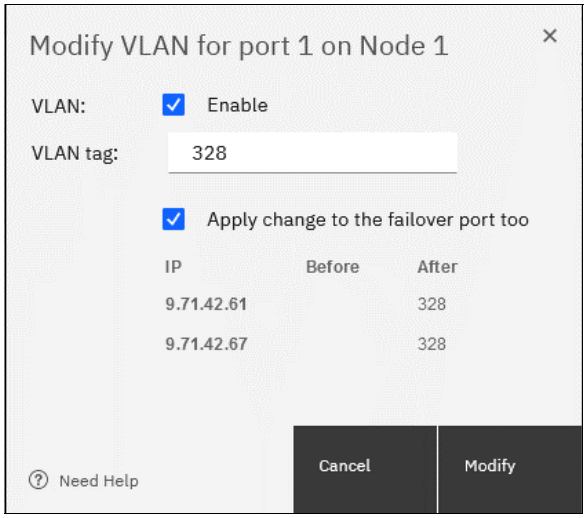


Figure 7-26 VLAN setting configuration details

The IBM SAN Volume Controller system is now configured and ready for iSCSI host use. Make a note of the initiator IQN names of the IBM SAN Volume Controller nodes (see Figure 7-19 on page 470) because they are necessary to configure access from host to the storage.

For more information about how to create volumes and map them to a host, see Chapter 6, “Creating and provisioning volumes” on page 331.

Creating iSCSI hosts

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 fails. This design reduces the need for multipathing support in the iSCSI host.
- The iSCSI Qualified Name (IQN) of the host is added to an IBM Spectrum Virtualize host object in the same way that you add FC WWPNs. How to get IQN from the host see practical examples 7.8.3, “iSCSI host connectivity and capacity allocation” on page 516.
- Host objects can have WWPNs and IQNs.
- Standard iSCSI host connection procedures can be used to discover and configure IBM Spectrum Virtualize as an iSCSI target.
- IBM Spectrum Virtualize supports the Challenge Handshake Authentication Protocol (CHAP) authentication methods for iSCSI.
- The name `iqn.1986-03.com.ibm:2145.<cluster_name>.<node_name>` is the IQN for an IBM Spectrum Virtualize node. Because the IQN contains the clustered system name and the node name, it is important not to change these names after iSCSI is deployed.
- Each node can be given an iSCSI alias as an alternative to the IQN.

To create iSCSI hosts, 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 7-27). Choose **iSCSI** in **Host Connections** list. Note that the CHAP authentication trigger is on; therefore, the fields for CHAP credentials are shown in the interface. If CHAP is not used, turn off the trigger.

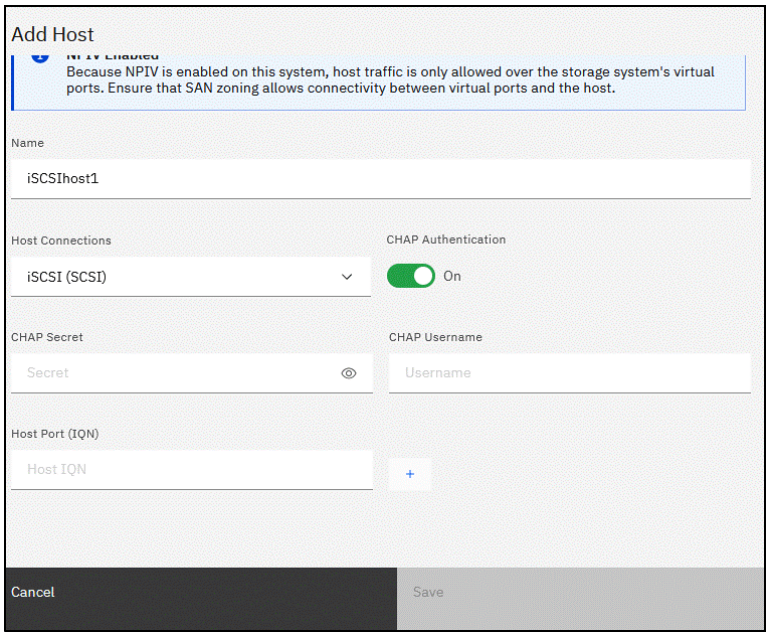


Figure 7-27 Creating iSCSI host object

2. Enter CHAP authentication credentials (if suitable), 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 being 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 7-28 shows the Hosts view window after creating the FC host and iSCSI host.

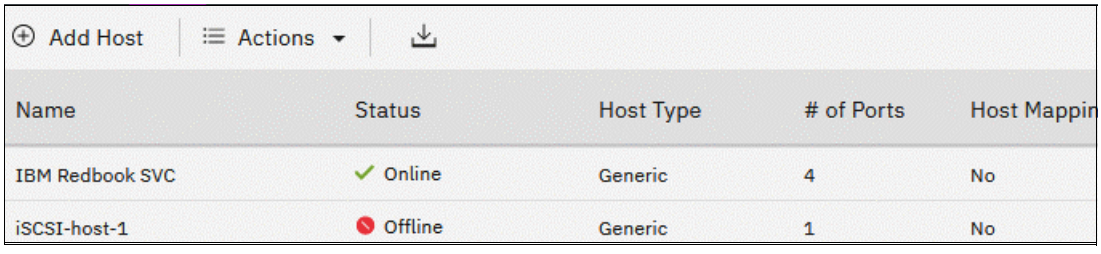


Figure 7-28 Defined Hosts list

Creating NVMe hosts objects

The process for creating NVMe hosts is similar to that of creating SCSI FC hosts, except that it uses host port NQN instead of WWPNs.

To configure an NVMe host object, complete the following steps:

1. Go to the Host view by selecting **Hosts** → **Hosts** and clicking **Add Host**. Then, in the **Host connections** menu, select **Fibre Channel (NVMe)**, as shown in Figure 7-29.

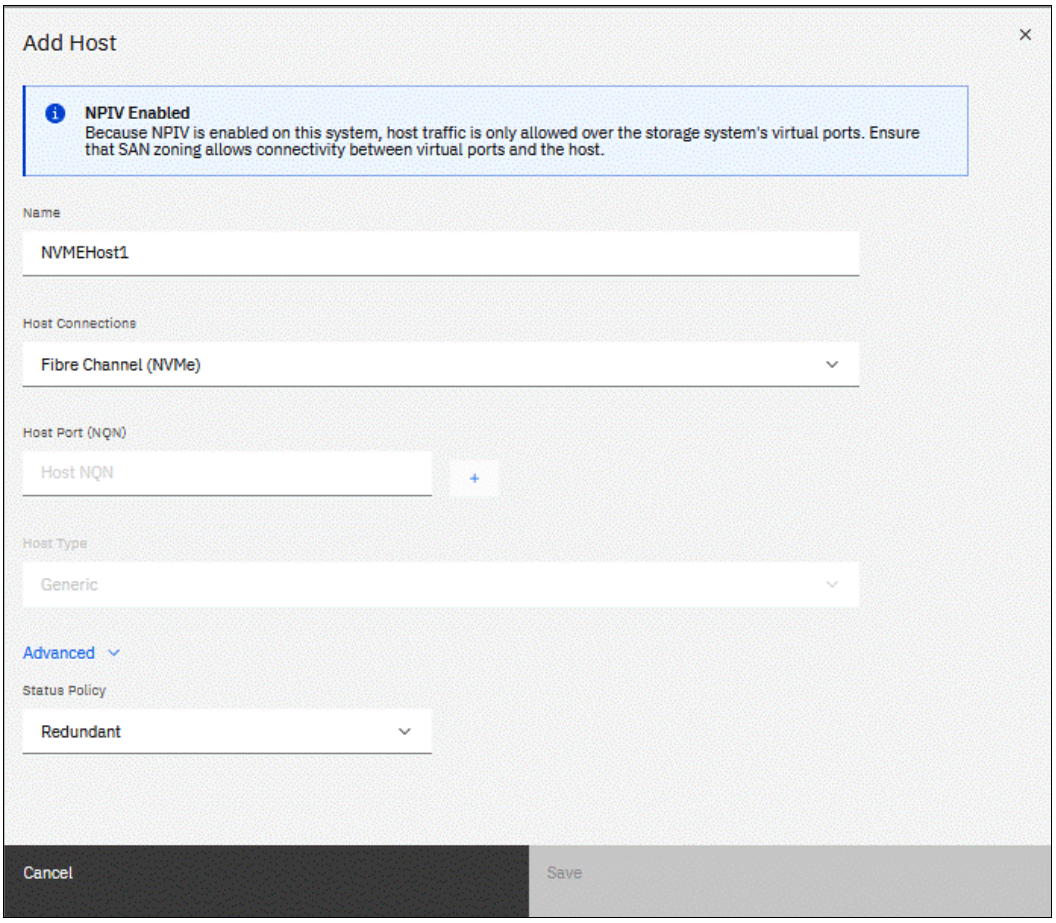


Figure 7-29 Host object configuration: NVMe over Fibre Channel connection

2. Enter the host name and the NVMe Qualified Name (NQN) of the host, as shown in Figure 7-30. For more information about how to obtain the host NQN, see 7.8.4, “NVMe over Fabric host connectivity example” on page 520.

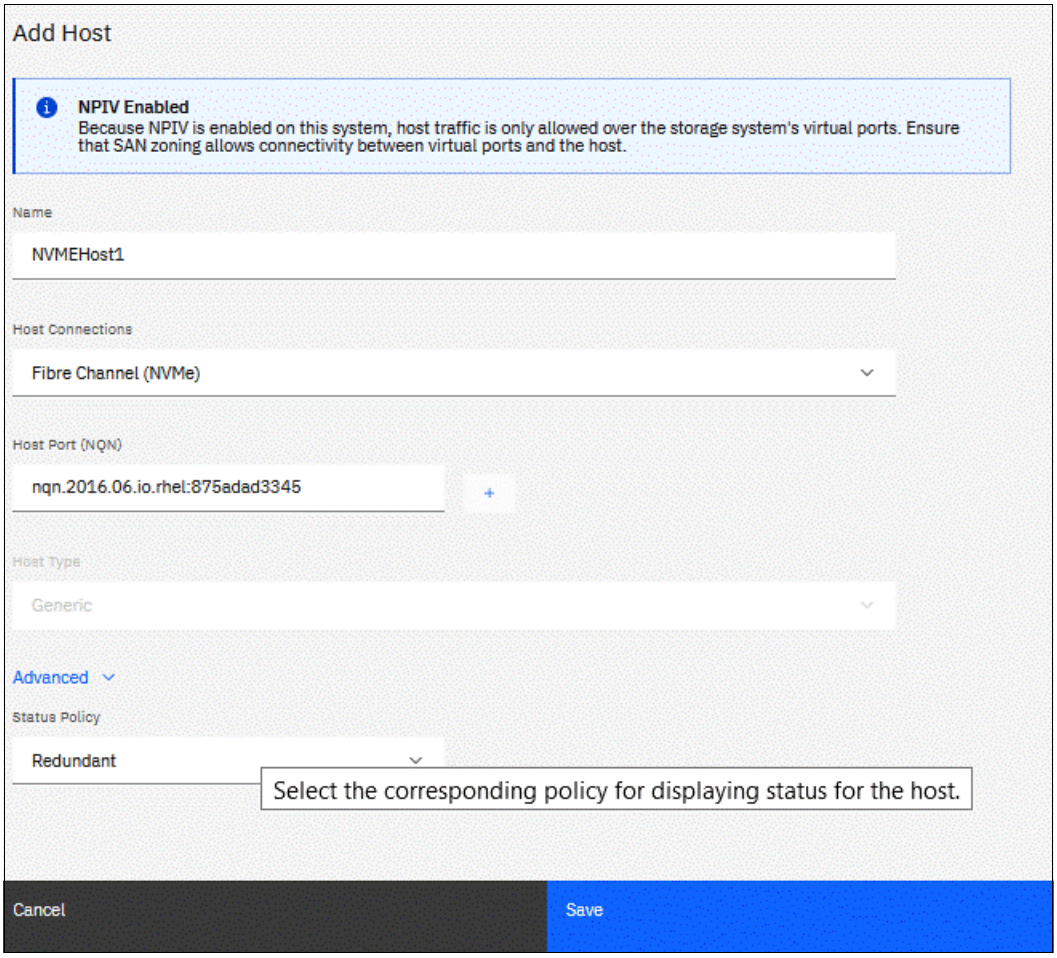


Figure 7-30 Defining NQN

3. Click **Save**. The configured host object appears in the defined host list.
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 7-14).

Example 7-14 The lsiogrp command

```
IBM_2145:ITSO-SV1:superuser>lsiogrp 0
id 0
name io_grp0
node_count 2
vdisk_count 8
host_count 3
flash_copy_total_memory 20.0MB
flash_copy_free_memory 19.9MB
remote_copy_total_memory 20.0MB
remote_copy_free_memory 20.0MB
mirroring_total_memory 20.0MB
```

```
mirroring_free_memory 19.9MB
raid_total_memory 40.0MB
raid_free_memory 40.0MB
maintenance no
compression_active no
accessible_vdisk_count 8
compression_supported no
max_enclosures 20
encryption_supported yes
flash_copy_maximum_memory 2048.0MB
site_id
site_name
fctargetportmode disabled
compression_total_memory 0.0MB
deduplication_supported yes
deduplication_active no
nqn nqn.1986-03.com.ibm:nvme:2145.0000020067214511.iogroup0
IBM_2145:ITSO-SV1:superuser>
```

The host can now be configured to use the IBM SAN Volume Controller system as a target.

Note: For more information about a compatibility matrix and supported hardware, see IBM Documentation and the IBM SSIC.

7.6.2 Host clusters

A *host cluster object* enables a user to group together multiple individual hosts and treat them 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

This section describes how to create a host cluster. It is assumed that individual hosts are already created, as described in 7.6.1, “Creating hosts” on page 465.

Complete the following steps:

1. From the menu on the left, select **Hosts** → **Host Clusters** (see Figure 7-31).

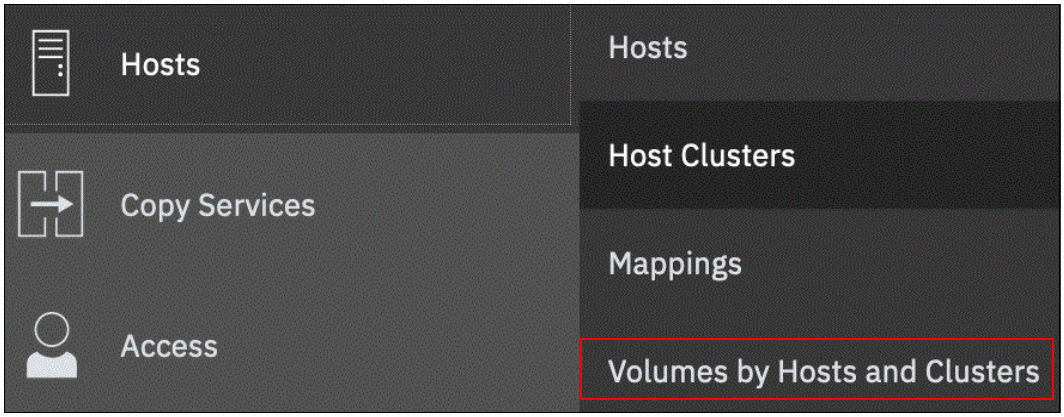


Figure 7-31 Host clusters menu

2. Click **Create Host Cluster** to open the wizard that is shown in Figure 7-32.

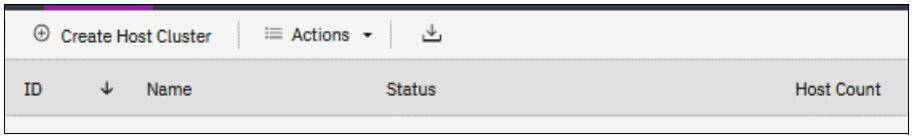


Figure 7-32 Host Clusters view and top menu

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 **Ctrl** or **Shift** and selecting them (see Figure 7-33).

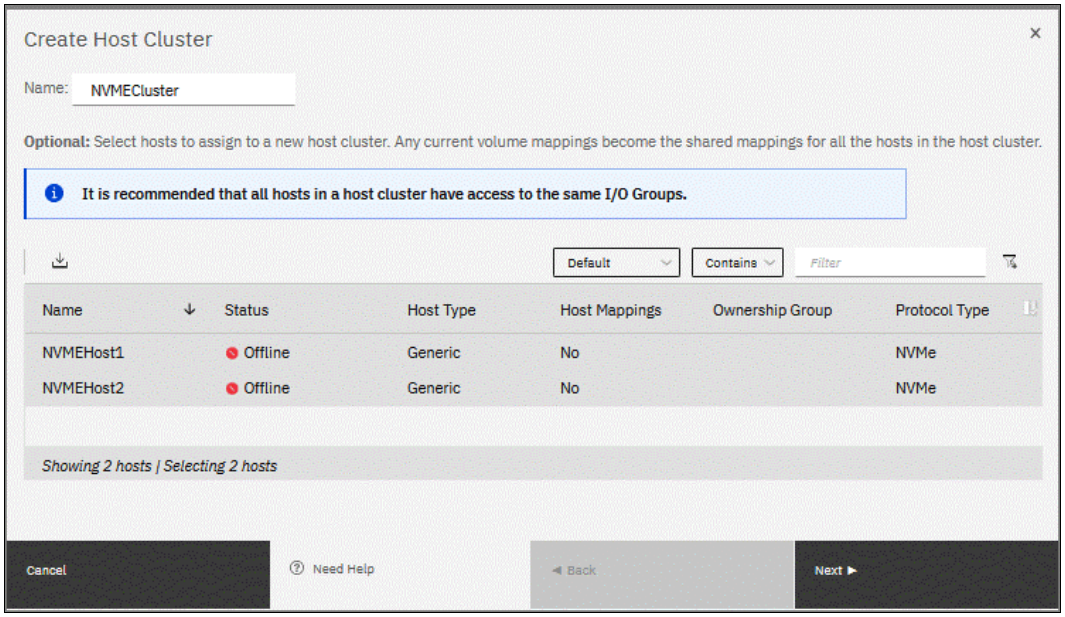


Figure 7-33 Creating Host Cluster object

4. Verify that the host selection is correct in Create Host Cluster Summary window and click **Make Host Cluster** (see Figure 7-34).

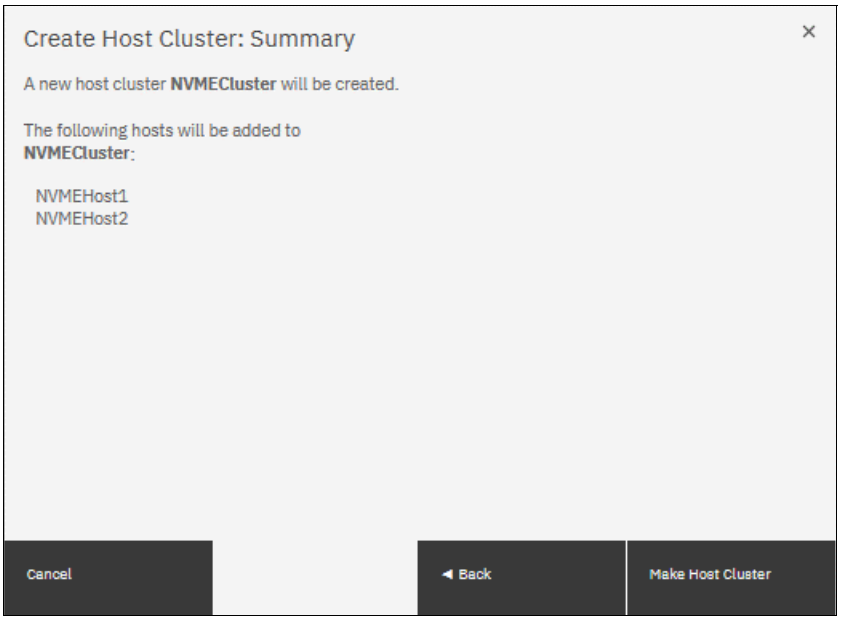


Figure 7-34 Create host cluster summary view

5. After the task completes, the newly created cluster is visible in the Host Clusters view (see Figure 7-35).

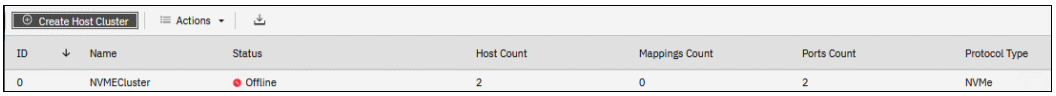


Figure 7-35 Host cluster view with the created cluster

Note: The host cluster status depends on its member hosts. One offline or degraded host sets the host cluster status as Degraded. If both host objects are offline, the host cluster status is 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 7-36).

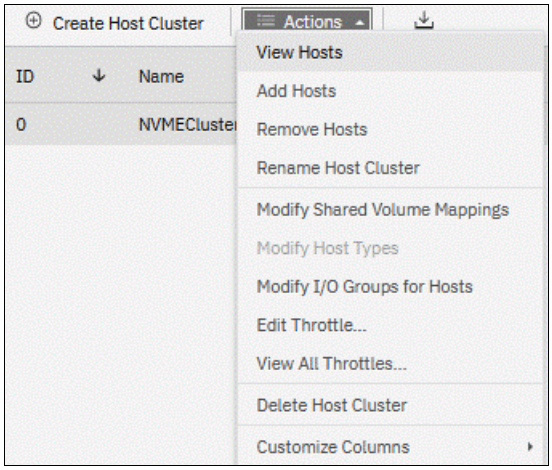


Figure 7-36 Action menu in Host Clusters view

From the Actions menu, the following tasks can be performed:

- 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 input/output operations per second (IOPS) bandwidth for the host cluster.
- View All Throttles: Shows all throttling settings and allows 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.

7.6.3 Actions on hosts

This section covers host administration, including host modification, host mappings, and deleting hosts. The basic host creation process is described in 7.6.1, “Creating hosts” on page 465.

Select **Hosts** → **Hosts** view and right-click one of the hosts, or expand the **Actions** menu. A list of available actions appears in the drop-down menu (see Figure 7-37).

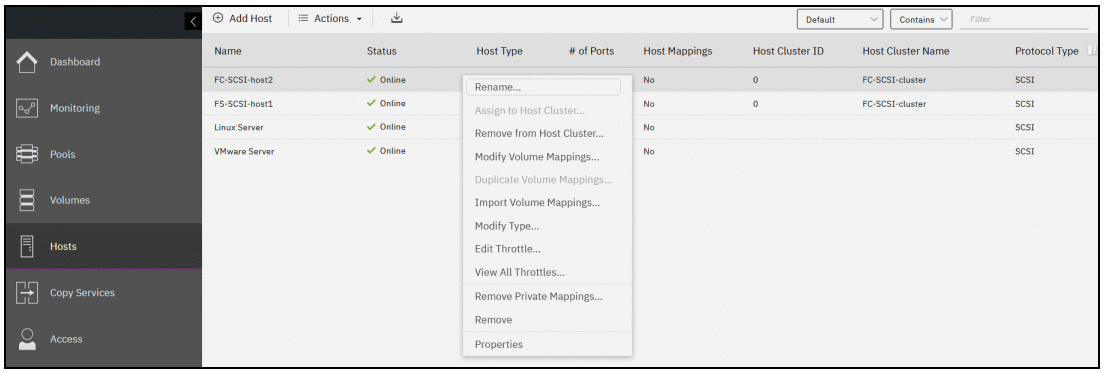


Figure 7-37 Actions on hosts

The following actions are available and described in this section:

- Rename host
- Assign host to cluster
- Modify volume mappings
- Duplicate and Import volume mappings
- Modify host type
- Edit throttle and view all throttles
- Remove private mappings
- Remove host
- View IP login information
- View 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 7-38). Clicking **Reset** reverts the changes to the original hostname.

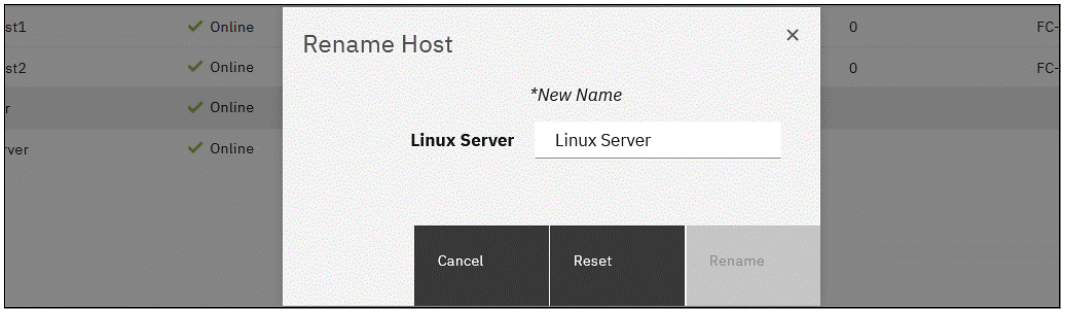


Figure 7-38 Rename Host window

Assign to host cluster and remove from host cluster

To assign host to a cluster, complete the following steps:

1. Right-click the host or a set of hosts you want to add and select **Assign to Host Cluster**. To select multiple objects, press and hold **Ctrl** and click each host to be assigned. Or, press and hold **Shift** and click the first and last objects to be selected.

2. Select the cluster in which the host is to be added (see Figure 7-39) and click **Next**.

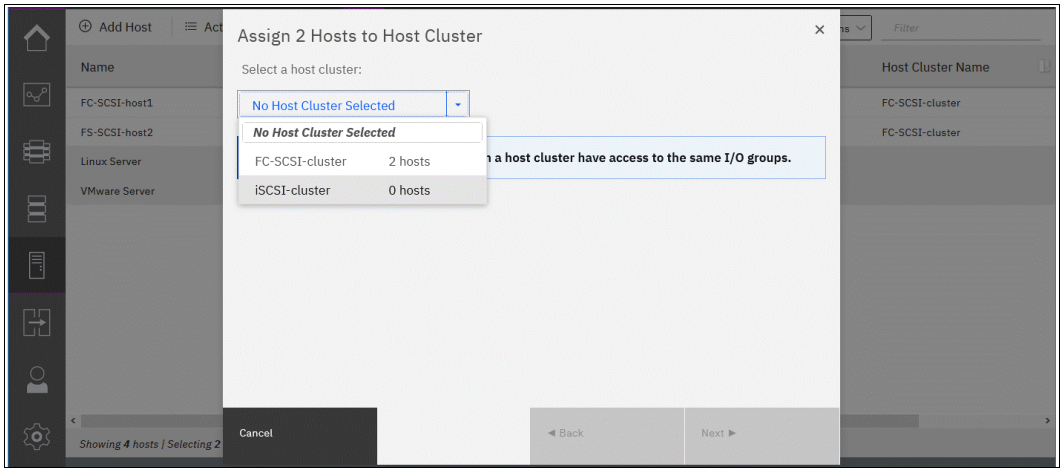


Figure 7-39 Assign host to cluster dialog

3. The IBM SAN Volume Controller system checks for SCSI ID conflicts. In a host cluster, all hosts must have 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 7-40). Click **Assign** to complete the operation. When the operation completes, the host is 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.

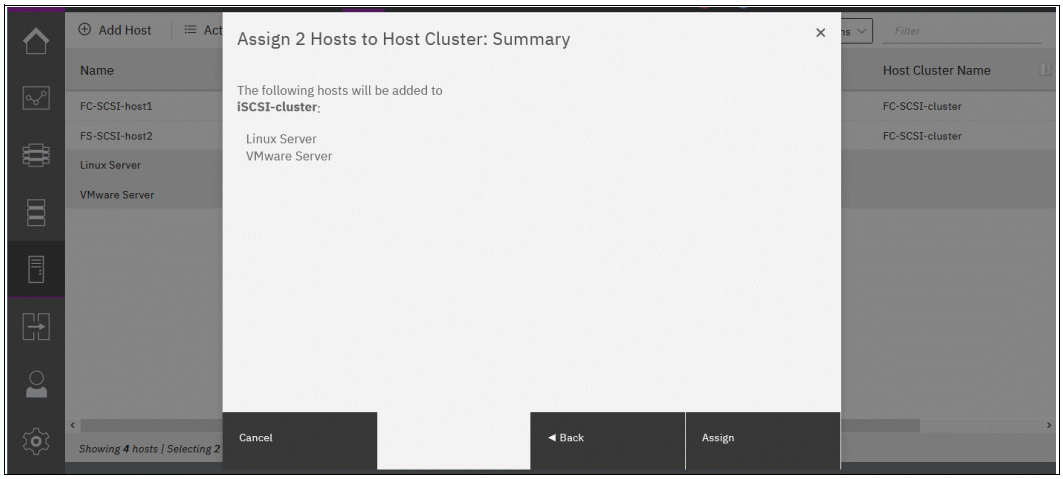


Figure 7-40 Assign host to host cluster confirmation

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 7-41). 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 7-41 Conflict between private and shared volume mappings

To remove a host from host cluster, complete the following steps:

1. Select a host or a group of hosts, right-click them, and choose **Remove from Host Cluster.**
2. Complete the information required in the dialog window (see Figure 7-42). Verify the list of hosts to be removed and determine what to do with the host mappings. Click **Remove Hosts** to complete the operation.

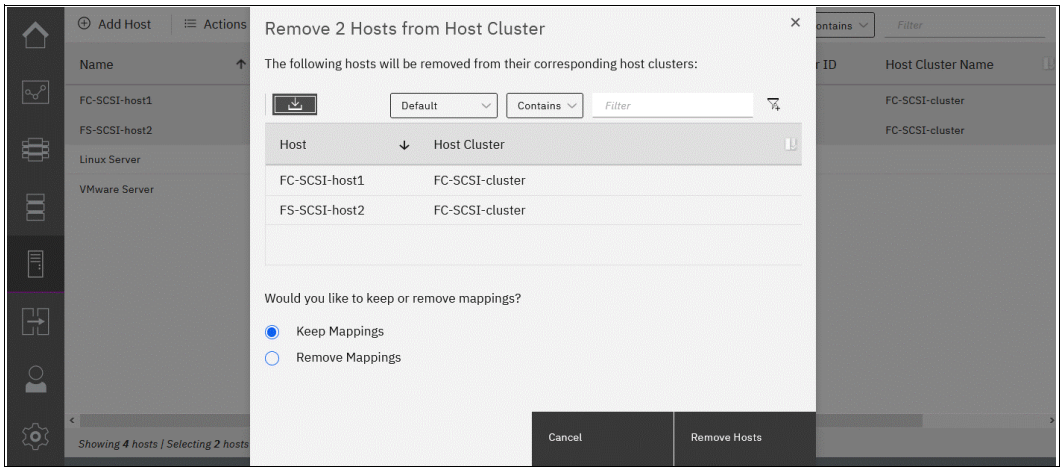


Figure 7-42 Remove host from cluster dialog

Modify volume mappings

Private volume mappings can be modified for a single host by using the **Modify Volume Mappings** action.

To do so, complete the following steps:

1. Right-click a host and select **Modify Volume Mappings**.
2. Verify whether the correct private volume mappings are displayed in the pop-up window (see Figure 7-43).

> **Note:** Host cluster shared mappings are not shown in this view. Only host private mappings are listed. To modify the share host cluster mappings, use another GUI view, as described in “Modify Shared Volume Mappings” on page 495.

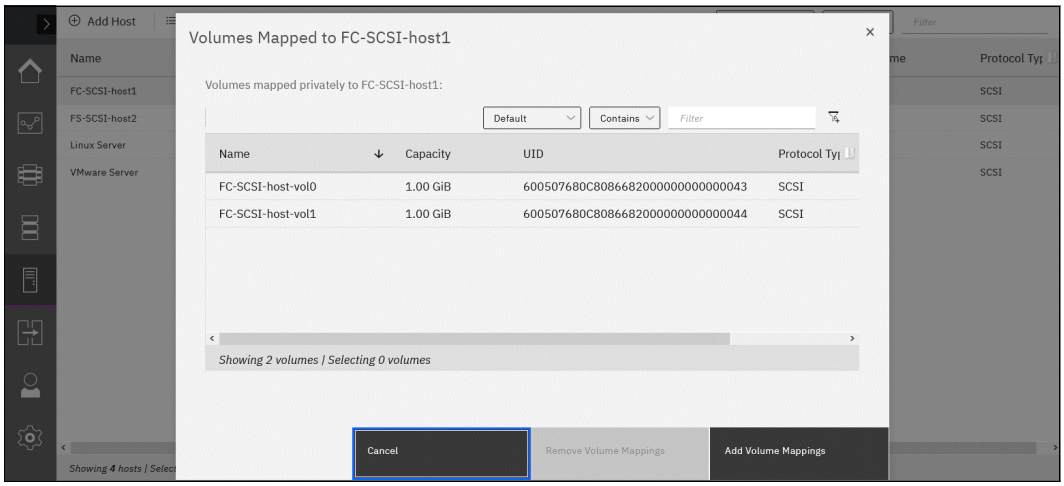


Figure 7-43 Modifying the host volume mapping

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. 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 on the list.

   If the volume is mapped to another host or host cluster, the word “Yes” appears next to it in Host Mappings column. Any attempt to map that volume to the host generates a warning message (see Figure 7-44 on page 486). However, the mapping can be added if access is coordinated at the host side.

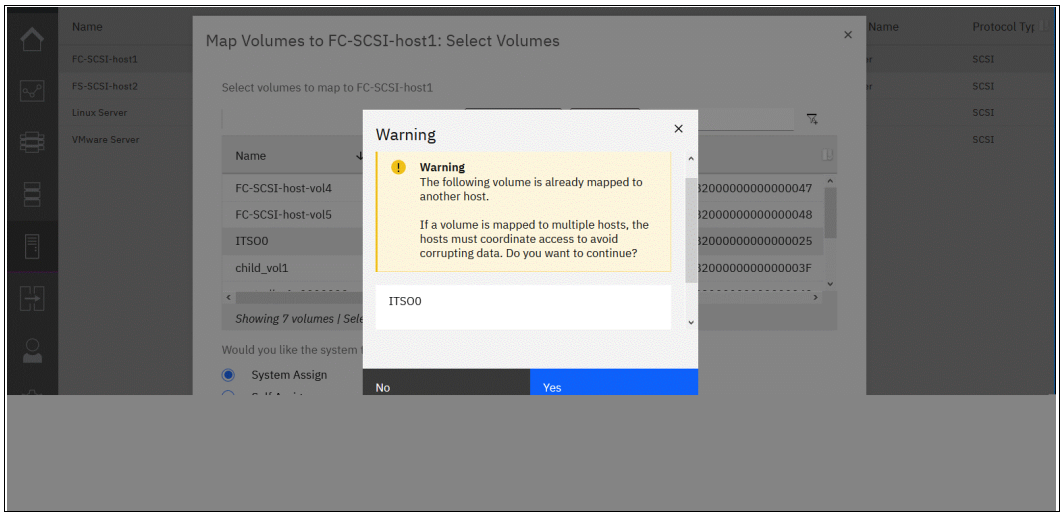


Figure 7-44 Warning indicates that mapping to another host exists

5. 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** (see Figure 7-45). Note that as shown in Figure 7-43 on page 485, only two mappings are shown for this host; however, Figure 7-45 shows five mappings because this window lists private and shared volume mappings.

Note: The SCSI ID of a volume can be changed only before the volume 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.

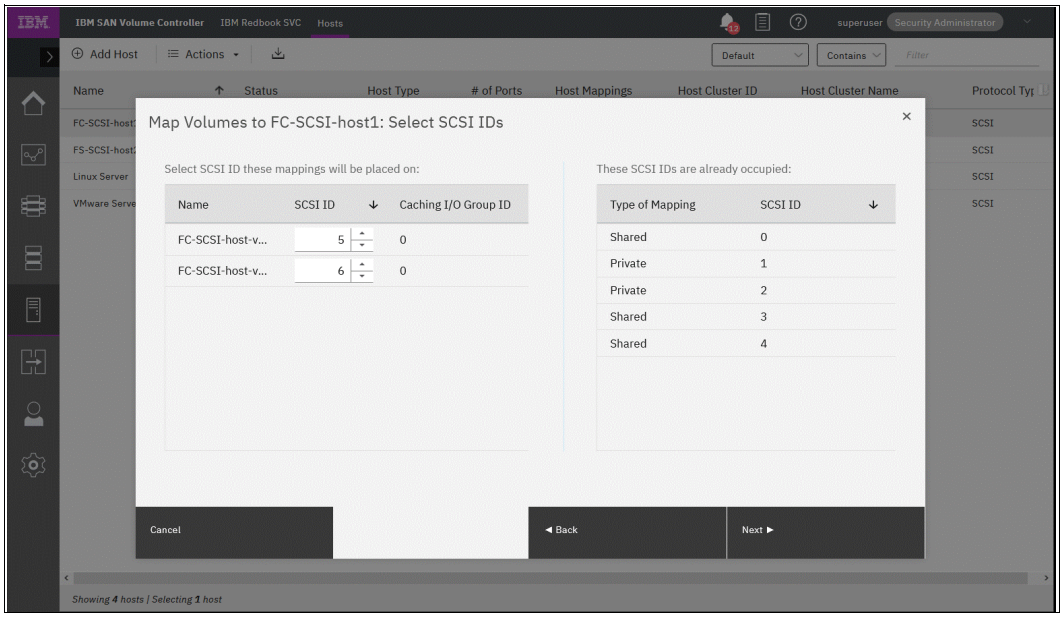


Figure 7-45 Choosing SCSI IDs manually

6. When assignments are done, click **Next** to verify prepared changes, and then, click **Map Volumes** to complete the process.

Duplicating and importing mappings

Volumes that are assigned to one host can be quickly and easily mapped to another host object.

This mapping can be done 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 already 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. The Duplicate Mappings window opens. Select a target host to which all the source host volumes are to be mapped (see Figure 7-46). In this example, the only target candidate is a host that has no mappings.

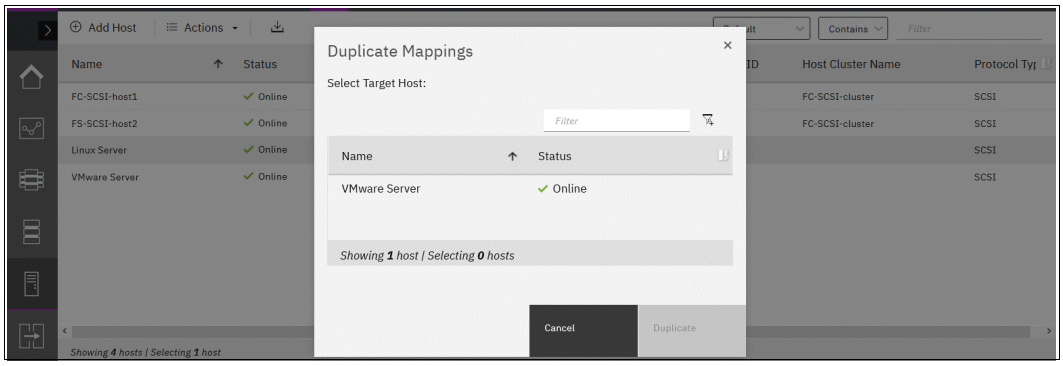


Figure 7-46 Duplicate mappings window

3. Select a target host and click **Duplicate**. After the operation completes, the target host has the same volume mappings as the source host. Both private and shared mappings are duplicated. Mappings on the source host also remain, and can be deleted manually if necessary.

To import hosts mappings from an existing host to the new 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 7-47), and click **Import**.

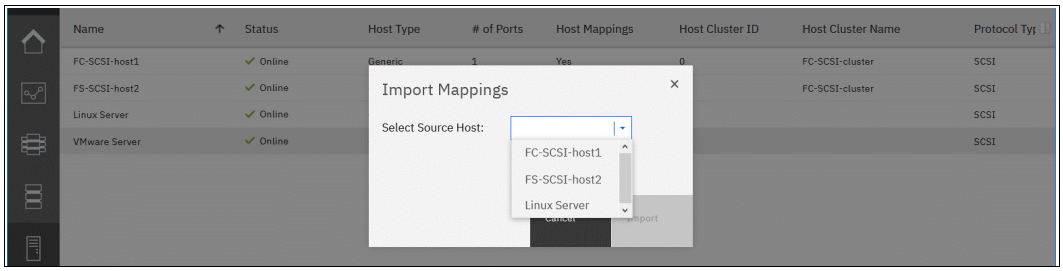


Figure 7-47 Import volume mappings source host selection

3. After the task completes, the target host has 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 and the error message is displayed:
>
> `The command failed because the objects are in different ownership groups.`

Modifying the host type

The host type often is specified during the host creation process. However, it can be changed by using the **Modify Type** option.

To change host type, complete the following steps:

1. Select the host or hosts to be modified, right-click, and then, select **Modify Type**.

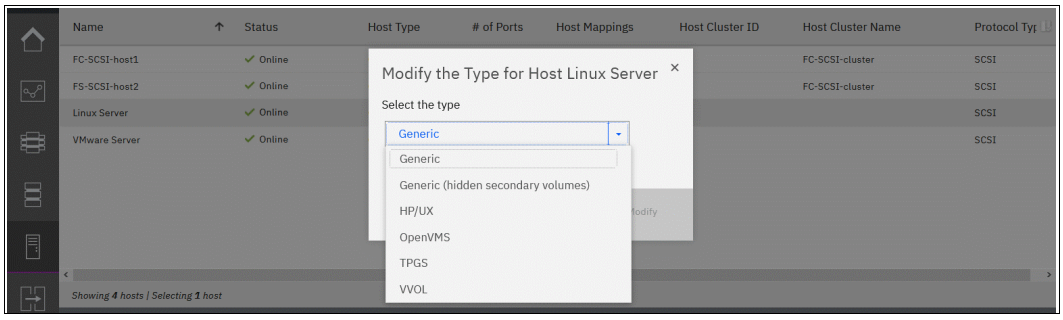


Figure 7-48 Changing host type

2. From the Modify Type dialog window (see Figure 7-48), select one of the available host types:
   - 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 appropriate host operating system types.
   - VVOL: Set if the host is configured to work with VVOLs.VVOL: Set if host is configured to work with VMware vSphere Virtual Volumes.

For more information about selecting host types, see IBM Documentation.

3. Click **Modify** to complete the task.

Viewing and editing throttles

A *throttle* is a mechanism that controls the amount of resources that is used when the system is processing I/O for a specific host or host cluster. If a throttle is defined, the system processes the I/O or delay 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, 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 7-49), specify the **IOPS limit**, **Bandwidth limit**, or both. Click **Create** to create a host throttle, change the throttle limit and click **Save** to edit a throttle, or click **Remove** to delete a host throttle.

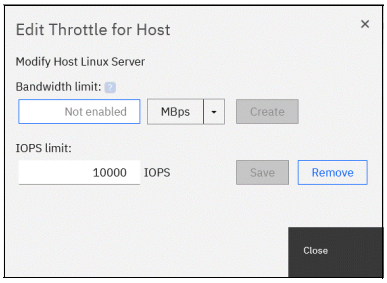


Figure 7-49 Edit throttle dialog

3. When finished editing or creating, click **Close**.
4. To view and edit all the throttles configured on the system, right-click any of the hosts and select **View All Throttles** (see Figure 7-50). Switch between throttle types by clicking the drop-down menu that is next to the Actions menu. It is also possible to change the view to see all the system's throttles on a single list.

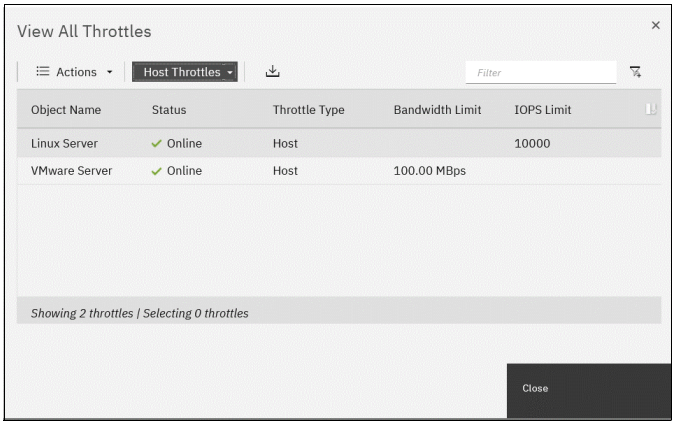


Figure 7-50 View all throttles window

From this view, you can delete or edit any throttle by right-clicking the throttle in the list and selecting the required 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 cluster, a window opens with a warning that shared mappings are not removed. Click **Yes** if you want to continue.
3. In the next window, to confirm this action by entering the number of volume mappings to be removed (see Figure 7-51). Click **Remove**.

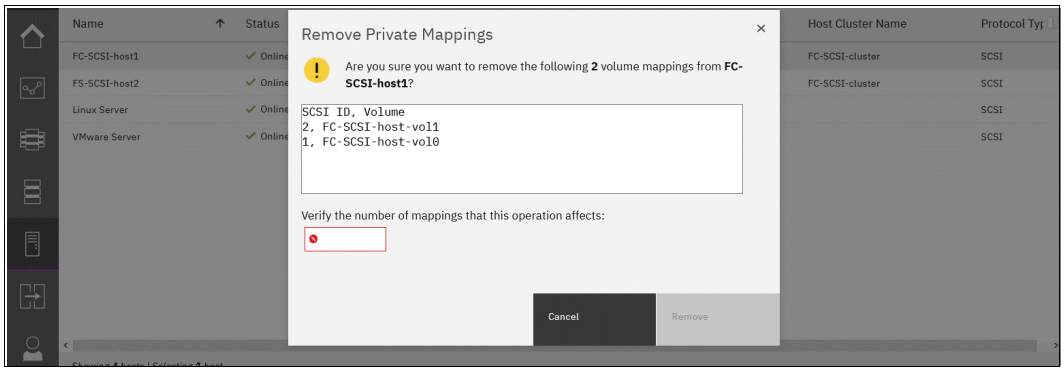


Figure 7-51 Confirming the number of mappings to be removed

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 it and 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 click **Remove** (see Figure 7-52).

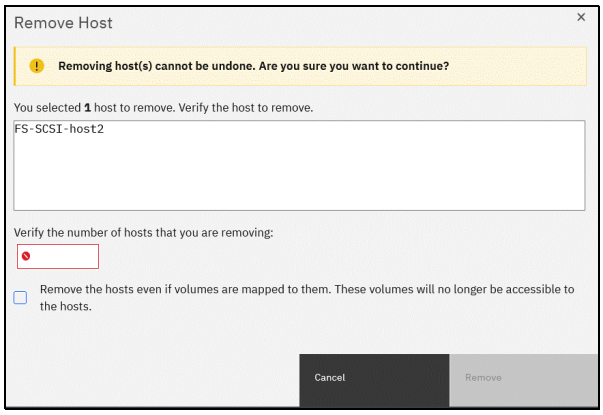


Figure 7-52 Confirming the removal of the host

3. If the host to be removed has volumes already mapped to it, select **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 7-53). The drop-down menu in the upper section of the window can be used to switch between the IQNs of the host.

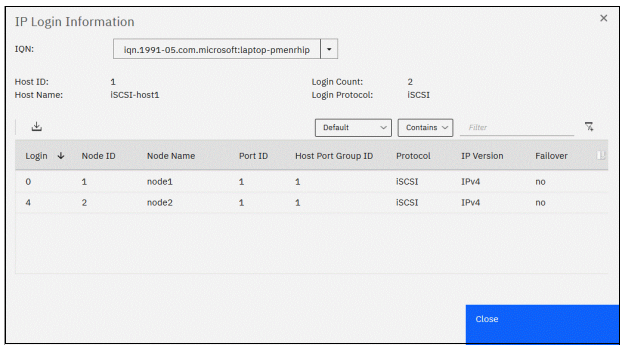


Figure 7-53 View IP login information

Viewing Host properties

To view a host object's properties, complete the following steps:

1. Right-click a host and select **Properties**.

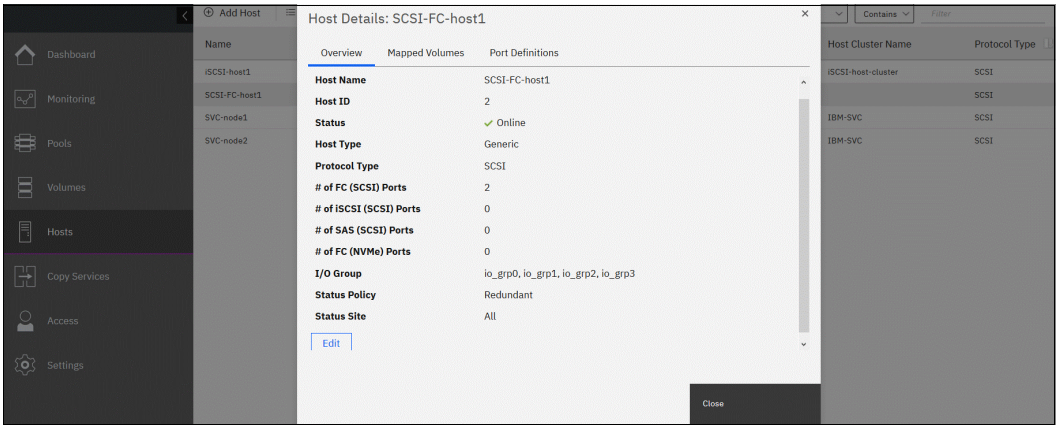


Figure 7-54 Host properties overview

2. The Host Details window (see Figure 7-54) opens and present the following tabs:
   - Overview: In this tab, click **Edit** to change hostname or the host type, select and clear the associate host I/O groups and modify the host status policy and status site.
   - Mapped Volumes: This tab lists all the volumes that mapped to the host (private and shared mappings are shown).
   - Port Definitions: This tab enables the user to view all the ports that belong to a specific host, and add ports or remove any assigned ports (see Figure 7-55 on page 492). This tab also includes the NQN of the NVMe host.

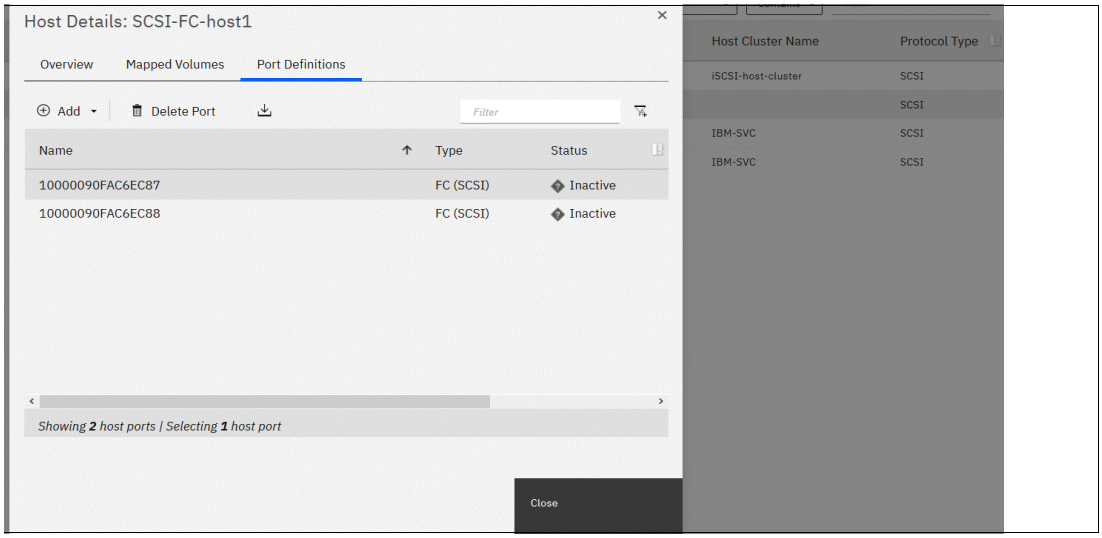


Figure 7-55 A list of port definitions

7.6.4 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 an overview of the Host Cluster feature and the actions that are required to create a host cluster, see 7.6.2, “Host clusters” on page 478.

Select **Hosts** → **Host Clusters** for a list of configured host clusters and their major parameters, like 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 7-56).

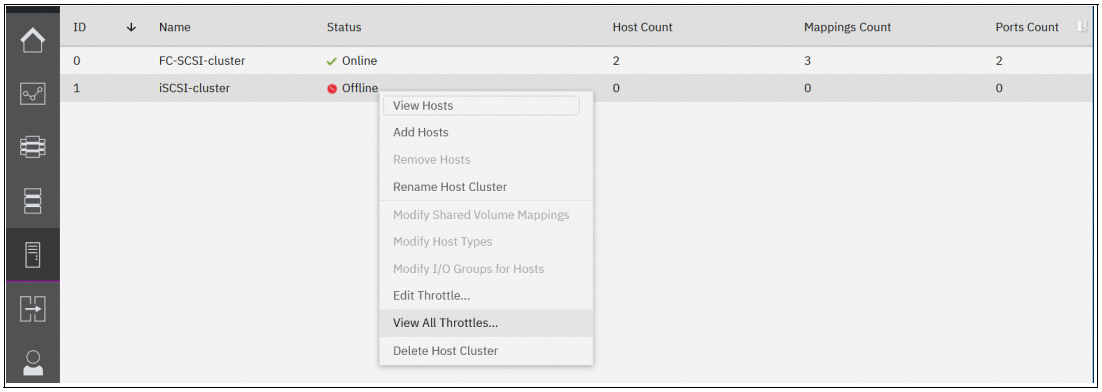


Figure 7-56 Actions available on host cluster object

View Hosts

The first block of actions allows you to check and modify cluster members list.

Click **View Host** to see a list of hosts that assigned to host cluster (see Figure 7-57). Click **Next** and **Previous** to switch display to other clusters in the cluster list.

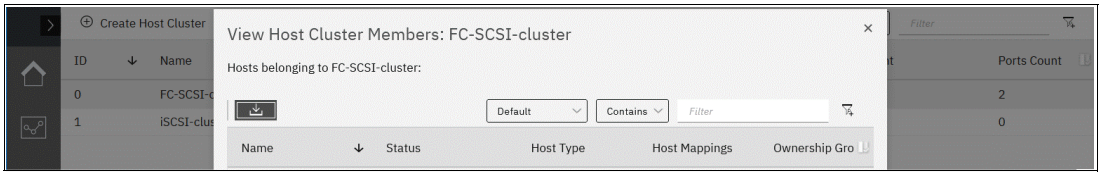


Figure 7-57 View Host Cluster Members

Add Hosts

Clicking **Add Hosts** opens a dialogue box that displays all stand-alone hosts not assigned to any of the clusters (see Figure 7-58). Select the specific host to be added to the cluster and click **Next**.

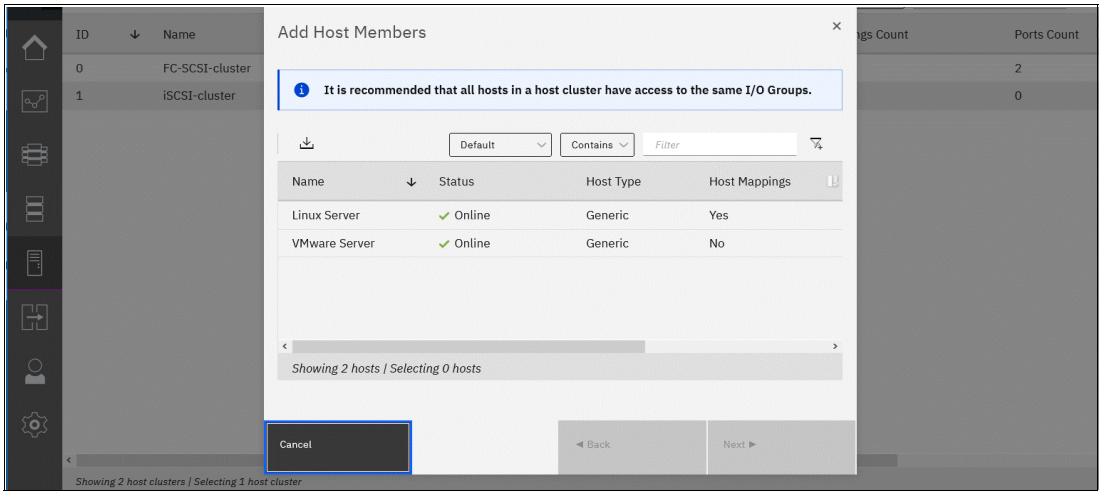


Figure 7-58 Adding host member

A prompt appears stating that the shared host cluster mappings are applied to the added host, which gains access to all volumes mapped to host cluster (see Figure 7-59).

If changes are correct, click **Add Hosts** to complete the operation.

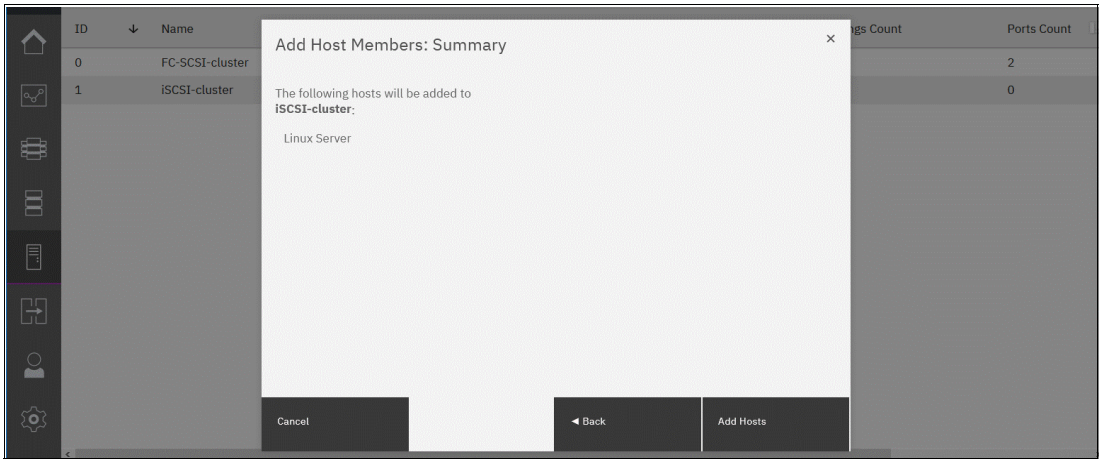


Figure 7-59 Confirm adding host to cluster

Remove Hosts

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 7-60).

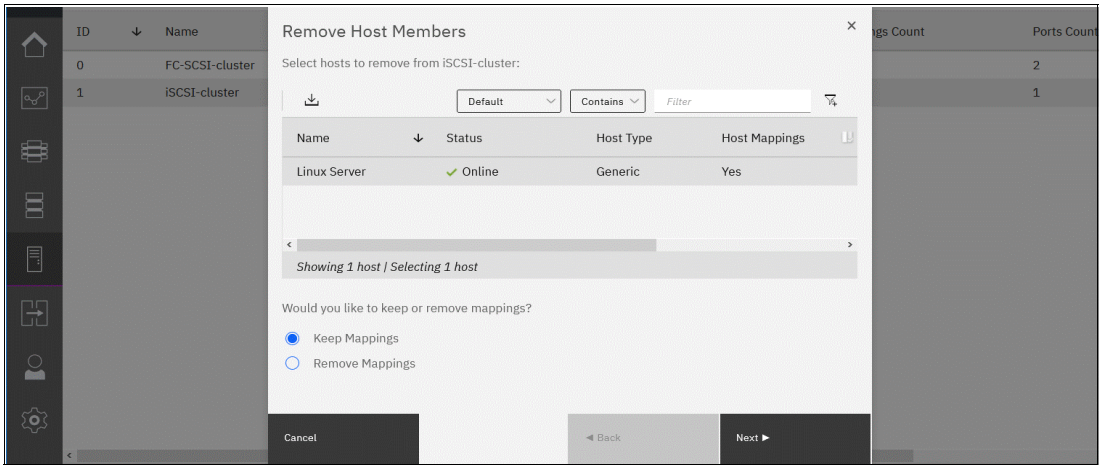


Figure 7-60 Remove Host from Cluster

The following options are available:

- A removed host can keep all the shared cluster mappings as private mappings and retain access to volumes.
- All shared cluster mappings can be removed from the host.

Select the appropriate action and click **Next.** Verify the changes and click **Remove Hosts** to complete the procedure.

Rename Host Cluster

This action provides an opportunity to change the host cluster object name.

Modify Shared Volume Mappings

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 (see Figure 7-61). From this view, select one or more of the shared mappings to be removed and click **Remove Volume Mappings**.

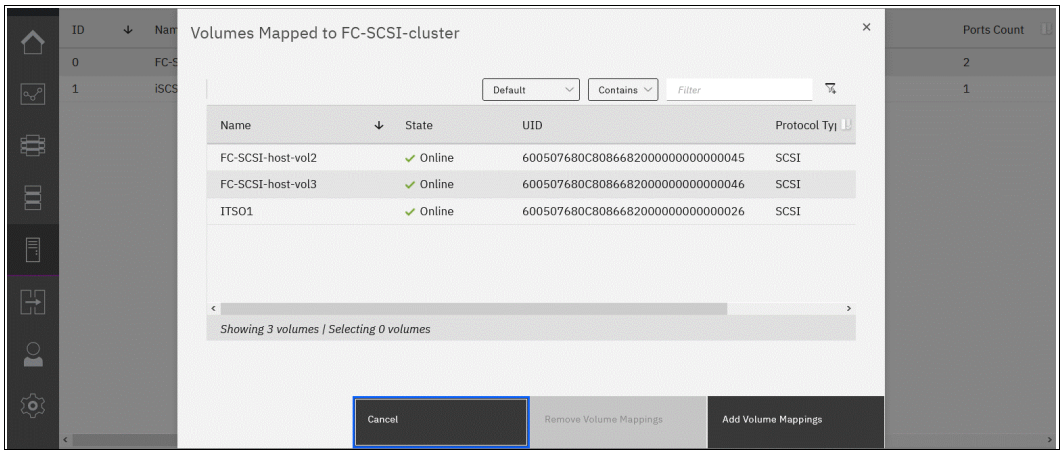


Figure 7-61 Existing shared mappings

3. If new shared mappings must be created, click **Add Volume Mappings** to open the next window (see Figure 7-62). A list appears displaying all of the volumes that are not yet mapped to the selected cluster.

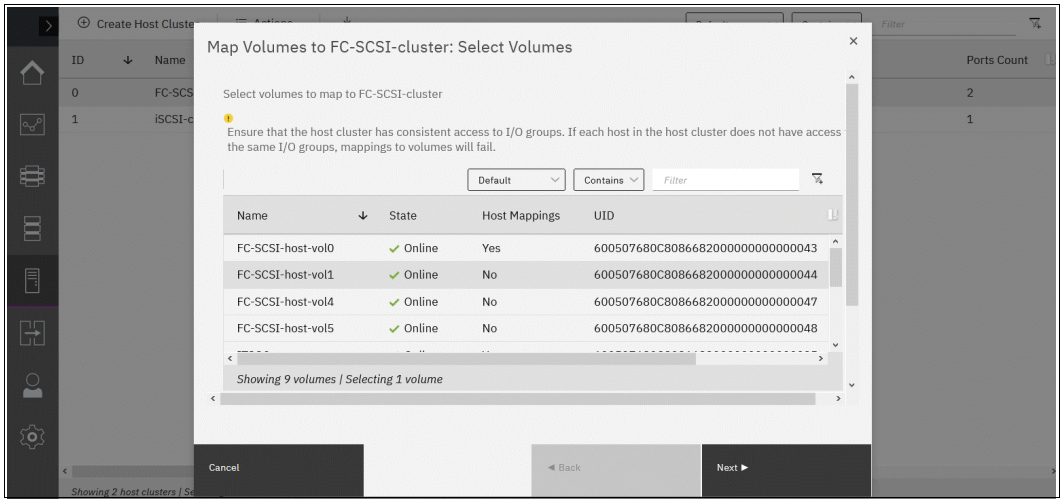


Figure 7-62 Adding shared mappings

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 7-63).

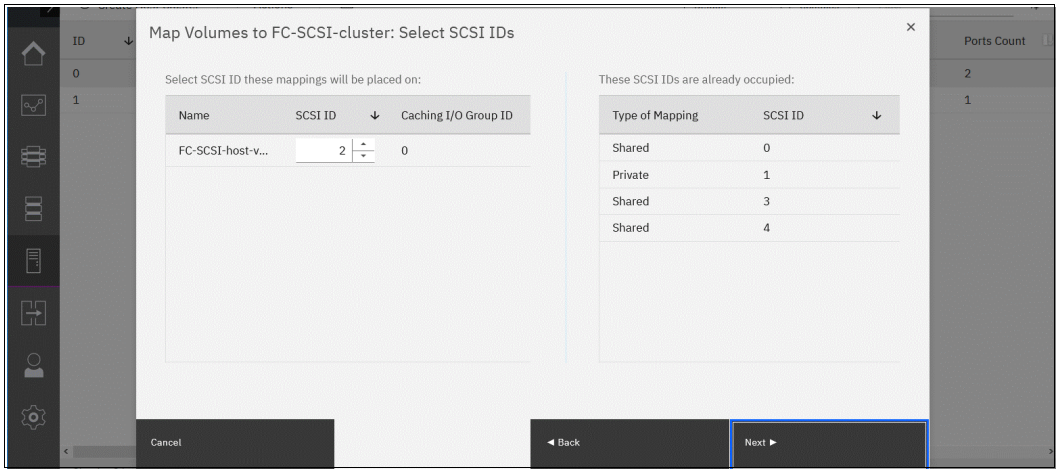


Figure 7-63 Assigning SCSI ID to mapped volumes manually

5. Click **Next** and a window appears asking for verification that the changes are correct (see Figure 7-64). 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.

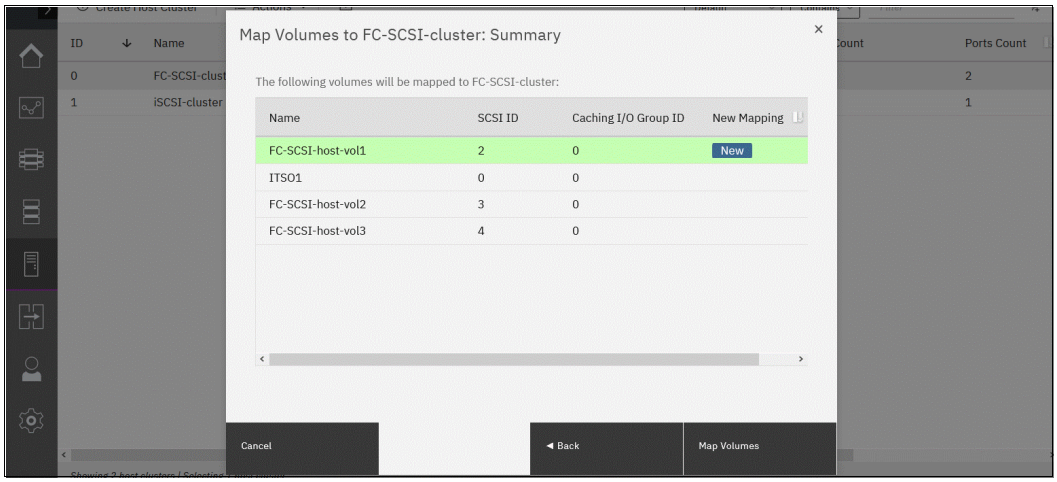


Figure 7-64 Summary of added shared mappings

Modify 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 separate host, as described in “Modifying the host type” on page 488, except that the changes are applied across all hosts that are assigned to the cluster.

Modify I/O groups for hosts

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.

Click the **Modify I/O groups for hosts** action on host cluster object to change the I/O group assignment for all hosts that are members of this cluster (see Figure 7-65).

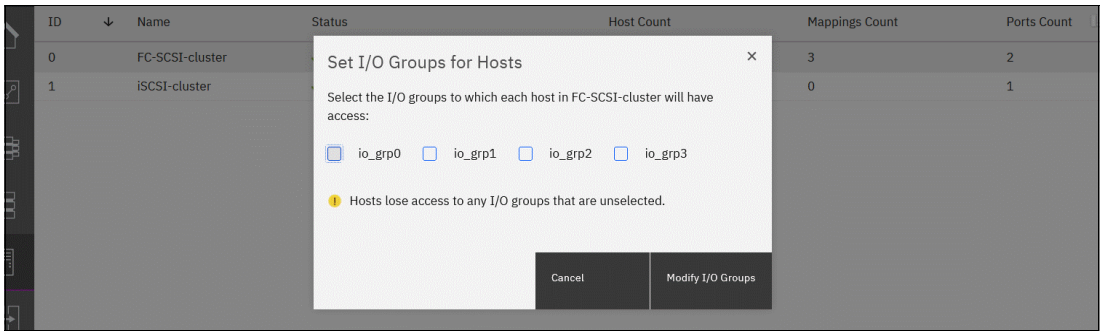


Figure 7-65 Setting I/O groups for hosts

Viewing and editing all throttles

As with the individual hosts (see *"Viewing and editing throttles" on page 489*), it is possible to similarly create and edit the throttle parameters for a host cluster with 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 host cluster throttle, complete the following steps:

1. Right-click a host cluster object and select **Edit Throttle**.
2. If the any of the hosts in the cluster featured defined individual throttles, those throttles must be removed or a warning appears (see Figure 7-66). Click **Remove Throttles**, or click **Cancel** to leave the individual throttles and stop the host throttle creation process.

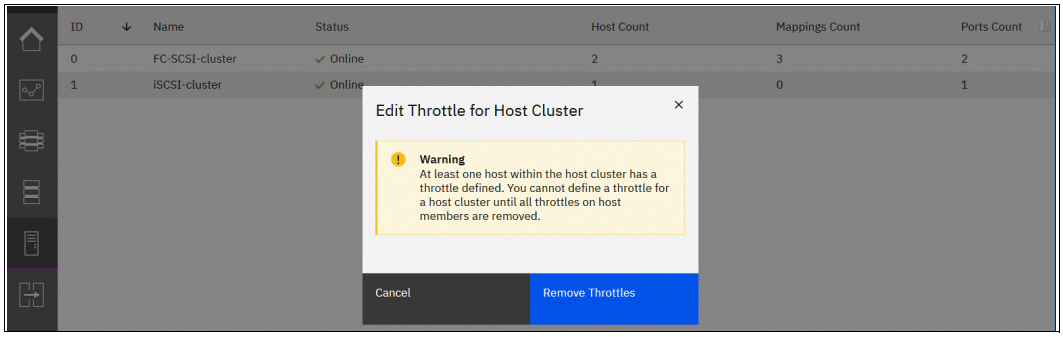


Figure 7-66 Warning that host throttles exist

3. If no individual throttles exist, set or edit I/O or data rate limits in the pop-up window (see Figure 7-67). Click **Create** to create a throttle, or click **IOPS limit** and click **Save** to change the throttle.

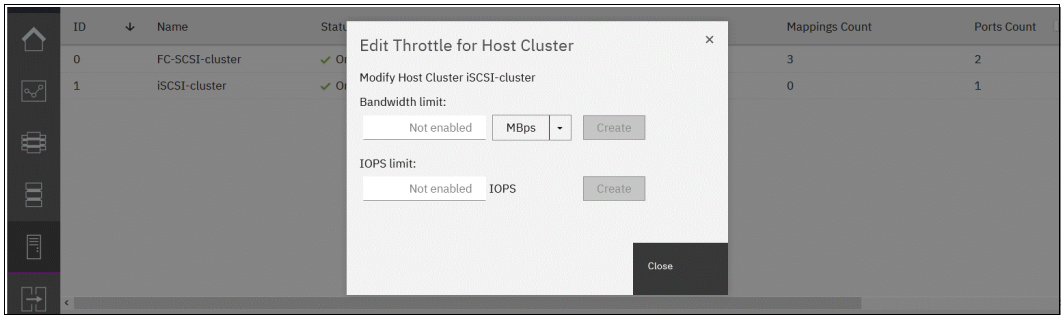


Figure 7-67 Creating host cluster throttle

Click **View All Throttles** 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 next to the Actions menu. It is also possible to change the view to see all the system's throttles in one list.

From this view, you also can delete or edit any throttle by right-clicking it in the list and selecting the suitable action (see Figure 7-50 on page 489).

Delete Host Cluster

With 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 host object.

An example of the Delete Host Cluster window is show in Figure 7-68. Hover over the question marks that are next to the suggested removal options to view more information.

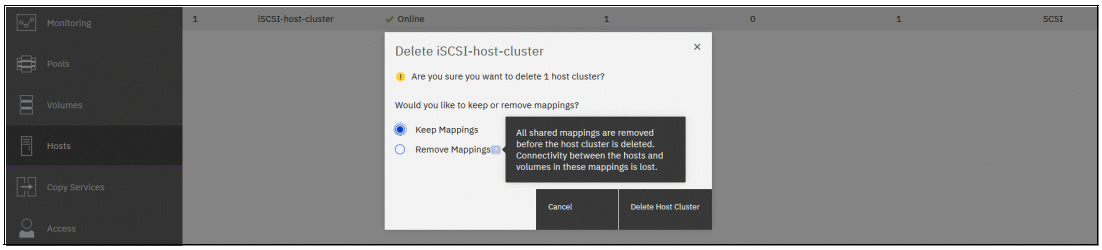


Figure 7-68 Removing host cluster

7.6.5 Host management views

The Hosts menu provides two more management views: Mappings and Volumes by Hosts and Clusters (see Figure 7-69).

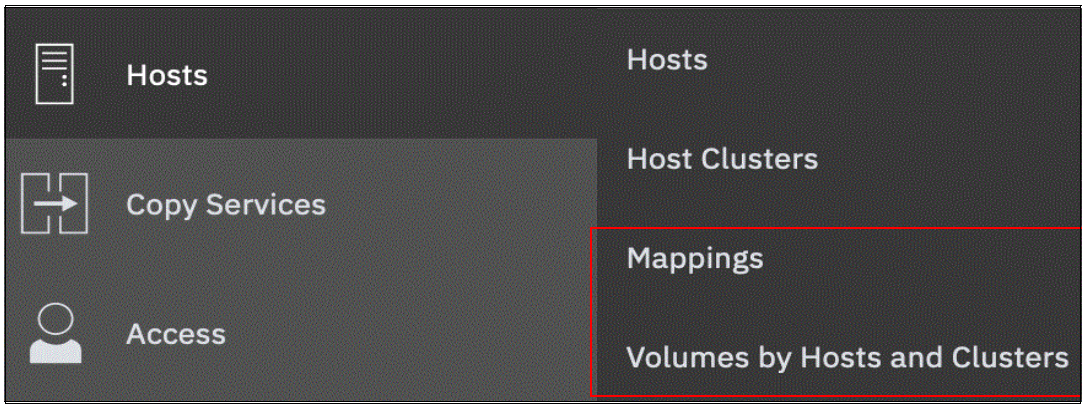


Figure 7-69 Host administration panels

These actions are the same as those actions that are found in the Hosts and Host Clusters views. However, this mapping 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. This overview also displays 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 7-70 and Figure 7-71).

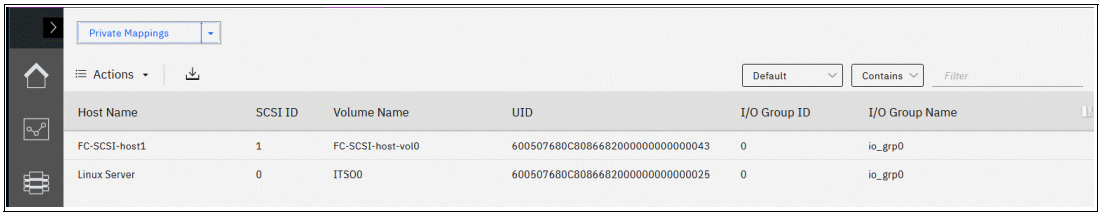


Figure 7-70 Private mappings list

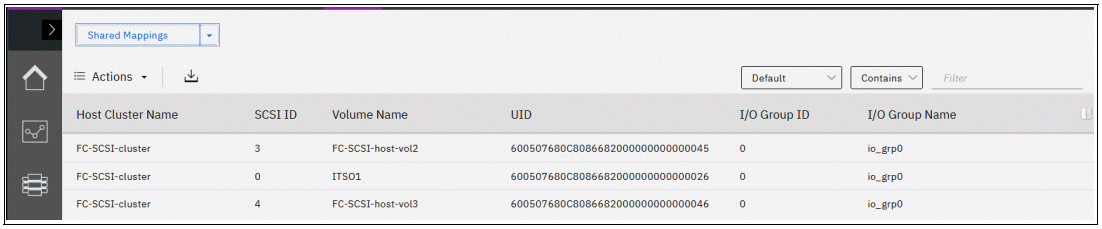


Figure 7-71 Shared mappings list

Click **Actions**, or right-click a mapping in the list for the following tasks:

- Unmap Volumes
- Host Properties
- Volume Properties

Unmapping a volume

The unmap action is available for shared mappings in the Shared mappings view (see Figure 7-71 on page 499). Similarly, the unmap action is available for private mappings while in the Private mappings or All Host mappings view.

To remove volume mapping or mappings, select the records to remove, right-click and select **Unmap volumes**, or choose **Unmap Volumes** from the **Actions** menu (see Figure 7-72).

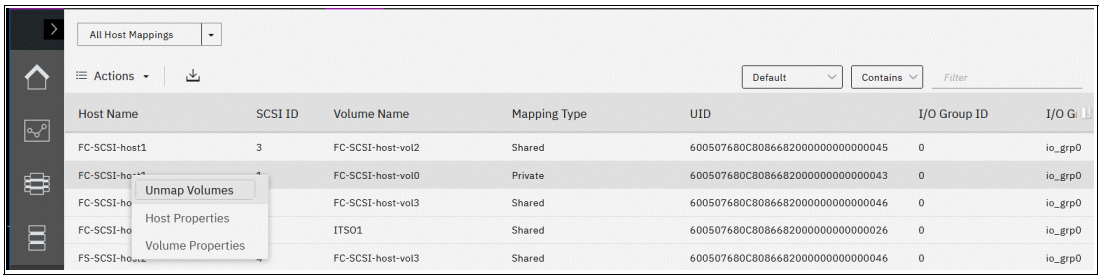


Figure 7-72 Removing private mapping

In the dialogue box, confirm how many volumes are to be unmapped by entering that number in the **Verify** field (see Figure 7-73). Then, click **Unmap**.

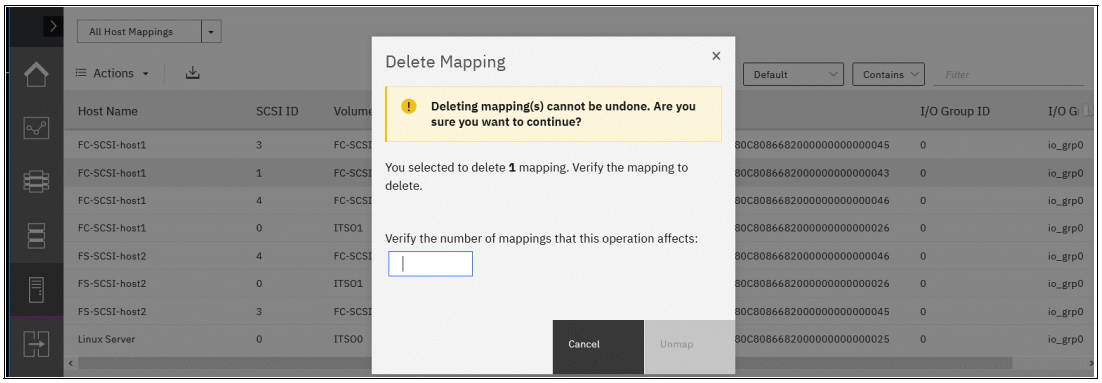


Figure 7-73 Confirm mapping deletion

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 Host properties” on page 491.

Volume Properties

Select an entry and click **Actions** → **Volume Properties**. The Volume Properties window opens and displays the same contents as described in Chapter 6, “Creating and provisioning volumes” on page 331.

Volumes by Hosts and Clusters

For a convenient way to manage volumes mapped to a particular 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. On top of the column, you can use a text input string to perform quick filtering by object names. Below the list, an **Add Host** (**Create Host Cluster**) button is available that opens a dialog, as described in 7.6.1, “Creating hosts” on page 465, and 7.6.2, “Host clusters” on page 478.

The main window shows a list of volumes and their parameters that are mapped to the selected object (see Example 7-15 on page 502). The **Volumes by Hosts Cluster** view shows only volumes with shared mappings.

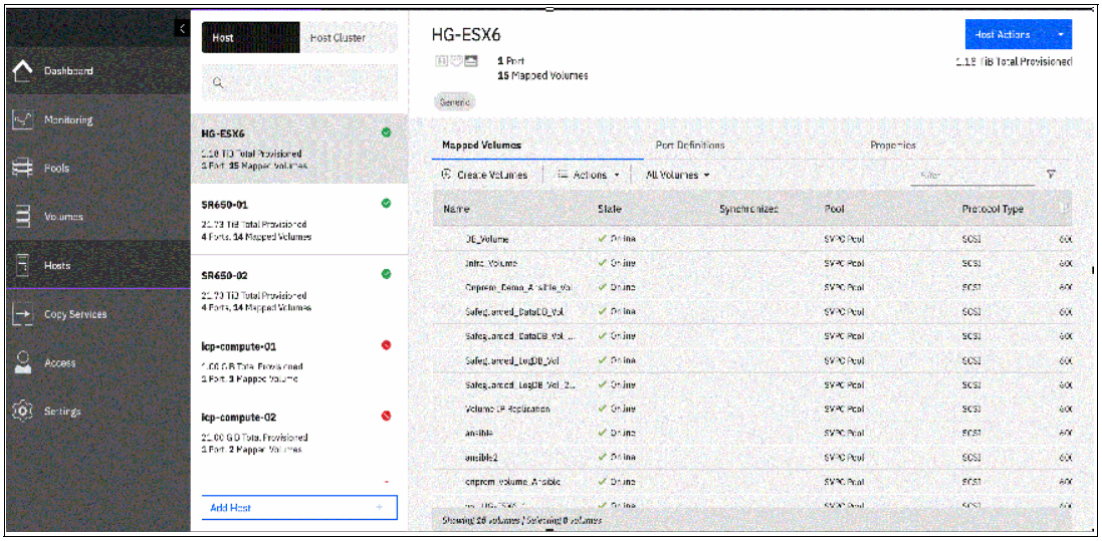


Figure 7-74 Volumes by Hosts Cluster view

Select one of options from the **Volumes** menu to filter the volume type:

- All Volumes
- Thin-Provisioned Volumes
- Compressed Volumes
- Deduplicated Volumes

Right-clicking a volume in the list opens the volume actions menu, which is described in Chapter 6, “Creating and provisioning volumes” on page 331.

Finally, create and map a volume by clicking **Create Volumes**.

7.7 Performing hosts operations by using CLI

This section describes of the host-related actions that can be taken within the system by using the command-line interface (CLI).

7.7.1 Creating a host by using the CLI

This section describes how to create FC and iSCSI hosts by using the system CLI. It is assumed that hosts are prepared for attachment as noted in the guidelines in the “Host Attachment” section of IBM Documentation.

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 (see Example 7-15).

Example 7-15 Rescanning the SAN

```
IBM_2145:ITSO-SVC: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 running the **detectmdisk** command.

2. List the candidate WWPNs and identify the WWPNs belonging to the new host, as shown in Example 7-16.

Example 7-16 Available WWPNs

```
IBM_2145:ITSO-SVC:superuser>lsfcportcandidate
fc_WWPN
2100000E1E09E3E9
2100000E1E30E5E8
2100000E1E30E60F
2100000E1EC2E5A2
2100000E1E30E597
2100000E1E30E5EC
```

3. Run the **mkhost** command with the required parameters, as shown in Example 7-17.

Example 7-17 Host creation

```
IBM_2145:ITSO-SVC:superuser>mkhost -name FC-SCSI-HOST-03 -fcwwpn
2100000E1E30E597:2100000E1E30E5EC
Host, id [3], successfully created
```

Creating iSCSI hosts

Before creating an iSCSI host in the system, determine the IQN address of the host. To find the IQN of the host, see the operating system documentation that is specific to that host.

To create a host, complete the following steps:

1. Create the iSCSI host by running the **mkhost** command (see Example 7-18).

Example 7-18 Creating an iSCSI host by running the mkhost command

```
IBM_2145:ITSO-SVC: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, as shown in Example 7-19.

Example 7-19 Verifying the iSCSI host by running the lshost command

```
IBM_2145:ITSO-SVC: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 method also can be used for FC hosts.

Creating NVMe hosts

Before creating an NVMe host in an IBM SAN Volume Controller system, determine the NQN address of the host. To find the NQN, consult the operating system documentation that is specific to that host.

Create a host by completing the following steps:

1. Create the NVMe host by running the **mkhost** command, as shown in Example 7-20.

Example 7-20 The mkhost command

```
IBM_2145:ITSO-SVC: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 NVMe host by running the **lshost** command, as shown in Example 7-21.

Example 7-21 The lshost command

```
IBM_2145:ITSO-SVC: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 preconfigured ownership group. Use the ownership group name or ID.

Here is an example command:

```
mkhost -name NVMe-Host-01 -nqn
nqn.2014-08.com.redhat:nvme:nvm-nvmehost01-edf223876 -protocol nvme -type
generic -ownershipgroup ownershipgroup0
```

7.7.2 Host administration by using the CLI

This section describes the following advanced host operations that can be performed by using the CLI:

- Mapping a volume to a host
- Mapping a volume already mapped to a different host
- Unmapping a volume from a host
- Renaming a host
- Host properties

Mapping a volume to a host

To map a volume, complete the following steps:

1. Run the **mkvdiskhostmap** command, as shown in Example 7-22.

Example 7-22 Mapping a volume

```
IBM_2145:ITSO-SVC: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 7-23).

Example 7-23 Checking the mapped volume

```
IBM_2145:ITSO-SVC: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 already mapped to a different host, complete the following steps:

1. Run the **mkvdiskhost -force** command, as shown in Example 7-24.

Example 7-24 Mapping the same volume to a second host

```
IBM_2145:ITSO-SVC: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 mapping is the 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, as shown in Example 7-25.

Example 7-25 Ensuring that the same volume is mapped to multiple hosts

```
IBM_2145:ITSO-SVC: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_2145:ITSO-SVC:superuser>
```

Unmapping a volume from a host

To unmap a volume from the host, run the **rmvdiskhostmap** command (see Example 7-26).

Example 7-26 Unmapping a volume from a host

```
IBM_2145:ITSO-SVC: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 7-27. In this example, the host RHEL-Host-06 is renamed to FC_RHEL_HOST).

Example 7-27 Renaming a host

```
IBM_2145:ITSO-SVC:superuser>chhost -name FC_RHEL_HOST RHEL-Host-06
```

Removing a host

To remove a host from the IBM SAN Volume Controller system, run the **rmhost** command (as shown in Example 7-28).

Example 7-28 Removing a host

```
IBM_2145:ITSO-SVC:superuser>rmhost RHEL-Host-07
```

Note: Before removing a host from an IBM SAN Volume Controller system, ensure that all of the volumes are unmapped from that host, as shown in Example 7-26 on page 505.

Host properties

To obtain more information about a host, run the **lshost** command with **hostname** or **host id** as a parameter (see Example 7-29).

Example 7-29 Host details

```
IBM_2145:ITSO-SVC: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_2145:ITSO-SVC:superuser>
```

Note: The **status_policy** property can include one of the following 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.

7.7.3 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 for FC-SCSI host ports

To add ports to a defined host, complete the following steps:

1. 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 7-30).

   Example 7-30 Listing the newly available WWPN

   ```
   IBM_2145:ITSO-SVC:superuser>lsfcportcandidate
   fc_WWPN
   2100000E1E09E3E9
   2100000E1E30E5E8
   2100000E1E30E60F
   2100000E1EC2E5A2
   ```

2. 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 (see Example 7-31).

   Example 7-31 Adding the newly discovered WWPN to the host definition

   ```
   IBM_2145:ITSO-SVC:superuser>addhostport -hbawwpn
   2100000E1E09E3E9:2100000E1E30E5E8 ITSO-VMHOST-01
   ```

   This command adds the WWPNs 2100000E1E09E3E9 and 2100000E1E30E5E8 to the ITSO-VMHOST-01 host.

3. If the new HBA is not connected or zoned, the **lshbaportcandidate** command does not display your WWPN. In this case, you can manually enter the WWPN of your HBA or HBAs and use the **-force** flag to create the host (see Example 7-32).

   Example 7-32 Adding a WWPN to the host definition by using the -force option

   ```
   IBM_2145:ITSO-SVC: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.

4. Verify the host port count by running the **lshost** command (see Example 7-33, which shows that the host ITSO-VMHOST-01 has a port count updated from 2 to 5, after two commands in previous examples were run).

   Example 7-33 Host with updated port count

   ```
   IBM_2145:ITSO-SVC:superuser>lshost
   id name           port_count iogrp_count status   site_id site_name
   0  ITSO-VMHOST-01 5          4           online
   ```

Adding ports to a defined host for iSCSI and FC-NVMe host ports

Complete the following steps:

1. If the host uses iSCSI or FC-NVMe as a connection protocol, the host port ID (iSCSI IQN or NVMe NQN) is used to add the port. Unlike FC-attached hosts, available candidate IDs cannot be checked. The host administrator can provide the IQN or NQN.
2. After getting the ID, run the **addhostport** command (see Example 7-34, which shows the command for iSCSI port).

Example 7-34 Adding an iSCSI port to the defined host

```
IBM_2145:ITSO-SVC:superuser>addhostport -iscsiname
iqn.1994-05.com.redhat:e6ddffaab567 RHEL-Host-05
```

Example 7-35 shows FC-NVMe port being added.

Example 7-35 The addhostport command

```
IBM_2145:ITSO-SVC: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 it is the correct port being removed by running the **lshost** command (see Example 7-36, which shows a check ID being performed to verify that WWPN to be removed belongs the host ITSO-VMHOST-01).

Example 7-36 Running the lshost command to check the WWPNs

```
IBM_2145:ITSO-SV1: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 the WWPN or iSCSI IQN that must be deleted is determined, run the **rmhostport** command to delete the host port (see Example 7-37).

Example 7-37 Running the rmhostport command to remove a WWPN

```
IBM_2145:ITSO-SVC:superuser>rmhostport -fcwwpn 2100000E1E30E597 ITSO-VMHOST-01
```

To remove the iSCSI IQN, run the **rmhostport** command with **iscsiname** argument (as shown in Example 7-38).

Example 7-38 Removing the iSCSI port from the host

```
IBM_2145:ITSO-SVC:superuser>rmhostport -iscsiname
iqn.1994-05.com.redhat:e6ddffaab567 RHEL-Host-05
```

3. To remove the NVMe NQN, run the **rmhostport** with **nqn** argument, as shown in Example 7-39.

Example 7-39 Removing the NQN port from the host

```
IBM_2145:ITSO-SVC:superuser>rmhostport -nqn nqn.2016-06.io.rhel:875adad3345
RHEL-Host-08
```

Note: Multiple ports can be removed at one time by using the separator or colon (:) between the port names, as shown in the following example:

```
rmhostport -hbawwpn 210000E08B054CAA:210000E08B892BCD ITSO-VMHOST-02
```

7.7.4 Host cluster operations

This section describes the following host cluster operations that can be performed by using the CLI:

- Creating a host cluster (`mkhostcluster`)
- Adding a member to the host cluster (`addhostclustermember`)
- Listing a host cluster (`lshostcluster`)
- Listing a host cluster member (`lshostclustermember`)
- Assigning a volume to the host cluster (`mkvolumehostclustermap`)
- Listing a host cluster for mapped volumes (`lshostclustervolumemap`)
- Removing volume mapping from the host cluster (`rmvolumehostclustermap`)
- Removing a host cluster member (`rmhostclustermember`)
- Removing the host cluster (`rmhostcluster`)

Creating a host cluster

To create a host cluster, run the **mkhostcluster** command (see Example 7-40).

Example 7-40 Creating a host cluster

```
IBM_2145:ITSO-SVC: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 7-41).

Example 7-41 Adding a host or hosts to a host cluster

```
IBM_2145:ITSO-SVC:superuser>addhostclustermember -host ITSO-VMHOST-01:ITSO-VMHOST-02
ITSO-ESX-Cluster-01
IBM_2145:ITSO-SVC:superuser>
```

In Example 7-41, 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 7-42).

Example 7-42 Listing host cluster members by running the lshostclustermember command

```
IBM_2145:ITSO-SVC: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_2145:ITSO-SVC: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 7-43).

Example 7-43 Mapping a volume to a host cluster

```
IBM_2145:ITSO-SVC:superuser>mkvolumehostclustermap -hostcluster ITSO-ESX-Cluster-01 VMware1
Volume to Host Cluster map, id [0], successfully created
IBM_2145:ITSO-SVC: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 7-44).

Example 7-44 Listing volumes that are mapped to a host

```
IBM_2145:ITSO-SVC: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 (see Example 7-45).

Example 7-45 Removing a volume mapping

```
IBM_2145:ITSO-SVC:superuser>rmvolumehostclustermap -hostcluster ITSO-ESX-Cluster-01 VMware3
```

In Example 7-45, volume VMware3 is unmapped from the host cluster ITSO-ESX-Cluster-01.

Note: Use the **-makeprivate** flag to specify which host or hosts 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 7-46).

Example 7-46 Removing a host cluster member

```
IBM_2145:ITSO-SVC:superuser>rmhostclustermember -host ITSO-VMHOST-02 -removemappings
ITSO-ESX-Cluster-01
```

In Example 7-46, 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 7-47).

Example 7-47 Removing a host cluster

```
IBM_2145:ITSO-SVC: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 flag converts the host volume mapping to `private` instead of `shared`.

7.7.5 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 7-48).

Example 7-48 Adding a host cluster to an ownership group

```
IBM_2145:ITSO-SVC: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 7-49, 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 7-49).

Example 7-49 Removing a host cluster from an ownership group

```
IBM_2145:ITSO-SVC:superuser>chhostcluster -noownershipgroup 0
```

This command removes host cluster 0 from the ownership group to which it is assigned.

7.8 Host attachment practical examples

This section demonstrates how to attach a Linux-based host by using the information that is provided in the previous sections of this chapter.

7.8.1 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 that is shown in Example 7-50.

Example 7-50 RHEL release check

```
20201028-09:50:34 root@redbookvm7-1:~ # cat /etc/redhat-release
Red Hat Enterprise Linux Server release 7.6 (Maipo)
```

7.8.2 Fibre Channel host connectivity and capacity allocation

To collect the necessary data, configure the host object in the storage system, and access the storage capacity, complete the following steps:

1. Obtain the necessary credentials for connectivity 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 7-51.

   The information about HBAs on the host with FC connectivity ability is available in `/sys/class/fc_host` directory for each HBA that is on board of the host machine. Checking `port_name` file in each of `hostN` directories provides the host's WWPNs (in bold in Example 7-51), which should be used in host object configuration on the storage system.

Example 7-51 Finding out 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 Fibre Channel hosts objects" on page 466. If zoning is done for the host, the host's WWPN should be available for selection in Host Port (WWPN) drop list. If the host is not yet zoned, ports can be added manually in the field.

3. After the host object is defined, it is visible in the hosts view and volumes (VDisks) can be mapped to it for use. 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 7-75) to verify that all the information is correct.

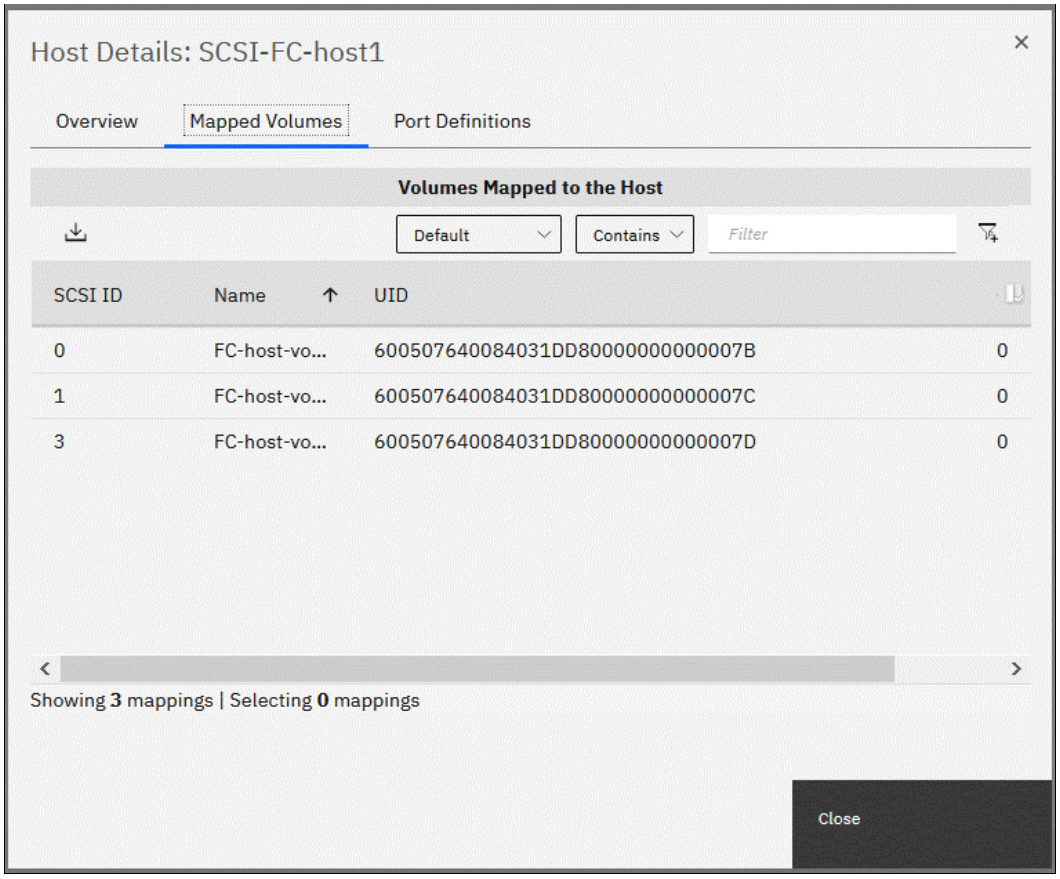


Figure 7-75 Host Details: Mapped volumes

4. Configure the host side to discover and use the mapped VDisks. Consider the following points:
   - RHEL has its own native multipath driver, which correctly maps the discovered drives and their paths to the `mpath n` device files accordingly in `/dev/mapper`. This feature makes it ready for other utilization. The multipath driver must be correctly configured, as described in **IBM Documentation** → **Host Attachment section for Linux hosts**. To check that volumes are detected correctly, run the command that is shown in Example 7-52.

Example 7-52 Scanning and rebuilding 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
```

- The **rescan-scsi-bus.sh -r** command (see Example 7-52 on page 513) rescans for new devices. In some cases, it might be necessary to run the **rescan-scsi-bus.sh -a** command because it scans all available devices. The **multipath -F** command flushes the configuration of multipath driver.

  The next command to run, **multipath**, builds a new configuration for new devices and paths. The **multipath -ll** command provides information about path states and to which `mpath n` device capacity was mapped for each mapped VDisk.

- To start using the capacities that are provided as logical volumes, use *only* `/dev/mapper/mpath n` device for access because they are created early in the RHEL boot process. For example, to use the VDisk `600507640084031dd80000000000007c` as a LVM physical volume, the name of the RHEL device (which was mapped by the multipath driver) can be found by running the **multipath -ll** command. Example 7-52 on page 513 shows the output. `mpathat` with `UID 3600507640084031dd80000000000007c` is marked in bold.

- This new physical volume can be added to volume group of the host and logical volumes can be created or extended and configured for any application on the host (see Example 7-53).

Example 7-53 Creating 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 points when capacity is provisioned 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 systems). 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.

7.8.3 iSCSI host connectivity and capacity allocation

The iSCSI protocol uses the initiator from host side to send SCSI commands to storage systems target devices. Therefore it is necessary to prepare the correct environment on the host side and correctly configure the storage system. For more information, see “Preparing for iSCSI connection” on page 469.

This example demonstrates a RHEL host configuration and how to obtain access to the dedicated volumes (VDisks) on the storage system.

The detailed steps to prepare a RHEL host for SCSI connectivity can be found on **IBM Documentation site for specific Storage System** → **Configuring** → **Host Attachment** → **iSCSI Ethernet host attachment**.

Complete the following steps:

1. Install the iSCSI initiator on the RHEL host by using **yum**, as shown in Example 7-54.

Example 7-54 Installing iscsi-initiator-utils

```
20201028-18:13:51 root@redbookvm7-1:/mnt/disc # yum install iscsi-initiator-utils
```

2. Configure the iSCSI initiator and set the connection credentials in `/etc/iscsi` files. Check or define IQN in `/etc/iscsi/initiatorname.iscsi` (see Example 7-55).

Example 7-55 Checking the initiators IQN

```
20201028-19:04:34 root@redbookvm7-1:/etc/iscsi # cat initiatorname.iscsi
InitiatorName=iqn.1994-05.com.redhat:f3de6ef11811
```

3. Restart iSCSI initiator service if IQN was modified.
4. After the host is ready and has the iSCSI initiator configured, define a host object with the storage system as is described in “Preparing for iSCSI connection” on page 469. Make sure that the IQN is set correctly in Host Details in the Port Definitions tab (see Figure 7-76 on page 517).

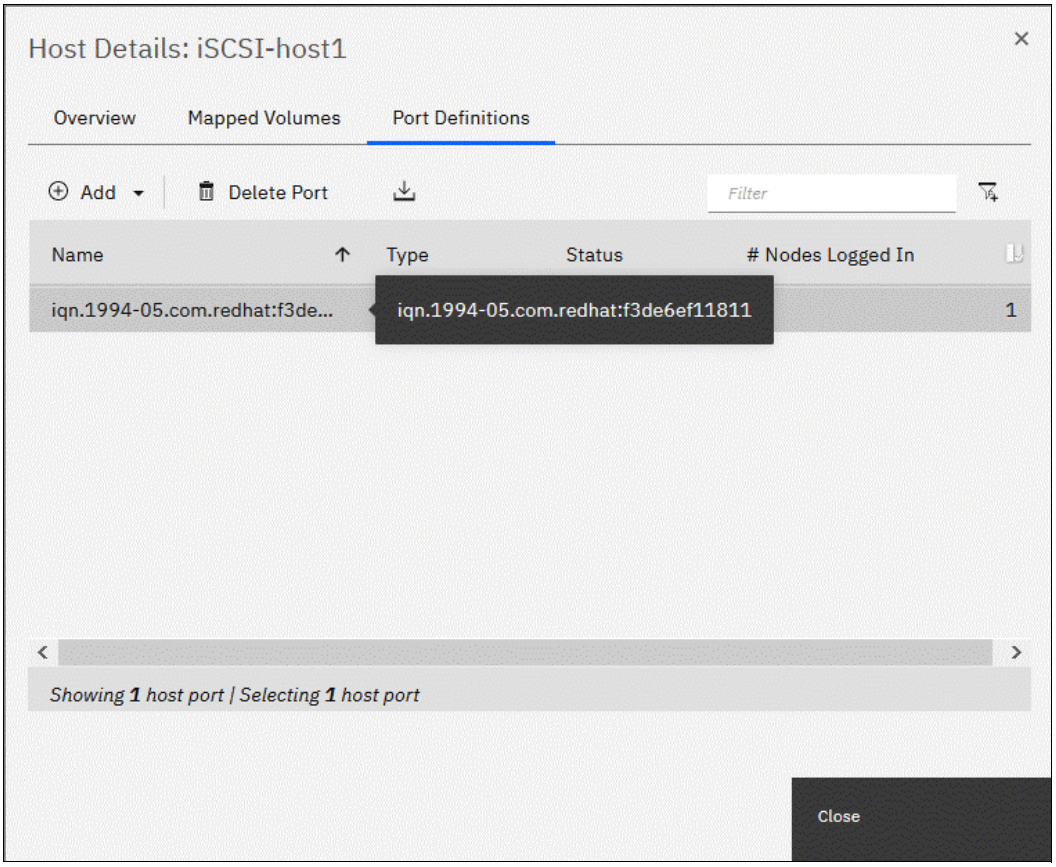


Figure 7-76 Host Details: Port Definitions Tab

5. Map the dedicated volumes (VDisks) to the host object.
6. When iSCSI host object is configured in the storage system and the wanted volumes (VDisks) are mapped to it, discover the iSCSI targets from the host by using one of the following methods:
   - Using send targets
   - Internet storage name service (iSNS)

Record the IP address under the IP column of the Ethernet Ports configuration tab (as shown in Figure 7-77), which was configured for an iSCSI connection as described in “Preparing for iSCSI connection” on page 469 step 4. It is used to determine the IQN of our target and for logging in.

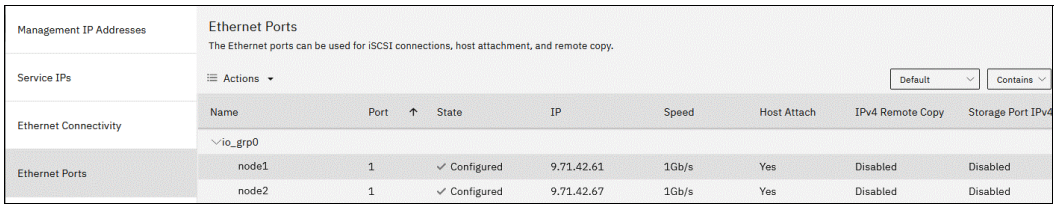


Figure 7-77 Ethernet Ports Configuration tab

Example 7-56 shows discovery by using the send targets method.

Example 7-56 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.ibmflashsystem7200.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.ibmflashsystem7200.node2
```

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 in 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 the **`iscsiadm --mode discovery --type isns`** command 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 7-57).

Example 7-57 Login 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.ibmflashsystem7200.node1 --portal 9.71.42.61 --login
Logging in to [iface: default, target:
iqn.1986-03.com.ibm:2145.ibmflashsystem7200.node1, portal: 9.71.42.61,3260]
(multiple)
Login to [iface: default, target:
iqn.1986-03.com.ibm:2145.ibmflashsystem7200.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.ibmflashsystem7200.node2 --portal 9.71.42.67 --login
Logging in to [iface: default, target:
iqn.1986-03.com.ibm:2145.ibmflashsystem7200.node2, portal: 9.71.42.67,3260]
(multiple)
Login to [iface: default, target:
iqn.1986-03.com.ibm:2145.ibmflashsystem7200.node2, portal: 9.71.42.67,3260]
successful.
```

5. After logging in is successful, ensure that the native multipath driver on the RHEL host is installed and configured correctly, similar to the example with FC connection, and check the output by running the **`multipath -ll`** command.

   An example of the output is shown in Example 7-58.

Example 7-58 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 bold in Example 7-58 on page 518 because they can be used in other configuration tasks, such as creating LVM physical volumes or file systems and mounting. Those devices are in `/dev/mapper/`.

Record the UID number after devices names, without the first digit (3); they correspond to the UID of the volume (VDisk) on the storage system.

Summary

Though the example in “iSCSI host connectivity and capacity allocation” on page 516 is provided specifically for RHEL host connectivity, the main principals can be followed when configuring the connectivity by way of iSCSI for other operating systems.

To summarize the process for host to storage iSCSI connectivity:

1. Install the iSCSI initiator software on the host.
2. Configure the iSCSI initiator software according with the requirements for storage system target and host’s operational system.
3. Get the host IQNs.
4. Define host object with iSCSI connectivity by using host IQNs.
5. Note and check the Ethernet ports IPs on the storage system, which are configured for iSCSI connectivity.
6. Discover iSCSI targets by using the storage system IP that was obtained in the previous step.
7. Log in to the storage system iSCSI targets.
8. Check and configure the native multipath driver to get the correct representation of the volumes that are on the host.

7.8.4 NVMe over Fabric host connectivity example

NVMe-oF uses different fabrics for transport by encapsulating NVMe protocol. In this example, we use Fibre Channel fabric for our NVMe connectivity from RHEL host to IBM SAN Volume Controller system.

Start by defining the necessary connectivity information and configuring our host and storage 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 7-59 to discover the WWPNs of the host as FC-NVMe connectivity is achieved through FC.

Example 7-59 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 7-60 and determine which to use. Because the FC-NVMe connectivity dedicated port is a virtualized port, NPIV must be enabled.

Example 7-60 Finding the NVMe FC ports

```
IBM_FlashSystem:FS9100-1: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 7-60 on page 520, the host is zoned to the ports that are marked in bold.

4. On the host, ensure that the driver is ready to provide NVMe connectivity (see Example 7-61). In this example, we use an Emulex HBA.

Example 7-61 Checking NVMe support for lpfc driver

```
[root@flashlnx4 fc_host]# cat /etc/modprobe.d/lpfc.conf
options lpfc lpfc_enable_fc4_type=3
```

If the `lpfc.conf` file is absent or does not contain the string that is marked in bold in Example 7-61, the file muse be created and populated with the string. Then, the `lpfc`

driver must be restarted by running the **modprob** commands (remove it 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` is installed on the host (see Example 7-62).

Example 7-62 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
```

If these packages are not installed, install them.

6. Obtain the NQN from the host because it is used to define host object on the storage system (see Example 7-63).

Example 7-63 Checking 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 storage system by using the obtained host NQN as described in “Creating NVMe hosts objects” on page 476. Check that the host object has correct NQN set (see Figure 7-78).

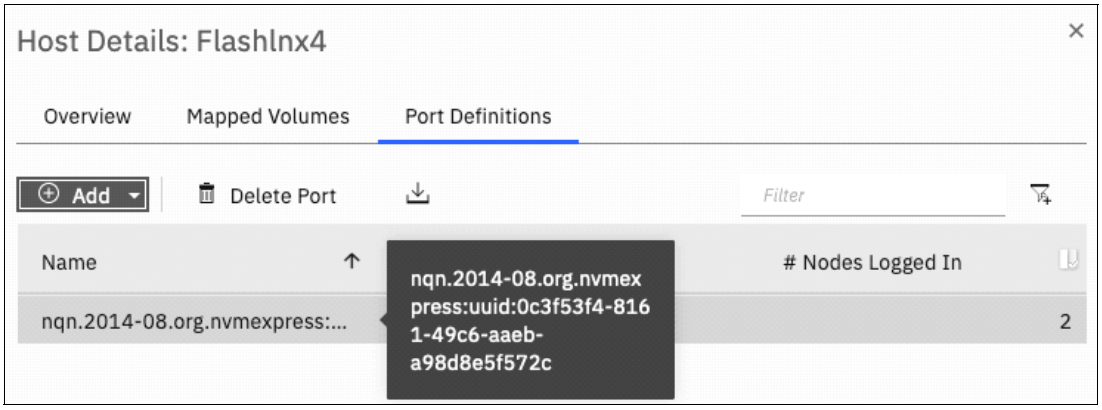


Figure 7-78 Checking host objects NQN on the storage system

8. On the storage system, map the volumes to the host object.
9. If the zoning is done correctly, a host object is created on the storage system and on the host side where the necessary utilities and drivers are configured (see Example 7-64). It is necessary to verify the target ports, which our host can see. This information is used in the discovery and connection process.

Example 7-64 Verifying remote/target ports and information about FC-NVMe connection

```
[root@flashlnx4 nvme]# cat /sys/class/scsi_host/*/nvme_info

NVMe Initiator Enabled
XRI Dist lpfc0 Total 6144 IO 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 OutIO 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 IO 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 OutIO 0000000000000003
   abort 00000001 noxri 00000000 nondlp 00000000 qdepth 00000000 wqerr 00000000
err 00000000
FCP CMPL: xb 00000001 Err 00000005
```

Tip: If remote ports (RPORT), which are presented from the storage system, are not visible, check whether the zoning was done correctly to the virtualized NVMe ports on the storage 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 7-65 can be used to automate this process. The commands from **`nvme-cli`** are in bold.

Example 7-65 Script which can be used 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
```

An example of the possible output is shown in Example 7-66.

Example 7-66 Discovery and Connection

```
[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
```

After the systems are successfully discovered and connected, record the ports that are marked in bold in Example 7-66 on page 523 and check the list of `nvme` devices that are visible from the host (see Example 7-67).

Example 7-67 NVMe devices list, visible from the host

```
[root@flashlnx4 tmp]# nvme list
Node             SN                   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 the **multipath -ll** command to see the paths and information about the volumes (see Example 7-68).

Example 7-68 Output of multipath -ll command

```
multipath -ll
...
eui.2800000000000053005076081 08a000f 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. Note the name of the volume and UID in bold in Example 7-68 for use in future actions, such as adding a partition or implementing this volume as part of LVM. In Figure 7-79, the same volume is shown on the storage system.

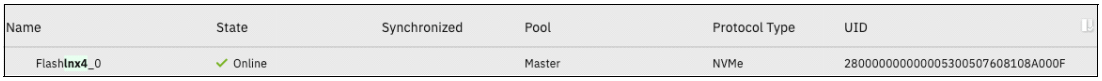


Figure 7-79 Volume as it is seen on the storage system

Summary

This section provided an example of using a RHEL host and an IBM FlashSystem 9100 system. Although other operating system distributions might include 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 process:

1. Ensure that the host is ready and meets the requirements for FC-NVMe connectivity, such as the following examples:
   - HBA supports FC-NVMe
   - Drivers are configured for NVMe connectivity
2. Ensure that the storage system supports the host HBA for FC-NVMe connectivity.
3. Obtain connectivity information from the host.
4. Create host object on the storage system by using connectivity information from the host.
5. Map the wanted volumes to the host object.
6. Perform discovery and connection from the host (some hosts' operating systems can perform this task automatically).
7. Use the obtained storage resources.

7.9 Container Storage Interface drivers

With IBM Spectrum Virtualize 8.4.2, support for containers is now available.

A *container* is a lightweight application package that is designed to 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 function, 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 hGitHub 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

Storage migration

This chapter describes the steps that are involved in migrating data from an external storage system to the IBM SAN Volume Controller by using the storage migration wizard. Migrating data from other storage systems to the IBM SAN Volume Controller consolidates storage and enables IBM Spectrum Virtualize features, such as IBM Easy Tier, thin provisioning, compression, encryption, storage replication, and the easy-to-use GUI to be used across all volumes.

Storage migration uses the volume mirroring functionality to allow reads and writes during the migration, and minimizes disruption and downtime. After the migration is complete, the existing external storage system can be retired.

IBM SAN Volume Controller supports migration through Fibre Channel (FC) and internet Small Computer Systems Interface (iSCSI) connections.

In addition to migrating data by way of external virtualization and volume mirroring that is used by the storage migration wizard, scenarios exist in which host-based mirroring is an alternative.

In environments where operating system administrators can and want to 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 Spectrum Virtualize system and the established storage systems are visible to the host concurrently. However, a reboot 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 Spectrum Virtualize, see this web page (log in required). Although this demonstration shows an IBM FlashSystem, it also is applicable to IBM SAN Volume Controller.

At the web page, click **Start** to explore the menu of options available. The Explore external virtualization menu demonstrates the steps involved in scenarios, such as third-party storage array virtualization and Storage migration.

© Copyright IBM Corp. 2021. All rights reserved.

With the IBM Spectrum Virtualize 8.4.2 code release, it is now possible to migrate data between two IBM Spectrum Virtualize systems non-disruptively, without the need to cluster or externally virtualize. This volume mobility method also can be used to offload some workloads from one IBM Spectrum Virtualize system cluster to another. For example, from one IBM SAN Volume Controller to another.

The non-disruptive system migration uses the remote-copy relationship type method to migrate volumes between the source and the target system; for example migrating from a FlashSystem to an IBM SAN Volume Controller or vice-versa. By using this method, the data migration of volumes between systems is done without any application downtime.

This chapter includes the following topics:

- 8.1, “Storage migration overview” on page 529
- 8.2, “Storage migration wizard” on page 531
- 8.3, “Migrating data between systems non-disruptively” on page 548

Note: This chapter does *not* cover migration outside of the storage migration wizard. To migrate data outside of the wizard, you must use the Import option in the GUI.

This chapter also does not cover virtualization of external storage.

For more information about these topics, see Chapter 5, “Using storage pools” on page 271.

8.1 Storage migration overview

To migrate data from a storage system to the IBM SAN Volume Controller, it is necessary to use the built-in External Virtualization capability. This capability places external connected logical units (LUs) under the control of the IBM SAN Volume Controller, which acts as a proxy while hosts continue to access them. The volumes are then fully virtualized in the IBM SAN Volume Controller.

Note: A license is required for external systems that are being virtualized based on storage capacity units. 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 purchasing 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 a new license.

Consider the following points about the storage migration process:

- Typically, storage systems divide storage into many Small Computer System Interface (SCSI) LUs that are presented to hosts.
- I/O to the LUs must be stopped and changes made to the mapping of the external storage system LUs and to the fabric or iSCSI configuration so that the original LUs are presented directly to the IBM SAN Volume Controller and not to the hosts. The IBM SAN Volume Controller discovers the external LUs as *unmanaged* managed disks (MDisks).
- The unmanaged MDisks are *imported* to the IBM SAN Volume Controller as *image-mode volumes* and placed into a temporary storage pool. This storage pool is now a logical container for the LUs.
- Each MDisk has a one-to-one mapping with an image-mode volume. From a data perspective, the image-mode volumes represent the LUs exactly as they were before the import operation. The image-mode volumes are on the same physical drives of the external storage system and the data remains unchanged. The IBM SAN Volume Controller is presenting active images of the LUs and is acting as a proxy.
- You might need to remove the storage system multipath device driver from the host, and reconfigure host attachment with this system. However, most current operating systems might not require vendor specific multipathing drivers and can access the established and new Spectrum Virtualize system by using native multipathing drivers, such as AIX AIXPCM, Linux device mapper, or Windows 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 Spectrum 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 LUs are now migrated and the now redundant storage can be decommissioned or reused elsewhere.

Important: If you are migrating volumes from an IBM Storwize or FlashSystem family product, be aware that the source system must be configured in the *storage* layer. Otherwise, the IBM SAN Volume Controller does not discover the source as a backend controller.

The default layer setting for Storwize and FlashSystem family systems is *storage*:

```
chsystem -layer replication
chsystem -layer storage
```

8.1.1 Interoperability and compatibility

Interoperability is an important consideration when a new storage system is set up in an environment that contains existing storage infrastructure. Before attaching any external storage systems to the IBM SAN Volume Controller, see the IBM System Storage Interoperation Center (SSIC).

Select **IBM System Storage IBM SAN Volume Controller** in Storage Family, then, **SVC Storage Controller Support** in Storage Model. You can then refine your search by selecting the external storage controller that you want to use in the **Storage Controller** menu.

The matrix results indicates the external storage that you want to attach to the IBM SAN Volume Controller, such as minimum firmware level or support for disks greater than 2 TB.

8.1.2 Prerequisites

Before the storage migration wizard can be started, the external storage system must be visible to the IBM SAN Volume Controller. You also need to confirm that the restrictions and prerequisites are met.

Administrators can migrate data from the external storage system to the system that uses iSCSI or FC. For more information about how to manage external storage, see Chapter 5, “Using storage pools” on page 271.

Common prerequisites

In VMware environments, Storage vMotion often is used to move guest data transparently to newly provisioned data stores from the IBM SAN Volume Controller, whereas storage virtualization might be a good option for Raw Device Mapping.

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 non-production server. The LUN has a different unique identifier (UID) after it is imported. It resembles a mirrored volume to the VMware server.

Prerequisites for Fibre Channel connection

The following prerequisites for FC connectivity must be met:

- Cable this system into the SAN of the external storage that you want to migrate. Ensure that your system is cabled into the same SAN as the external storage system that you are migrating.
- If you use FC, connect the FC cables to the FC ports in *both* nodes of your system, and then, to the FC network. If you use FCoE, connect Ethernet cables to the 10 Gbps Ethernet ports.

For more information, see 2.6, “Fibre Channel SAN configuration planning” on page 65. Alternatively, directly attach the external storage system to the IBM SAN Volume Controller nodes instead of using a switched fabric.

Prerequisites for iSCSI connections

The following prerequisites for iSCSI connections must be met:

- Cable the system to the external storage system with a redundant switched fabric. Migrating iSCSI external storage requires that the system 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.
- 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 storage system to fabric in the same way as the 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, extra redundant iSCSI connection can be established for increased throughput.

8.2 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 what is performed throughout the process.

Attention: The risk of losing data when the storage migration wizard is used correctly is low. However, it is prudent to avoid potential data loss by creating a backup of all the data that is stored on the hosts, the existing storage systems, and the IBM SAN Volume Controller before the wizard is used.

Complete the following steps to complete the migration by using the storage migration wizard:

1. Select **Pools** → **System Migration**, as shown in Figure 8-1. The System Migration pane provides access to the storage migration wizard and displays information about the migration progress.

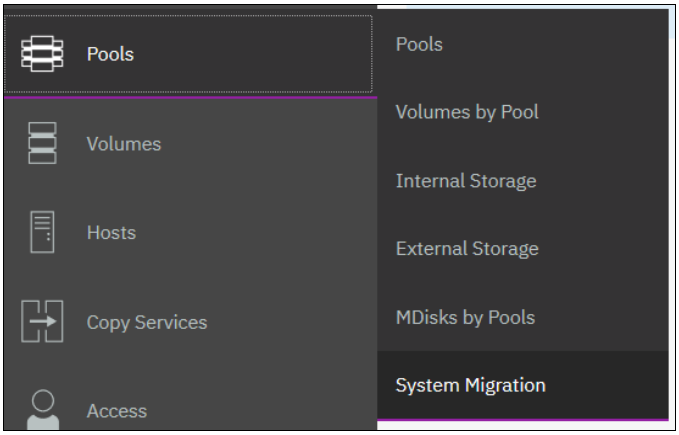


Figure 8-1 Navigating to System Migration

2. Click **Start New Migration** to begin the storage migration wizard, as shown in Figure 8-2.

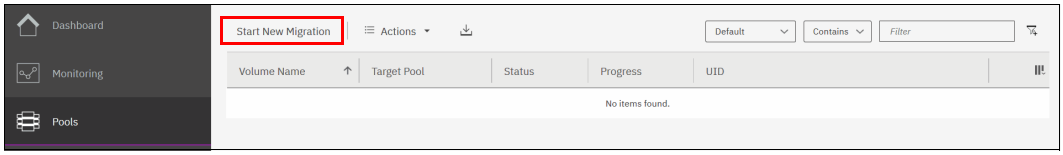


Figure 8-2 Starting a migration

Note: Starting a new migration adds the volume to be migrated to the list displayed in the pane. 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 dialogue is shown prompting you to decide which protocol to use. Select the type of attachment between the IBM SAN Volume Controller and the external system from which you want to migrate volumes and click **Next**. (If only one type of attachment is detected, this dialogue is not displayed.)

   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.

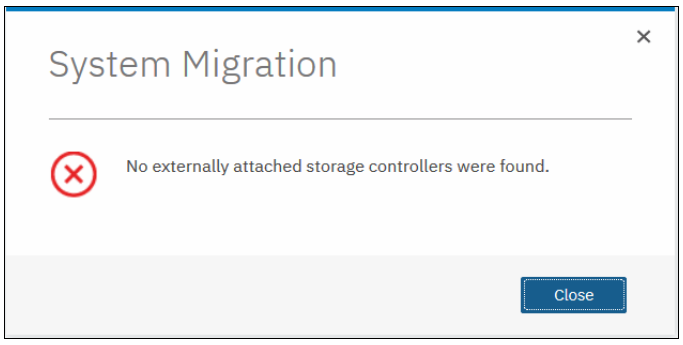


Figure 8-3 Error message if no external storage is detected

4. When the wizard starts, you are prompted to verify the restrictions and prerequisites that are listed in Figure 8-4 on page 534. Address the following restrictions and prerequisites:
   - Restrictions:
     - You are not using the storage migration wizard to migrate clustered hosts, including clusters of VMware hosts and Virtual I/O Servers (VIOS).
     - You are not using the storage migration wizard to migrate SAN boot images.

     If you have either of these two 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 information, see this web page.
   - Prerequisites:
     - IBM SAN Volume Controller nodes and the external storage system 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 8.1.2, “Prerequisites” on page 530.

If all restrictions are satisfied and prerequisites are met, select all of the boxes and click **Next**, as shown in Figure 8-4.

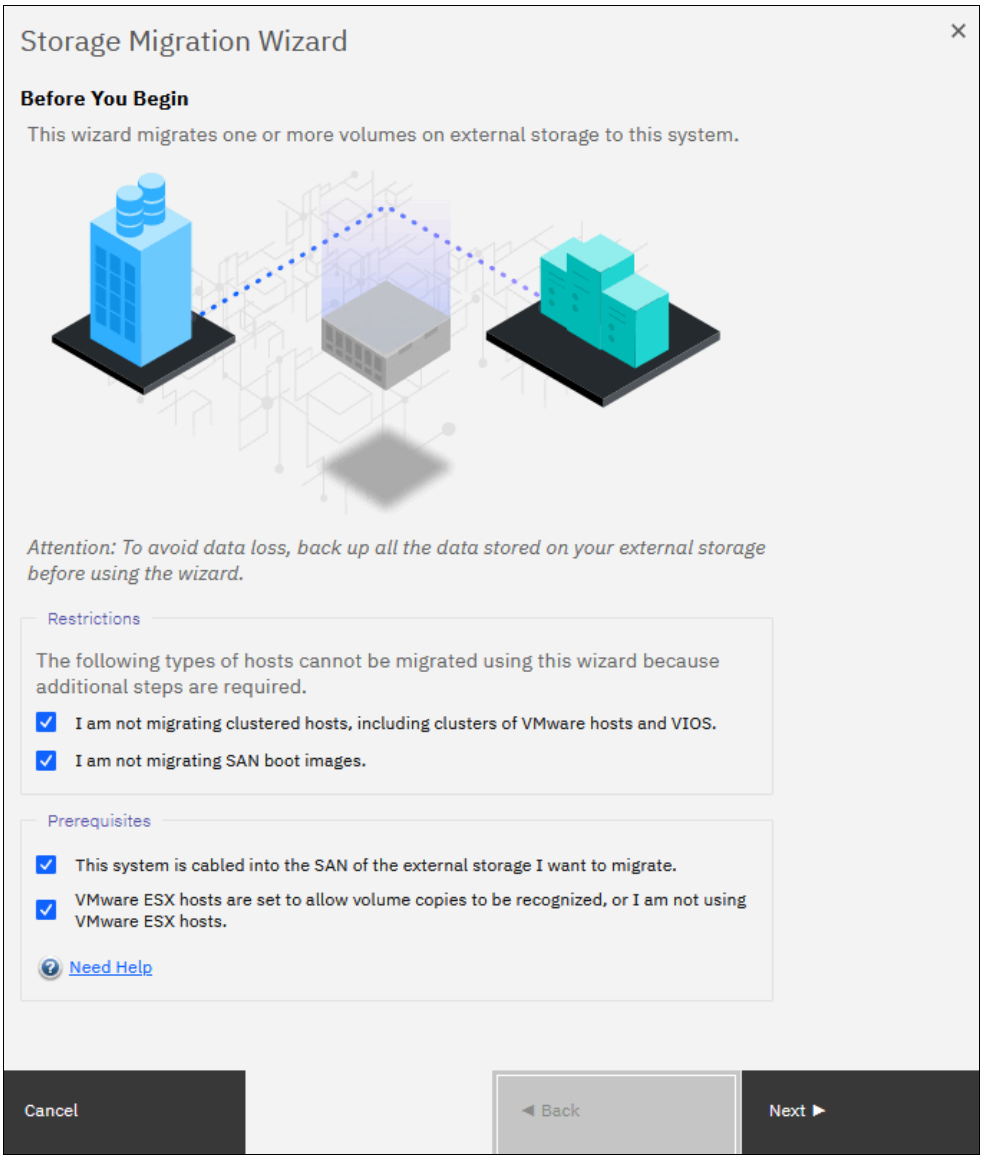


Figure 8-4 Restrictions and prerequisites confirmation

5. Prepare the environment migration by following the instructions that are shown in Figure 8-5.

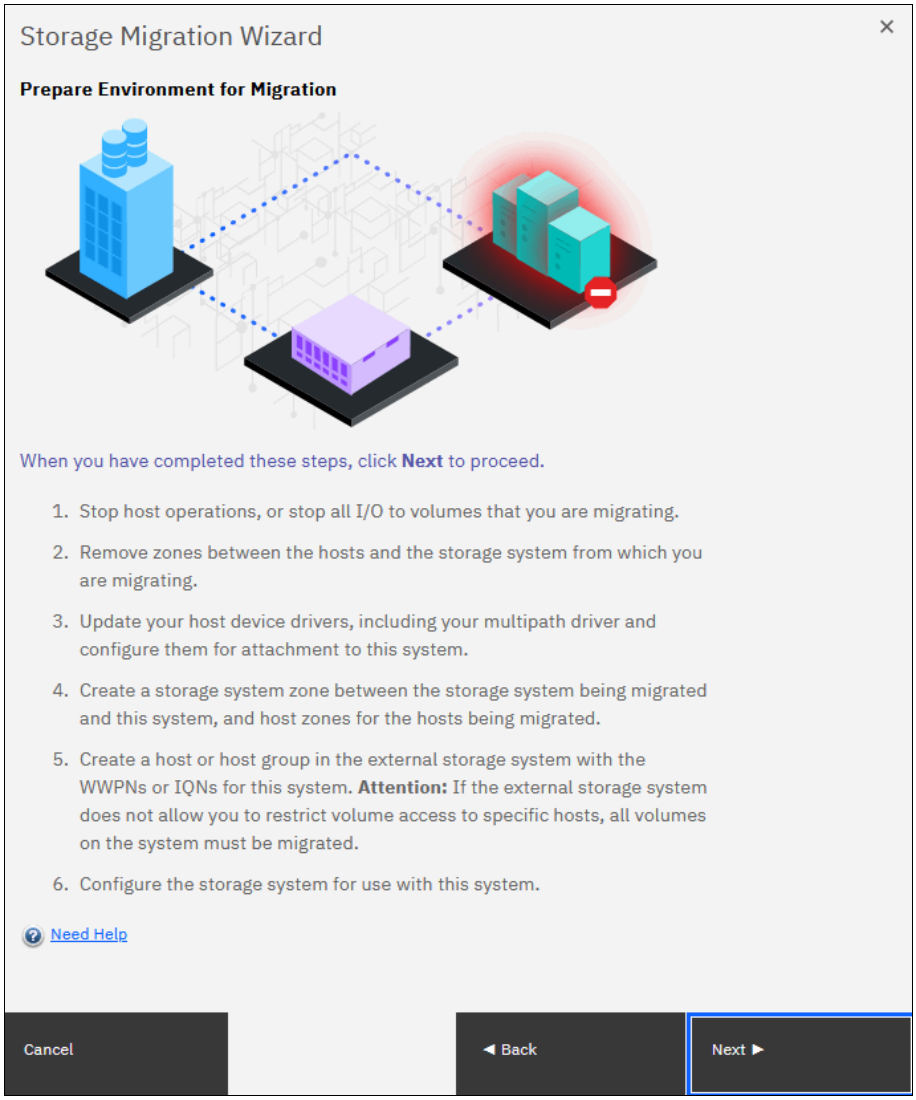


Figure 8-5 Preparing your environment for storage migration

The preparation phase includes the following steps:

a. Before migrating storage, ensure that all host operations are stopped to prevent applications from generating I/Os to the migrated system.

b. Remove all existing zones between the hosts and the system you are migrating.

c. Hosts usually do not support concurrent multipath drivers at the same time. You might need to remove drivers that are not compatible with the IBM SAN Volume Controller, from the hosts and use the recommended device drivers. For more information about supported drivers, see the IBM SSIC.

d. If you are migrating external storage systems that connect to the system that uses FC or FCoE connections, ensure that you complete suitable zoning changes to simplify migration. Refer to 2.6.2, “Zoning” on page 66 for more information.

e. Because the IBM SAN Volume Controller should now be seen as a host cluster from the external system to be migrated, you must define the IBM SAN Volume Controller as a host or host group by using the WWPNs or IQNs on the system to be migrated. Some established systems do not permit LUN-to-host mapping and then present all the LUs to the IBM SAN Volume Controller. In that case, all the LUs must be migrated.

6. If the previous steps are followed, the IBM SAN Volume Controller is now seen as a host from the system to be migrated. LUs can then be mapped to the IBM SAN Volume Controller. Map the external storage system by following the instructions that are shown in Figure 8-6.



Figure 8-6 Steps to map the LUs to be migrated to the IBM SAN Volume Controller

Before you migrate storage, record the hosts and their WWPNs or IQNs for each volume that is being migrated, and the SCSI LUN when mapped to the IBM SAN Volume Controller.

Table 8-1 shows an example of a table that is used to record information that relates to the external storage system LUs.

Table 8-1 Example table for capturing external LU information

Volume name or ID	Hosts accessing this LUN	Host WWPNs or IQNs	SCSI LUN when mapped
1 IBM DB2 logs	DB2server	21000024FF2...	0
2 DB2® data	DB2Server	21000024FF2...	1
3 file system	FileServer1	21000024FF2...	2

Note: Make sure to record the SCSI ID of the LUs 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 issues a **`detectmdisk`** command, as shown in Figure 8-7.

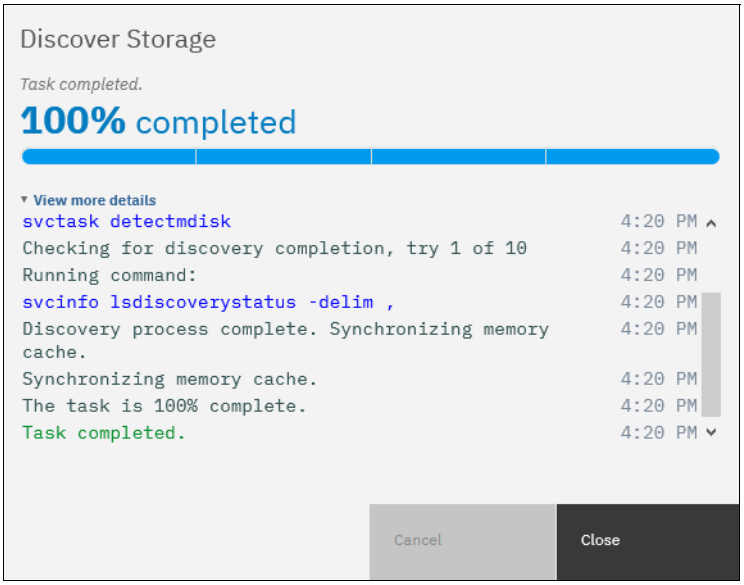


Figure 8-7 Storage Migration external storage discovery detectmdisk detail

7. The next window shows all of 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 the previous step to trigger the discovery procedure again.

Select the MDisks that you want to migrate, as shown in Figure 8-8. In this example, the following MDisks were found and will be migrated: `mdisk18` and `mdisk16`. Detailed information about an MDisks are available by double-clicking it.

To select multiple elements from the table, use the standard Shift+left-click or Ctrl+left-click actions. 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**.

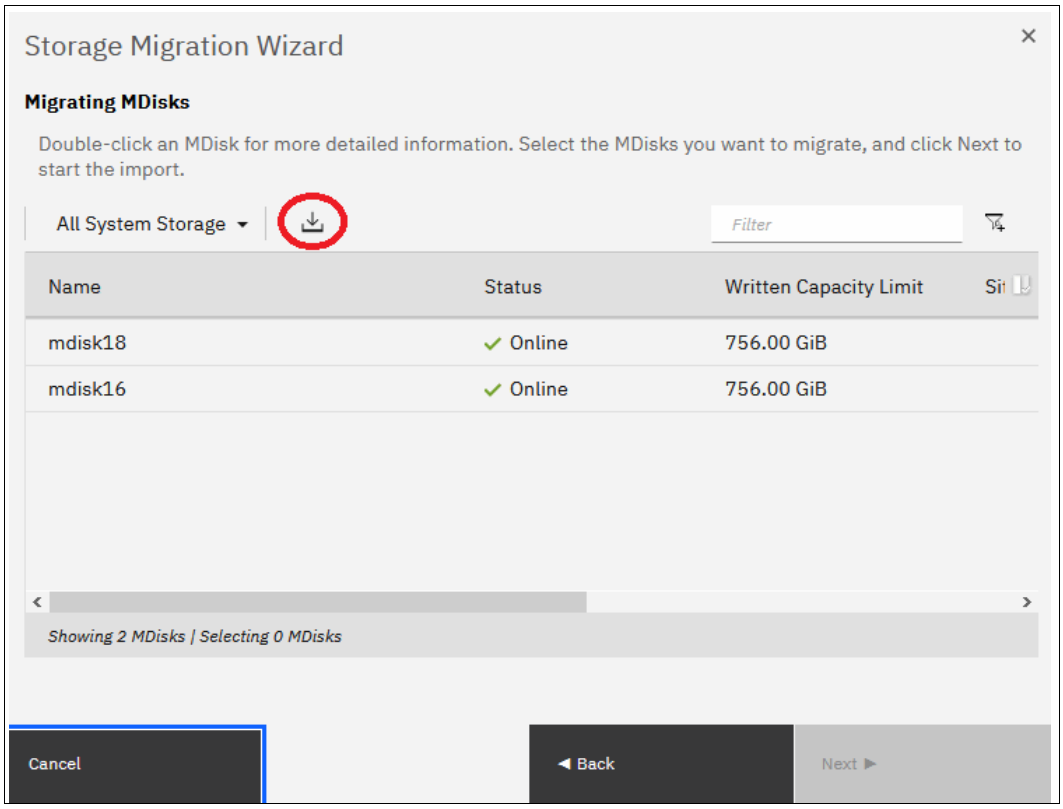


Figure 8-8 Discovering mapped LUs from external storage

Note: Select only the MDisks that are applicable to the current migration plan. After step 15 on page 545 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 system creates a new 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 on page 539.

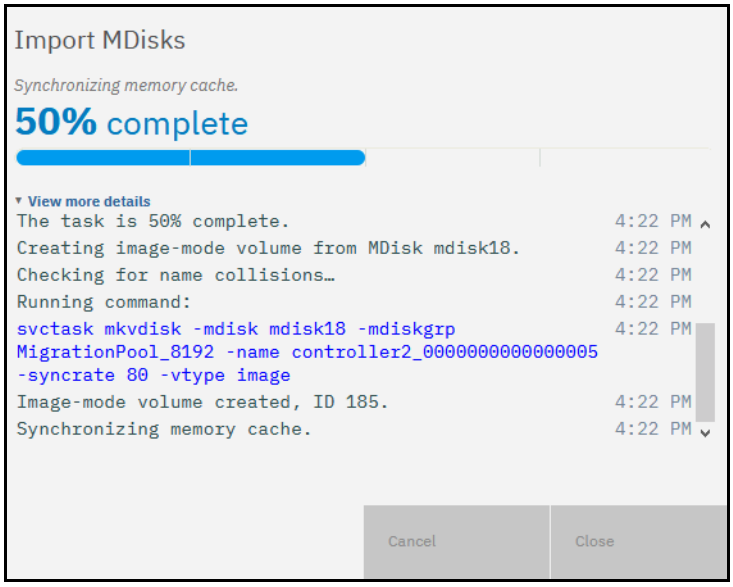


Figure 8-9 Storage Migration image mode volume creation detail

9. The next window lists the hosts that are configured on the system and enables you to configure new hosts. This step is optional and can be bypassed by clicking **Next**. In this example, the host `X366-SLES-12SP2` is configured, as shown in Figure 8-10. If no host is selected, you can create a host after the migration completes and map the imported volumes to it.

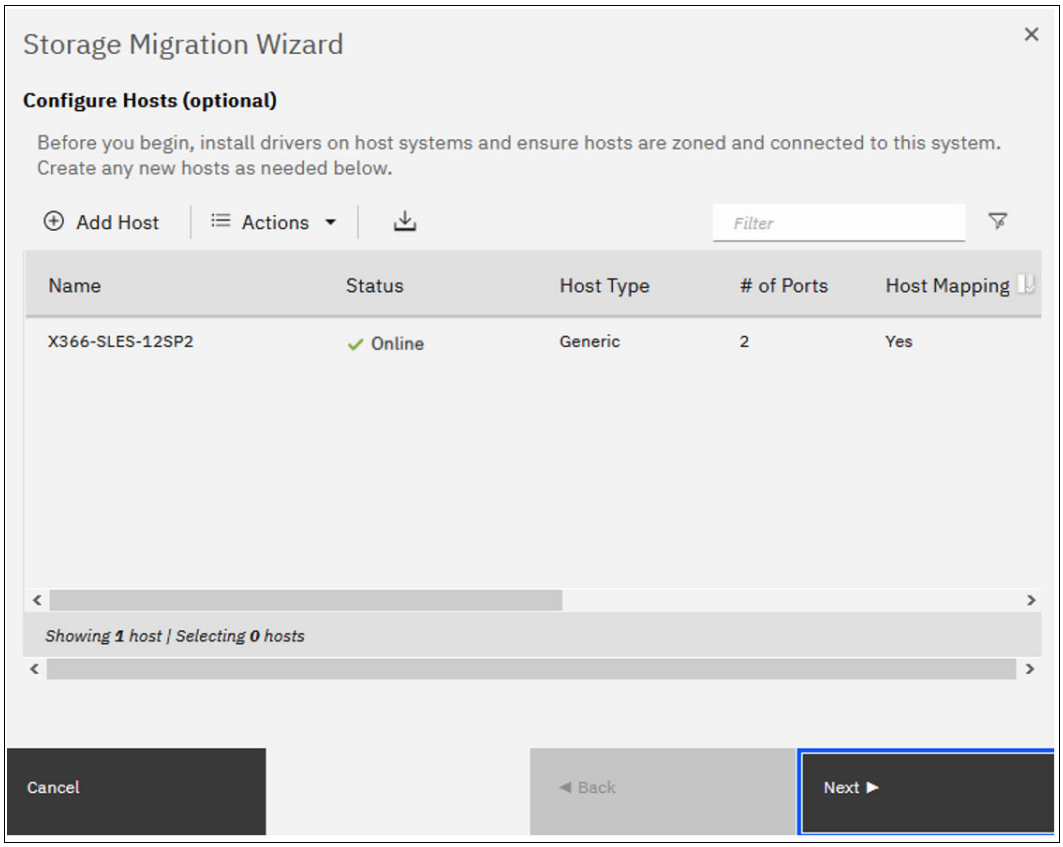


Figure 8-10 Listing of configured hosts to map the imported Volume to

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 shows the Add Host wizard with the details completed.

    For more information about the Add Host wizard, see Chapter 7, “Hosts” on page 443.

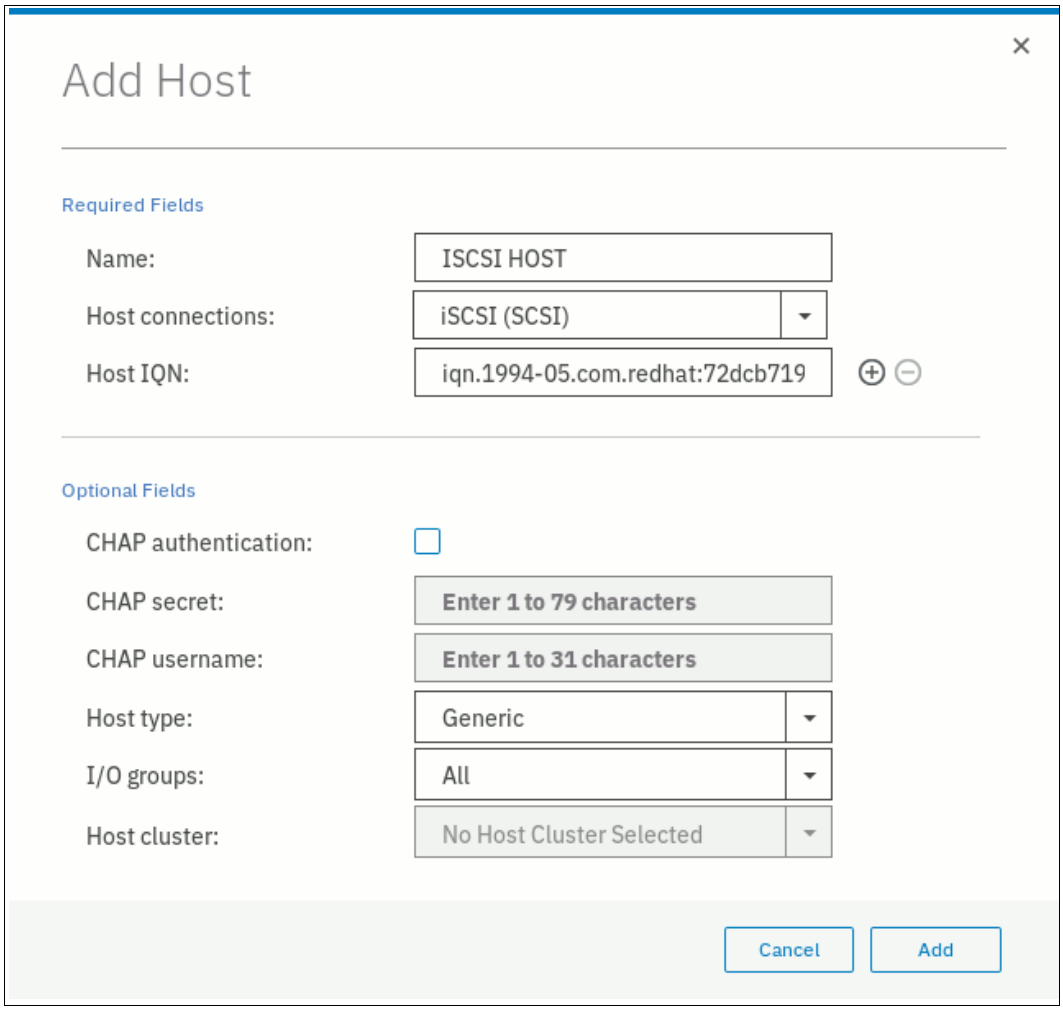


Figure 8-11 If not already defined, you can create a host during the migration process

11. Click **Add**. The host is created and is now listed in the Configure Hosts window, as shown in Figure 8-10 on page 539. Click **Next** to proceed.

12. The next window lists the new volumes and enables you to 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.

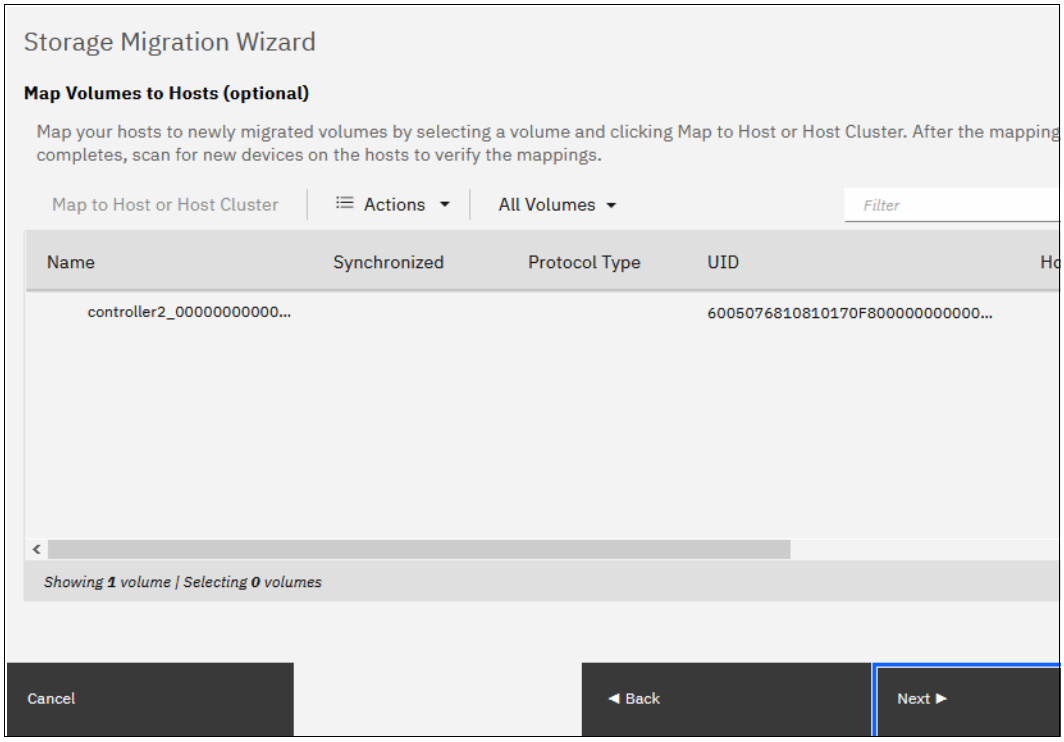


Figure 8-12 Map volumes to hosts

13. Map the volumes to hosts by selecting the volumes and clicking **Map to Host or Host Cluster**, as shown in Figure 8-13. This step is optional and can be bypassed by clicking **Next**.

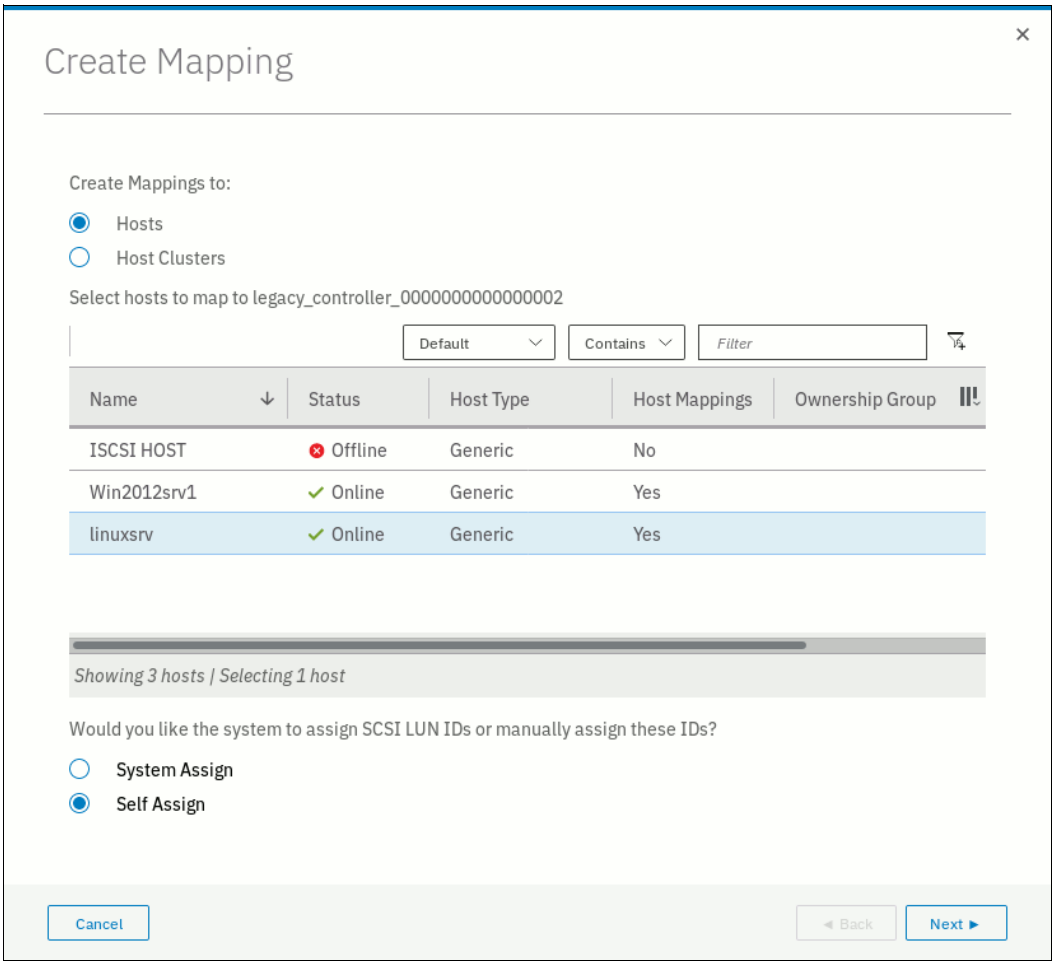


Figure 8-13 Select the host to map the new Volume to

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 543.

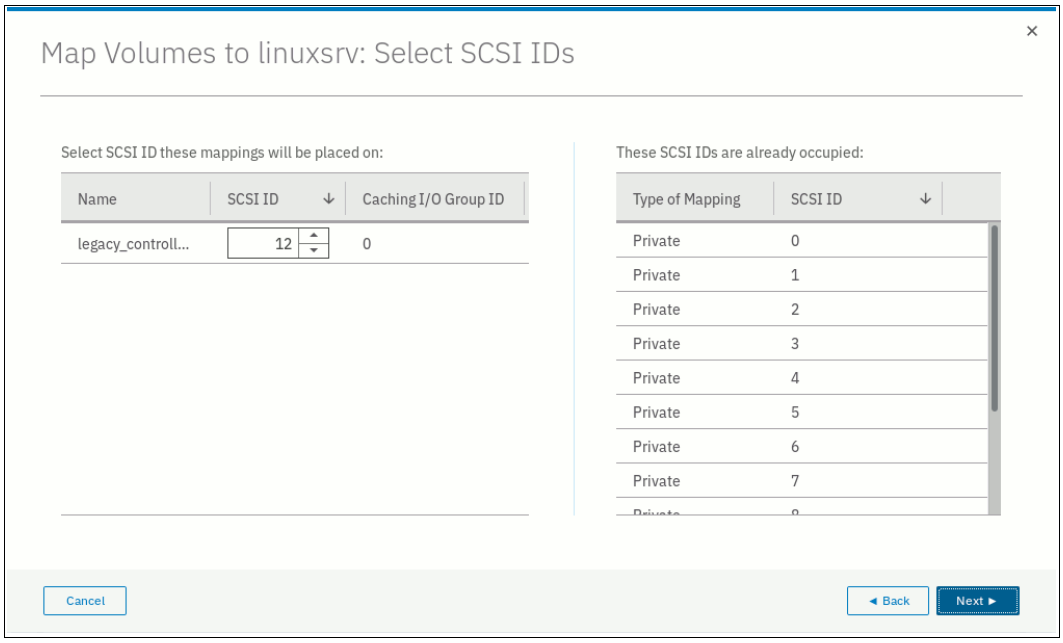


Figure 8-14 Manually assign a LUN SCSI ID to mapped Volume

When your LUN mapping is ready, click **Next**. A new window is displayed with a summary of the new and existing mappings, as shown in Figure 8-15.

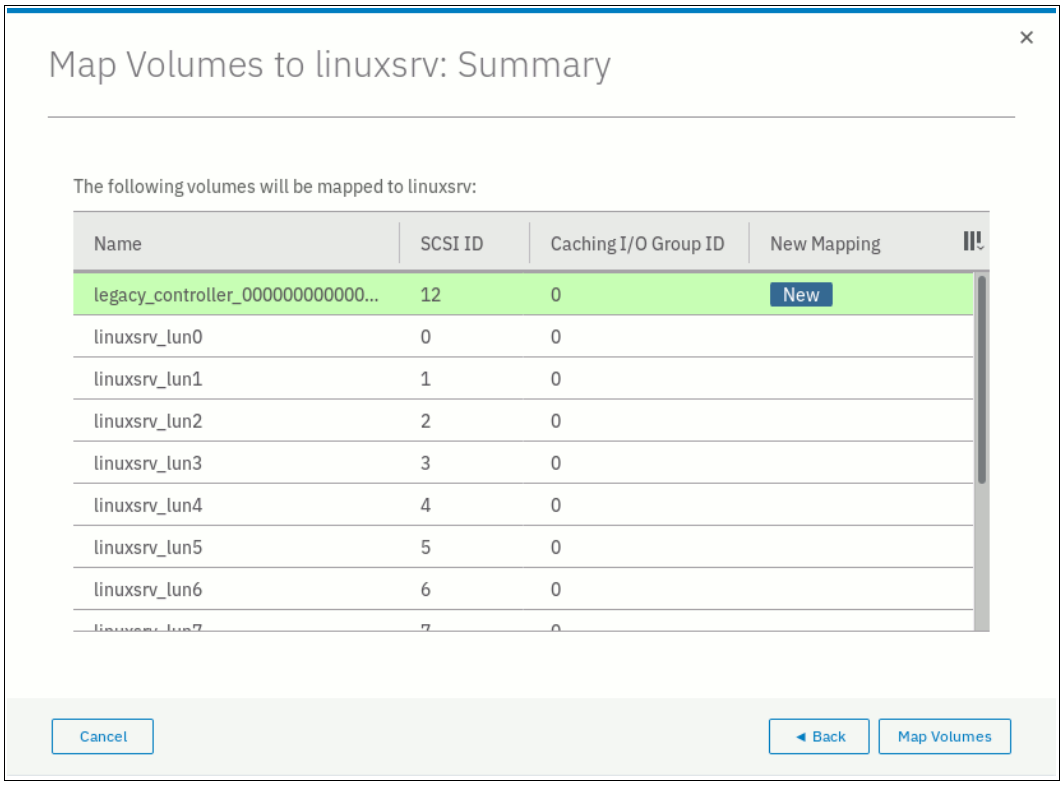


Figure 8-15 Volumes mapping summary before migration

Click **Map Volumes** and wait for the mappings to be created. Continue to map volumes to hosts until all mappings are created. Then, click **Next** to continue with the next migration step.

14. Select the storage pool that you want to migrate the imported volumes into. Ensure that the selected storage pool has enough space to accommodate the migrated volumes before you continue. This step is optional. 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 storage system to be migrated to the system. Therefore, although it is acceptable to delay the mirroring at some point, it should be done. Click **Next**, as shown in Figure 8-16.

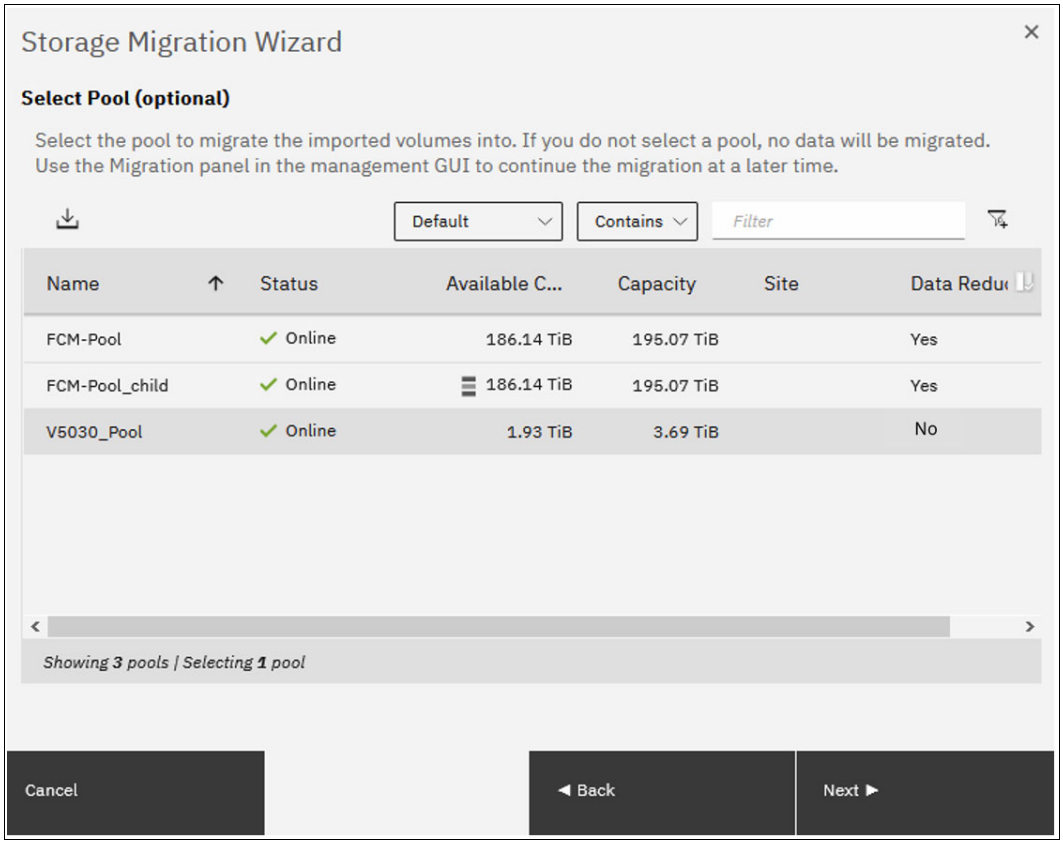


Figure 8-16 Select the target pool for the migration of the image mode MDisk

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.

For more information about Volume Mirroring, see Chapter 6, “Creating and provisioning volumes” on page 331.

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 at the same time 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.

For more information about volume mirroring synchronization rate, see Chapter 6, “Creating and provisioning volumes” on page 331.

15. Click **Finish** to end the storage migration wizard, as shown in Figure 8-17.

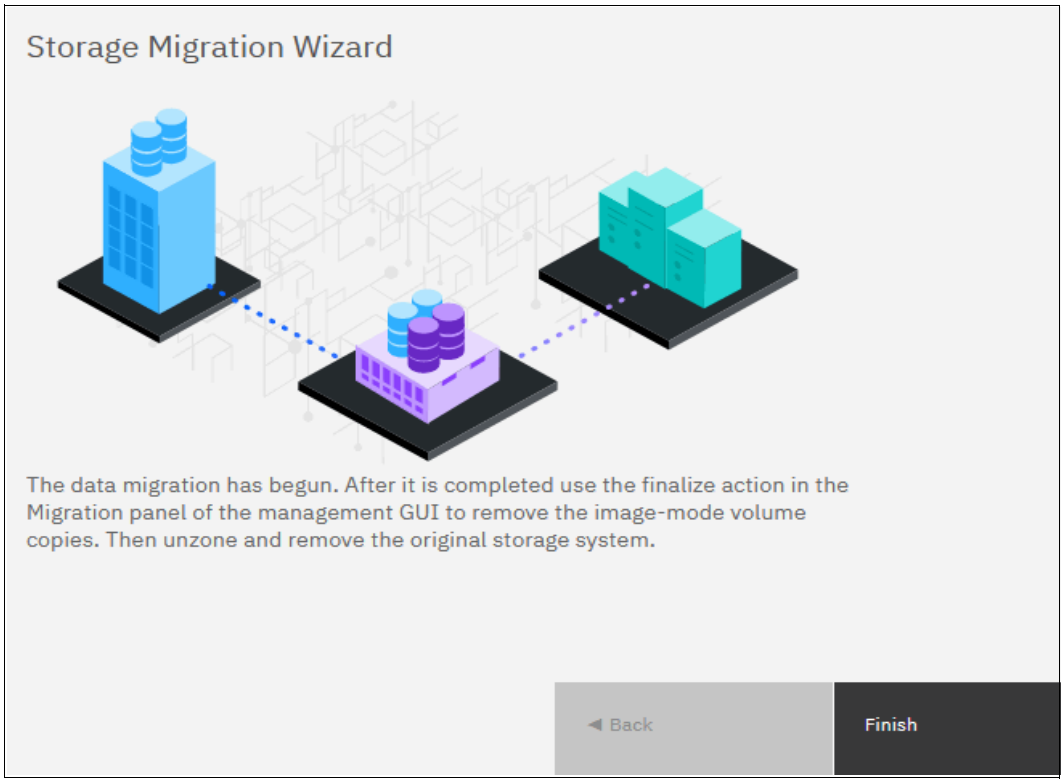


Figure 8-17 Migration is started

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.

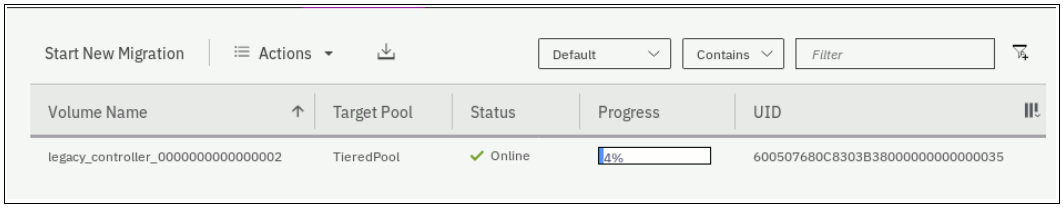


Figure 8-18 The ongoing Migration is listed in the Storage Migration window

16. If you want to check the progress by using the 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 CLI

```
IBM_2145:ITSO-SV1:superuser>lsvdisksyncprogress
vdisk_id vdisk_name                          progress estimated_completion_time
20       legacy_controller_0000000000000002 1        191021123932
```

17. When the migration completes, select all of the migrations that you want to finalize, right-click the selection, and click **Finalize**, as shown in Figure 8-19.

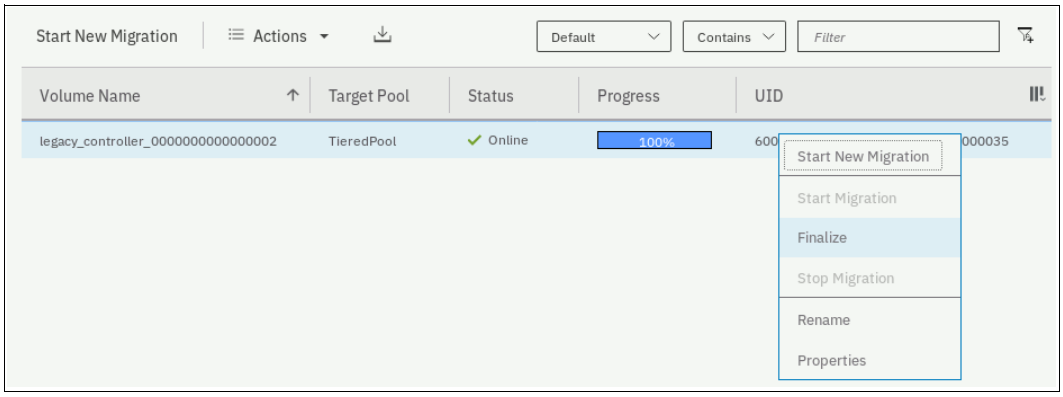


Figure 8-19 Finalizing a migration

You are prompted to confirm the Finalize action because this action 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.

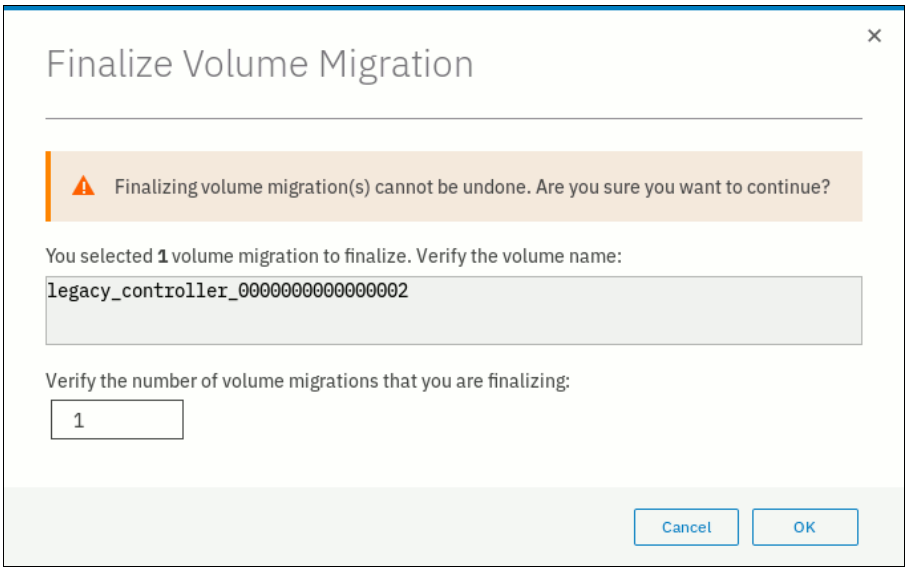


Figure 8-20 Migration finalization confirmation

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, and it displays as `unmanaged` in the external storage window.

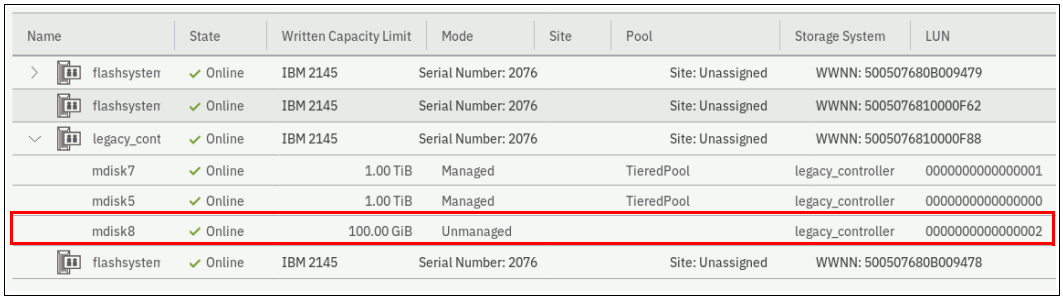


Figure 8-21 External Storage MDisks window

All the steps that are described in the Storage Migration wizard can be performed manually by using the CLI, but it is highly recommended to use the wizard as a guide.

8.3 Migrating data between systems non-disruptively

The volume mobility feature in IBM Spectrum Virtualize 8.4.2 release enables migrating volume data between systems non-disruptively.

A subtle difference exists between the non-disruptive 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 we have clusters that are reaching maximum limits from a cluster perspective, such as maximum VDisks, this method provides the ability to move volumes non-disruptively while maintaining I/O to the host application through the entire duration.

Using the existing remote copy (Metro Mirror) function and enhancements to the Asymmetric Logical Unit Access (ALUA) path state of features of SCSI made this new migration method available. This non-disruptive volume mobility feature is an alternative, but not a replacement for the image mode and clustering based methods, as discussed in Chapter 10, “Advanced Copy Services” on page 605.

In its first release, the following restrictions exists regarding the use of this method (for more information about the latest updates, see this IBM Documentation web page):

- 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 Fibre Channel or IP partnerships between them. For more information about validating whether the systems are suitable to be upgraded to 8.4.2, see this IBM Support web page.
- Current host support and restrictions:
  - RHEL, SLES, VMware ESXi, Solaris, HP-UX. SCSI only:
    - Solaris 10, 11 and HP-UX 11iV4 - Fibre Channel connectivity
    - RHEL 7.x, 8.x and SLES 12.x, 15.x - iSCSI and Fibre Channel connectivity
    - VMware ESXi 7.x - iSCSI, iSER and Fibre Channel connectivity
    - VMware ESXi 6.x - iSCSI and Fibre Channel connectivity
  - IBM i, AIX, Hyper-V and Windows are not supported.
  - SAN-boot volumes not supported.
  - Volumes mapped to NVMe-attached hosts is not supported.

– 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 re-enabled on the 8.4.2 software. As of this writing, ODX support is not available on 8.4.2.

8.3.1 Non-disruptive volume migration

In this section, we present an example of a host with one volume that is non-disruptively migrated from a source IBM SAN Volume Controller named `vvolsftw-sv1` to the target IBM SAN Volume Controller named `vs03-svc` by using the graphical user interface (GUI).

For more information about the GUI and CLI instructions, see this IBM Documentation web page.

For our example, Figure 8-22 shows one host attached volume `vdisk0` from the source IBM SAN Volume Controller with a specific UUID `600507680CD48000DC00000000000000`.

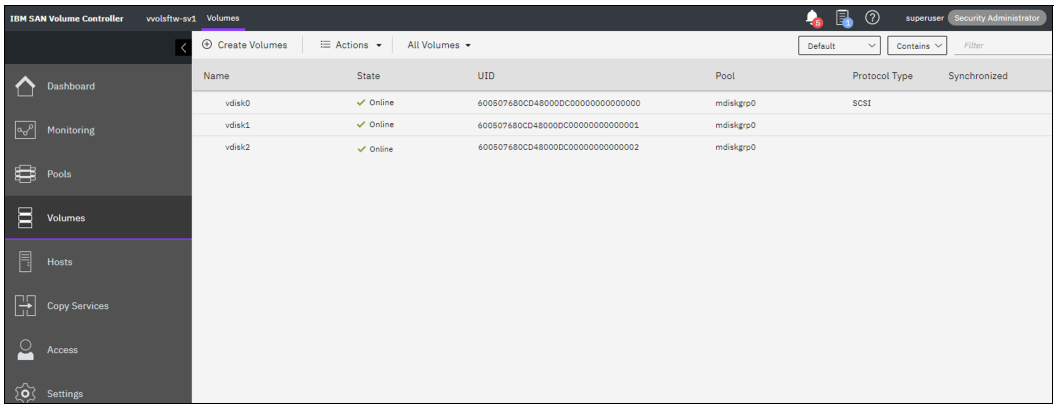


Figure 8-22 Existing volume on source system

Note: With the current restrictions, consistency groups are not supported. If multiple volumes are used, each of those volumes are individually migrated.

Complete the following steps to non-disruptively migrate a volume:

1. Create a partnership if it does not yet exist between the source and target system in scope, as shown in Figure 8-23. This partnership is required so that the volume mobility feature can use the remote copy replication. It supports IP or FC partnerships, which can be created by using the CLI and GUI methods.

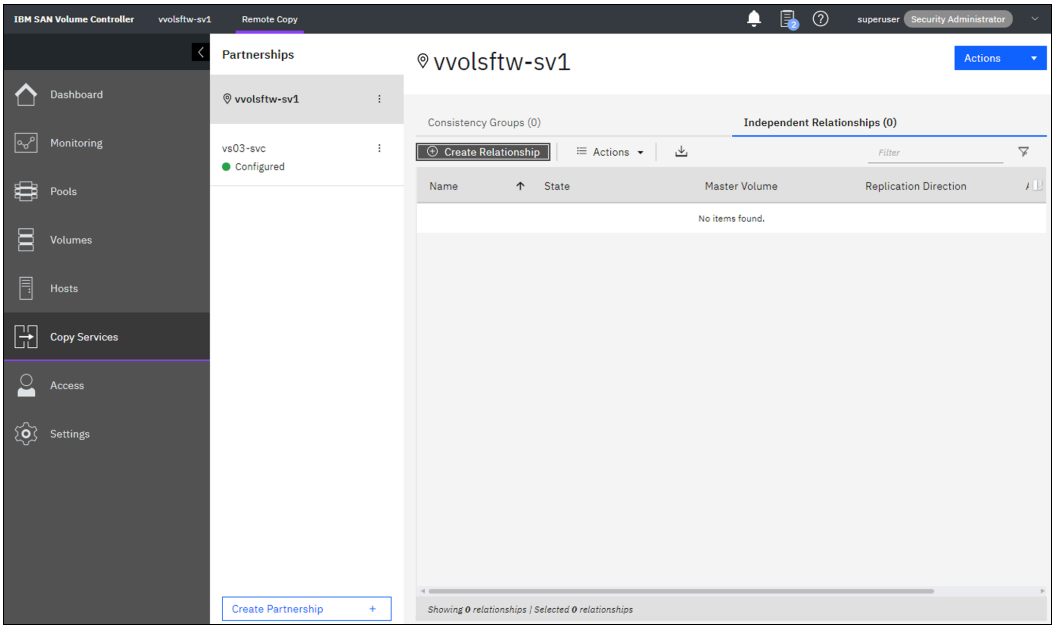


Figure 8-23 Establish partnership between source and target systems

2. Ensure that a volume that matches the size of the source system is created on the target system, but with a different UUID 600507680C8F00004C00000000000000 than the source system, as shown in Figure 8-24.

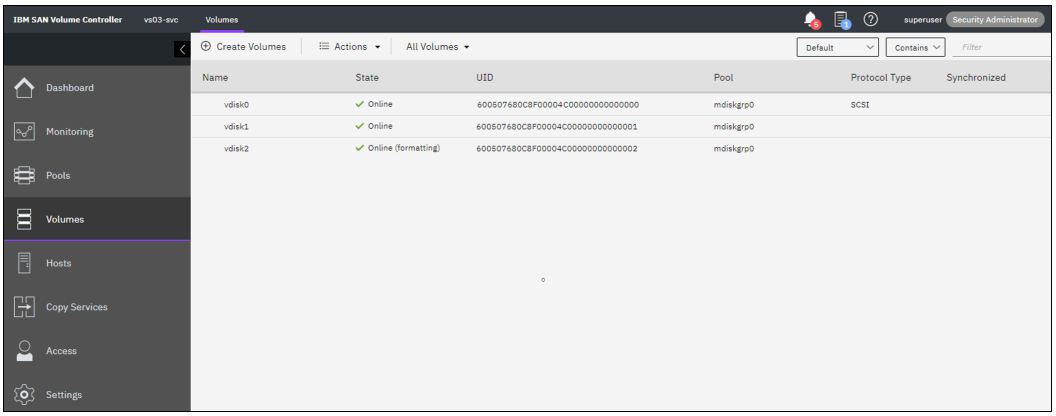


Figure 8-24 New volume on target system

The Metro Mirror relationship is created manually with a dedicated **`migration`** switch parameter through the command-line interface (CLI). Select **Migration** → **Non-disruptive System Migration** while creating the relationship (see Figure 8-25).

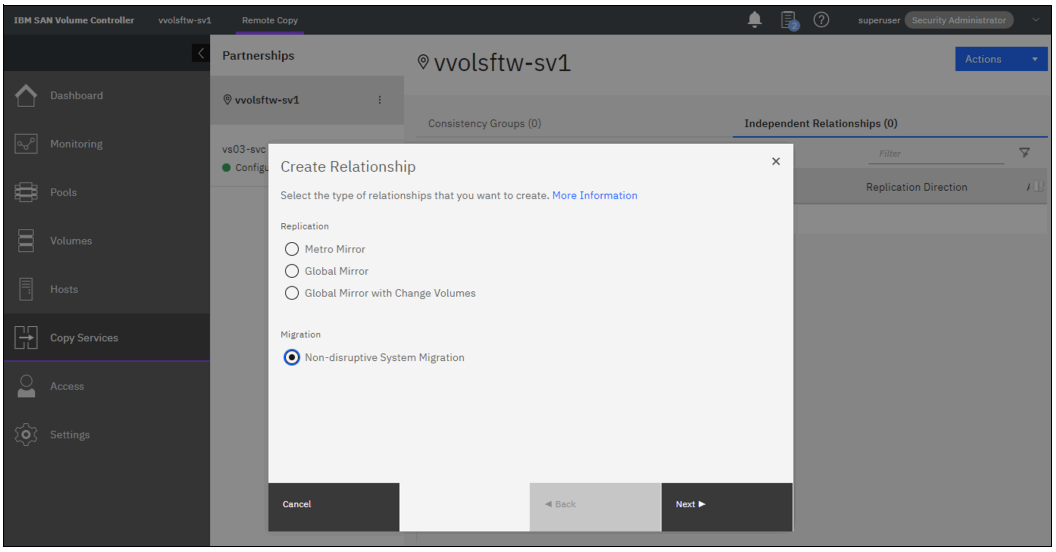


Figure 8-25 Create non-disruptive system migration relationship

3. Select the target system where the auxiliary volume is installed, as shown in Figure 8-26.

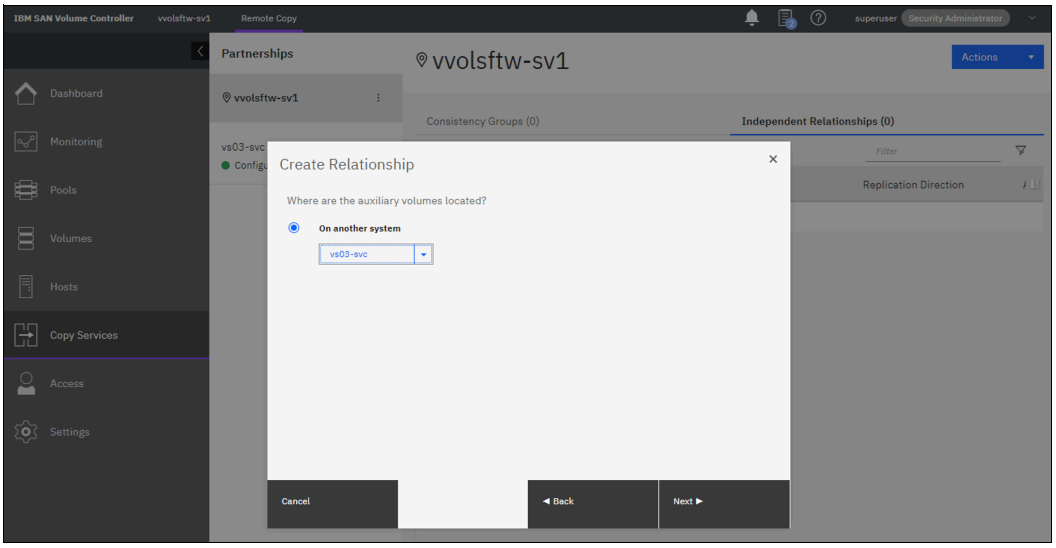


Figure 8-26 Choose target system

4. Select the master and auxiliary volume in scope for the migration, as shown in Figure 8-27.

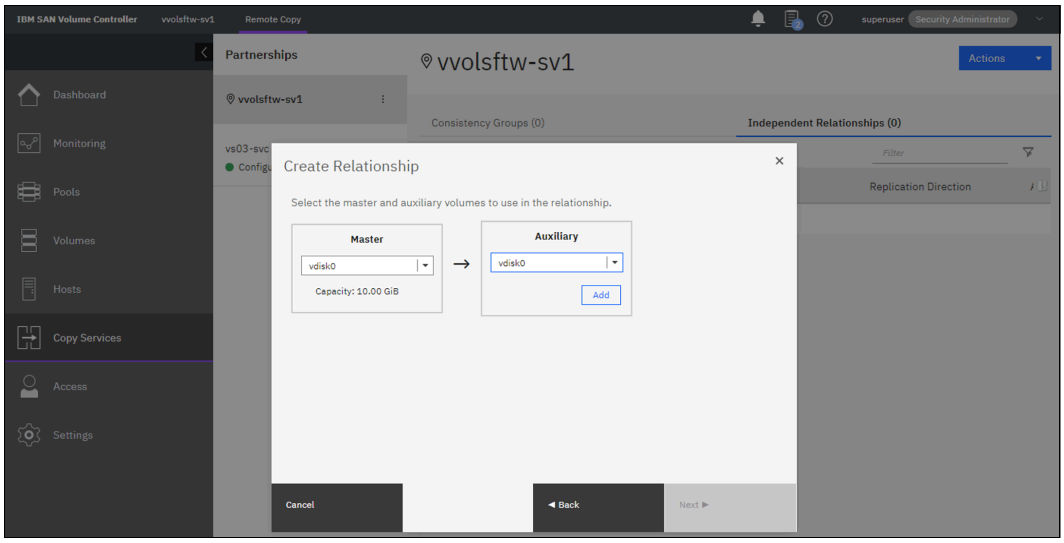


Figure 8-27 Choose master and auxiliary volumes

5. Follow the same process to add any other volumes in the scope. Select **Yes, start copying now** and then, click **Finish** to create the remote copy relationship (see Figure 8-28).

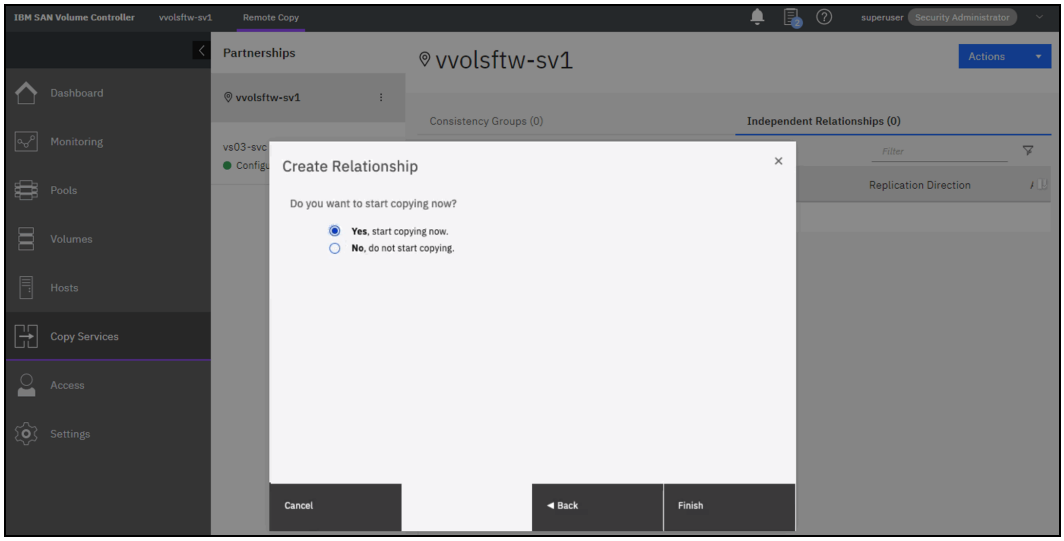


Figure 8-28 Complete remote copy relationship

Note: You can select the **No, do not start copying** option (see Figure 8-28 on page 552) to create the relationship, which results in volume relationship in an `Inconsistent Stopped` state.

With the current restrictions, consistency groups are not supported; therefore, if multiple volumes are used, each of those volumes must be selected individually and started manually.

To enable the copying process, select each volume individually, right click, and then, select **Start.** If multiple volumes are selected, the **Start** option 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 in the `Consistent Synchronized` state.

6. After the remote copy relationship is created, as shown in Figure 8-29 on page 553, the state of the volumes is `Inconsistent Copying`. The UUID of the source volume is migrated over to the target from the source system.

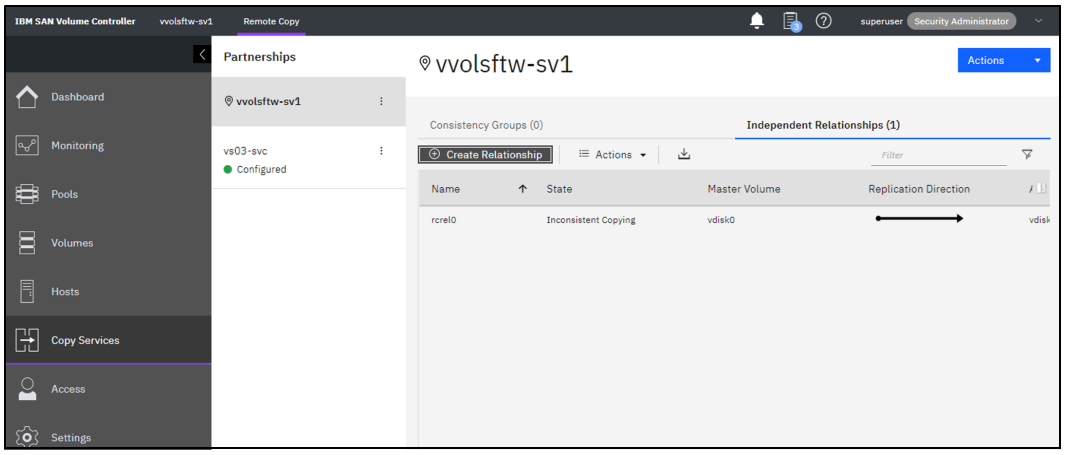


Figure 8-29 Volumes in Inconsistent Copying state

After the remote copy relationship completes the synchronization, the migrated volume is in the `Consistent Synchronized` state, as shown in Figure 8-30. Therefore, it is fully replicated and at this point is similar to a traditional Metro Mirror relationship, whereby I/O from the source is replicated to the target before notification of completion to the host.

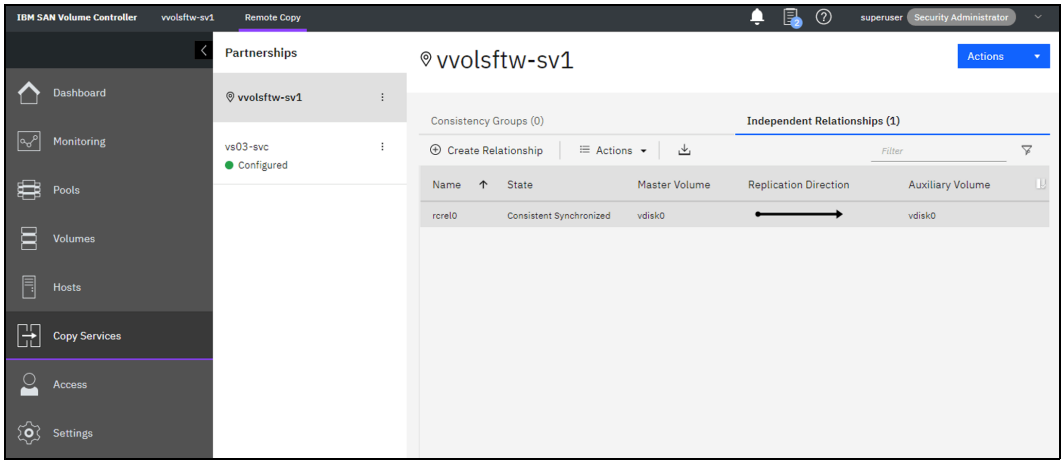


Figure 8-30 Volumes in Consistent Synchronized state

7. Create VDisk host maps or volume host cluster maps on the target system. Ensure that the host that is connected to the source system also can 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 used and presented are in standby state. Therefore, this non-disruptive volume migration is the new SCSI feature available in IBM Spectrum Virtualize 8.4.2 that is implemented here. Now, the host is still performing active I/O through the primary system and no I/O is sent through the secondary system. 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 added 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.

8. Run the **`switchrelationship`** command through the CLI. As a result, the direction of the relationship is changed so that the target system becomes the master and the source system becomes the auxiliary in this relationship. This step also can be done through the GUI (see Figure 8-31). Select the individual volume and right click to choose the **Switch** option.

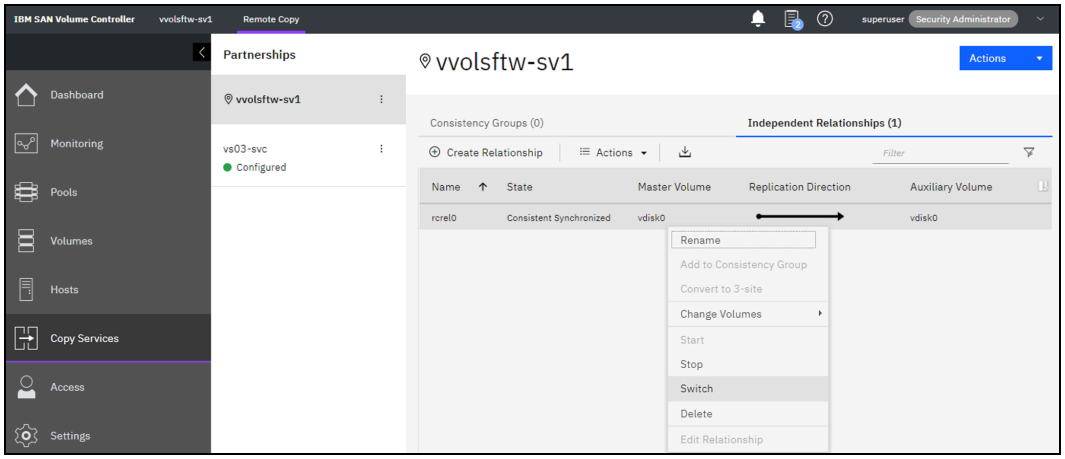


Figure 8-31 Switch master and auxiliary relationship between systems

A warning message is displayed (see Figure 8-32).

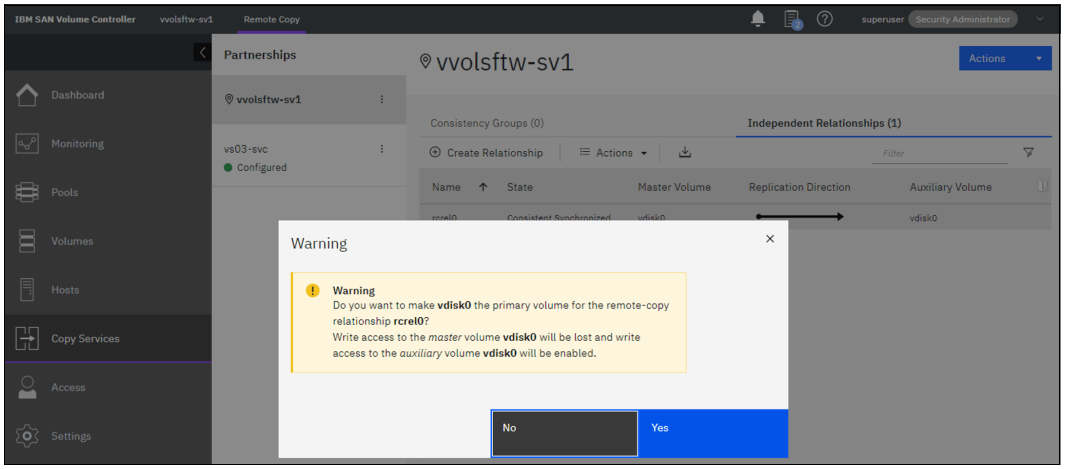


Figure 8-32 Approval state to switch relationship

9. Click **Yes**. The successful completion of the role reversal process is confirmed, as shown in Figure 8-33.

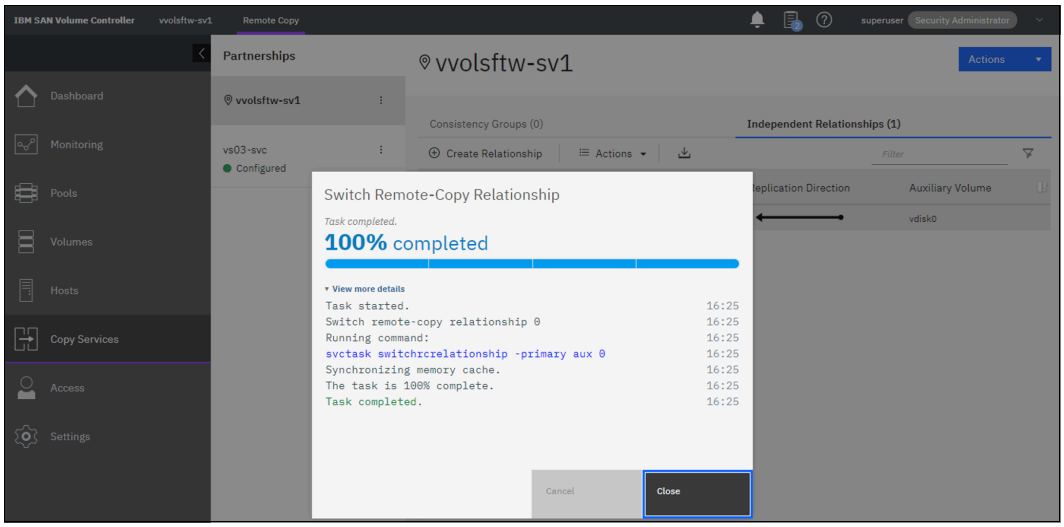


Figure 8-33 Successful completion of switch relationship for one volume

The standby (or ghost) paths that were presented from the target system are now converted to active paths, and the previously active paths 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 8-34.

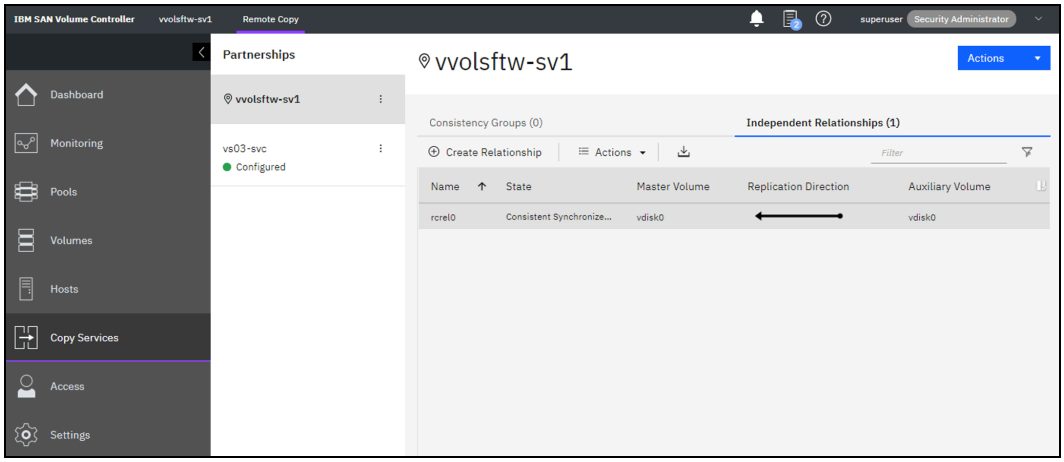


Figure 8-34 Switch relationship completion for all volumes

Note: Validate that the host is using the volumes from the target system for I/O. It is recommended to reboot the host to ensure that it comes up clean and no issues exist before the volume from the source system is unmapped and deleted, as described next.

10. Now, the host is performing I/O to the target system (which is the new system) and the standby paths are connected to the originating source system. Unmap the host from the volume on the source system.

11. Delete the volume that was migrated from the source system, as shown in Figure 8-35 and Figure 8-36.

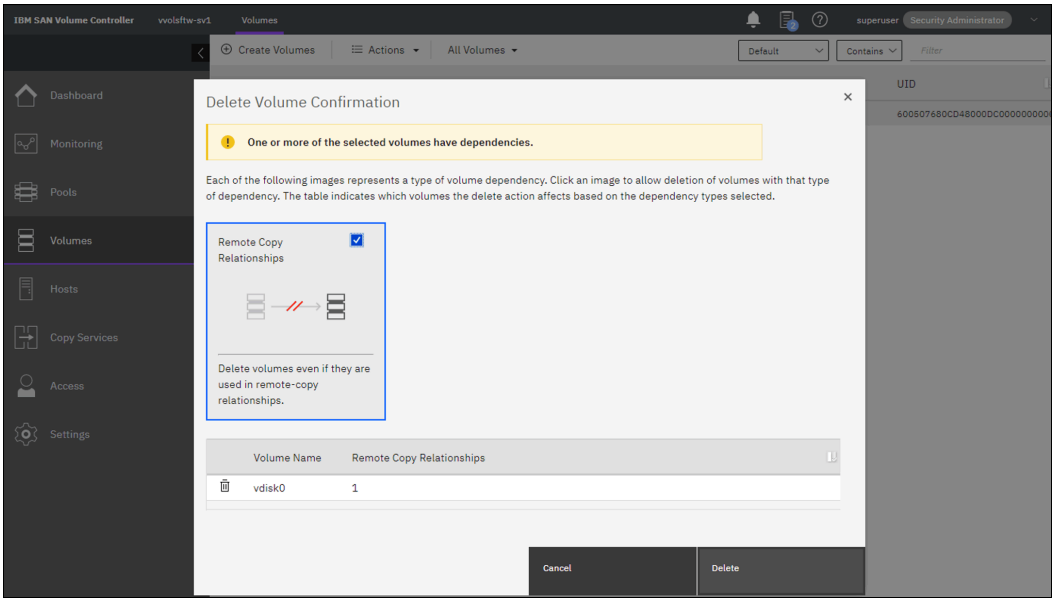


Figure 8-35 Delete Volume from source system

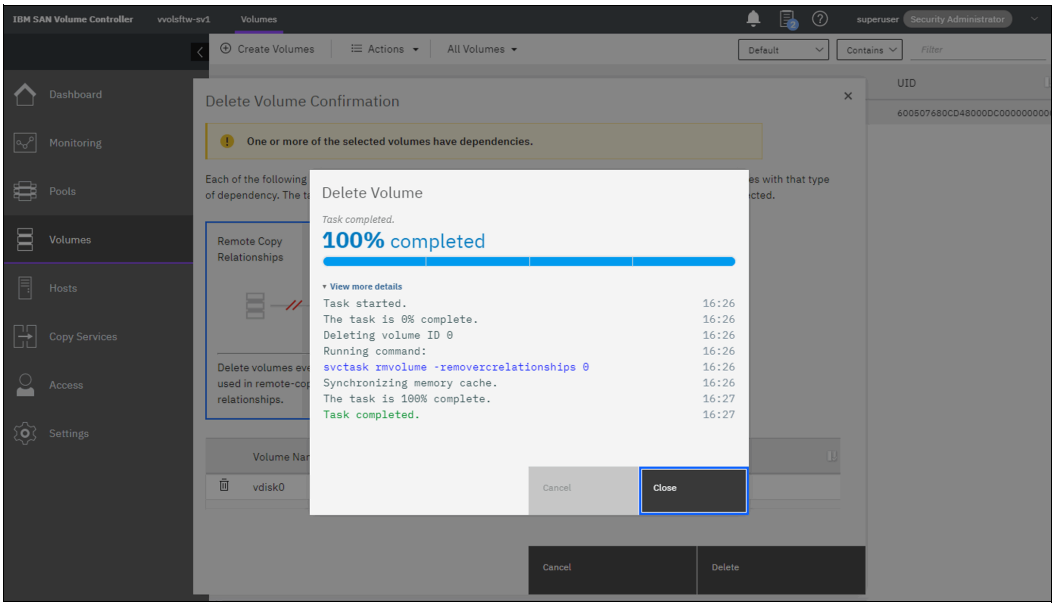


Figure 8-36 Volume deleted from source system

12. We now have a volume that was migrated to the target system and retains the same UUID 600507680CD48000DC00000000000000 from the source system as shown in Figure 8-37. This entire process is completely transparent to the host that is accessing the volume.

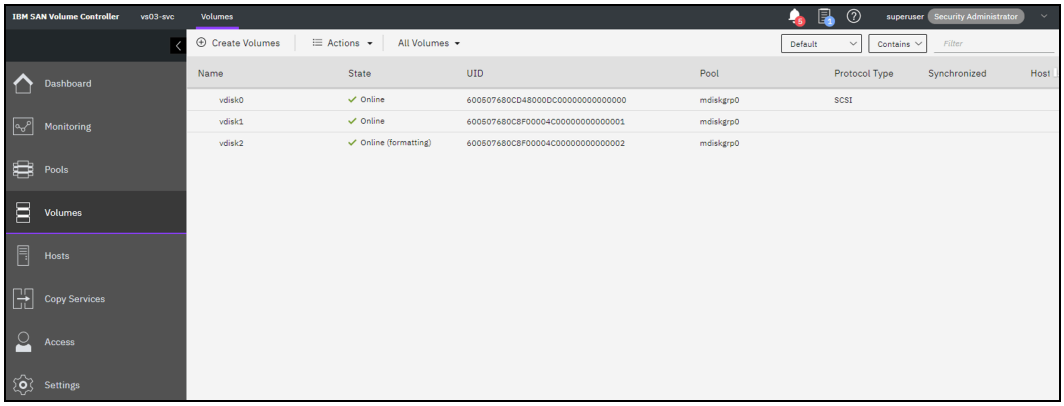


Figure 8-37 Volume migration completed with source UUID on target system

13. Remove the partnership that was defined between the two systems in step 1 on page 550. This removal can be done after all the volumes in scope are migrated.

9

Advanced features for storage efficiency

IBM Spectrum Virtualize running inside the IBM SAN Volume Controller offers several functions for storage optimization and efficiency.

This chapter introduces the basic concepts of those functions. It also provides a short technical overview and implementation recommendations.

For more information about planning and configuration of storage efficiency features, see the following publications:

- *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521
- *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430
- *IBM Real-time Compression in IBM SAN Volume Controller and IBM Storwize V7000*, REDP-4859
- *Implementing IBM Real-time Compression in SAN Volume Controller and IBM Storwize V7000*, TIPS1083
- *Implementing IBM Easy Tier with IBM Real-time Compression*, TIPS1072

This chapter includes the following topics:

- 9.1, “IBM Easy Tier” on page 560
- 9.2, “Thin-provisioned volumes” on page 579
- 9.3, “Unmap” on page 581
- 9.4, “Data Reduction Pools” on page 584
- 9.5, “Saving estimation for compression and deduplication” on page 596
- 9.6, “Overprovisioning and data reduction on external storage” on page 599

© Copyright IBM Corp. 2021. All rights reserved.

9.1 IBM Easy Tier

IBM Spectrum Virtualize includes the IBM System Storage Easy Tier function. It enables automated subvolume data placement throughout different storage tiers and automatically moves extents within the same storage tier to intelligently align the system with current workload requirements. This works with all available storage tiers and drive modules: Storage Class Memory, Flash drives, and spinning disk drives.

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.

Easy Tier acts to identify this skew and to 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 significantly reduced. By servicing most of the application workload from the fastest storage, Easy Tier acts to accelerate application performance and increase overall server utilization, which results in reduced costs in 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. Therefore, depending on the workload pattern, a large portion of the capacity can be provided by a lower and less expensive tier without affecting application performance.

Note: Easy Tier is a licensed function, but it is included in base code. No actions are required to activate the Easy Tier license on IBM SAN Volume Controller.

9.1.1 Easy Tier concepts

Easy Tier is a performance optimization function that automatically migrates (or moves) extents that belong to a volume between different storage tiers, based on their I/O load. Movement of the extents is online and unnoticed from the host perspective.

As a result of extent movement, the volume no longer has all of its data in one tier; rather, the data is in two or more tiers, and each tier provides optimal performance for the extent, as shown in Figure 9-1 on page 561.

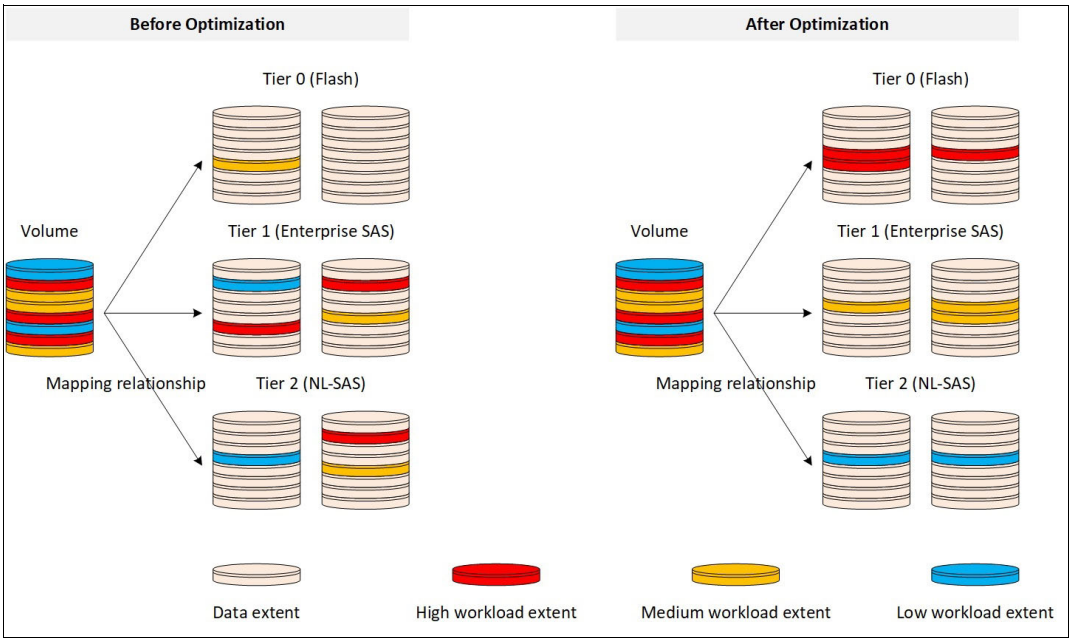


Figure 9-1 Easy Tier

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 managed disk (MDisk) back to a lower tier MDisk.

If a pool contains only MDisks of a single tier, Easy Tier only operates in balancing mode. Extents are moved between MDisks in the same tier to balance 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 IBM SAN Volume Controller might have different performance attributes because of their technology type, such as Flash or spinning drives and other characteristics.

The system divides available storage into the following tiers:

- Storage-class memory (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 all 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 lower performance and write endurance characteristics.

- Enterprise tier

  Enterprise tier exists when the pool contains MDisks on enterprise-class hard disk drives (HDDs), which are disk drives that are optimized for performance.

- Nearline (NL) tier

  NL tier exists 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, or modules. External MDisks need manual tier assignment when they are added to a storage pool.

Note: The tier of MDisks that are mapped from specific types of IBM System Storage Enterprise Flash is fixed to tier0_flash and cannot be changed.

Although IBM SAN Volume Controller 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.

Table 9-1 Storage Tier to Easy Tier mapping

Configuration	ET Top Tier	ET Middle Tier	ET 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)*	
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)*

Configuration	ET Top Tier	ET Middle Tier	ET Bottom Tier
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

The table represents all of the possible pool configurations. Some entries in the table contain optional tiers (in *italic*s), the configurations without the optional tiers are also valid.

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 the measure mode, which means no extent migrations are performed.

For more information about planning and configuration considerations or best practices, see *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

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 located and whether extent relocation is required. Once per day, Easy Tier analyzes the statistics to work out which data should be sent to a higher performing tier or might be sent to a lower tier.

Four times per day, it analyzes the statistics to identify if any data needs to be rebalanced between MDisks in the same tier. Once every 5 minutes, Easy Tier checks the statistics to identify if any of the MDisks is overloaded.

Based on this information, Easy Tier generates a migration plan that must be run for optimal data placement. The system then spends as long as needed running the migration plan. The migration rate is limited to ensure that host I/O performance is not affected while data is relocated.

The migration plan can consist of the following data movement actions on volume extents, as shown in Figure 9-2. Although each action is shown only once, all movement actions can be performed between any pair of adjacent tiers:

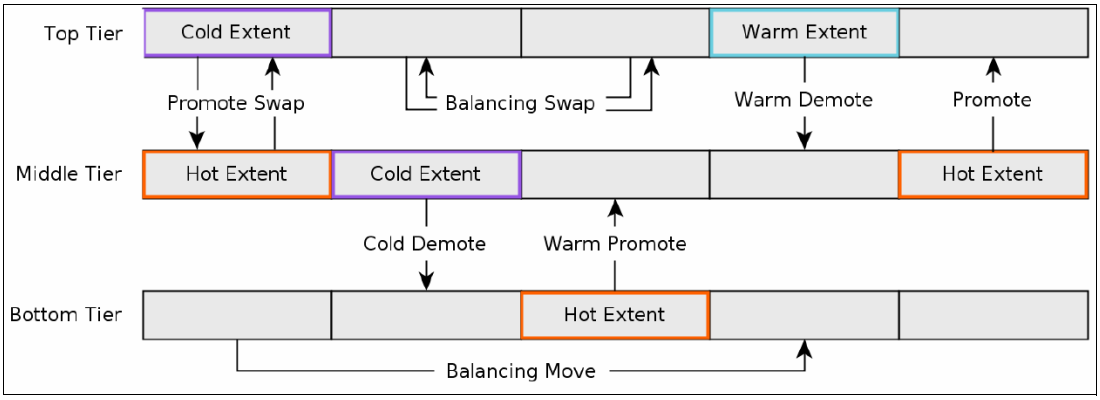


Figure 9-2 Actions on extents

- 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. This action 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 up space on the higher tier. In that way, 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. This process 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 actions as follows:

- Promote and rebalance get equal priority
- Demote is guaranteed 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 protection, 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, it is important to follow the best practice guidelines when configuring external storage, particularly the ratio between physical disks and MDisks presented to the system.

Each MDisk has an Easy Tier load parameter (low, medium, high, 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 only collects usage statistics for each extent in a storage pool (if it is enabled on the volume and 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 there is more than one pool in a tier). Also, auto-balance between MDisks in each tier is performed.

The default operation mode is `Enabled`. Therefore, the system balances storage pools. If the required licenses are installed, it also optimizes performance.

Note: The auto-balance process automatically balances data when MDisks are added into a pool. However, it does not migrate extents from 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, not 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 SAN Volume Controller:

- Volumes that are added to storage pools use extents from the “middle” tier of three-tier model, if available. Easy Tier then collects usage statistics to determine which extents to move to “faster” or “slower” tiers. If there are no free extents 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 extent I/O activity of each copy of a mirrored volume. Easy Tier works with each copy independently of the other copy in volume mirroring, 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.

- For compressed volumes in standard pools, only reads are analyzed by Easy Tier.
- 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 storage pools and volumes are created. It also starts extent migration when the necessary I/O statistics are collected.

A few parameters can be adjusted. Also, Easy Tier can be disabled on selected volumes in storage pools.

MDisk settings

The tier for internal (array) MDisks is detected automatically and depends on the type of drives that 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 then, click **Assign**.

Note: The tier of MDisks mapped from certain 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.

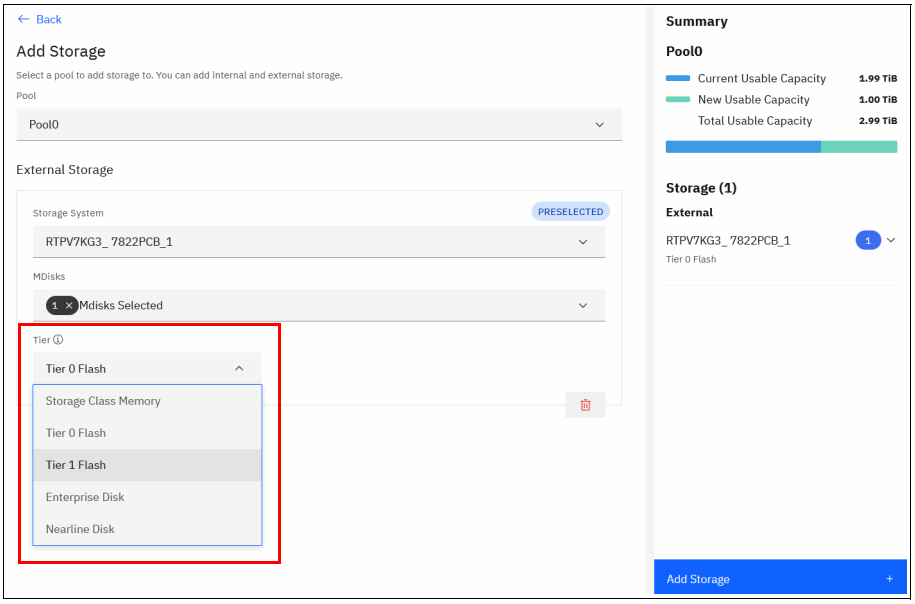


Figure 9-3 Choosing tier when assigning MDisks

To change the storage tier for an MDisk that is assigned, in **Pools** → **External Storage**, right-click one or more selected MDisks and choose **Modify Tier**, as shown in Figure 9-4.

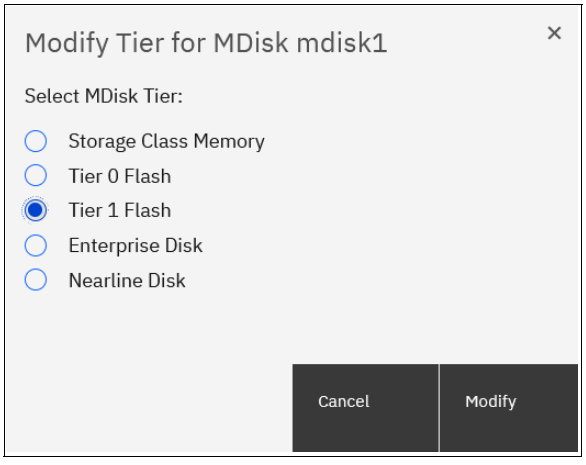


Figure 9-4 Changing MDisk tier

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, on **Pools** → **External Storage**, select **Actions** → **Customize columns** and select **Tier**. This selection includes the current tier setting into a list of MDisk parameters that are shown in the External Storage pane. You can also find this information in MDisk properties. To view it, right-click the MDisk, select **Properties**, and then, expand the **View more details** section, as shown in Figure 9-5.

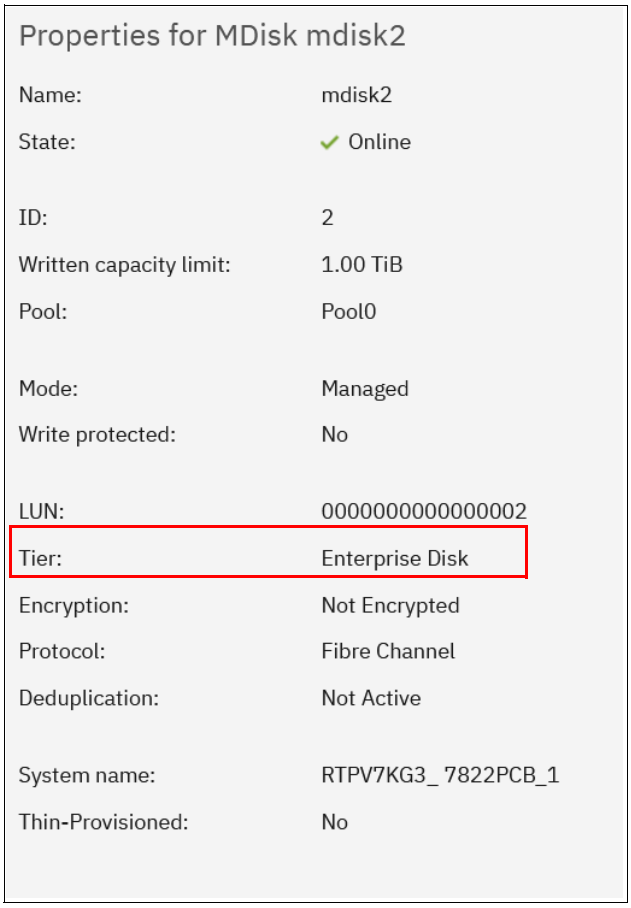


Figure 9-5 MDisk properties

To list MDisk parameters with the CLI, run the **lsmdisk** command. The current tier for each MDisk is shown. To change the external MDisk tier, run **chmdisk** command with the **-tier** parameter, as shown in Example 9-1.

Example 9-1 Listing and changing tiers for MDisks (partially shown)

```
IBM_2145:ITSO-SV1: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_2145:ITSO-SV1:superuser>chmdisk -tier tier1_flash mdisk2
IBM_2145:ITSO-SV1:superuser>
```

For an external MDisk, the system cannot calculate its exact performance capabilities; therefore, it has several predefined levels. In rare cases, statistics analysis might show that Easy Tier is overusing or under-utilizing an MDisk. If so, levels can be adjusted only by using the CLI. Run the **chmdisk** command with **-easytierload** parameter. To reset Easy Tier load to system-default for the chosen MDisk, use **-easytier default**, as shown in Example 9-2.

Note: Adjust Easy Tier load settings only if instructed to do so by IBM Technical Support or your solution architect.

Example 9-2 Changing Easy Tier load

```
IBM_2145:ITSO-SV1:superuser>chmdisk -easytierload default mdisk2
IBM_2145:ITSO-SV1:superuser>
IBM_2145:ITSO-SV1:superuser>lsmdisk mdisk2 | grep tier
tier tier_enterprise
easy_tier_load high
IBM_2145:ITSO-SV1:superuser>
```

To list the current Easy Tier load setting of an MDisk, run **lsmdisk** with MDisk name or ID as a parameter.

Storage pool settings

When a storage pool (standard pool or DRP) is created, Easy Tier is enabled 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 any moment later. This process is not possible with the GUI, only with the system CLI.

To check the current Easy Tier function state on a pool, select **Pools** → **Pools**, right-click the selected pool, choose **Properties**, and then, expand the **View more details** section, as shown in Figure 9-6. This window also displays the amount of data that is stored on each tier.

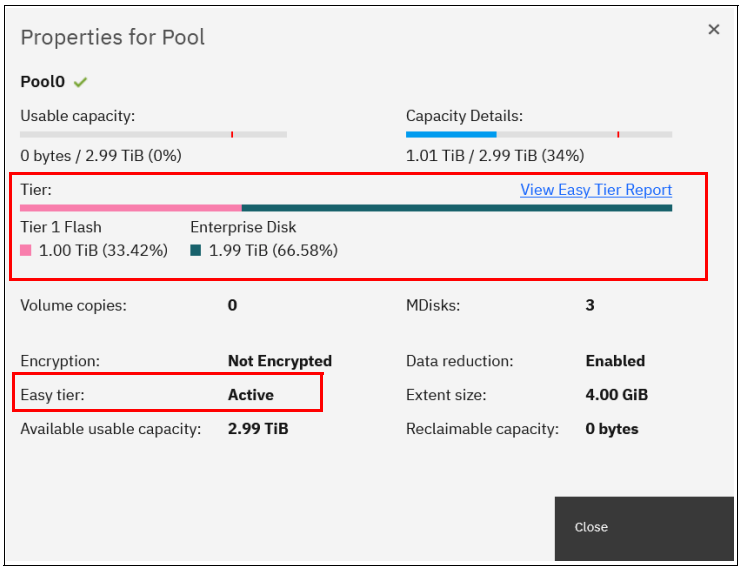


Figure 9-6 Pool properties

Easy Tier can be in one of the following statuses:

- 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 also is enabled. This state is expected 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 expected 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 status of the Easy Tier function on the pools with the CLI, run the **lsmdiskgrp** command without any parameters. To switch Easy Tier off or back on, run the **chmdiskgrp** command, as shown in Example 9-3. By running **lsmdiskgrp** with pool name and ID as a parameter, you also can determine how much storage of each tier is available within the pool.

Example 9-3 Listing and changing Easy Tier status on pools

```
IBM_2145:ITSO-SV1:superuser>lsmdiskgrp
id name       status mdisk_count ... easy_tier easy_tier_status
0  TieredPool online 1           ... auto      balanced
IBM_2145:ITSO-SV1:superuser>chmdiskgrp -easytier measure TieredPool
IBM_2145:ITSO-SV1:superuser>chmdiskgrp -easytier auto TieredPool
IBM_2145:ITSO-SV1:superuser>
```

Volume settings

By default, each striped-type volume allows 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 switched off. It is possible to switch off Easy Tier only at a pool level.

This process can be done by using the CLI only. Run the **lsvdisk** command to check and **chvdisk** command to modify Easy Tier function status on a volume copy, as shown in Example 9-4.

Example 9-4 Checking and modifying Easy Tier settings on a volume

```
IBM_2145:ITSO-SV1:superuser>lsvdisk  vdisk0 |grep easy_tier
easy_tier on
easy_tier_status balanced
IBM_2145:ITSO-SV1:superuser>chvdisk -easytier off vdisk0
IBM_2145:ITSO-SV1:superuser>
```

System-wide settings

A system-wide setting is available that is called *Easy Tier acceleration* that is disabled by default. Enabling this setting 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, whereas 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 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 recommended to keep Easy Tier acceleration mode off during normal system operation to avoid performance effects 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 state, as shown in Example 9-5.

Example 9-5 The chsystem command

```
IBM_2145:ITSO-SV1:superuser>lssystem |grep easy_tier
easy_tier_acceleration off
IBM_2145:ITSO-SV1:superuser>chsystem -easytieracceleration on
IBM_2145:ITSO-SV1: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 they 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 (STAT) application.

Monitoring Easy Tier by using the GUI

Select **Monitoring** → **Easy Tier Reports** to view the most recent Easy Tier statistics. Select the storage pool you want to see reports for in the filter section on the left, as shown in Figure 9-7.

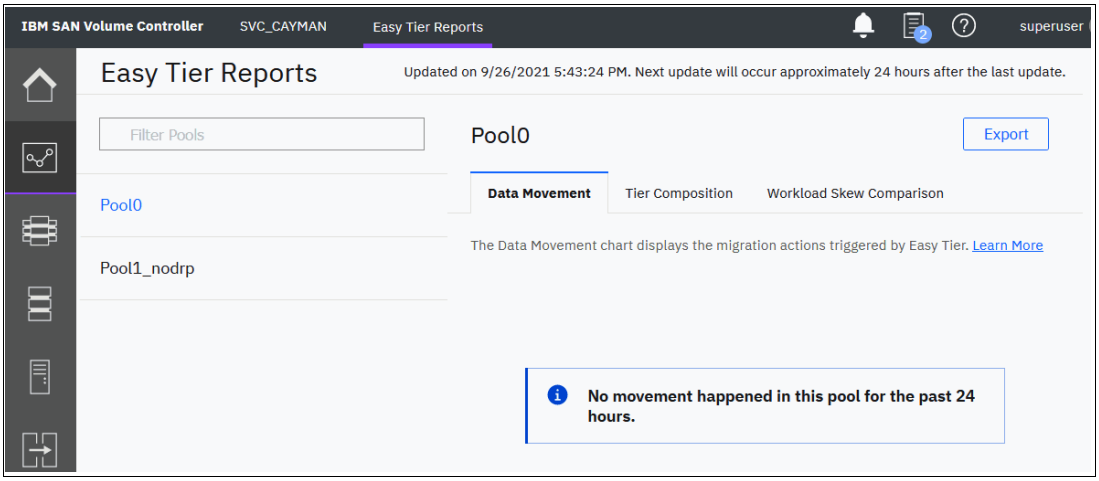


Figure 9-7 Easy Tier reports not available

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 that is shown in Figure 9-7 is displayed. In this case, wait until new reports were generated and then, revisit the GUI.

Types of reports

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, use the **Export** or **Export All** buttons to download the reports in comma-separated value (CSV) format.

Data Movement report

The Data Movement report shows the extent migrations Easy Tier performed to relocate data between different tiers of storage and within the same tier for optimal performance, as shown in Figure 9-8. 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.

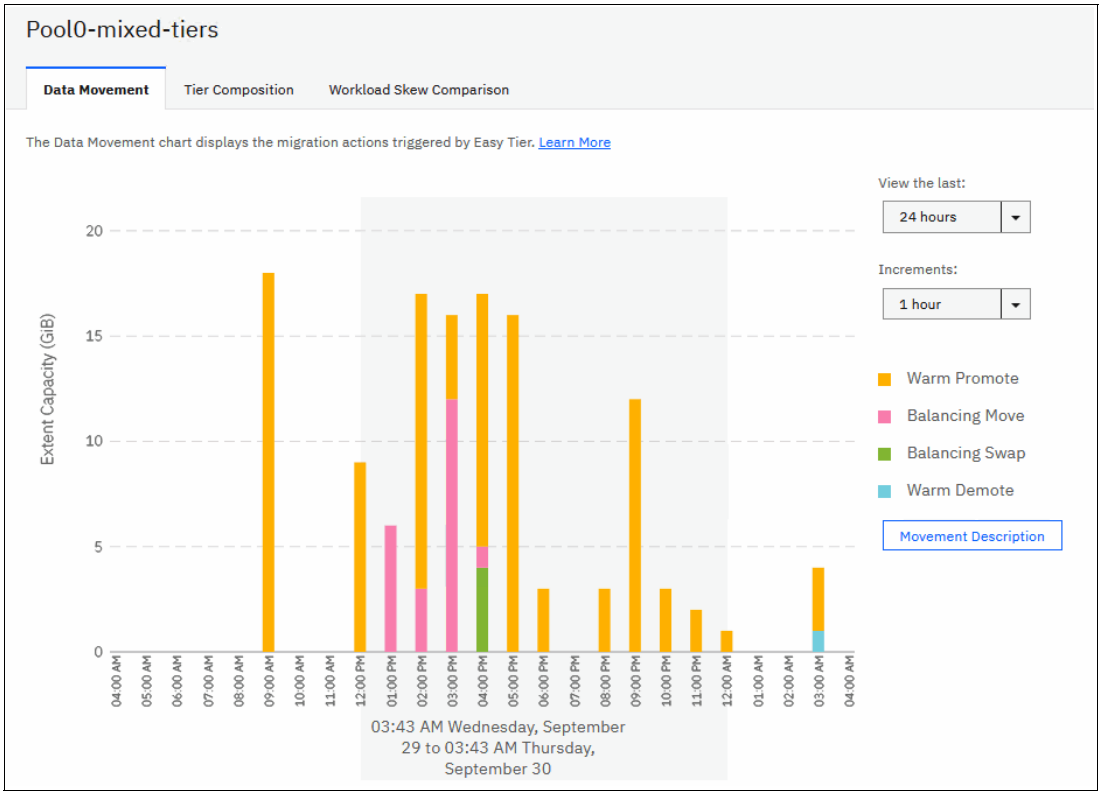


Figure 9-8 Easy Tier Data Movement report

The X-axis shows a time line for the selected period by using the selected increments. The Y-axis indicates the amount of extent capacity that was moved. For each time increment, a color-coded bar displays the amount of data 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 563, or click the **Movement Description** button next to the chart to see an explanation in the GUI as shown in Figure 9-9 on page 574.

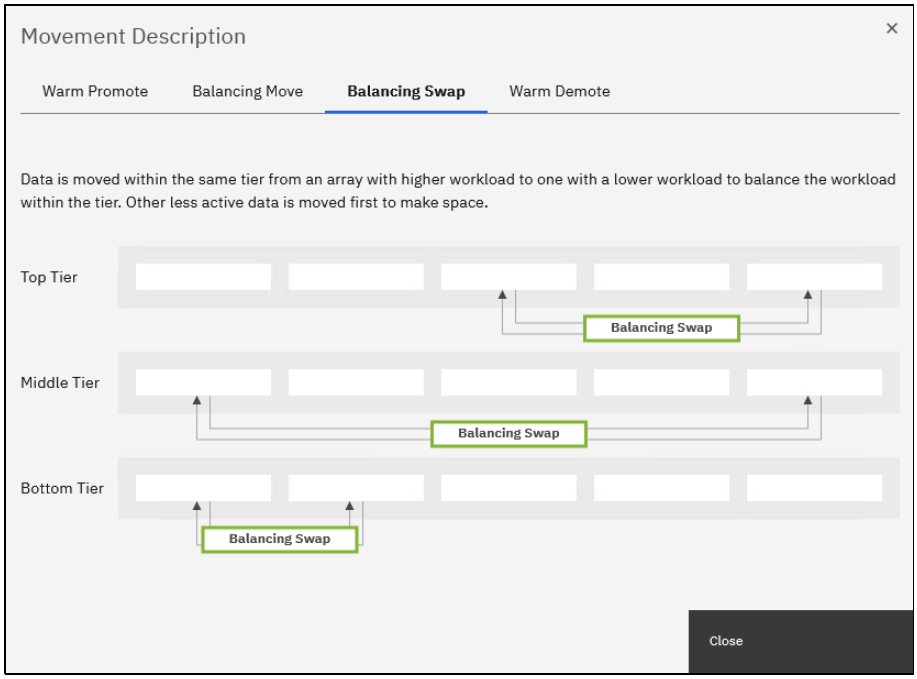


Figure 9-9 Easytier movement description

Tier Composition chart

The Tier Composition chart (see Figure 9-10) shows how different types of workloads are distributed between top, middle, and bottom tiers of storage in the selected pool.

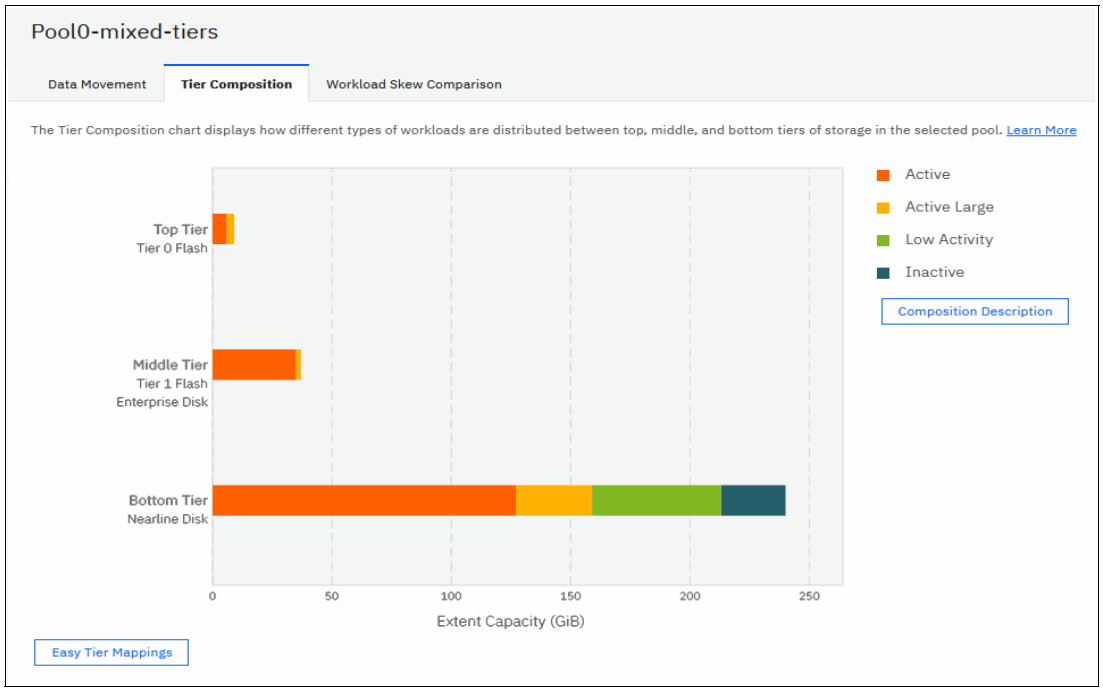


Figure 9-10 Tier Composition chart

A color-coded bar for each tier shows which workload types are present in that tier and to how much of the extent capacity in that tier they can be attributed to. Easy Tier distinguishes between the following workload types. Click the **Composition Description** button to show a short explanation for each workload type as shown in Figure 9-11.

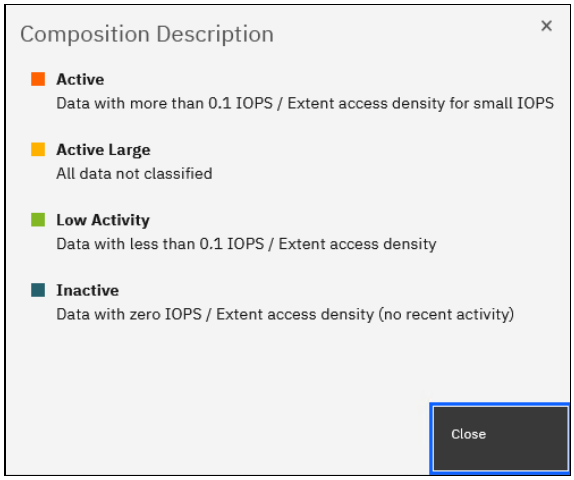


Figure 9-11 Composition description

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 9-12.

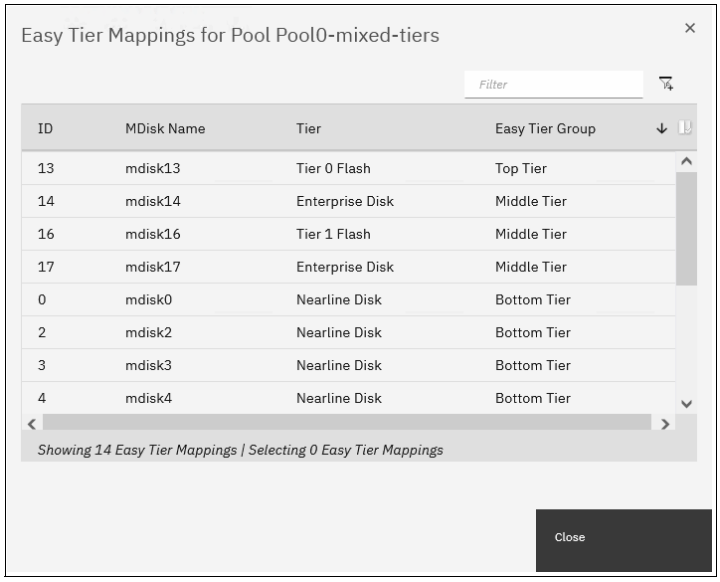


Figure 9-12 Easy Tier Mappings

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 562.

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 9-13.

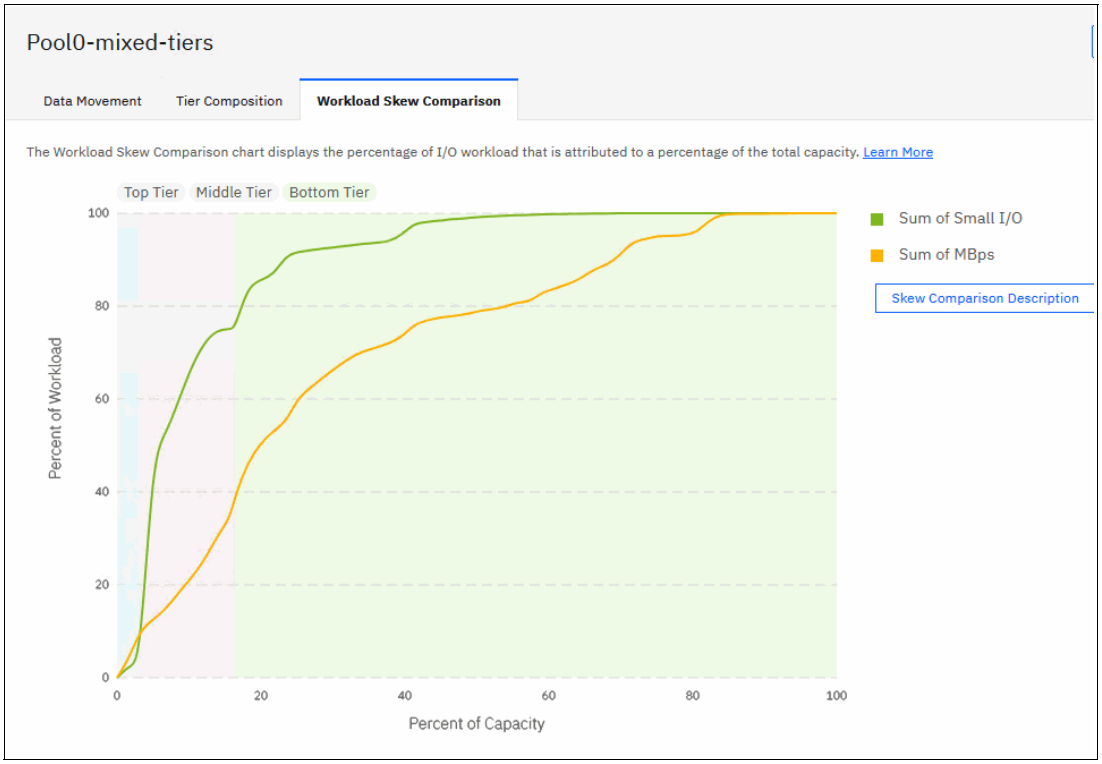


Figure 9-13 Easy Tier Workload Skew Comparison

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 MBps (sum of small and large bandwidth). The portion of capacity and workload attributed to a specific tier is color-coded in the chart with a legend above the chart.

Figure 9-13 on page 576 shows that the top tier (Tier1 Flash) contributes only a tiny percentage of capacity to the pool, but is handling approximately 85% of the IOPS and more than 40% of the bandwidth in that pool. The middle tier (enterprise disk) handles almost of the remaining IOPS and another 20% of the bandwidth. The bottom tier (NL disk) provides the most 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 9-14.

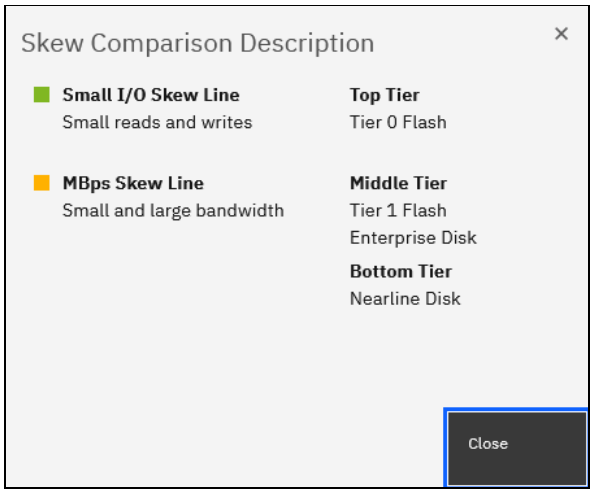


Figure 9-14 Skew Comparison Description

Monitoring Easy Tier by using the IBM Storage Tier Advisor Tool

The IBM Storage Tier Advisor Tool (STAT) is a Windows console application that can analyze heat data files that are generated by Easy Tier. It produces a graphical display of the amount of “hot” data per volume and predictions of the performance benefits of adding capacity to a tier in a storage pool.

By 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 this IBM Support web page.

You can download the 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 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 seven 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 STAT in Windows command prompt

```
C:\Program Files (x86)\IBM\STAT>stat dpa_heat.78DXRY0.191021.075420.data
```

The 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 9-15.

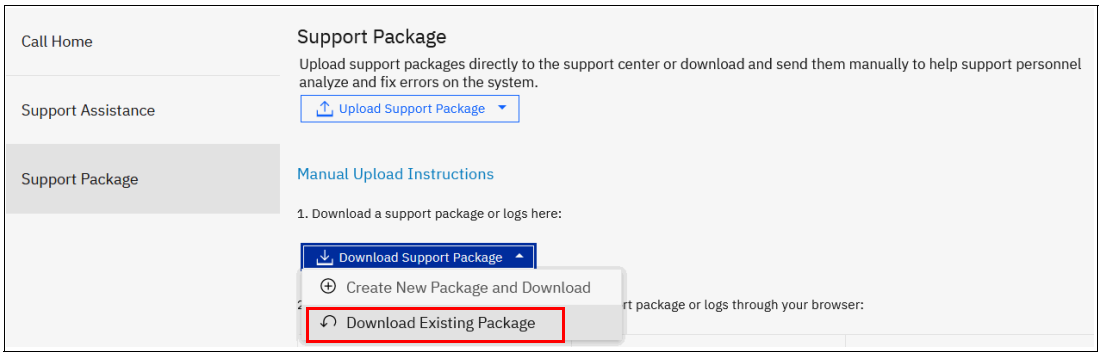


Figure 9-15 Downloading support package

The download window opens and shows all files in the `/dumps` directory and its subfolders on a current configuration node. You can filter the list by using the `easytier` keyword, select the `dpa_heat` file or files that are to be analyzed, and click **Download**, as shown in Figure 9-16. Save the files in a convenient location (for example, to a subfolder that holds the IBM STAT executable file).

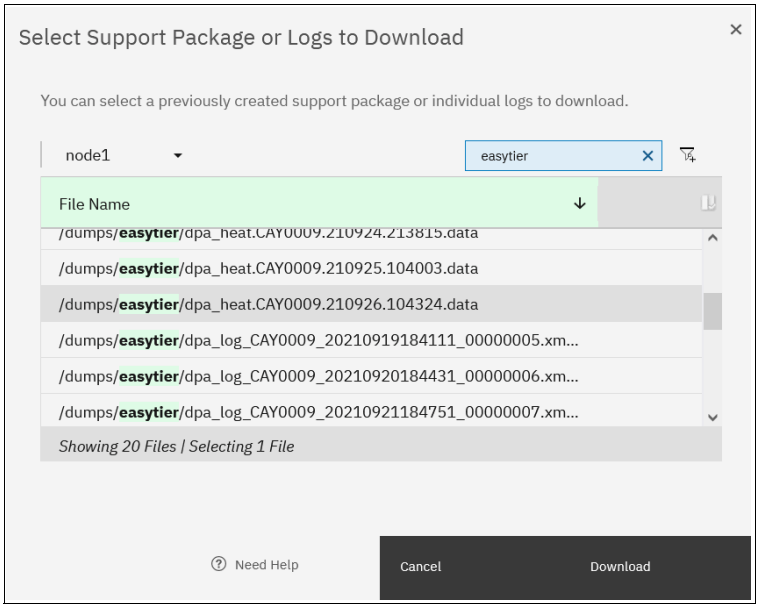


Figure 9-16 Downloading Easy Tier heat data file: dpa_heat files

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 analyzing CSV files, see *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

9.2 Thin-provisioned volumes

In a shared storage environment, thin provisioning is a method for optimizing the use of available storage. It relies on the allocation of capacity on demand as opposed to the traditional method of allocating all of the capacity at the time of initial provisioning. The use of this principle means that storage environments can achieve significantly 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). This storage allocation 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

IBM SAN Volume Controller 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. This capacity 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 typically 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 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, 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 called *grains*. Write I/O to grains that were not written to cause 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 affect performance. When you use thin-provisioning with 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 they are 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, “Creating and provisioning volumes” on page 331.

Metadata

In a standard pool, the system uses the 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 less than 0.1% of the provisioned capacity.

Therefore, if the host uses 100% of the provisioned capacity, some extra space is required on your storage pool to store thin provisioning metadata. In a 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 is not reflected in the volume 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 9-17.

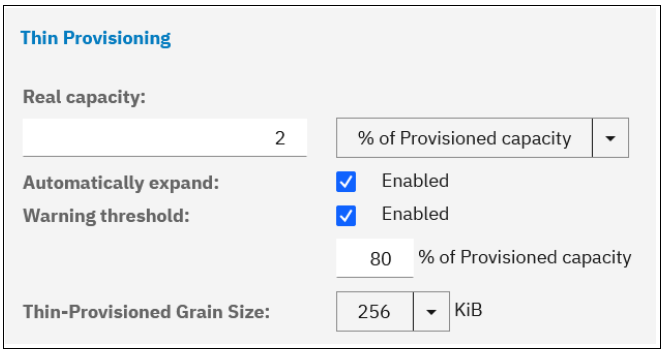


Figure 9-17 Volume parameters for thin-provisioning

Real capacity defines initial volume real capacity and the amount of contingency capacity. When autoexpand is enabled, the system attempts to always maintain the contingency capacity by allocating more real capacity when hosts write to the volume.

The warning threshold can be used to be notified 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 Spectrum Virtualize systems running V8.1.0 and later support the SCSI **unmap** command. The use of 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 SCSI unmap command

Unmap is a set of Small Computer System Interface (SCSI) primitives that allow hosts to indicate to a storage system that space allocated to a range of blocks on a storage volume is no longer required. This command allows 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 host operating systems to unprovision storage on the storage system, which means that the resources can automatically be freed up in the storage pools and used for other purposes.

One of the most common use cases is a host operating system, such as VMware ESXi, freeing storage within a file system. The storage system can then 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 being 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 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 backend controllers:

- IBM FlashSystem A9000
- IBM Storwize V5000 family, V5100, 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. This means **unmap** can improve Flash performance by erasing blocks in advance.

This feature is turned on by default. It is recommended to keep back-end **unmap** enabled, especially if a system is virtualizing an overprovisioned storage controller.

To verify that sending **unmap** commands to back-end is enabled, run the CLI command **lssystem**, as shown in Example 9-7.

Example 9-7 Verifying back-end unmap support status

```
IBM_2145:ITSO-SV1:superuser>lssystem | grep backend_unmap
backend_unmap on
```

9.3.3 Host SCSI unmap

The IBM Spectrum 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, by running SCSI **unmap** commands to the entire 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 that are on a heavily loaded NL serial-attached SCSI [SAS] array), this issue 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 advertise support the SCSI **unmap** command.

If you use DRPs or an overprovisioned back-end that supports **unmap**, it is recommended to turn on SCSI **unmap** support. Host **unmap** support is disabled by default.

If only standard pools are configured and back-end is traditional (fully provisioned), consider keeping host **unmap** support switched off because it does not provide any benefit.

To check and modify current setting for host SCSI **unmap** support, run the **lssystem** and **chsystem** CLI commands, as shown in Example 9-8.

Example 9-8 Turning host unmap support on

```
IBM_2145:ITSO-SV1:superuser>lssystem | grep host_unmap
host_unmap off
IBM_2145:ITSO-SV1:superuser>chsystem -hostunmap on
IBM_2145:ITSO-SV1:superuser>
```

Note: You can switch host **unmap** support on and off non-disruptively on the system side. However, hosts must rediscover storage, or (in the worst case) be restarted for them to stop sending **unmap** commands.

9.3.4 Offload IO 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 up hosts and speed up the copy process by offloading the operations of specific 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. These workloads can take bandwidth from production volumes and affect 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. This delay can improve performance, but limits the rate at which host features, such as VMware VMotion, can copy data. It can also increase the time it takes to format file systems on a host.

Note: For systems that are managing any NL storage, it might be recommended to set the offload throttle to 100 MBps.

To implement offload throttle, you can run the **mkthrottle** command with **-type offload** parameter. In the GUI, select **Monitoring** → **System Hardware** and click **System Actions** → **Edit System Offload Throttle**, as shown in Figure 9-18.

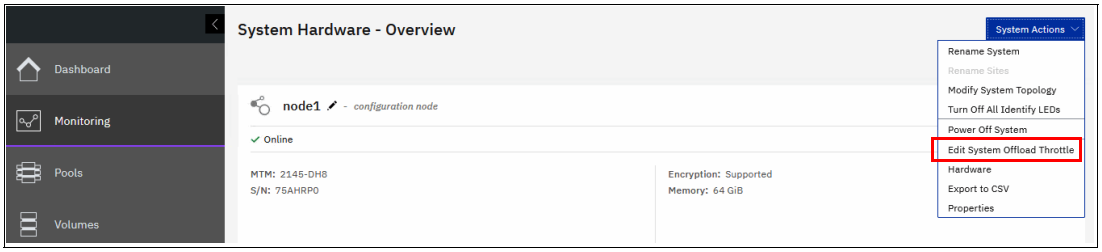


Figure 9-18 Setting offload throttle

9.4 Data Reduction Pools

Data Reduction pools (DRPs) 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 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 lack of a separate license requirement does not apply to real-time compression, where a specific capacity based license is needed.

Note: This book provides only an overview of DRP aspects. 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 being thin-provisioned, duplicates of data blocks are detected and are replaced with references to the first copy.

- Thin and 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 and 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 that is 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 allows a tree-like directory to be used 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 allows 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 address 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 backend storage.

This directory often is too large to store in memory; therefore, 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 have to read some directory metadata in addition to the data. A write I/O request might have to read directory metadata write updated directory metadata, journal metadata, and the data.

Conversely, data reduction by design reduces the size of data by using compression and deduplication, so less data is written to the backend storage.

Starting with Spectrum Virtualize V8.4, child pools for DRPs were introduced.

A *child pool* is a folder-like object within a parent DRP and 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 allows to segment administrative domains. Parent pools and associated child pools share the managed disk, 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 kind of volume migration is available to move volumes within a single DRP and its affiliated child pools. This feature allows moving 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 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 of an All-Flash system to be competitive with HDD.

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 so by maintaining a fingerprint database that contains hashes of data blocks that are written to the pool. If new data that is written by hosts matches an entry in this database, 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 same Lempel-Ziv (LZ) based IBM Real-time Compression (RtC) and decompression algorithm that is used by Random Access Compression Engine (RACE) in standard pools. However, DRP offers a new implementation of data compression that is fully integrated into the IBM Spectrum Virtualize I/O stack. It makes optimal use of node resources, such as memory and CPU cores, and uses hardware acceleration on supported platforms more efficiently. DRP compression also operates on smaller block sizes, which results in more consistent and predictable performance.

Deduplication and compression can be combined, in which case data is first deduplicated and then compressed. Therefore, deduplication references are created on the compressed data stored on the physical domain.

DRPs support end-to-end SCSI **unmap** functions. Hosts use the set of SCSI **unmap** commands to indicate 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 reduction techniques were 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 again, designed to minimize latency because Flash is well-suited for small block workload with high IOPS. All write operations are coalesced into 256 KB sequential writes to simplify the garbage collection on Flash devices and to gain full stride writes from RAID arrays.

DRP is designed to work well alongside Easy Tier. The directory metadata of DRPs does not fit in memory and is stored on disk that uses dedicated metadata volumes, which are separate from the 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 are large but frequently rewritten data is grouped to consolidate “heat”. Easy Tier can work as usual to accurately identify active data.

RAID Reconstruct Read (3R) is a technology to increase reliability and availability of data stored in DRPs and is introduces in Spectrum Virtualize V8.4.

Any read is evaluated and in case of a mismatch the data is reconstructed while using the parity information. To eliminate re-reading corrupted data, the affiliate cache block is marked invalid. This marking works for internal and external backend devices.

9.4.3 Planning for DRPs

Before configuring and using DRPs in production environments, it is important to plan for capacity and performance. Because DRPs have different performance characteristics than standard pools, sizing models cannot be used directly without modifications.

For more information about how to estimate capacity savings that are achieved by compression and deduplication, see 9.5, “Saving estimation for compression and deduplication” on page 596.

The following software and hardware requirements must be met for DRP compression and deduplication:

- The system runs V8.1.3.2 or higher.
- Avoid Global Mirror with Change Volumes to or from a deduplicated volume.

For more information about recommended firmware versions, see this IBM Support web page.

In most cases, it is recommended 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. An exception are systems without support for hardware compression.

In systems with compressing backend storage controllers, certain system configurations make determining accurate physical capacity usage difficult. If the system contains compressing backend storage controllers and DRPs with thin-provisioned volumes without compression, it is difficult to monitor capacity usage.

In this case, overcommitting and losing access to write operations is possible. To prevent this issue from occurring, use compressed volumes (with or without deduplication) or fully allocated volumes. Separate compressed volumes and fully allocated volumes by using separate pools.

For more information, see Chapter 9.6, “Overprovisioning and data reduction on external storage” on page 599.

A maximum of four DRPs can be 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. The maximum size of a customer data volume is limited to 128,000 extents per I/O group. This size 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 causes 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 and without affecting I/O performance. 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.

To ensure garbage collection is working properly, a DRP has a minimum capacity limit, depending on extent size and number of IO groups, as listed in Table 9-3. Even when there are no volumes 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. This issue 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, consider the use of a smaller extent size for a smaller metadata overhead and a lower minimum capacity limit.

For more information about the use of data reduction on the system and the backend storage, see 9.6, “Overprovisioning and data reduction on external storage” on page 599.

9.4.4 Implementing DRP with compression and deduplication

To use all data reduction technologies on the system, you must create a DRP and volumes within the DRP, and then, map these volumes to hosts that support SCSI **unmap** commands. Also, consider enabling support for host SCSI **unmap** commands, as described in 9.3.3, “Host SCSI unmap” on page 582.

The implementation process for DRP is similar to standard pools, but has its own specifics.

Creating pools and volumes

To create a DRP, enable **Data Reduction** in the Create Pool window, which is available by selecting **Pools** → **Pools**. For more information about how to create a storage pool and populate it with MDisks, see Chapter 5, “Using storage pools” on page 271.

To create a volume within a DRP, select **Volumes** → **Volumes**, and then, click **Create Volumes**.

Figure 9-19 shows the Create Volumes window. 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.

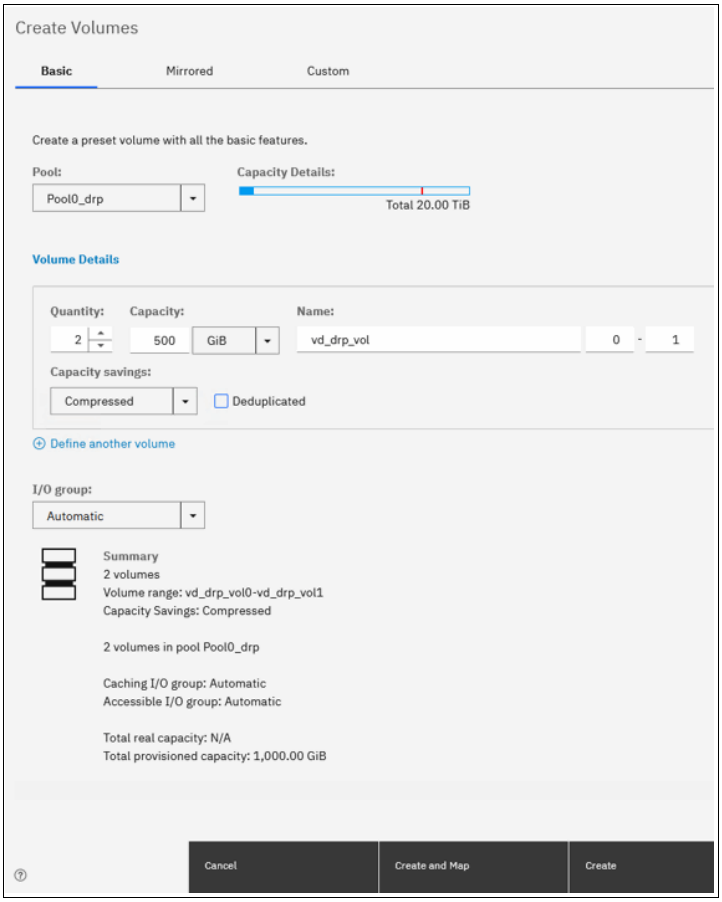


Figure 9-19 Create compressed volume

Capacity monitoring

Capacity monitoring in DRPs is done mostly on the system level and on storage pool level. 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 the 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. Select **Pools** → **Pools** to monitor this capacity, as shown in Figure 9-20.

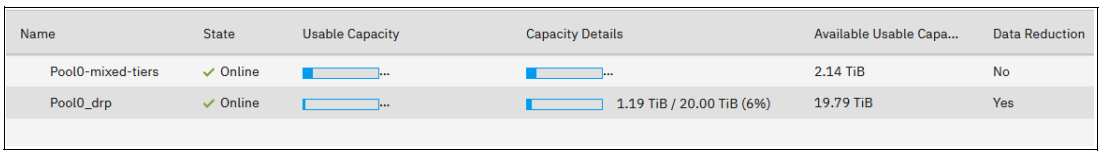


Figure 9-20 DRP capacity overview

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 achieved by thin-provisioning, compression, deduplication, and the total data reduction savings in the pool, as shown in Figure 9-21.

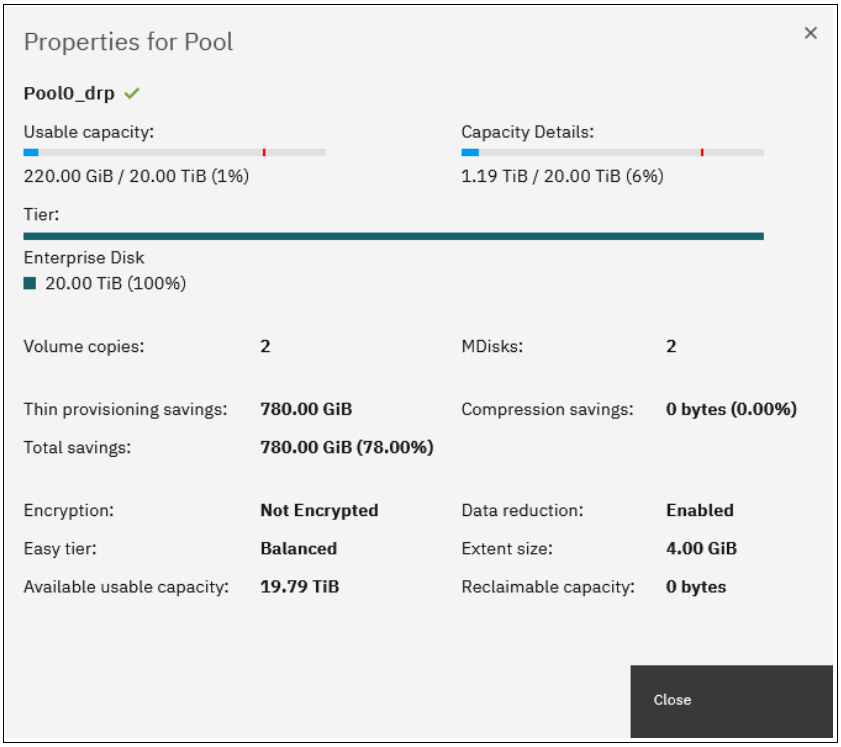


Figure 9-21 Capacity reporting in a DRP

The Reclaimable capacity also is shown. This unused capacity 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 is decreasing. 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 a detailed per-volume capacity reporting, as shown in the Volumes by Pool pane in Figure 9-22. Only the Capacity (which is the provisioned capacity that is available to hosts) is shown. Real capacity, Used capacity, and Compression savings are not applicable for volumes with capacity savings. Only fully allocated volumes display those parameters.

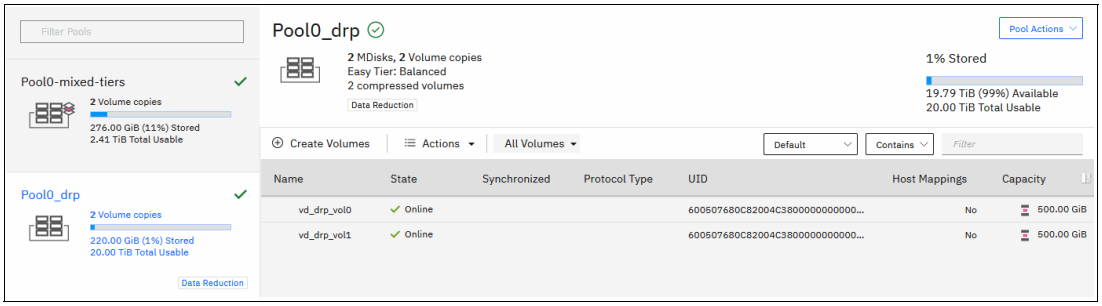


Figure 9-22 Volumes in DRP pool

Per-volume compression savings are not visible directly, but they can be accurately estimated by using the IBM Comprestimator. For more information, see 9.5.1, “Evaluate compression savings by using IBM Comprestimator” on page 596. This tool can be used for 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` 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 that are not in a DRP.

To find this value, run the **`lsvdisk`** command with 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 provisioned to the host. A 53 GB file was written from the host.

Example 9-9 DRP volume capacity reporting on the CLI

```
IBM_2145:ITSO-SV1: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, 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 CLI command **lsmdiskgrp** with the pool ID or name as a parameter, as shown in Example 9-10.

Example 9-10 DRP capacity reporting on the CLI

```
IBM_2145:ITSO-SV1: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
```

Note: Compression-related properties are not valid for DRPs. They are used exclusively for RACE compression in standard pools.

For more information about every reported value, see IBM Documentation.

Migrating to and from 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 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.

Also, Real-time compressed volumes cannot coexist with data-reduced DRP volumes in a single I/O group. Therefore, migrating such volumes needs some extra considerations. One possible solution might be to inflate real-time compressed volumes in standard pools and migrate these volumes in a second step to data-reduced volumes in a DRP pool.

Depending on the system configration and the type of migration, a one-step migration or a two-step migration is necessary. The reason is that compressed volumes in standard pools cannot coexist with deduplicated volumes in DRPs.

Therefore, a two-step migration is required in the following scenarios:

- Migrating to a deduplicated volume in a DRP if compressed volumes in standard pools exist in the same I/O group.
- Migrating to a compressed volume in a standard pool if deduplicated volumes exist in the same I/O group.

Note: The two-step migration requires all volumes that cannot coexist with the wanted volume type to be migrated in one step.

The migration processes includes the following steps:

1. One-step migration process

   To create a second copy, right-click the source volume and choose **Add Volume Copy**, as shown in Figure 9-23 on page 594. Choose the target pool of the migration for the second copy and select the wanted capacity savings.

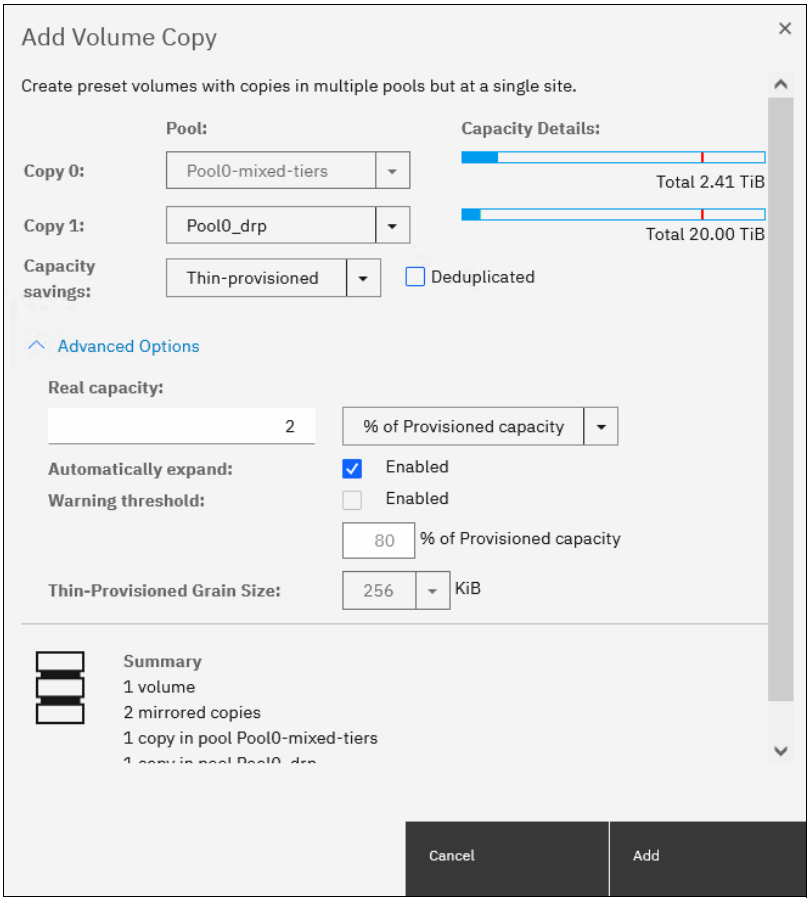


Figure 9-23 Add volume copy dialog

2. After you click **Add**, synchronization starts. The time synchronization takes to complete depends on the size of the volume, system performance, and the configured copy 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 pane. You can track the synchronization process by using the Running tasks pane, as shown in Figure 9-24. After it reaches 100% and the copies are synchronized, you can complete migration by deleting the source copy.

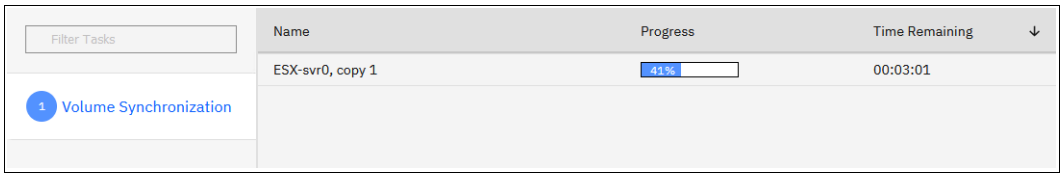


Figure 9-24 Synchronization progress

Note: You can increase the synchronization rate by right-clicking the volume and selecting **Modify Mirror Sync Rate**.

3. 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:

   a. 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 wanted 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.

   b. For all of the volumes that are handled in the previous step, add a second copy in the target pool, but select the wanted 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 also searches and moves deduplication references that other volumes might have to the deleting volume copies. This process is done to ensure that deduplicated data that is on deleted copies continues to be available for other volumes in the system.

After this process is complete, the volume copies are deleted and they 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. Therefore, used capacity of a removed volume is not available for reuse immediately after the removal.

The time it takes to delete a thin-provisioned or compressed volume copy depends on the size of the volume, system configuration, and workload. For deduplicated copies, the duration also depends on the amount and size of other deduplicated copies in the pool. Therefore, it might take a much time to delete a small deduplicated copy if many other deduplicated volumes are 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`** CLI 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 much time to complete. Used capacity is not immediately returned 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 new copy to the volume before the deletion finishes. If a new copy must be added without waiting for the deletion to complete, 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 estimation for compression and deduplication

This section describes the tools that are used for sizing the environment for compression and deduplication.

9.5.1 Evaluate 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 Spectrum 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 Storage Insights.

Comprestimator provides a quick and accurate estimation of compression and thin-provisioning benefits. The utility performs read-only operations and 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.

In the **Volumes** pane, right-click any volume, and select **Space Savings** → **Estimate Compression Savings**. This action shows you the results and the date of the latest estimation cycle, as shown in Figure 9-25 on page 597. 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.

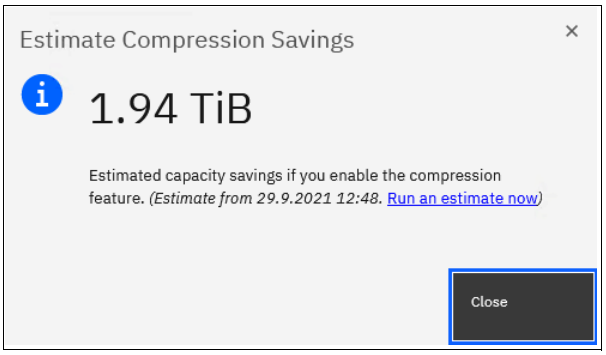


Figure 9-25 Estimate Compression Savings

To analyze all the volumes on the system from the CLI, run the **analyzevdiskbysystem** command.

The use of the command analyzes all the 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 it takes to analyze all the volumes on 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 analyze a single volume by specifying its name or ID as a parameter in 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 estimated time of completion.

The **lsvdiskanalysis** command is run to display information for thin provisioning and compression estimation analysis report for all volumes.

If you use a version of IBM Spectrum Virtualize that is older than V7.6 or if you want to estimate the compression savings of another IBM or non-IBM storage system, the separate IBM Comprestimator Utility can be installed on a host that is connected to the device that must be analyzed. For more information and the latest version of this utility, see this IBM Support IBM Comprestimator Utility Version 1.5.3.1.

Consider the following recommended best practices for using Comprestimator:

- Run the Comprestimator utility before implementing an IBM Spectrum Virtualize solution and DRPs.
- Download the latest version of the Comprestimator utility if you are not using one that is included in your IBM Spectrum Virtualize solution.
- Use Comprestimator to analyze volumes that contain as much active data as possible rather than volumes that are nearly empty or newly created. This use ensures more accuracy when sizing your environment for compression and DRPs.

Note: 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 deduplication and compression. The tool operates by scanning target workloads on any established array (from IBM or 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 written to a disk.

The DRET utility uses advanced mathematical and statistical algorithms to perform an analysis with low memory footprint. The utility runs on a host that can access 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, DRET in many cases 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, 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. This method 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 this IBM Support web page.

9.6 Overprovisioning and data reduction on external storage

Starting from IBM Spectrum Virtualize V8.1.x, overprovisioning on selected back-end controllers is supported. Therefore, if back-end storage performs data deduplication or data compression on LUs provisioned from it, they still can be used as external MDisks on the system. However, other configuration and monitoring considerations must be taken into account.

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 same as for fully allocated storage controllers. The system detects if the MDisk is overprovisioned, its total physical capacity, used, and remaining physical capacity. It detects if 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 A9000 V12.1.0 and above
- FlashSystem 900 V1.4 and above
- FlashSystem V9000 AE2 and AE3 expansions
- IBM Storwize or FlashSystem family system with code V8.1.0 and above
- Pure Storage
- HPE Nimble

Extra caution is required when planning and monitoring capacity for such configurations. See Table 9-4 for an overview of configuration guidelines when overprovisioned external storage controllers are used.

Table 9-4 Using data reduction at two levels

System	Backend	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. The system can report physical use, but does not manage to avoid out-of-space. No visibility of each application's use at the system layer. If the backend runs out-of-space, there is limited ability to recover. Consider creating a sacrificial emergency space volume.
DRP with compression	Overprovisioned	*Recommended with appropriate precautions*. Assume 1:1 compression in backend storage and do not overcommit capacity in the backend. 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 and it tends to fill the top tier with hottest data. Changes in compressibility of data in the top tier can overcommit the storage leading to out-of-space.

System	Backend	Comments
DRP with thin-provisioned or fully allocated volumes	Overprovisioned	*Avoid*. Very difficult to understand physical capacity use of the uncompressed volumes. High risk of overcommitting the backend. If a mix of DRP and fully allocated volumes is required, use separate pools.
DRP	DRP	*Avoid*. Creates two levels of IO amplification and capacity overhead. DRP at the bottom layer provides no benefit.

When DRPs are used with a compressing backend controller, use compression in DRP and avoid overcommitting the backend by assuming a 1:1 compression ratio in backend storage. Small, extra savings are realized from compressing metadata.

If the backend controller uses FCM devices that are always compressing with hardware acceleration, the same methodology should be used. Eventual capacity saving should be used by creating other managed disks 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 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, all of 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 of these MDisks.

Provisioning groups are used differently depending on the back-end storage, as shown in the following examples:

- FlashSystem A9000 and FlashSystem 900: The entire subsystem forms one provisioning group.
- Storwize and FlashSystem family systems: The storage pool forms a provisioning group, which allows 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 primarily be monitored on the overprovisioned backend storage controller itself.

From the system, capacity usage can be monitored on over-provisioned MDisks by using one of the following methods:

- GUI Dashboard, as shown in Figure 9-26.

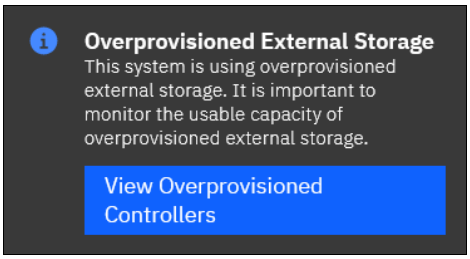


Figure 9-26 Dashboard button for overprovisioned storage monitoring

Click **View Overprovisioned Controllers** to show an overview of overprovisioned MDisks used by the system, as shown in Figure 9-27.

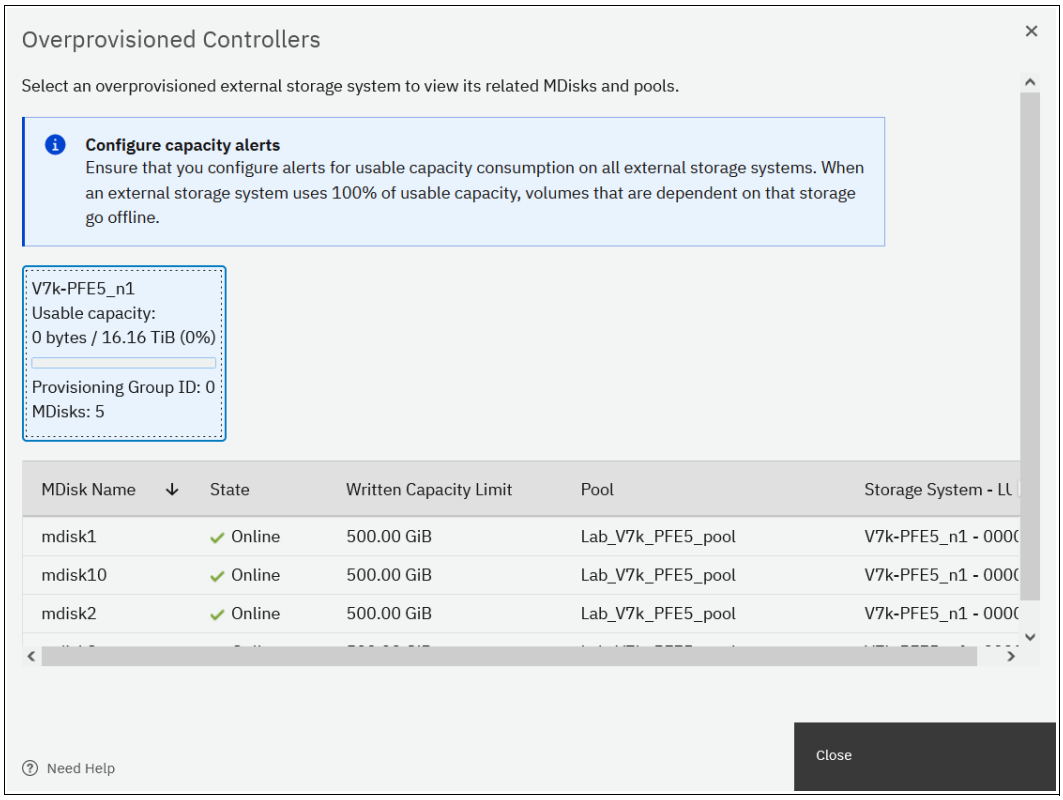


Figure 9-27 View overprovisioned controllers and mdisks

- MDisk properties window that opens by right-clicking an MDisk in **Pools** → **MDisks by Pools** and **Properties** menu option, as shown in Figure 9-28.

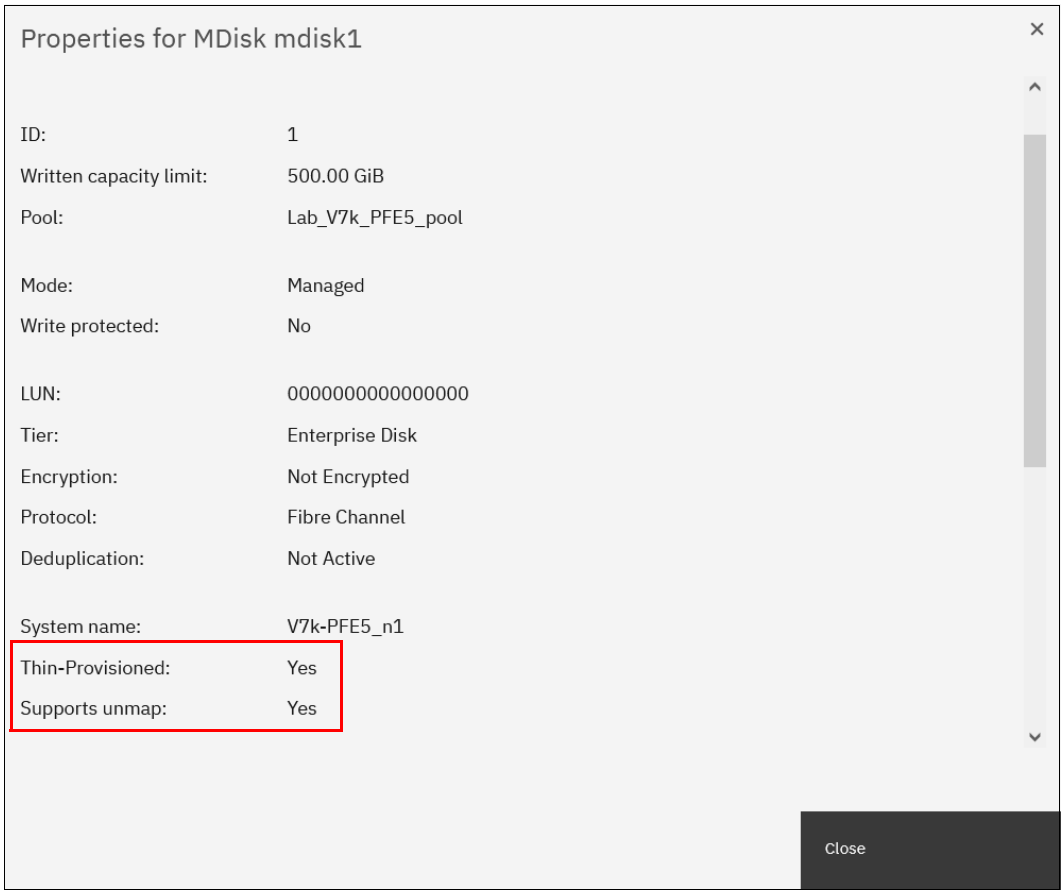


Figure 9-28 Thin-provisioned MDisk properties

- By using the CLI command **`lsmdisk`** with an MDisk name or ID as a parameter, you can display the full properties of a thin provisioned volume, as shown in Example 9-11.

Example 9-11 lsmdisk parameters for thin provisioned MDisks

```
IBM_2145:ITSO-SV1: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 recommended to create multiple storage pools from MDisks in a single provisioning group.

10

Advanced Copy Services

This chapter describes the Advanced Copy Services that are 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 Spectrum Virtualize software.

This chapter includes the following topics:

- 10.1, “Maximum configuration updates” on page 606
- 10.2, “IBM Safeguarded Copy” on page 607
- 10.4, “Managing FlashCopy by using the GUI” on page 640
- 10.5, “Transparent Cloud Tiering” on page 678
- 10.6, “Implementing Transparent Cloud Tiering” on page 681
- 10.7, “Volume mirroring and migration options” on page 691
- 10.8, “Remote Copy” on page 693
- 10.9, “Remote Copy commands” on page 722
- 10.10, “Native IP replication” on page 729
- 10.11, “Managing Remote Copy by using the GUI” on page 751
- 10.12, “Remote Copy memory allocation” on page 775
- 10.13, “Troubleshooting Remote Copy” on page 776

© Copyright IBM Corp. 2021. All rights reserved.

10.1 Maximum configuration updates

Spectrum Virtualize 8.4.2 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**
Number of VDisks	10,000	10,000	15,864[a]
Host-Mappable Volumes	10,000	10,000	15,864[a]
FlashCopy Mappings	5,000	10,000	15,864[a]
Number of Asynchronous Volumes	5,000	5,000	793[a]
Maximum Asynchronous Capacity	720TiB	720TiB	2 PiB[a]
Number of HA Volumes	1250	2000	2000
Maximum HA Capacity	720TiB	720TiB	2 PiB[b]
HA + 3-Site Volumes	1250	2000	2000

a. Only applies to storage systems that support 10,000 volumes
b. Only applies to IBM SAN Volume Controller SA2 and SV2 Node

10.2 IBM Safeguarded Copy

Traditional disaster recovery is designed around recovering from or preventing data loss. Data loss can occur as a result of catastrophic or accidental events. Strategies to prevent data loss include:

- Backup or archive of data to various media including disk and tape.
- Offsite transport of backup media (or directly backing up data off-site).
- Replication of data to another site. This approach ensures 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, in spite of any adverse cyber events. This concept brings information security, organizational resilience, and business continuity together.

Cyber events can include the following examples:

- Cyber attacks:
  - Phishing, web attacks, 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 Spectrum Virtualize 8.4.2 prevents point in time data from being deleted or changed. It keeps user error, malicious destruction, or ransomware attacks from modifying this data.

Safeguarded Copy provides the following benefits:

- Separation of duties: 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.
- Data reduced (thin) copies can limit growth requirements.
- Restore data instantly to any point in time before any attack.

For more information about supported storage systems and implementation, see *Implementation Guide for SpecV/FlashSystem Safeguarded Copy*, REDP-5654.

10.3 IBM FlashCopy

Through the IBM FlashCopy function of the IBM Spectrum 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. 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. This feature means that the copy can be immediately 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 purposes and data mining, and for quality assurance

Regardless of your business needs, FlashCopy within the IBM Spectrum Virtualize is flexible and offers a broad feature set, which makes it applicable to several 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, such as these examples, “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 often is 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 has 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 do that, you make the target available on a host. Is suggested to not make the target available to the source host because seeing duplicates of disks causes problems for most host operating systems. 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 10.3.12, “Reverse FlashCopy” on page 630.

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 then 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 that is using actual production data. This testing ensures the highest quality possible for your environment. FlashCopy makes this type of testing easy to accomplish without putting the production data at risk or requiring downtime to create a constant copy.

You can create a FlashCopy of your source 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.

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.

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.

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 are 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 later in this chapter.

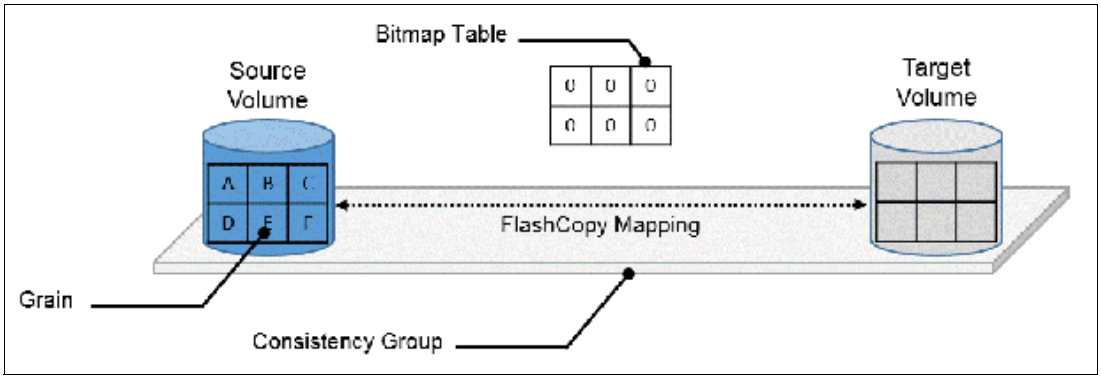


Figure 10-1 FlashCopy terminology

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.

The source and target volumes must be the same size. The minimum granularity that IBM Spectrum 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 Spectrum 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.

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 method, the IBM Spectrum 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 that spans 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 completed successfully. In fact, it did not complete successfully 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 actions that the user can perform on a FlashCopy mapping are the same for consistency groups.

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 have various levels and methods of caching data to provide better speed. Therefore, because the IBM Spectrum 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 is required following a host crash. FlashCopies that are crash consistent often can 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.

The target volumes are overwritten with a complete image of the source volumes. 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, it is sufficient in most cases to 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 as opposed to flushing data from the application and backing database, which is always the suggested method because it is safer. However, this 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 Spectrum 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 IBM Documentation.

10.3.6 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.

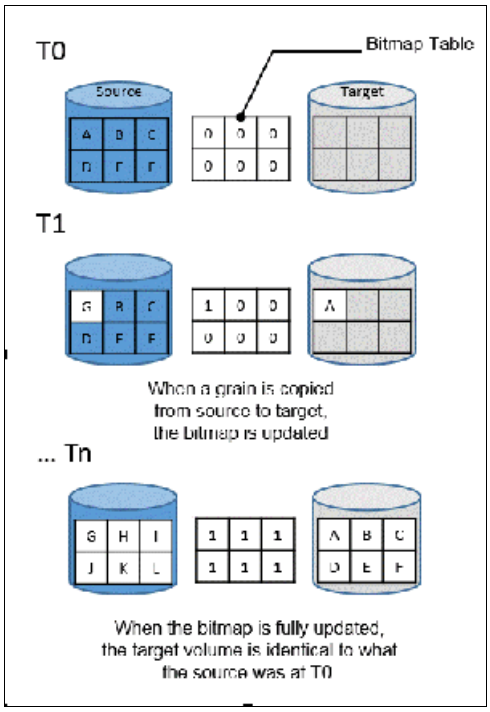


Figure 10-2 A simplified representation of grains and bitmap

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 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 Spectrum Virtualize Release 8.4 or higher, 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). In prior releases, it uses CoW only, regardless of the pool type.

With CoW, as shown in Figure 10-3, when data is written on a source volume, the grain where the to-be-changed blocks is stored is first copied to the target volume and then modified on the source volume. The bitmap is updated to track the copy.

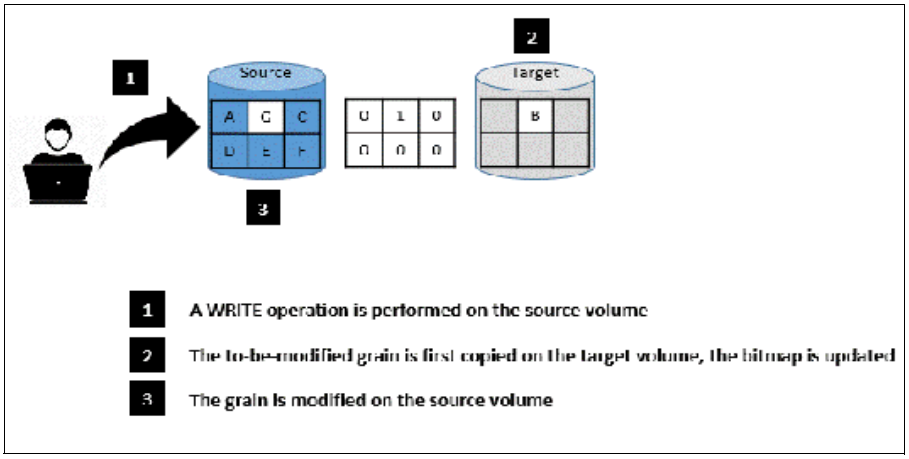


Figure 10-3 CoW steps

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 Spectrum 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 code 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 on page 616, 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 (*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.

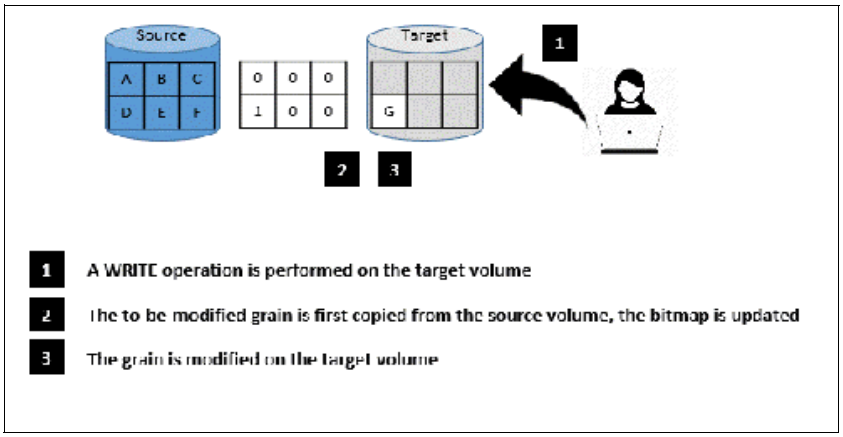


Figure 10-4 Copy on-Demand steps

Starting with IBM Spectrum Virtualize release 8.4 or higher, 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 on page 617.

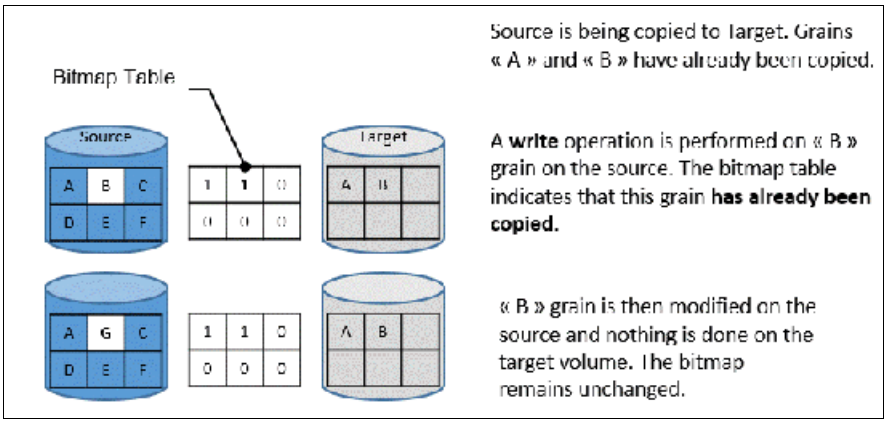


Figure 10-5 Modifying an already copied grain on the Source

- 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 is true for Standard Pools in IBM Spectrum Virtualize release 8.4 or higher, or any pool type if the code version is lower than 8.4.

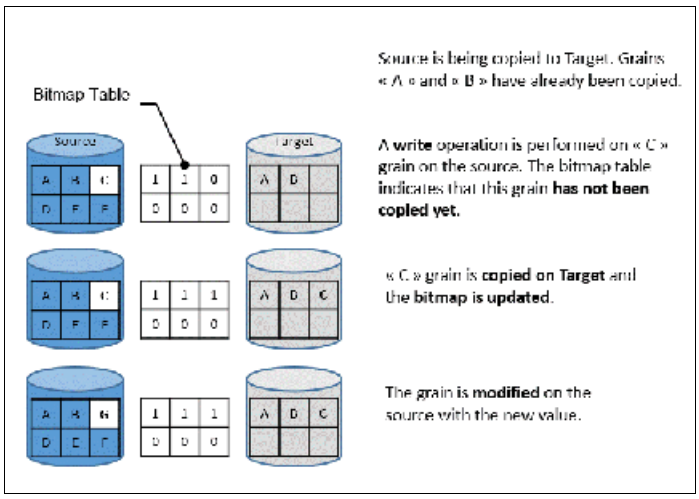


Figure 10-6 Modifying a non-copied grain on the Source

- If this is a DRP in a system running IBM Spectrum Virtualize release 8.4 or higher, the system performs a RoW operation, as described in “Copy on Write, Redirect on Write, and Copy on Demand” on page 614.

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 618. The source volume remains unchanged.

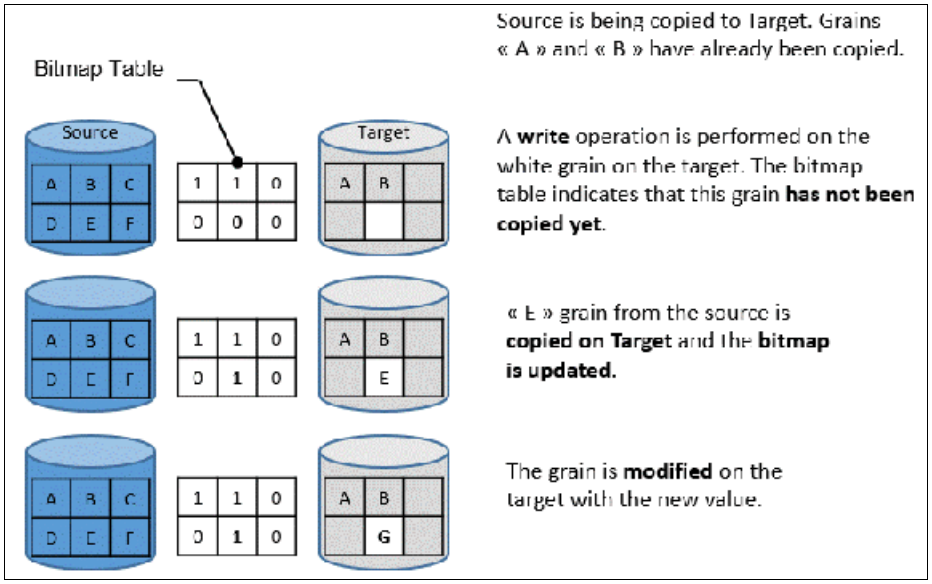


Figure 10-7 Modifying a non-copied grain on the target

Note: If the entire grain is to be modified and not only part of it (some blocks only), the copy on demand is bypassed. The bitmap is updated, and the grain on the target is modified but not copied first.

- 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.

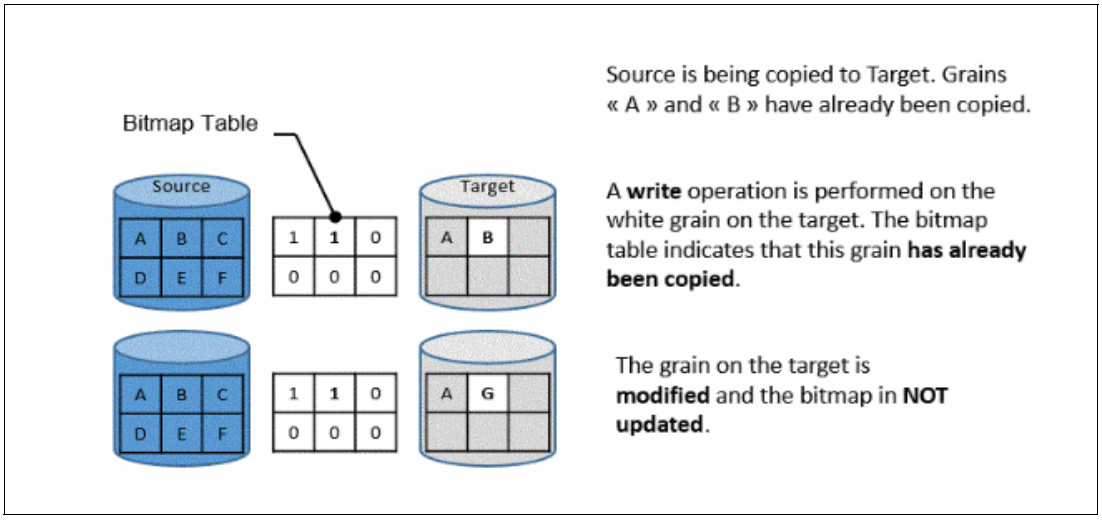


Figure 10-8 Modifying an already copied grain on the Target

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.

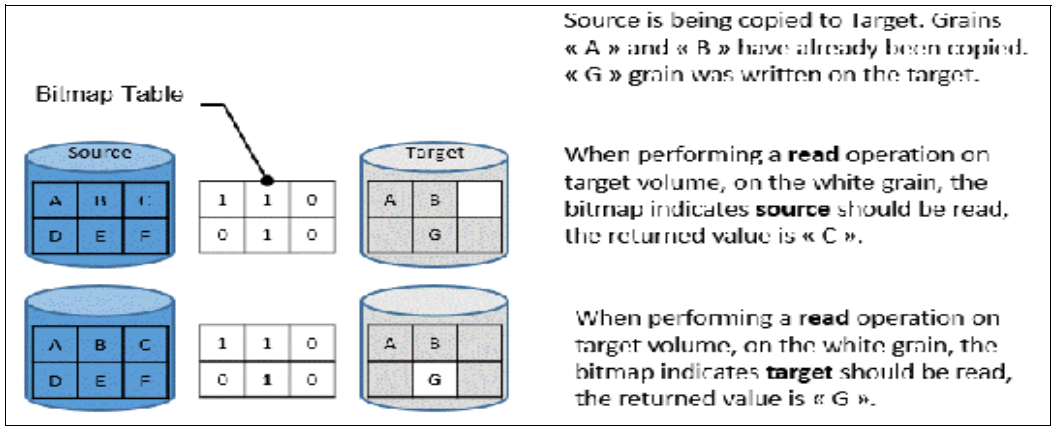


Figure 10-9 Reading a grain on target

If source has multiple targets, the Indirection layer algorithm behaves differently on Target I/Os. For more information about multi-target operations, see 10.3.11, “Multiple target FlashCopy” on page 625.

10.3.7 Interaction with cache

IBM Spectrum Virtualize based systems have their cache 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 shows the IBM Spectrum Virtualize software stack, including cache architecture.

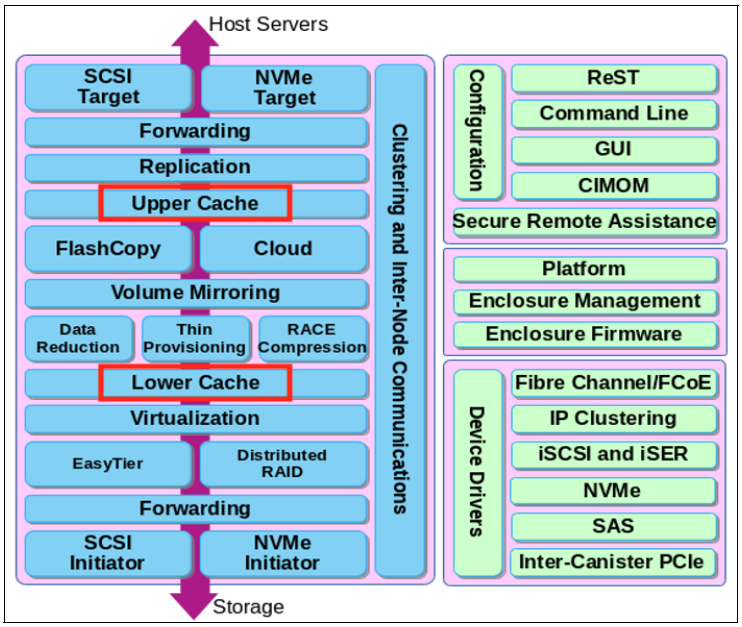


Figure 10-10 IBM Spectrum Virtualize software architecture

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 Spectrum Virtualize release 8.4 or higher, the RoW mechanism used for DRPs aims to reduce the overhead introduced by CoW.

Also, the two-level cache described above provides more performance improvements to the FlashCopy mechanism. Because the FlashCopy layer is above the lower cache in the IBM Spectrum 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 lower cache layer instead.

10.3.8 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 Spectrum Virtualize copies per second. This amount assumes that the bandwidth to the managed disks (MDisks) can accommodate this rate.

If the IBM Spectrum 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. Background copy I/O and I/O that is arriving from the hosts tend to see an increase in latency and a consequential reduction in throughput.

Background copy and foreground I/O continue to progress, and do not stop, hang, or cause the node to fail.

The background copy is performed by one of the nodes that belong to the I/O group in which the source volume is stored. This responsibility is moved to the other node in the I/O group if the node that performs the background and stopping copy fails.

10.3.9 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, which are time needed to recover data from a previous state), it does not help with the initial RTO.

For example, as shown in Figure 10-11 on page 622, a FlashCopy mapping was defined between a source volume and a target volume by using the **-incremental** option.

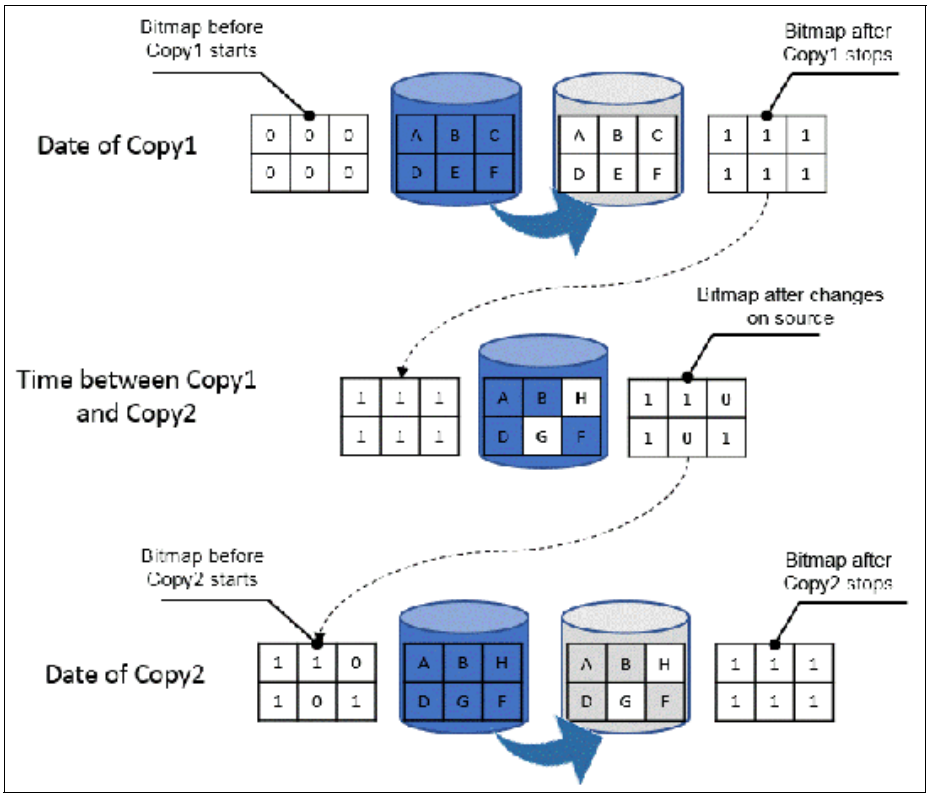


Figure 10-11 Incremental FlashCopy example

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.

10.3.10 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 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 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 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

10.3.11 Multiple target FlashCopy

A volume can be the source of multiple target volumes. A target volume can also be the source of another target volume. However, 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 626 shows these different possibilities.

Every source-target relation is a FlashCopy mapping and is maintained with its own bitmap table. No consistency group bitmap table exists.

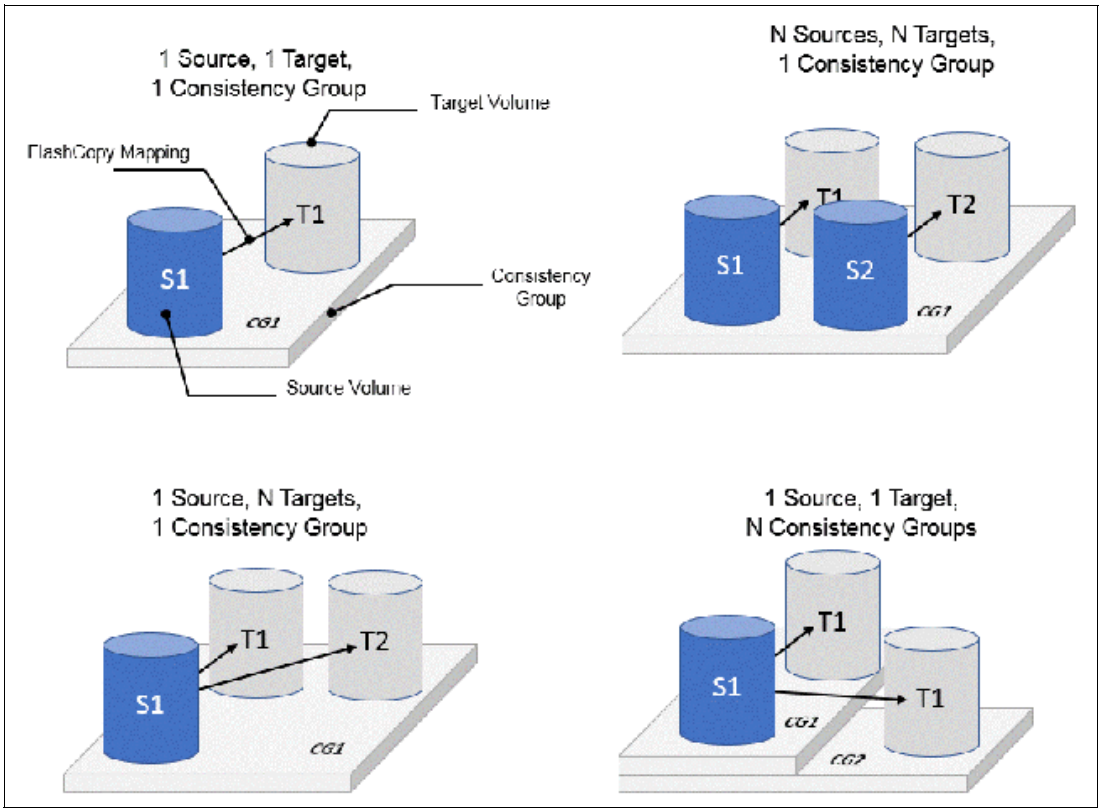


Figure 10-12 Consistency groups and mappings combinations

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 is 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 Spectrum Virtualize release 8.4 or higher 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 614.

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.

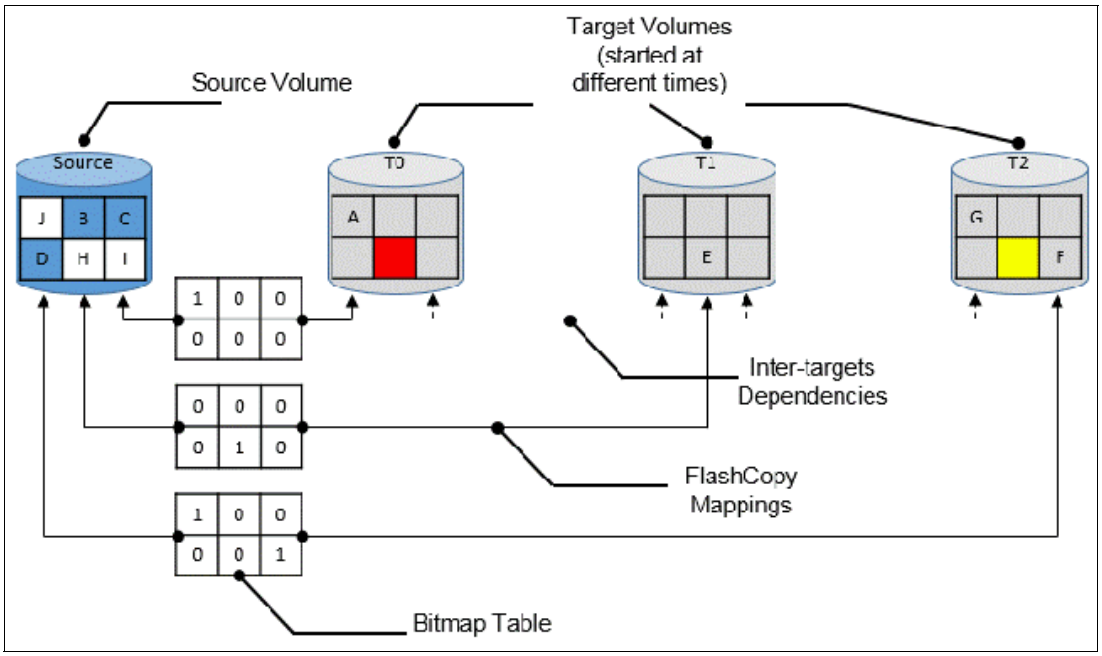


Figure 10-13 FlashCopy dependencies example

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 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 kind of 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 (see Figure 10-13 on page 627), 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 627), 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 627, 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 627, 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

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 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 627, 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 627, 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 per second	256 KB grains per second	64 KB grains per second
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

10.3.12 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. A FlashCopy source supports multiple targets (up to 256), and therefore, multiple rollback points.

A key advantage of the IBM Spectrum 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 Spectrum Virtualize also can 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).

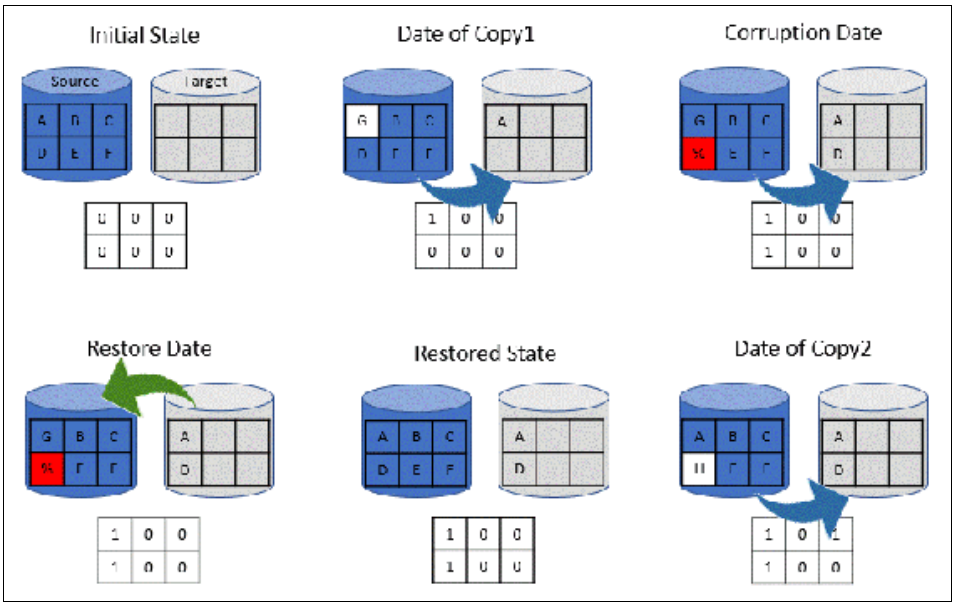


Figure 10-14 A reverse FlashCopy example for data restoration

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 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 below). In this last case, because “D” grain was copied, it is not copied again on 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.

10.3.13 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.

10.3.14 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 completes 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.

10.3.15 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.

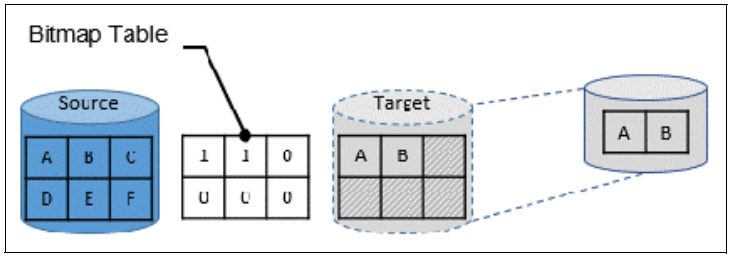


Figure 10-15 Thin-provisioned target volume

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.

10.3.16 Serialization of I/O by FlashCopy

In general, the FlashCopy function in the IBM Spectrum 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.

10.3.17 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 Spectrum 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.

10.3.18 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.

10.3.19 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 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.

10.3.20 FlashCopy attributes and limitations

The FlashCopy function in IBM Spectrum 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 Spectrum 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 Spectrum 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 Spectrum Virtualize V8.4.2 are listed in Table 10-9.

Table 10-9 FlashCopy limitations in V8.4.2

Property	Maximum number
FlashCopy mappings per system	15864[a]
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

a. Applies to IBM SAN Volume Controller DH8, SV2, SV2, SA2, FS7XXX, FS9XXX

10.3.21 Expanding Volumes In a FlashCopy Mapping

Volumes that are part of a FlashCopy mapping that are user-defined can now be expanded. User-defined mappings are those that are manually created. The limitations on volume expansion are:

- 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 reasonable number of host mappings. However, in enterprise data centers with many host mappings, use the CLI to run your FlashCopy commands.

10.4.1 FlashCopy presets

The IBM Spectrum 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`
- Incremental: `No`
- Delete after completion: `No`
- Cleaning rate: `No`
- Primary copy source pool: `Target pool`

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 641, 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 stopped or deleted), the target volumes become unusable.

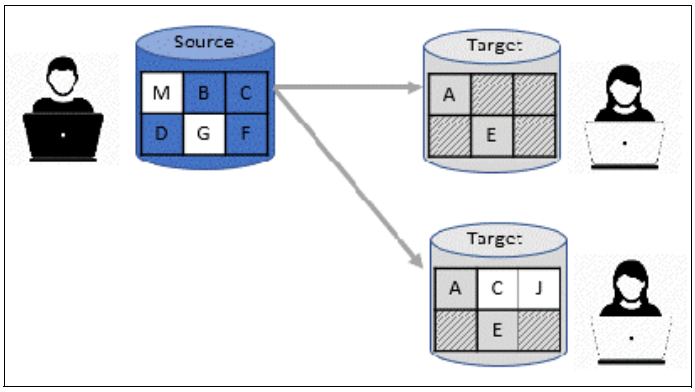


Figure 10-16 FlashCopy snapshot preset example

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`
- Incremental: `No`
- Delete after completion: `Yes`
- Cleaning rate: `50`
- Primary copy source pool: `Target pool`

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`
- Incremental: `Yes`
- Delete after completion: `No`
- Cleaning rate: `50`
- Primary copy source pool: `Target pool`

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 Spectrum 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 Spectrum Virtualize GUI.

When the IBM Spectrum 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.

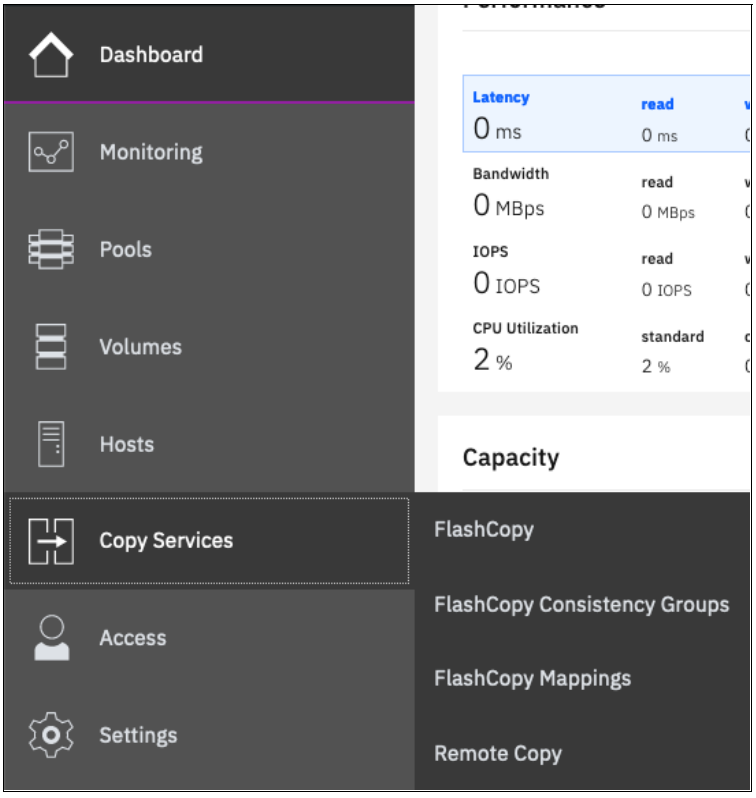


Figure 10-17 Copy Services menu

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.

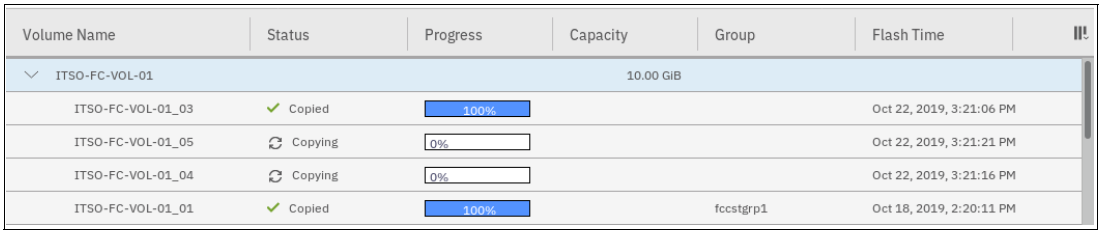


Figure 10-18 Source and target volumes displayed in the FlashCopy window

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.

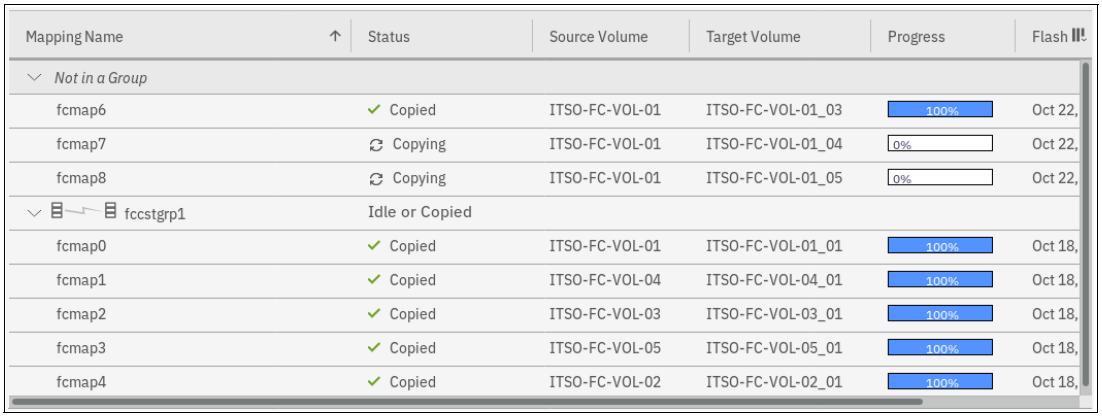


Figure 10-19 FlashCopy Consistency Groups window

- The FlashCopy Mappings window is accessible by clicking **Copy Services** → **FlashCopy Mappings**. Use the FlashCopy Mappings window (as shown in Figure 10-20) to display the list of mappings between source volumes and target volumes.

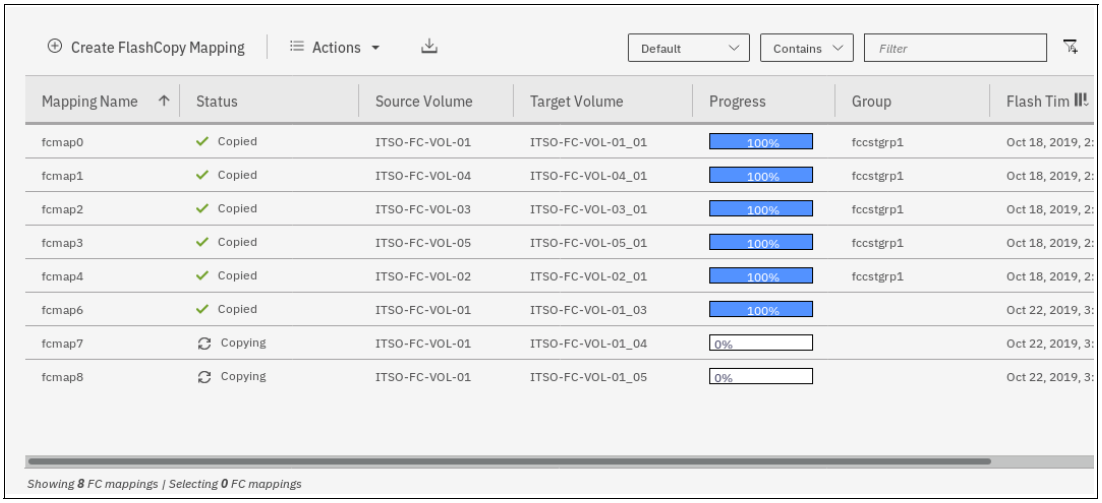


Figure 10-20 FlashCopy Mapping window

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. Select the volume for which you want to create the FlashCopy mapping. Right-click the volume or click the **Actions** menu.

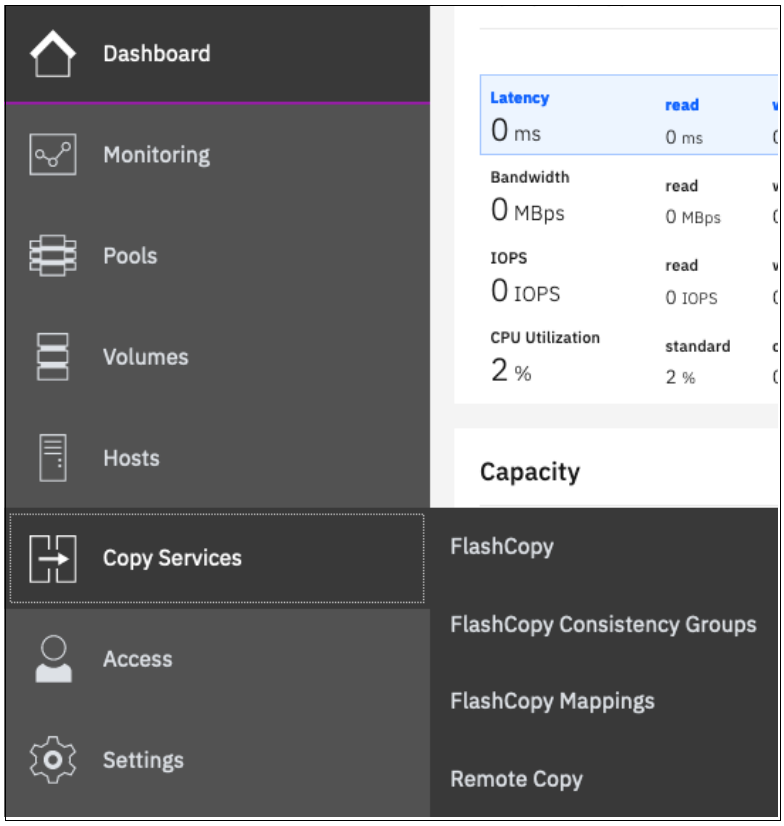


Figure 10-21 FlashCopy window

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 646.
- If you want the system to create the target volumes for you, see “Creating a FlashCopy mapping and target volumes” on page 651.

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.

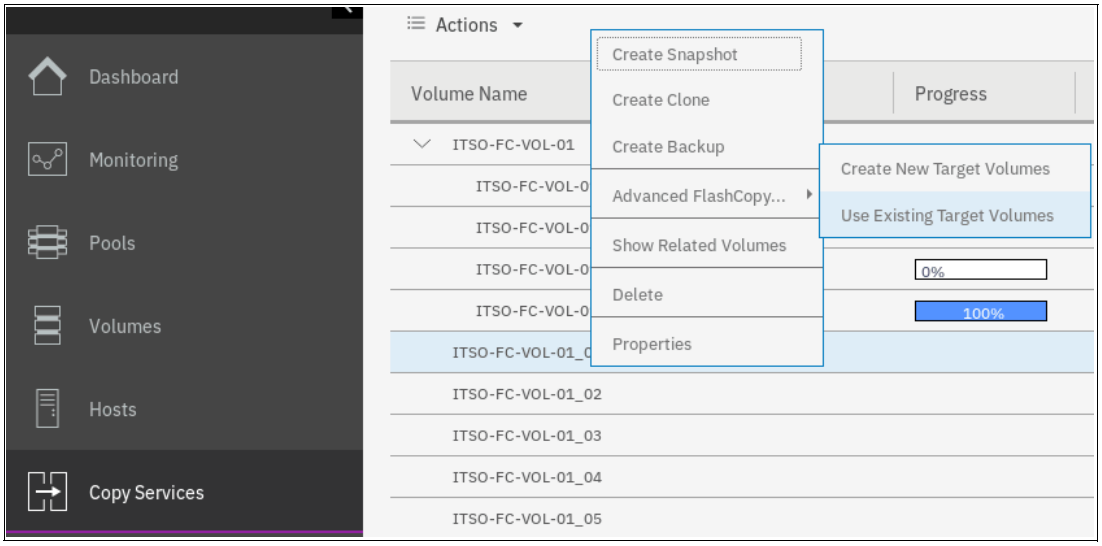


Figure 10-22 Creating a FlashCopy mapping with an existing target

The Create FlashCopy Mapping window opens, as shown in Figure 10-23. In this window, you create the mapping between the selected source volume and the target volume you want to create a mapping with. Then, 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.

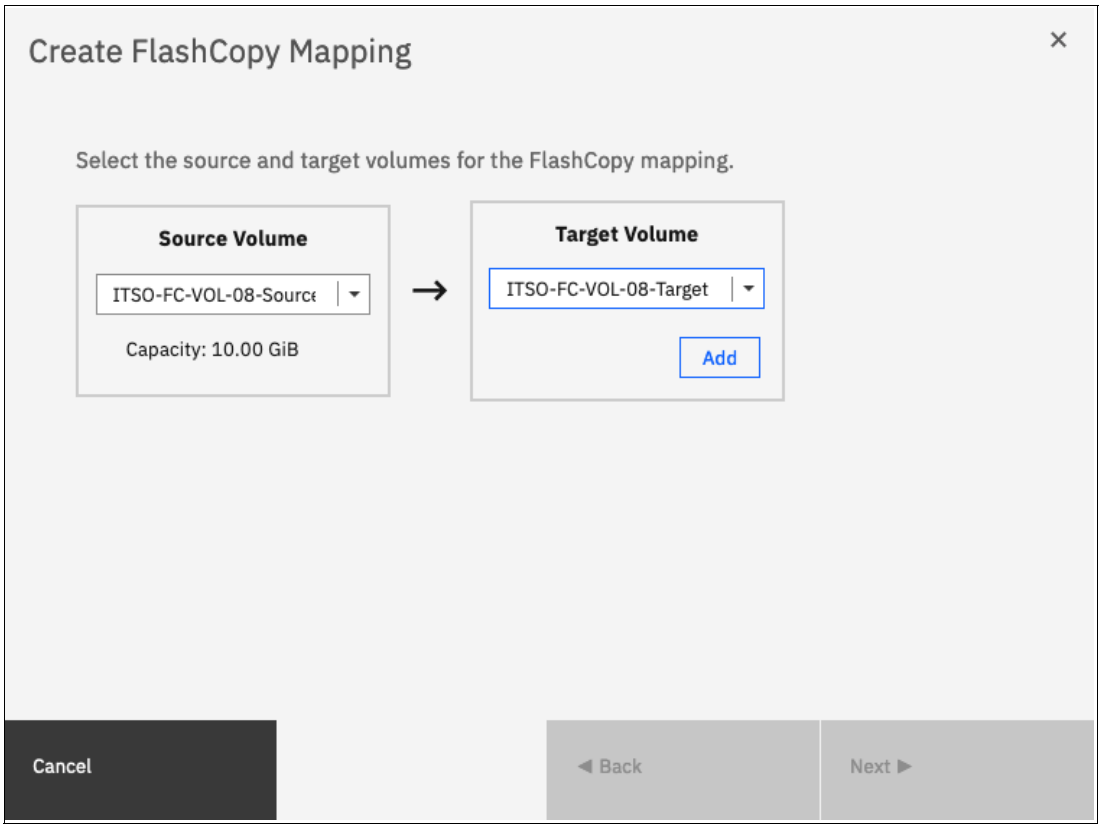


Figure 10-23 Selecting source and target for a FlashCopy mapping

To remove a mapping that was created, click ✖ (see Figure 10-24).

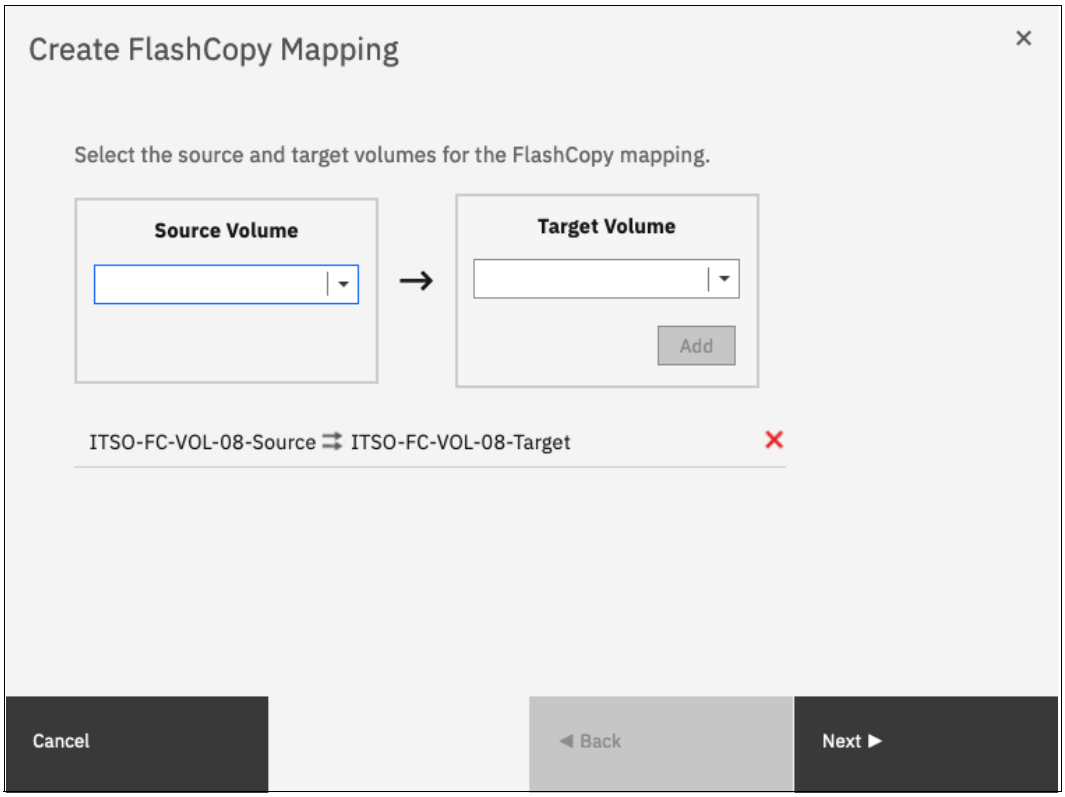


Figure 10-24 Viewing source and target at creation time

2. Click **Next** after you create all of the mappings that you need (see Figure 10-24).
3. 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 649. For more information about the presets, see 10.4.1, “FlashCopy presets” on page 640:
   - 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.

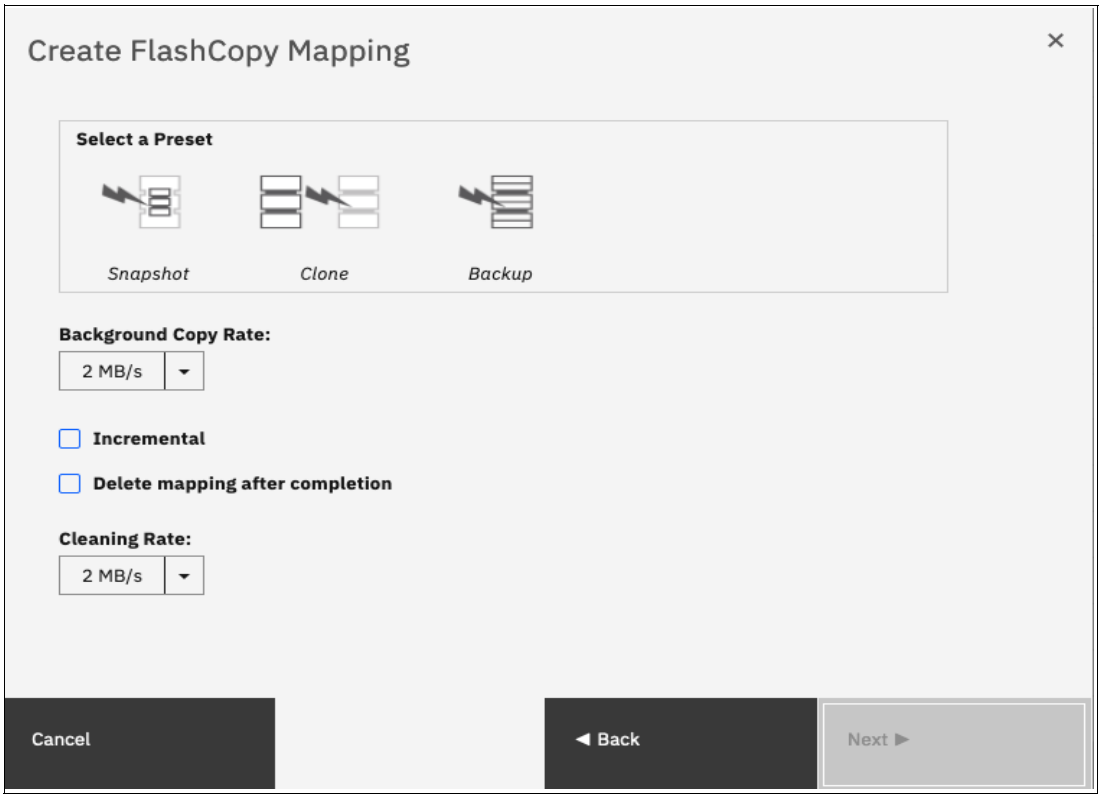


Figure 10-25 FlashCopy mapping preset selection

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 621, or Table 10-6 on page 630.

When your FlashCopy mapping setup is ready, click **Next**.

4. 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.

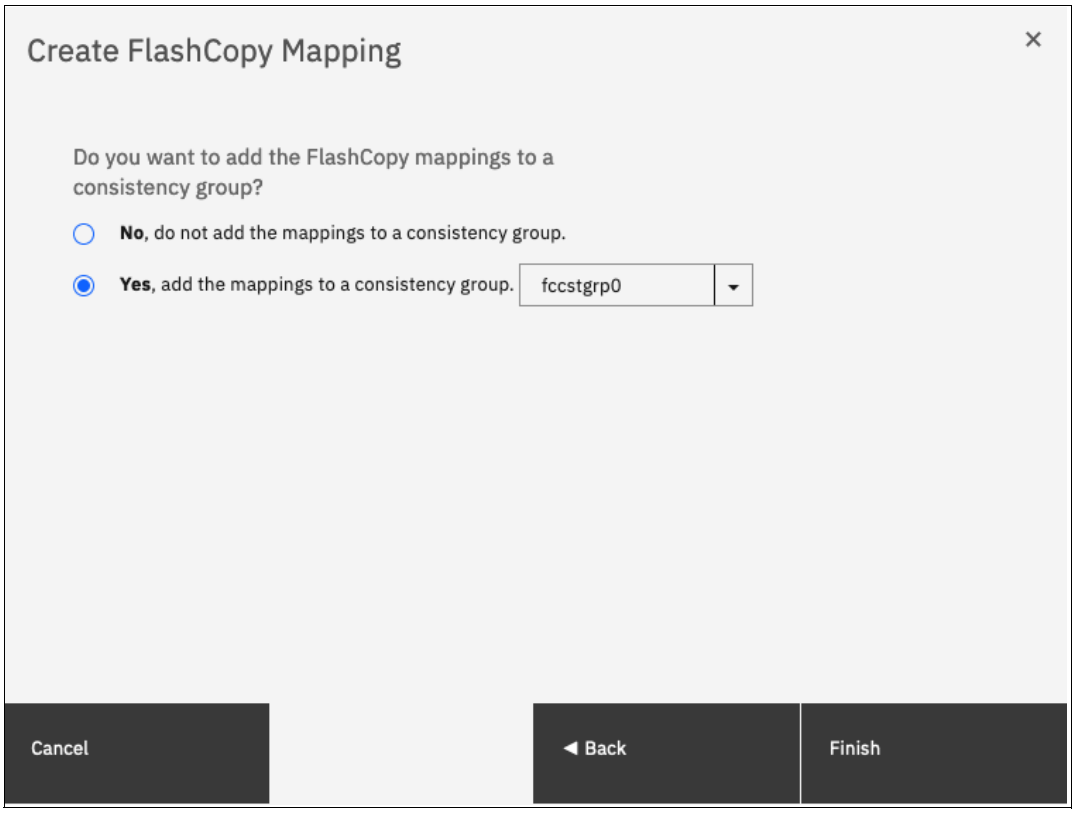


Figure 10-26 Select or not a consistency group for the FlashCopy mapping

5. 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.

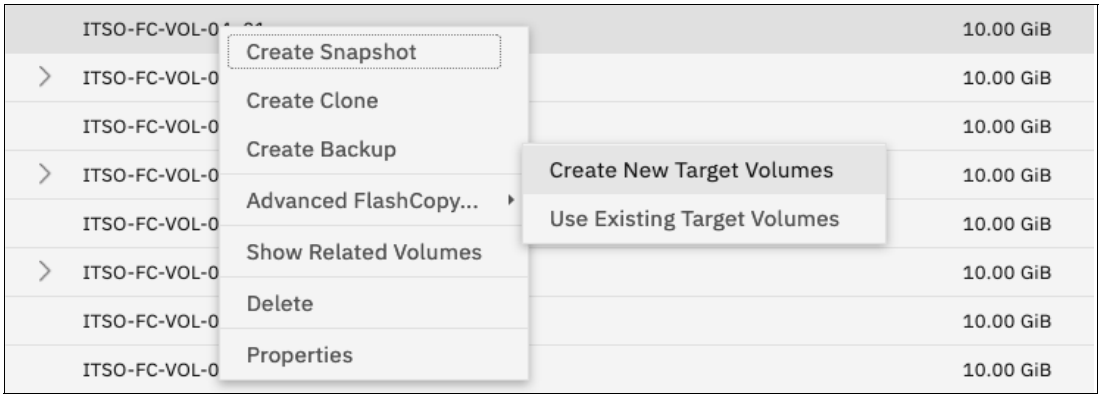


Figure 10-27 Creating a FlashCopy mapping and creating targets

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 652. For more information about the presets, see 10.4.1, “FlashCopy presets” on page 640:
   - 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.

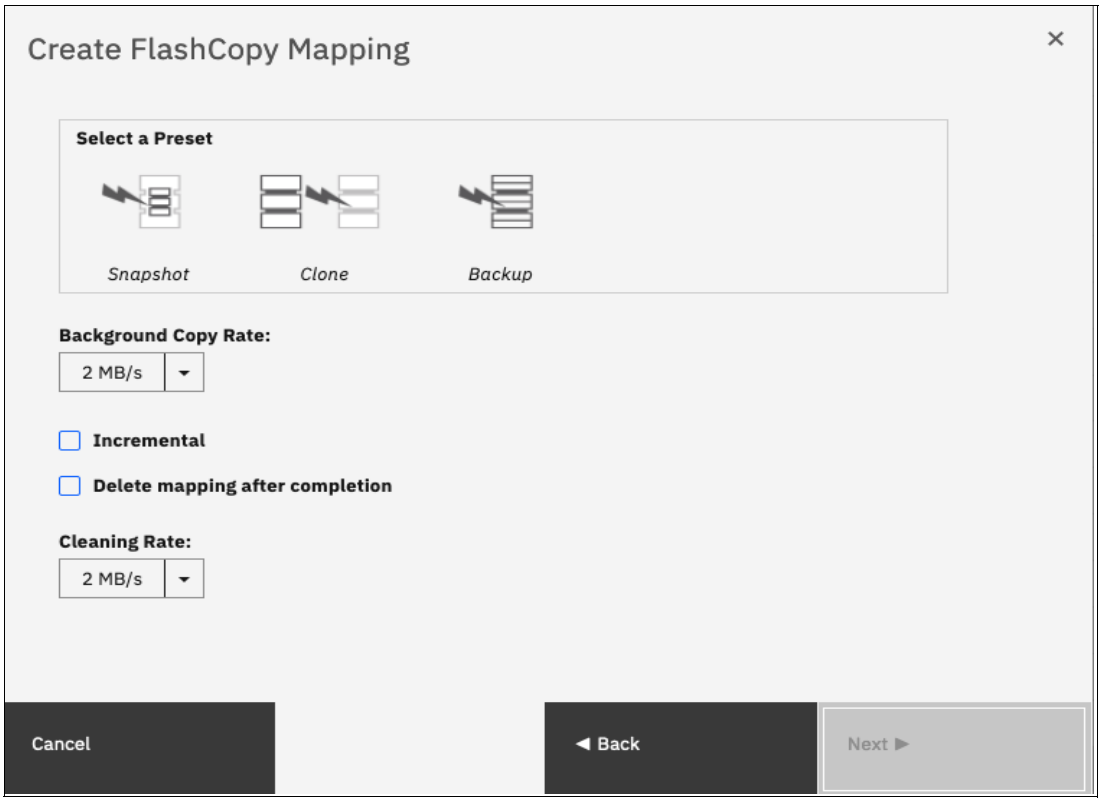


Figure 10-28 FlashCopy mapping preset selection

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 621, or Table 10-6 on page 630.

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.

   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.

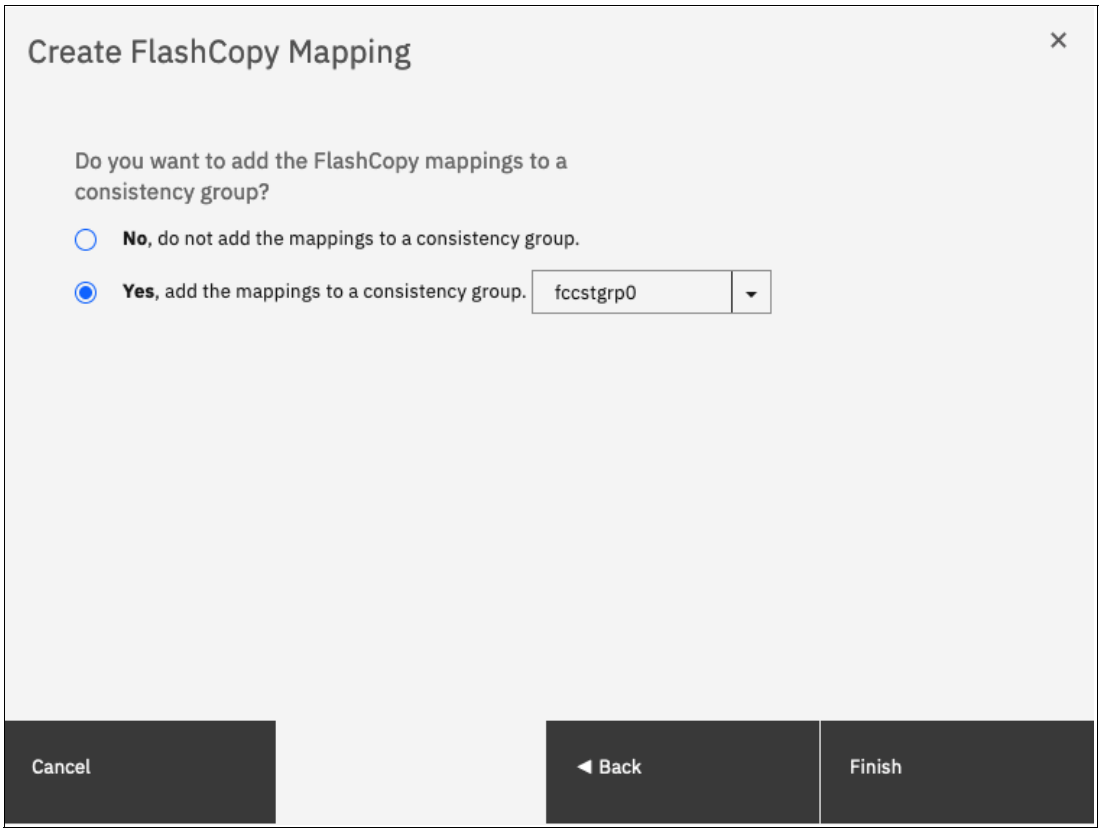


Figure 10-29 Select a consistency group for the FlashCopy mapping

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. Click **Next**.

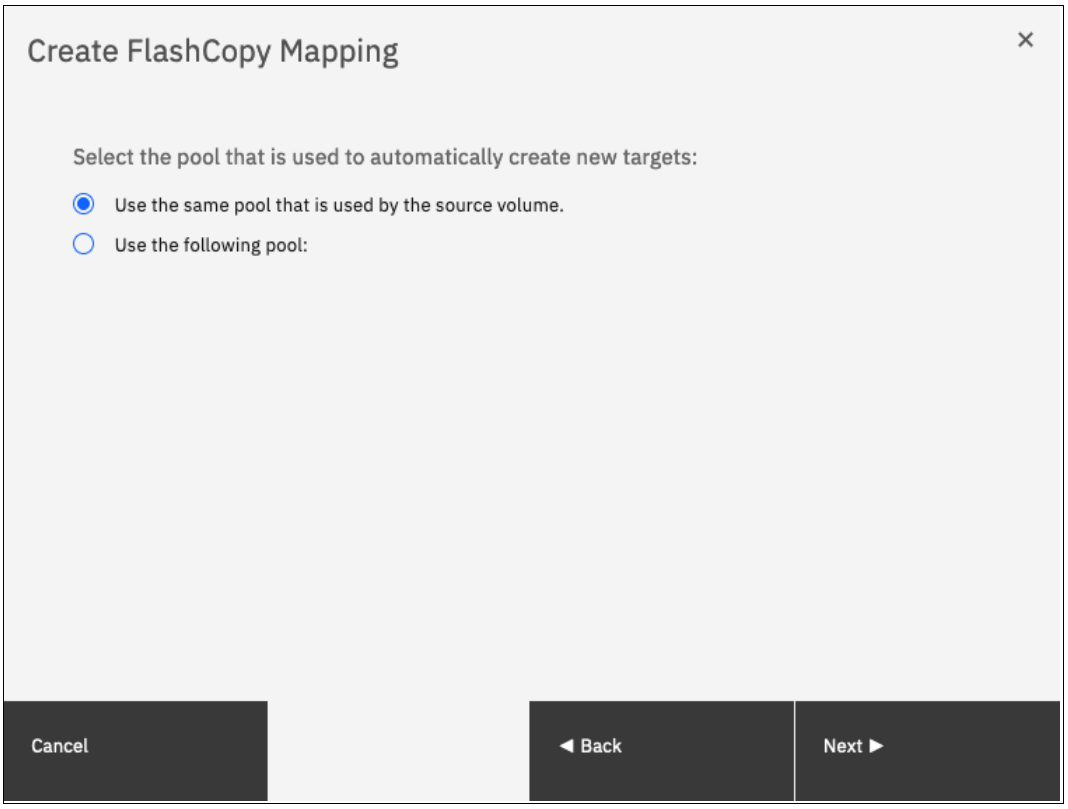


Figure 10-30 Select the pool

6. The system prompts the user how to define the new volumes that are created, as shown in Figure 10-31 on page 655. 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**.

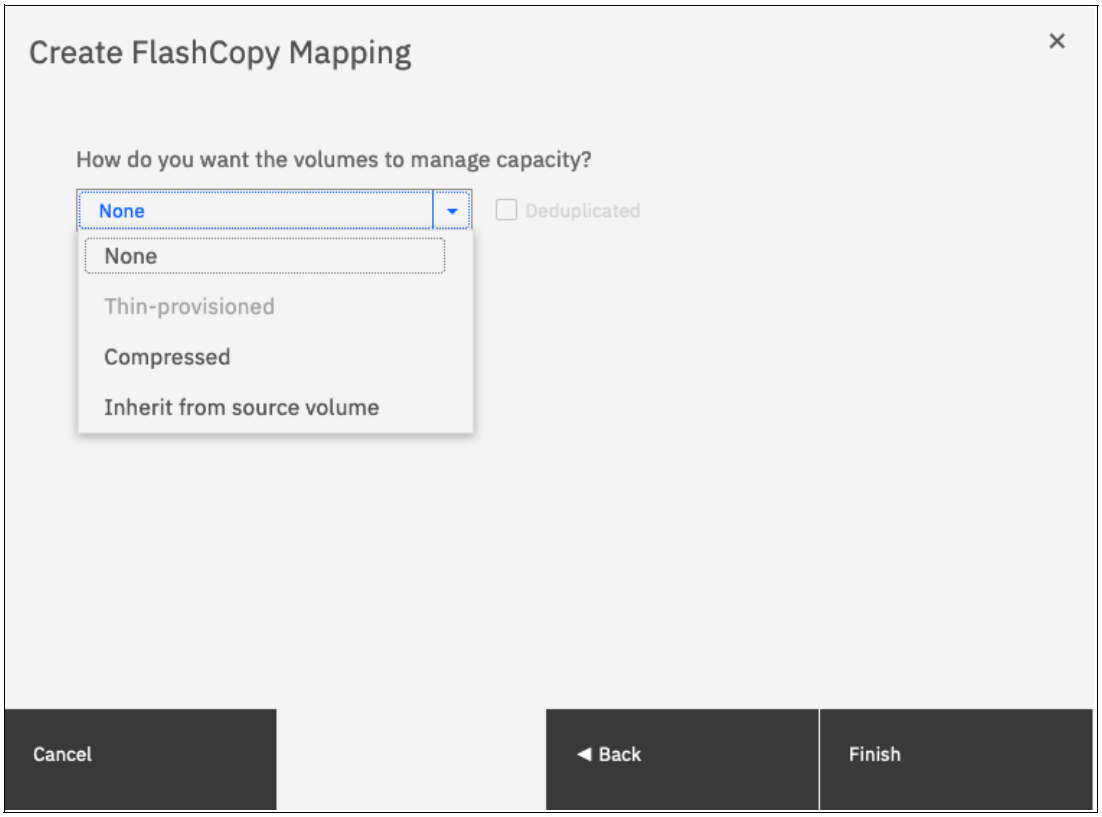


Figure 10-31 Select the type of volumes for the created targets

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`

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.

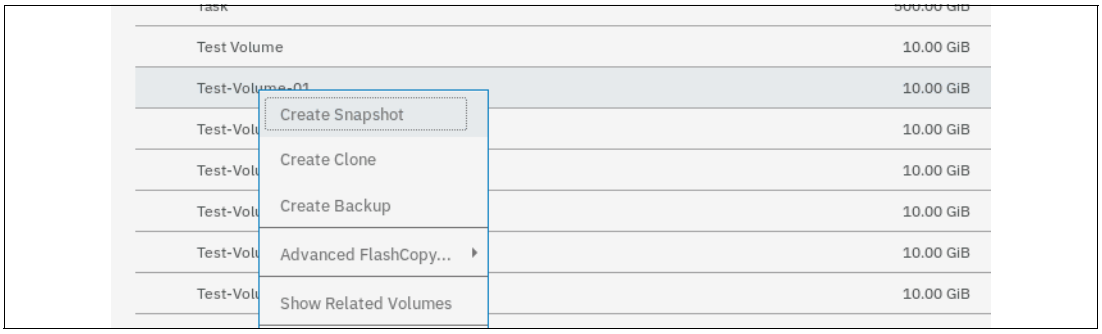


Figure 10-32 Single-click snapshot creation and start

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.

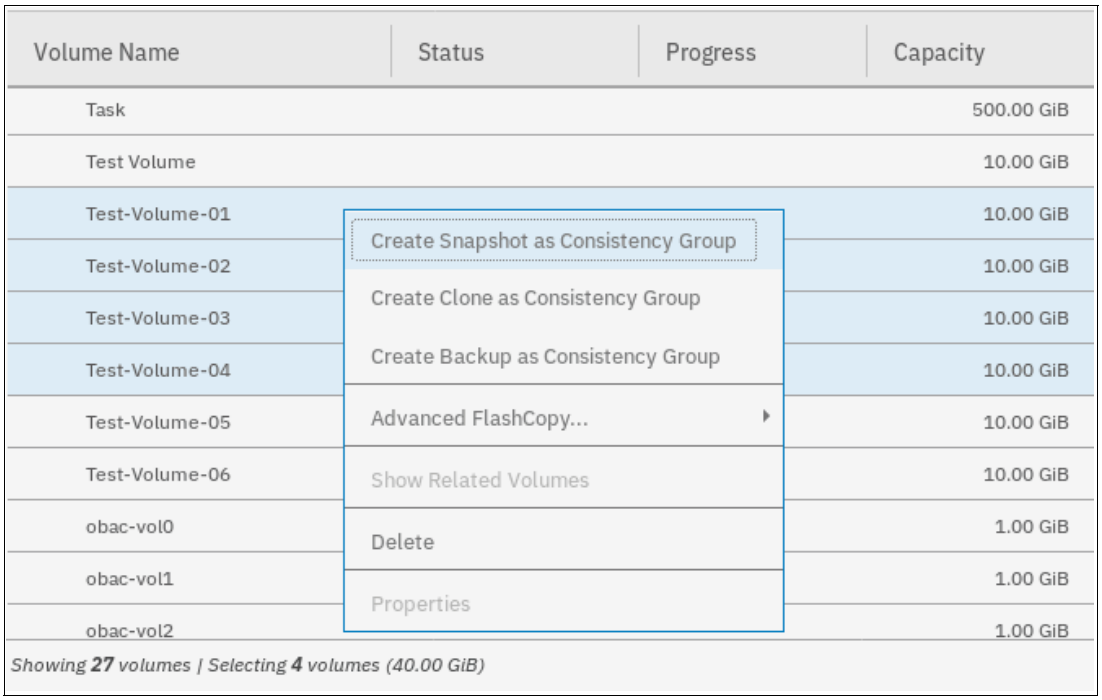


Figure 10-33 Selection single-click snapshot creation and start

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`

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.

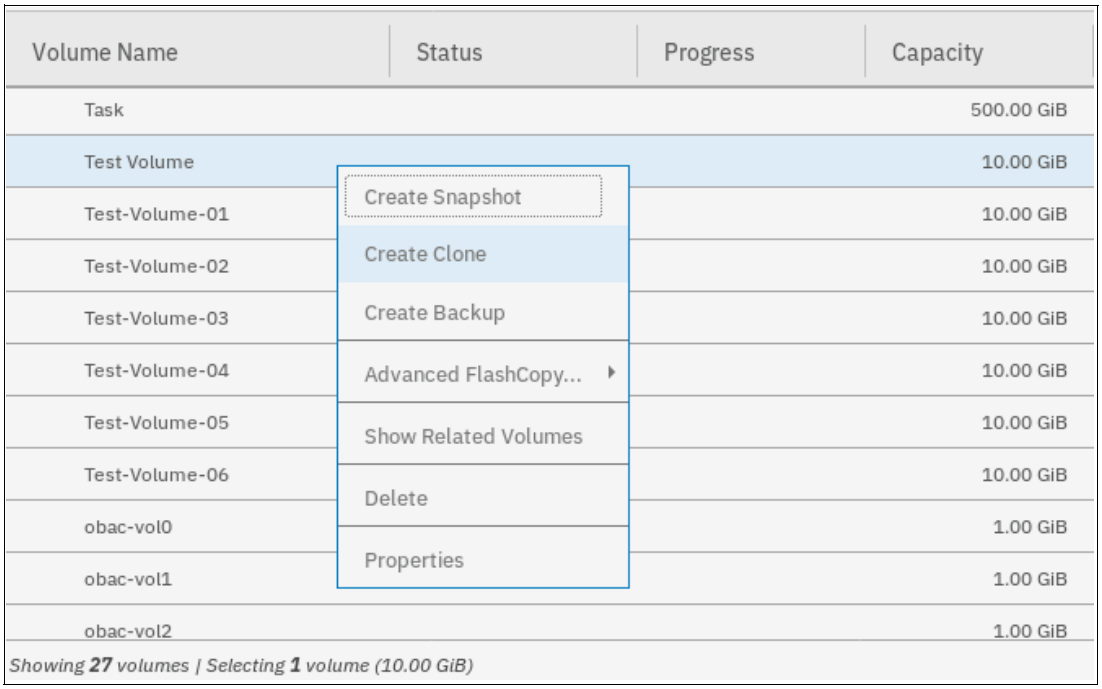


Figure 10-34 Single-click clone creation and start

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.

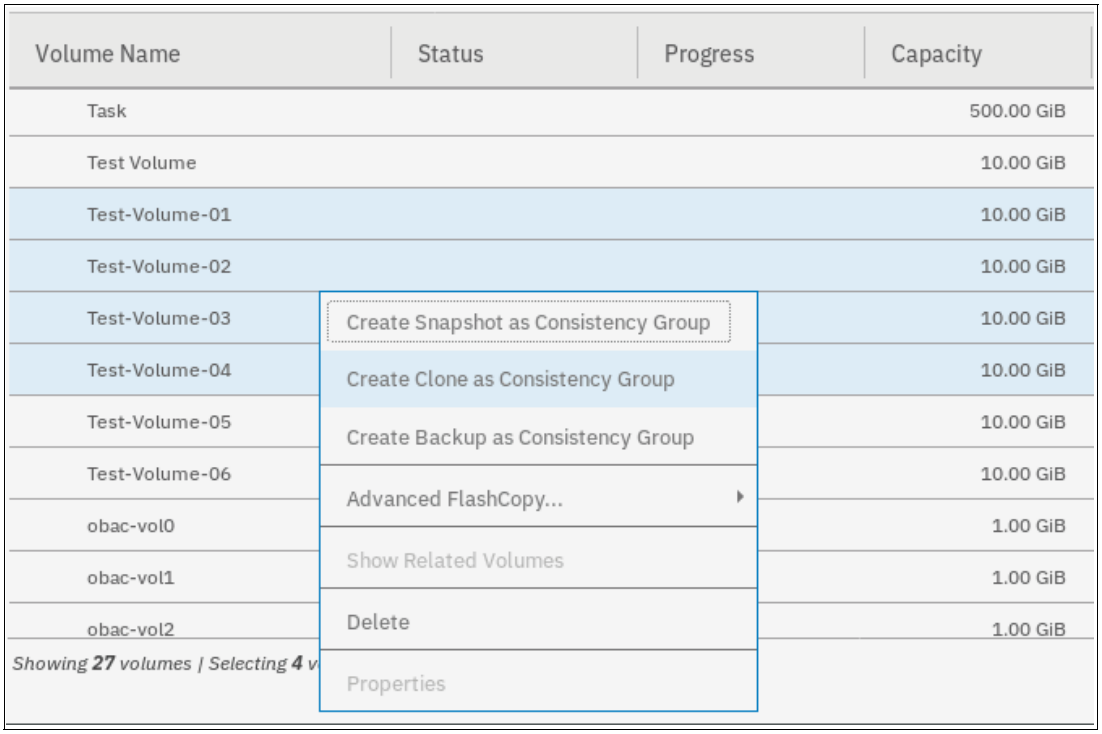


Figure 10-35 Selection single-click clone creation and start

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`

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.

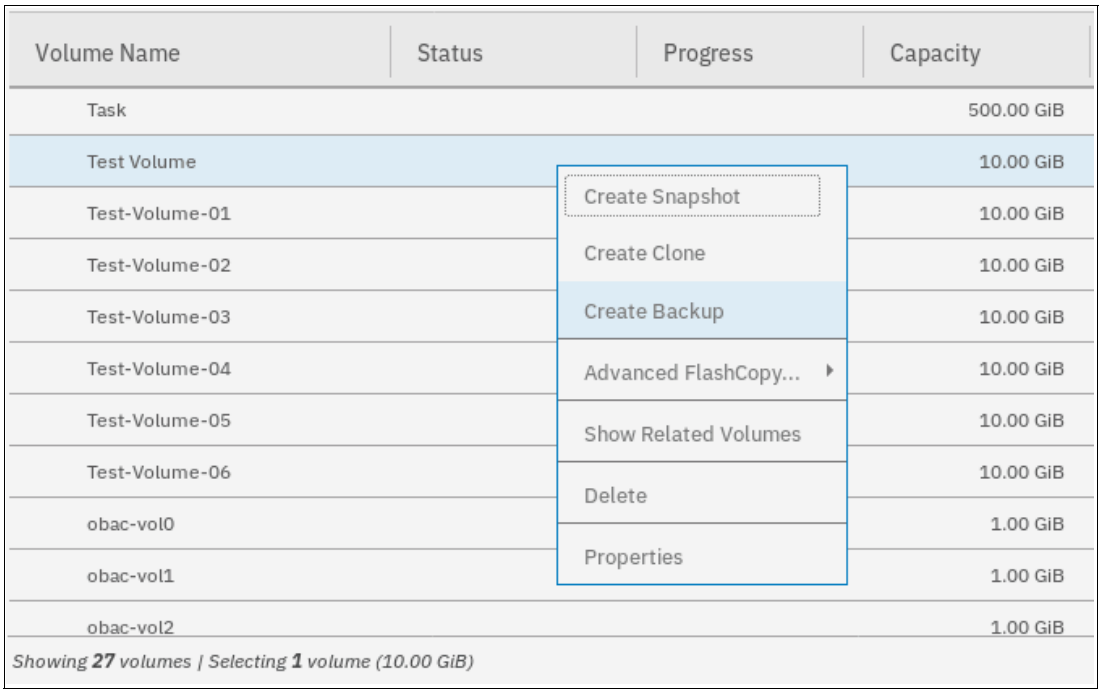


Figure 10-36 Single-click backup creation and start

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 660.

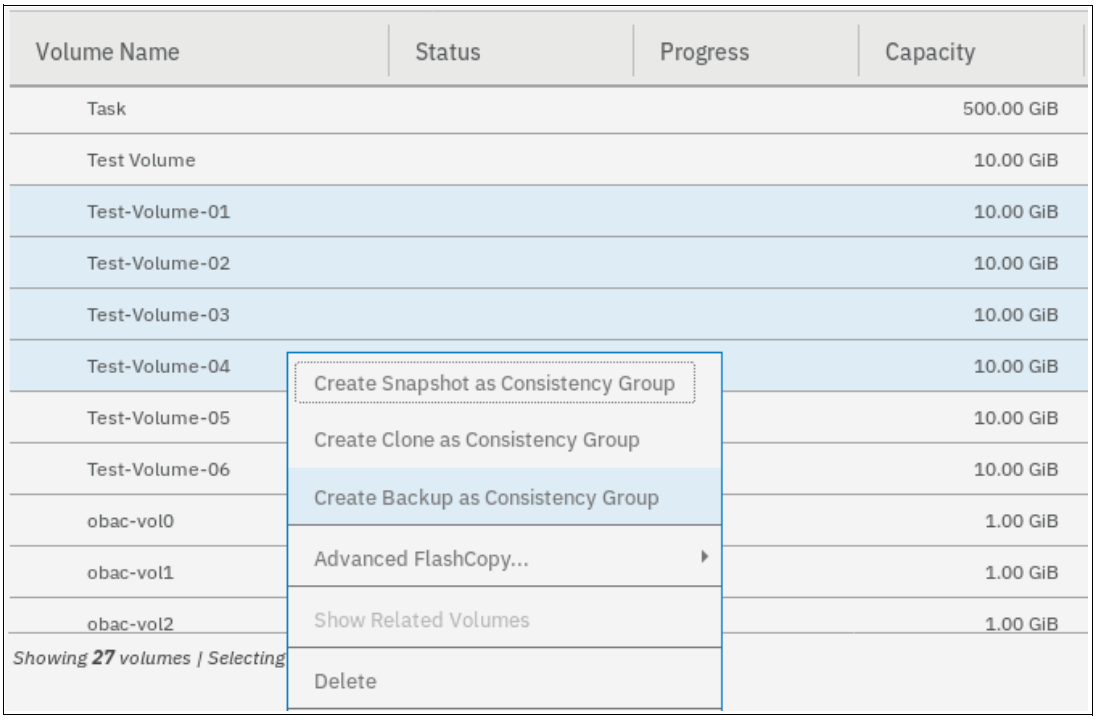


Figure 10-37 Selection single-click backup creation and start

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.

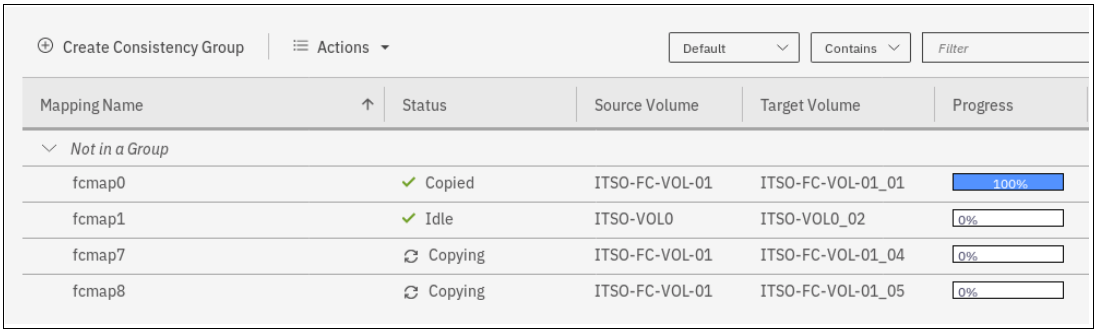


Figure 10-38 Creating a FlashCopy Consistency group

2. Enter the FlashCopy Consistency Group name that you want to use then, click **Create**, as shown in Figure 10-39.

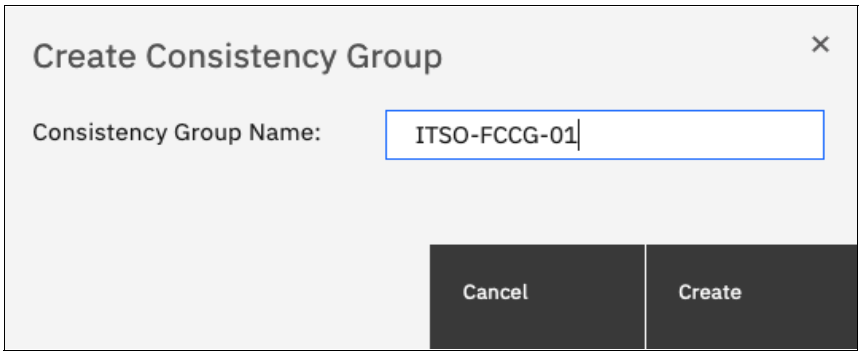


Figure 10-39 Enter the name and ownership group of new consistency group

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.

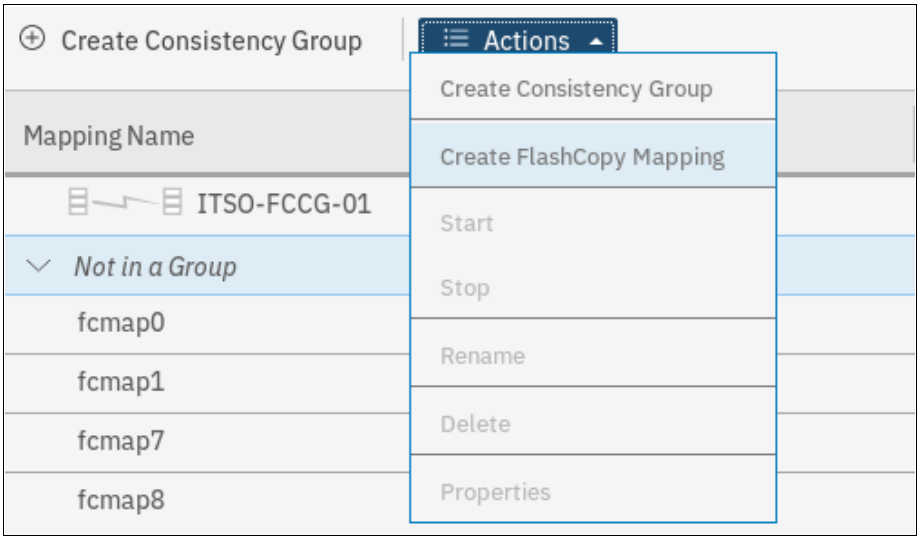


Figure 10-40 Creating a FlashCopy mapping

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.

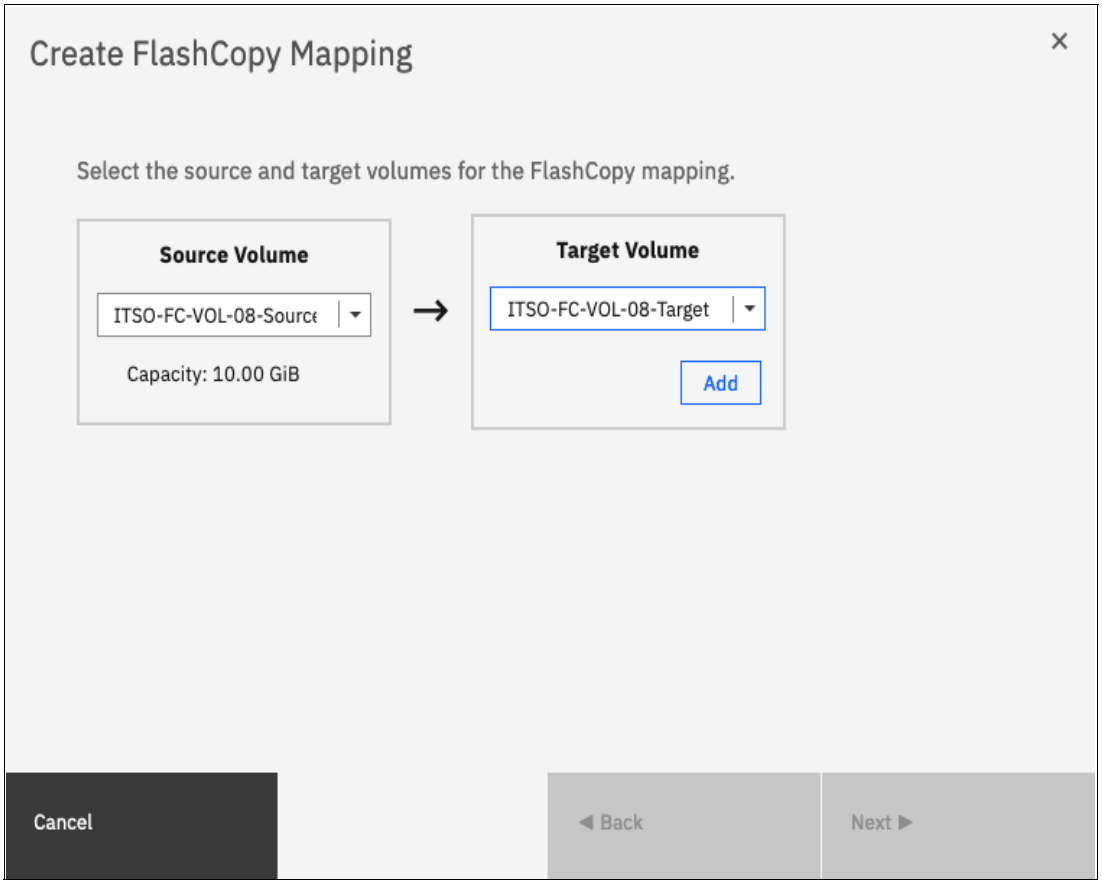


Figure 10-41 Select source and target volumes for the FlashCopy mapping

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 663. For more information about the presets, see 10.4.1, “FlashCopy presets” on page 640:
   - 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.

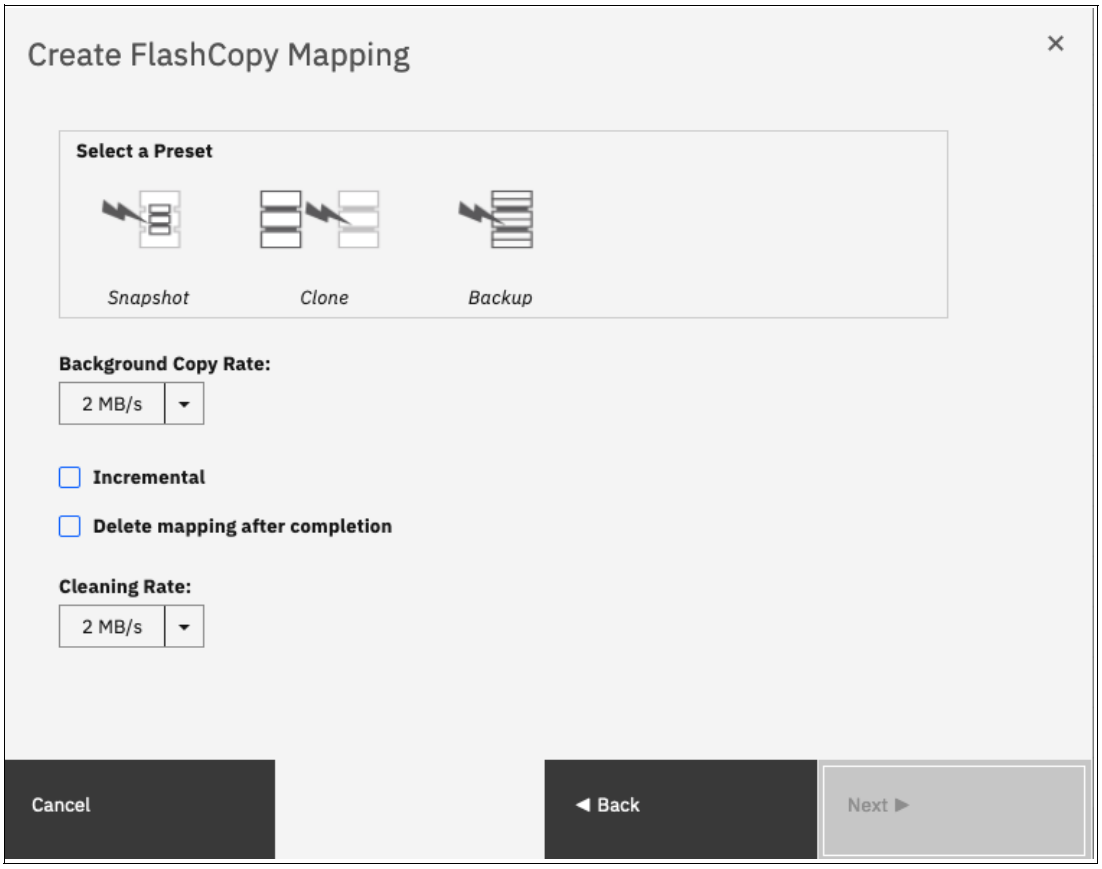


Figure 10-42 FlashCopy mapping preset selection

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. 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 621, or Table 10-6 on page 630.

5. 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 10-43. Also, depending on which window you are inside Copy Services, you can right-click at mappings and select **Show Related Volumes**.

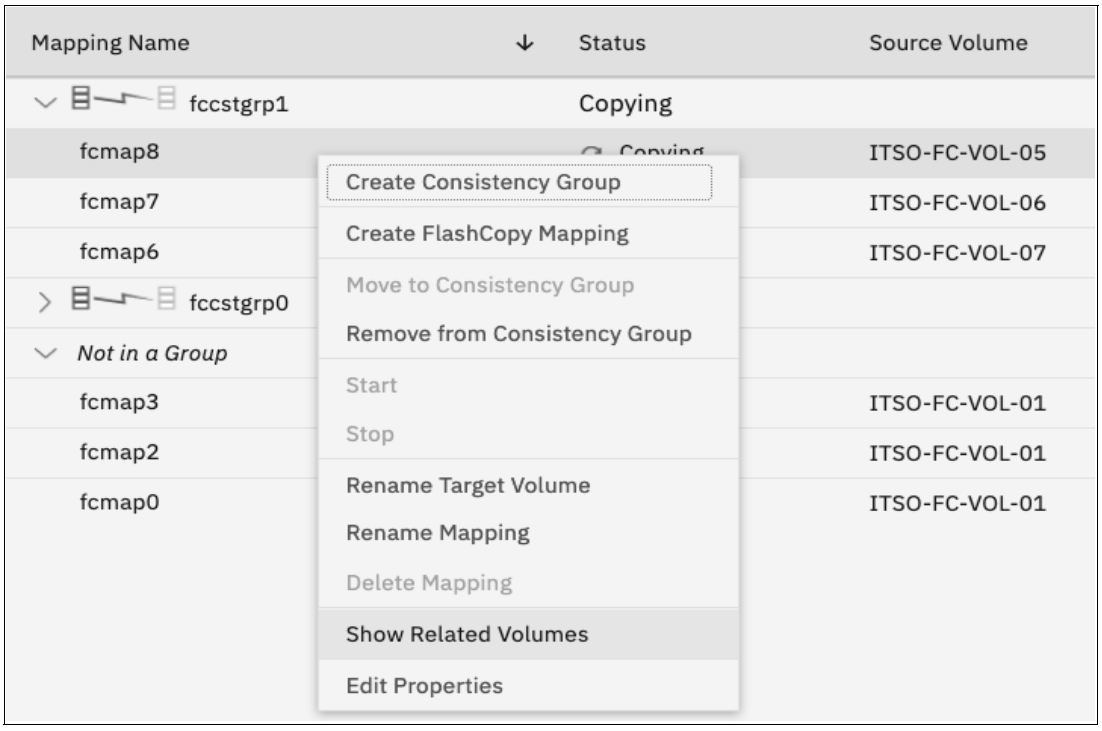


Figure 10-43 Showing related volumes for a mapping, a consistency group or another volume

3. In the related volumes window, you can see the related mapping for a volume, as shown in Figure 10-44. If you click one of these volumes, you can see its properties.



Figure 10-44 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 10-45.

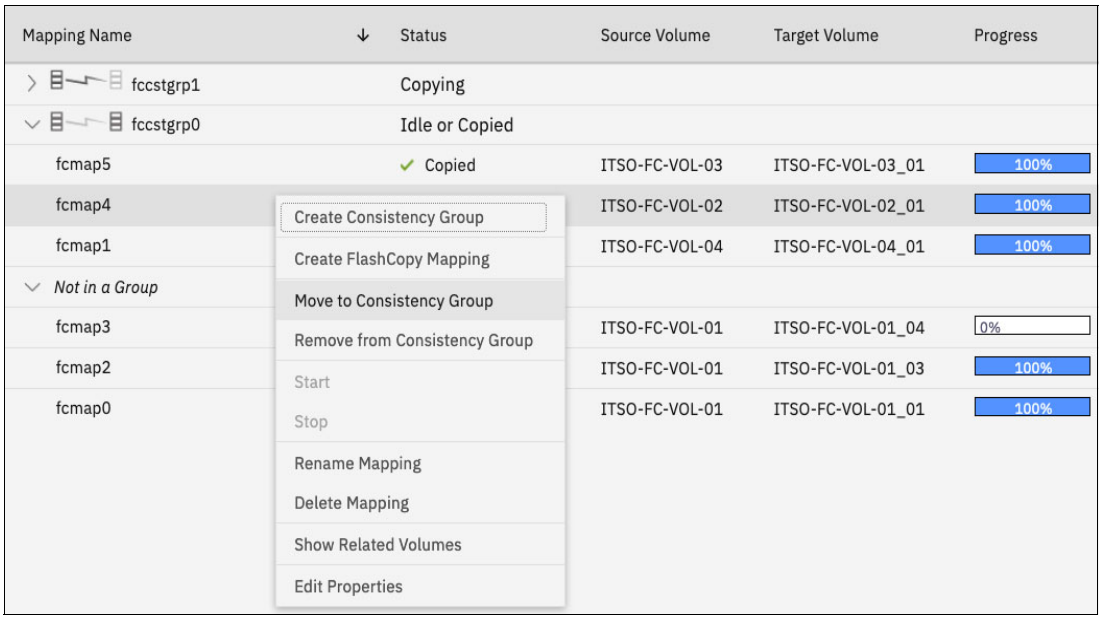


Figure 10-45 Moving a FlashCopy mapping to a consistency group

> **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 10-46.

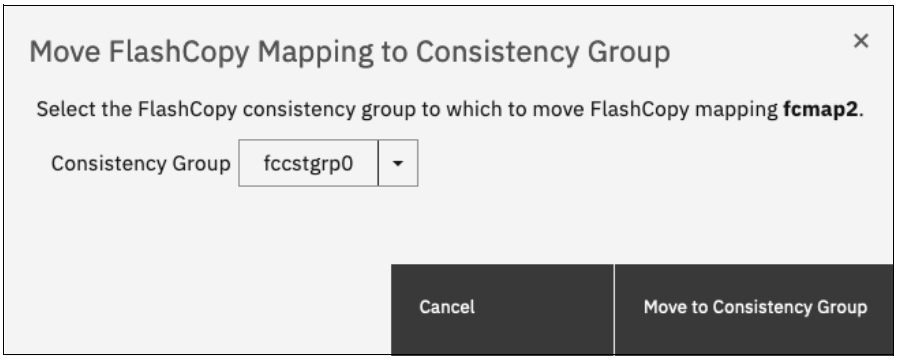


Figure 10-46 Selecting the consistency group where to move the FlashCopy mapping

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 10-47.

> **Note:** Only FlashCopy mappings that belong to a consistency group can be removed.

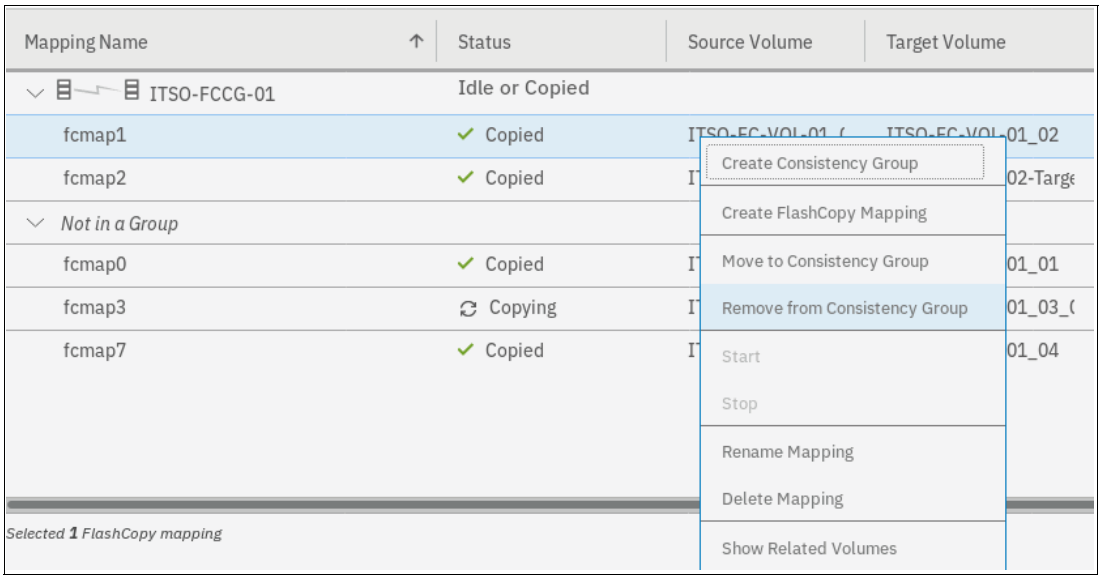


Figure 10-47 Removing FlashCopy mappings from a consistency group

3. In the Remove FlashCopy Mapping from Consistency Group window, click **Remove**, as shown in Figure 10-48.

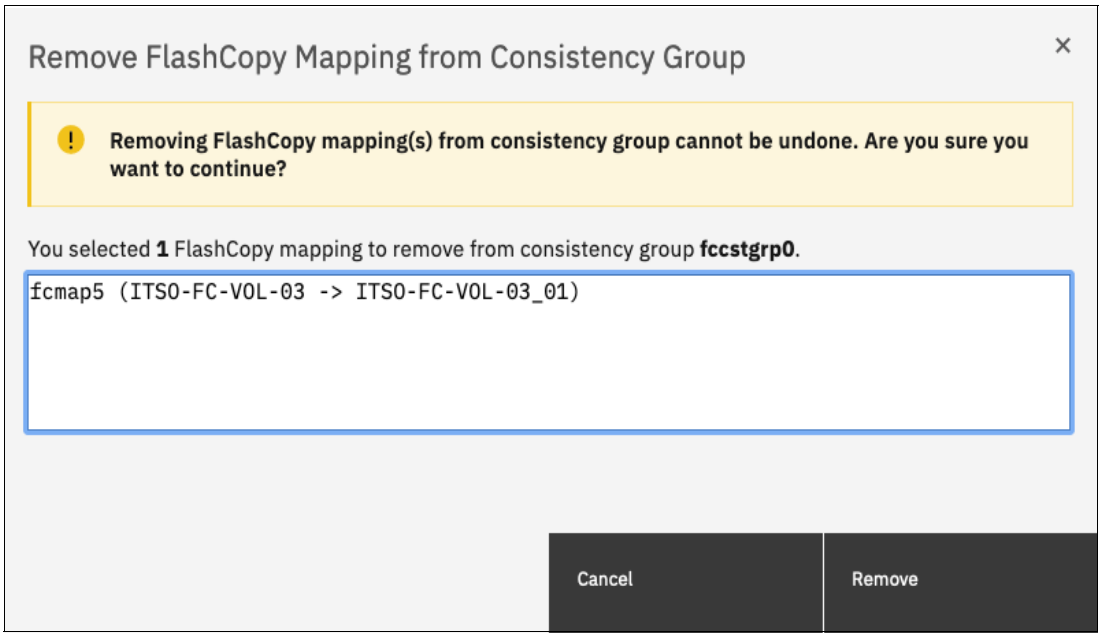


Figure 10-48 Confirm the selection of mappings to be removed

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 10-49.

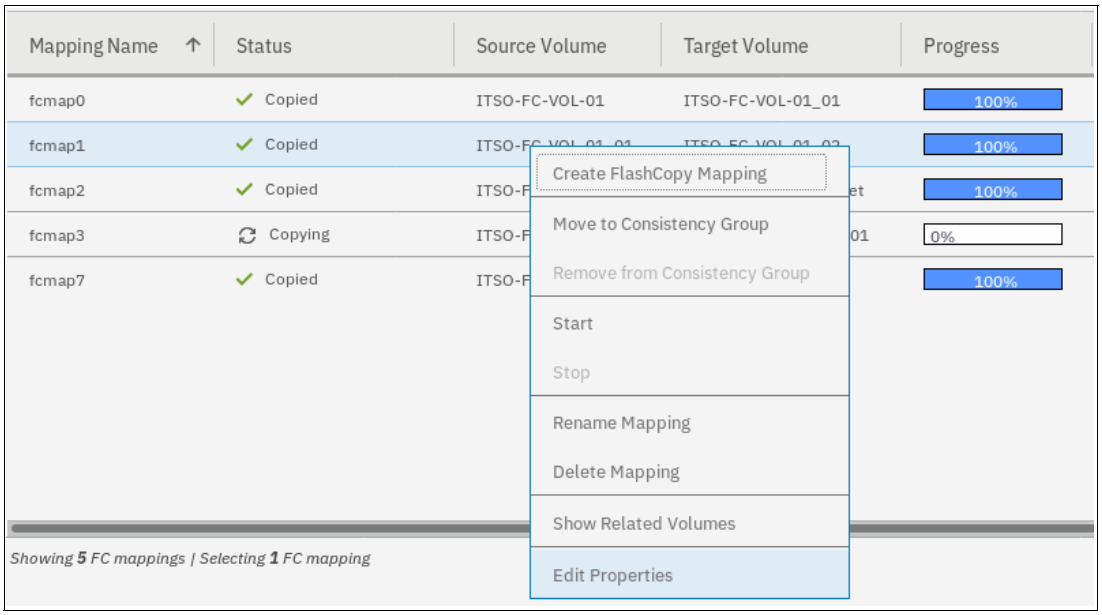


Figure 10-49 Editing a FlashCopy mapping properties

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 10-50.

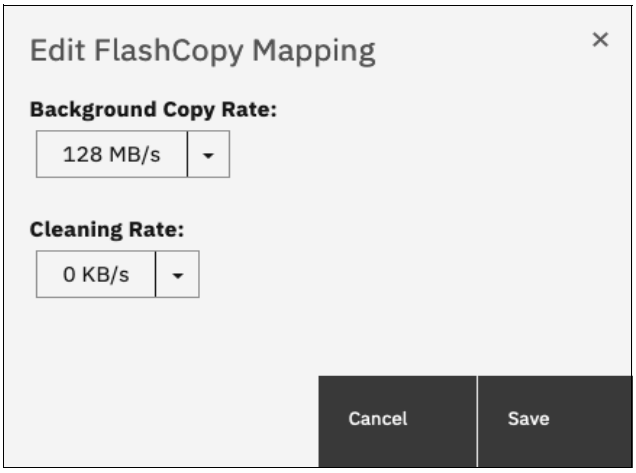


Figure 10-50 Editing copy and cleaning rates of a FlashCopy mapping

For more information about the Background Copy Rate and the Cleaning Rate, see Table 10-2 on page 621, or Table 10-6 on page 630.

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 10-51 on page 669.

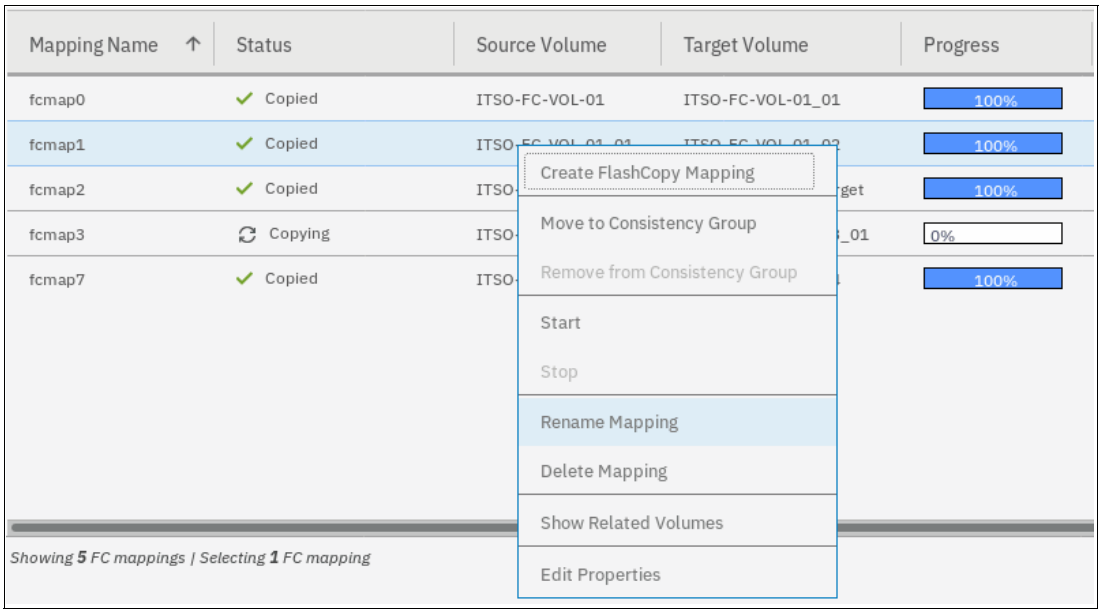


Figure 10-51 Renaming FlashCopy mappings

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 10-52.

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.

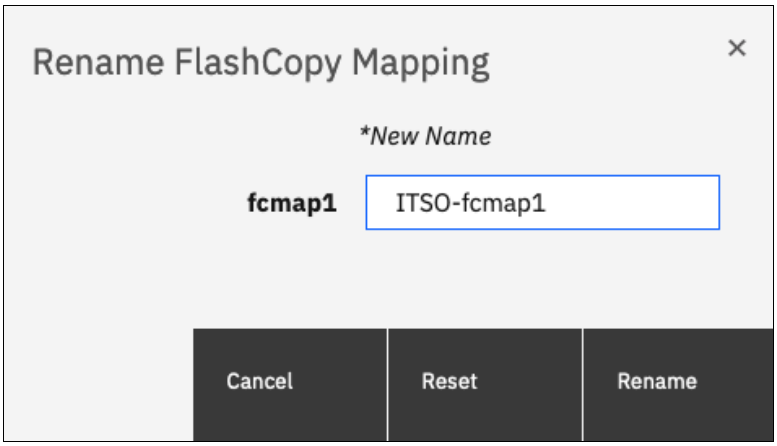


Figure 10-52 Renaming the selected FlashCopy mappings

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 you want to rename and select **Rename**, as shown in Figure 10-53.

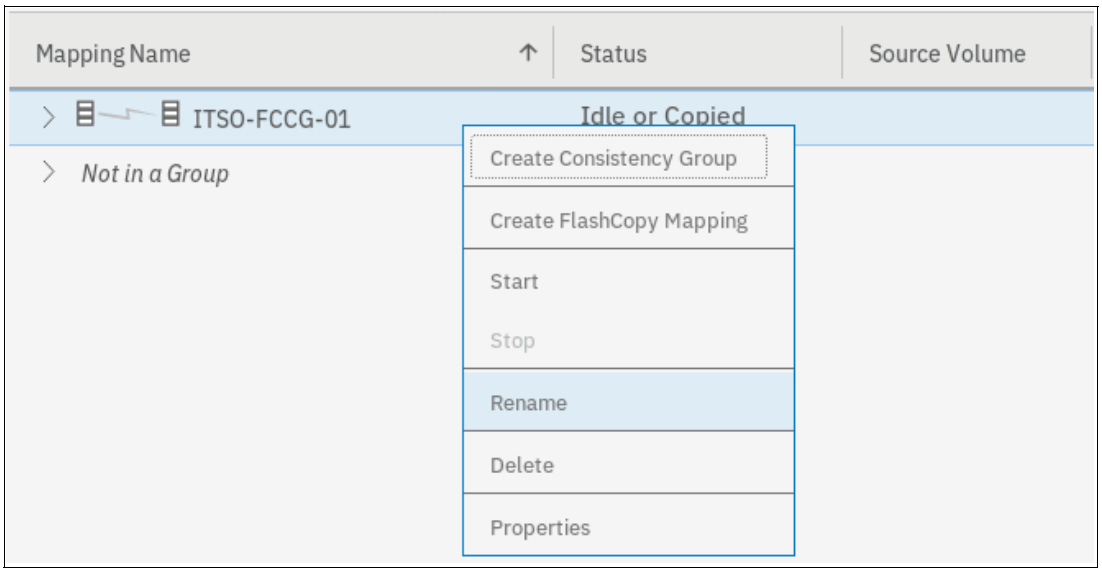


Figure 10-53 Renaming a consistency group

3. Enter the new name that you want to assign to the consistency group and click **Rename**, as shown in Figure 10-54.

> **Note:** It is not possible to select multiple consistency groups to edit their names all at the same time.



Figure 10-54 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 10-55.

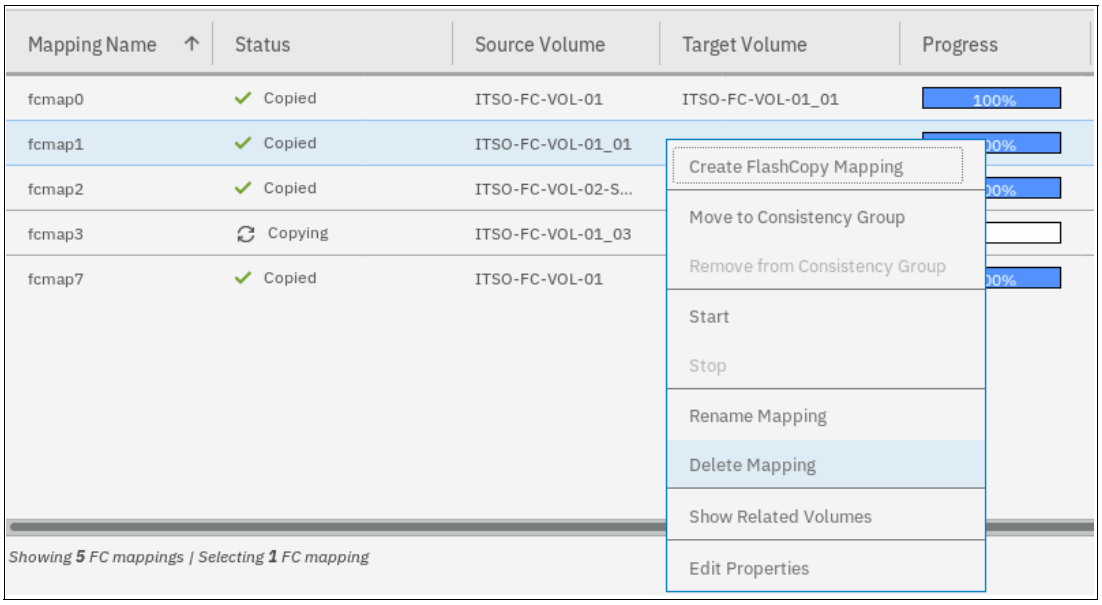


Figure 10-55 Deleting FlashCopy mappings

3. The Delete FlashCopy Mapping window opens, as shown in Figure 10-56. 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.

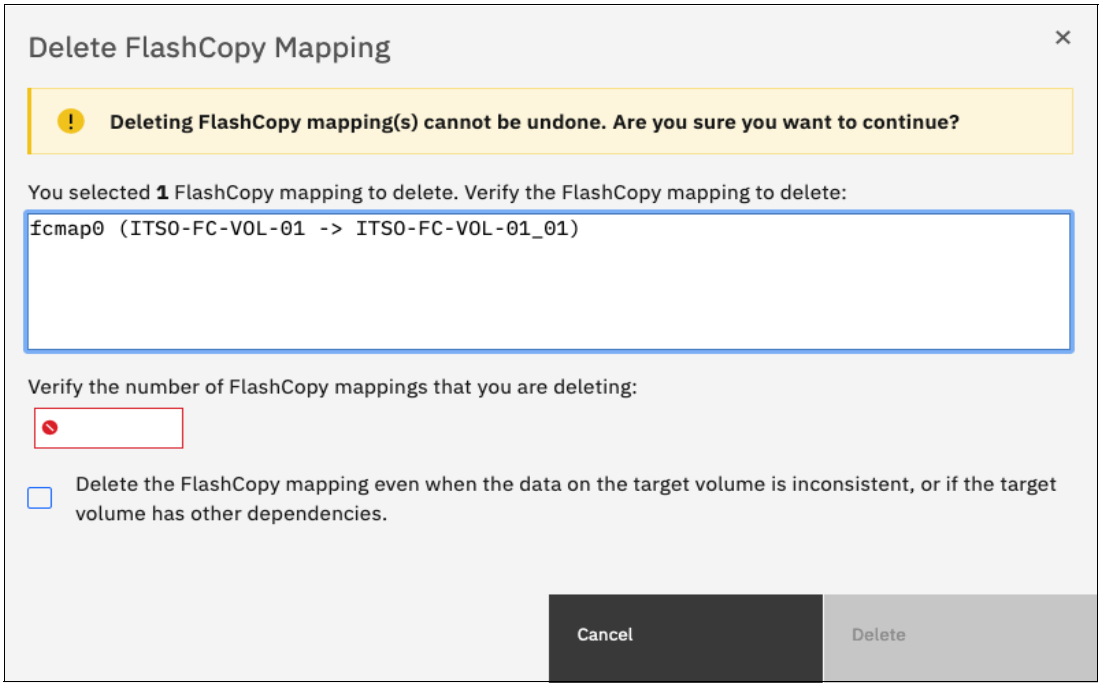


Figure 10-56 Confirming the selection of FlashCopy mappings to be deleted

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 10-57.

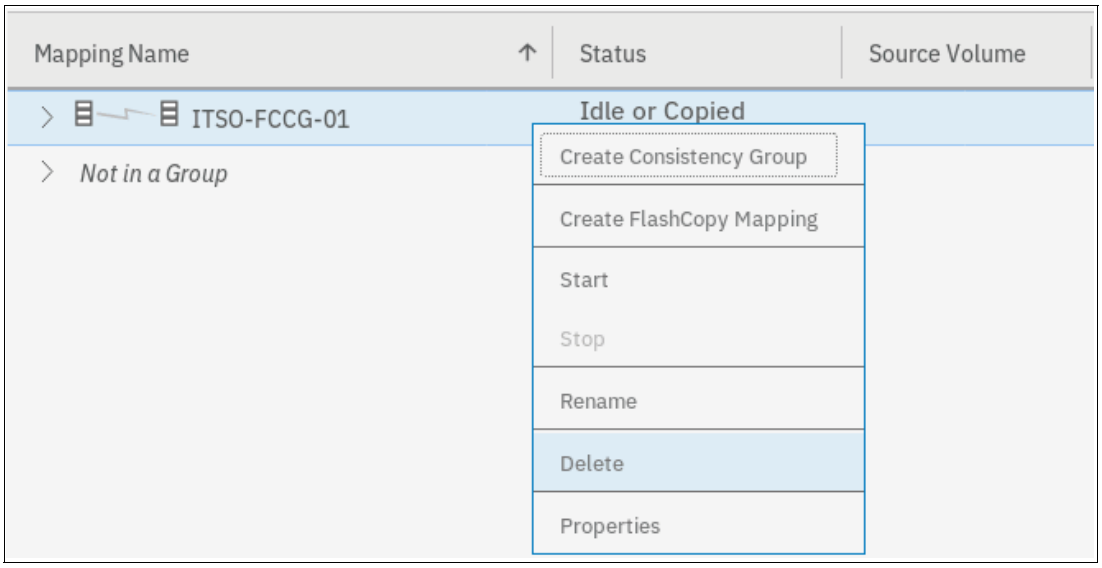


Figure 10-57 Deleting a consistency group

3. A warning message is displayed, as shown in Figure 10-58. Click **Yes**.

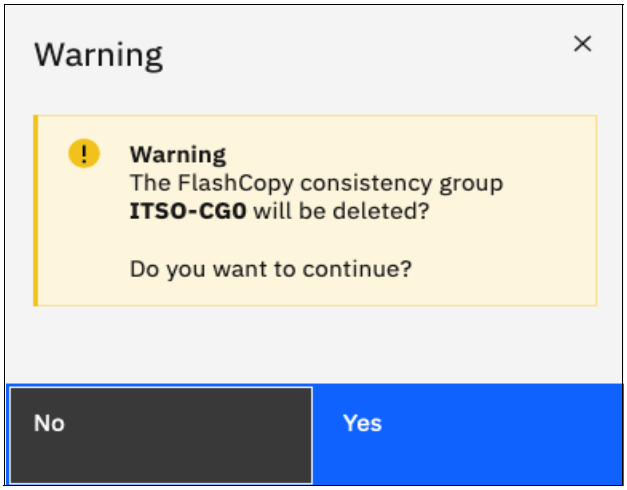


Figure 10-58 Confirming the consistency group deletion

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 10-59.

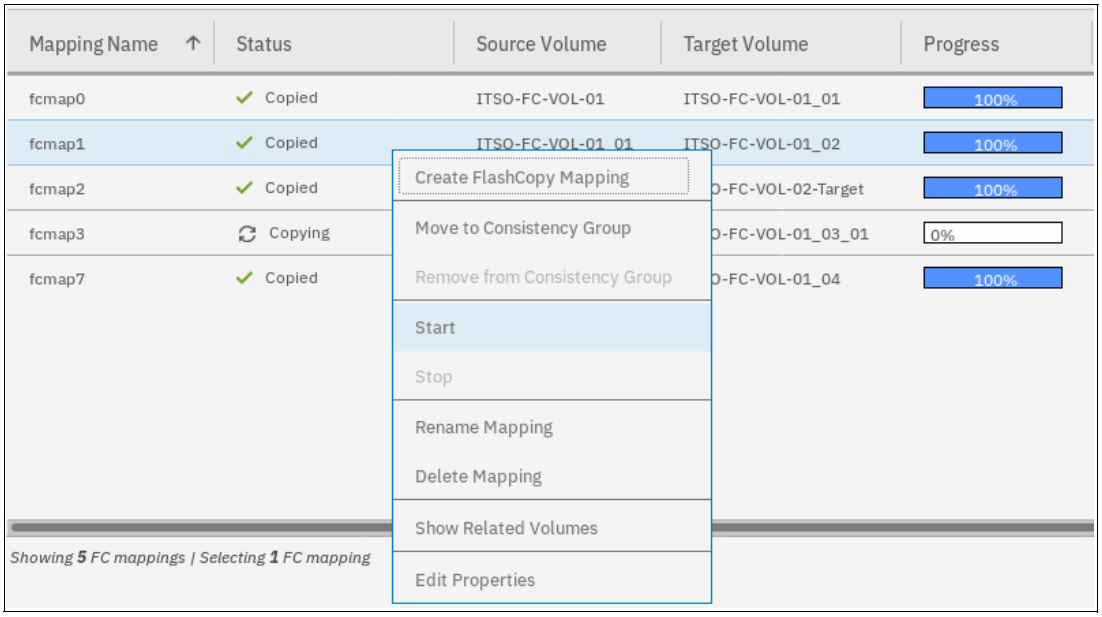


Figure 10-59 Starting FlashCopy mappings

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 10-60.

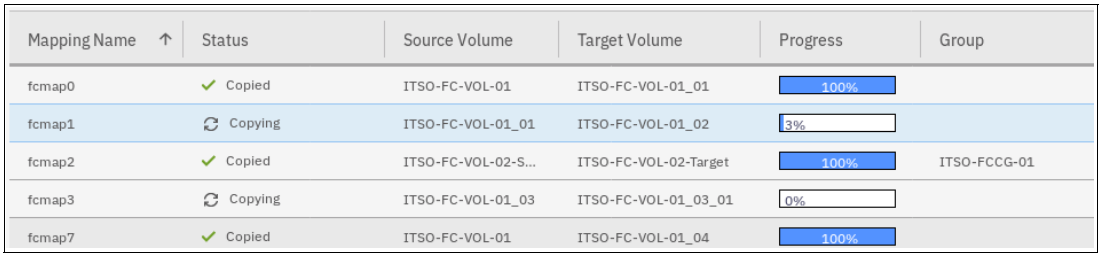


Figure 10-60 FlashCopy mappings status and progress examples

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 10.3.10, “Starting FlashCopy mappings and consistency groups” on page 623.

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 10-61.

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 10-61 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 10-62.

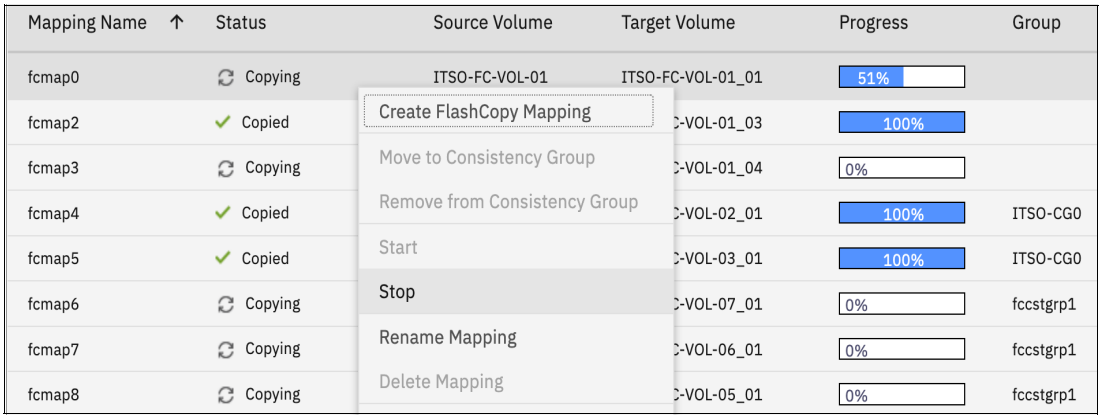


Figure 10-62 Stopping FlashCopy mappings

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 process in a multiple target FlashCopy: Cleaning Mode” on page 629.

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 Spectrum 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 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 Spectrum Virtualize system by selecting **Settings** → **System** and clicking the **Resources** menu, as shown in Figure 10-63 on page 678. This value can also be adjusted using the `chiogrp` CLI command.

For more information about the syntax, see IBM Documentation.

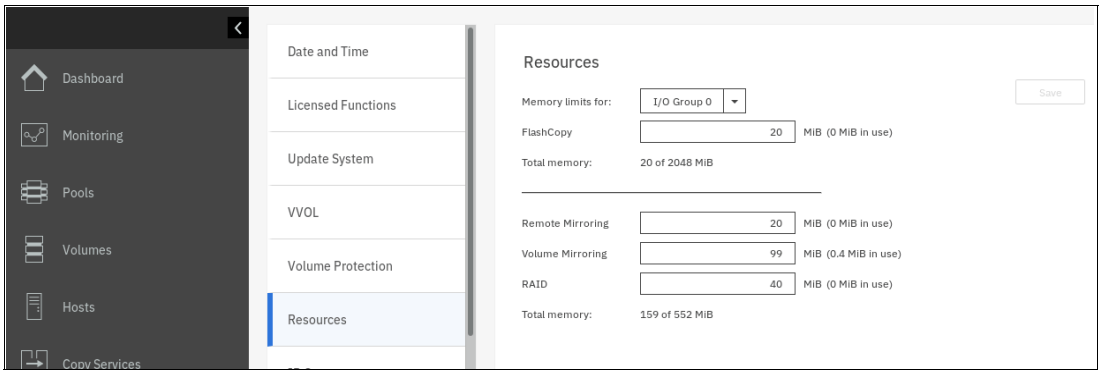


Figure 10-63 Modifying resources allocation per I/O group

10.5 Transparent Cloud Tiering

Introduced in V7.8, *Transparent Cloud Tiering* (TCT) is a function of IBM Spectrum 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 Spectrum 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.

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 Spectrum 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 Spectrum 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 about the IBM Cloud Object Storage portfolio, see this web page.

Demonstration: The IBM Client Demonstration Center has a demonstration available at this web page (log in required).

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
- Creating a consistent snapshot of 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
- Creating a snapshot of production data sets for 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 Spectrum Virtualize can provide opportunities to manage the exponential data growth and to manipulate data at low cost.

Today, many CSPs offers several *storage-as-services* solutions, such as content repository, backup, and archive. Combining all of these services, your IBM Spectrum 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 backup and data migration solution. In certain conditions, can be easily applied to eliminate the downtime that is associated with the needs to import and export data.

When TCT is applied as your backup strategy, IBM Spectrum 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 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 Spectrum Virtualize requires interaction with the correct architecture and specific properties. Conversely, CSPs offer attractive prices for Object Storage in 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 Spectrum 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 new 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 Spectrum 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.

Using your IBM Spectrum Virtualize management GUI, click **Settings** → **Network** → **DNS** and insert your DNS Internet Protocol Version 4 (IPv4) or Internet Protocol Version 6 (IPv6). 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 10-64.

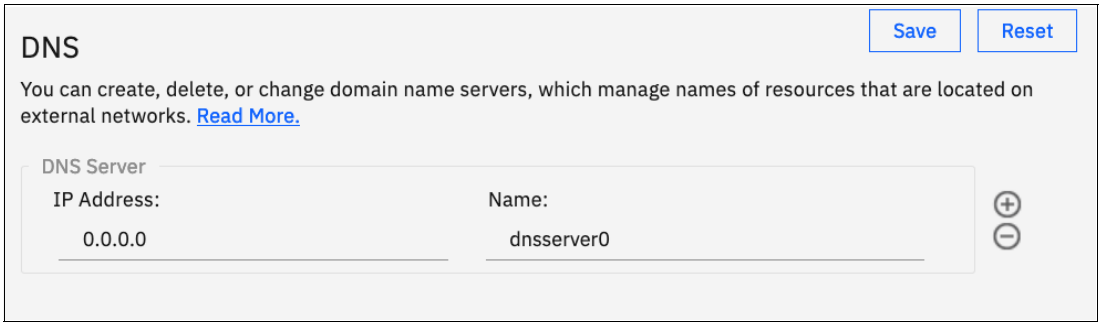


Figure 10-64 DNS settings

10.6.2 Enabling Transparent Cloud Tiering

After you complete the DNS settings, you can enable the TCT function in your IBM Spectrum Virtualize system by completing the following steps:

1. Using the IBM Spectrum Virtualize GUI, click **Settings** → **System** → **Transparent Cloud Tiering** and then, click **Enable Cloud Connection**, as shown in Figure 10-65. The TCT wizard starts and shows the welcome warning.

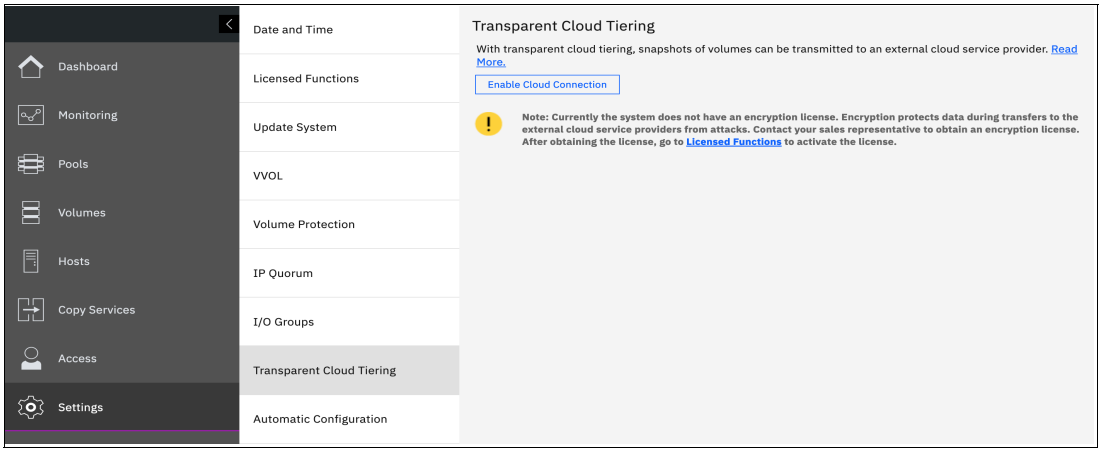


Figure 10-65 Enabling Cloud Tiering

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 purposes of this writing, our system does not have encryption enabled.

2. Click **Next** to continue. You must select one of three CSPs:
   - IBM Cloud
   - OpenStack Swift
   - Amazon S3

   Figure 10-66 shows the available options.

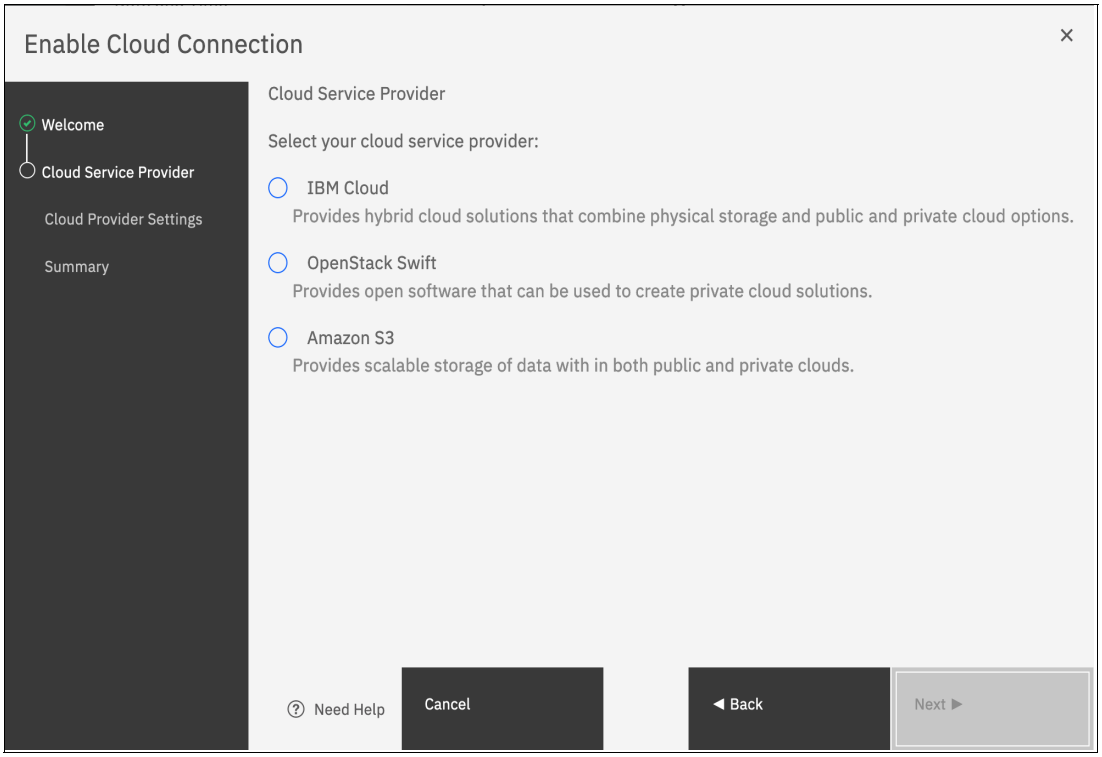


Figure 10-66 Selecting cloud service provider

3. 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 10-67 on page 683.

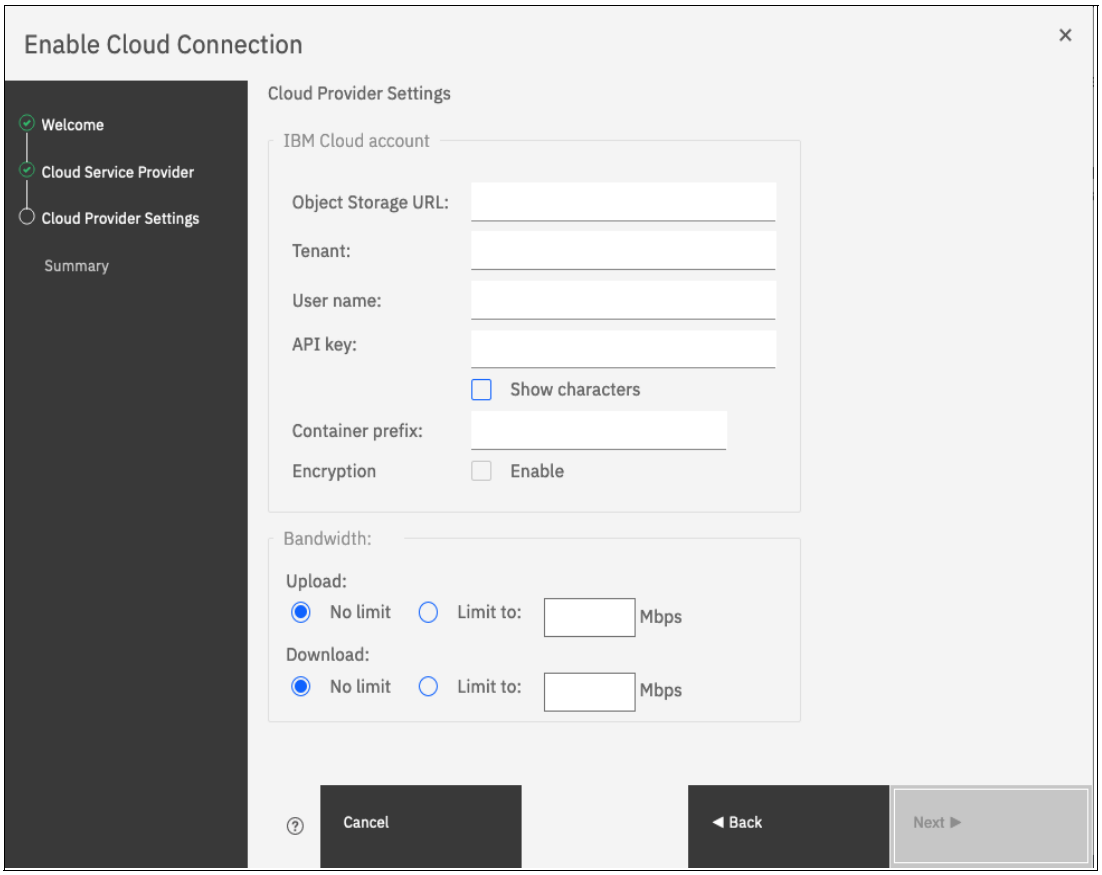


Figure 10-67 Entering cloud service provider information

4. Review your settings and click **Finish**, as shown in Figure 10-68.

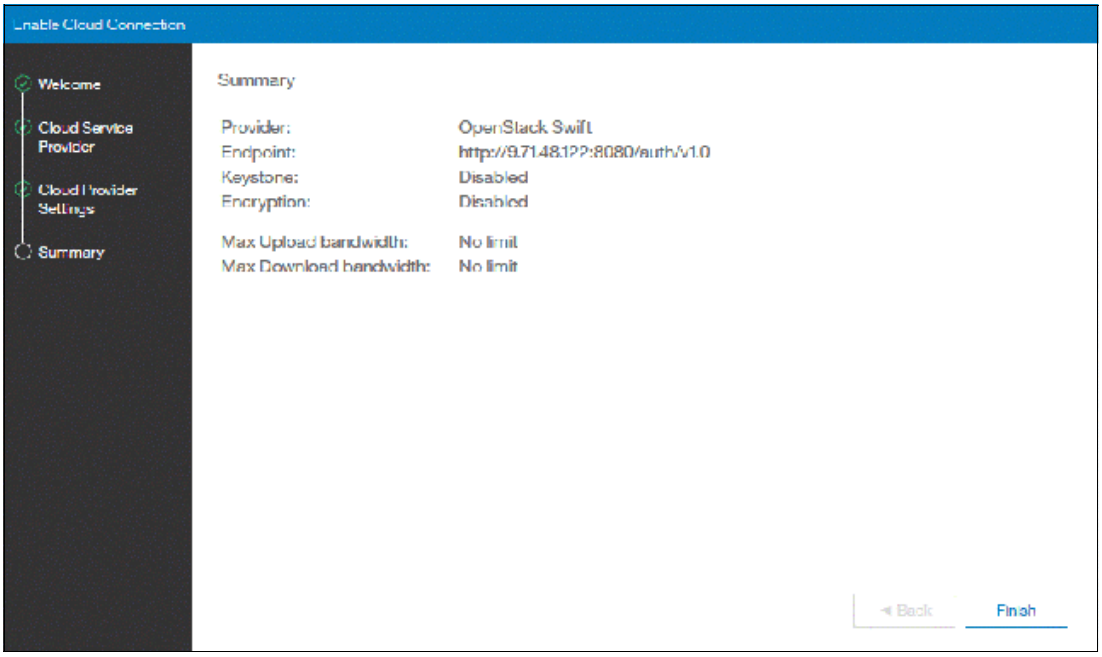


Figure 10-68 Cloud Connection summary

5. The cloud credentials can be viewed and updated at any time by using the function icons in 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 10-69.

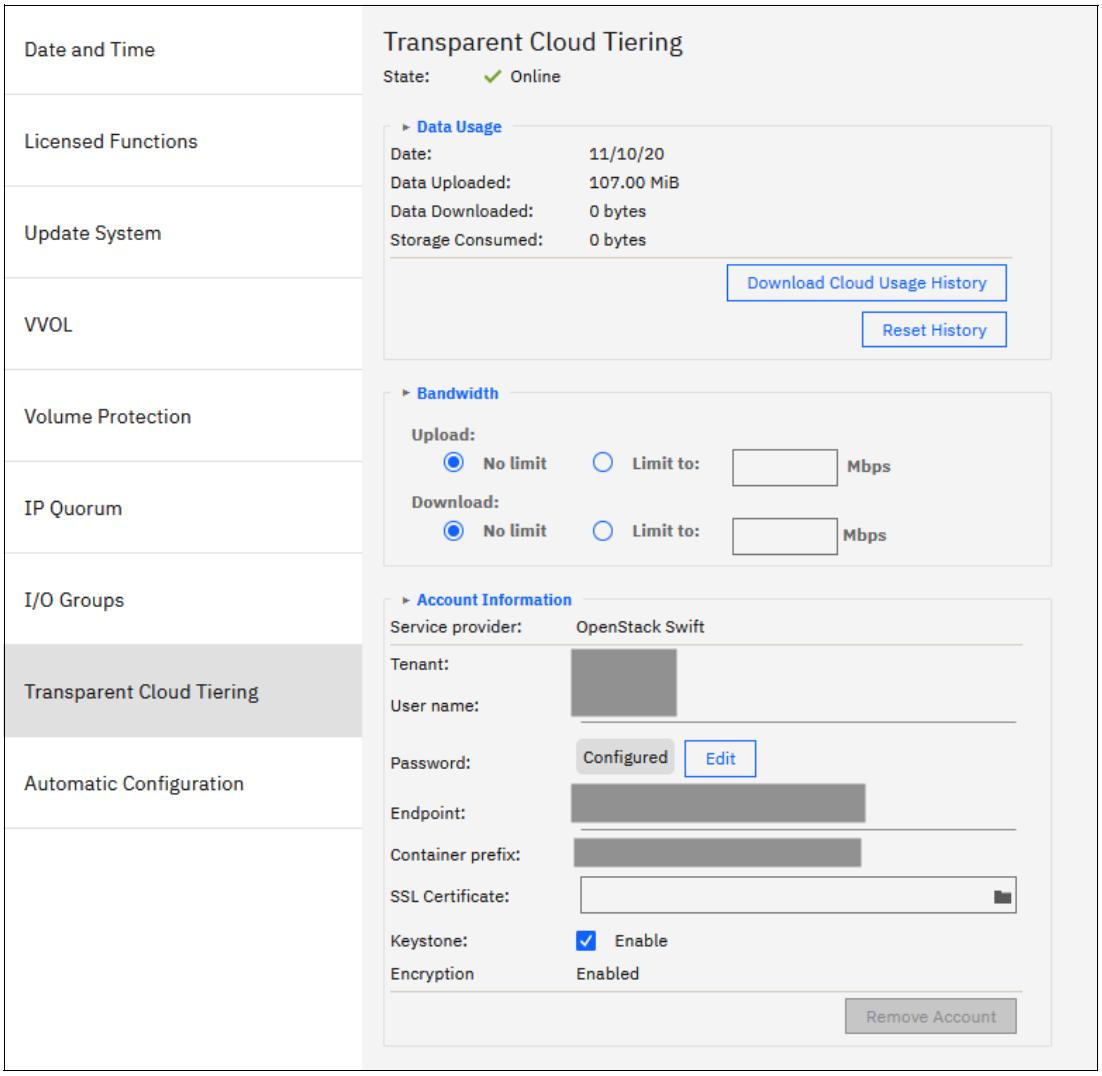


Figure 10-69 Enabled Transparent Cloud Tiering window

10.6.3 Creating cloud snapshots

To manage the cloud snapshots, the IBM Spectrum 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 680, 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 10-70.

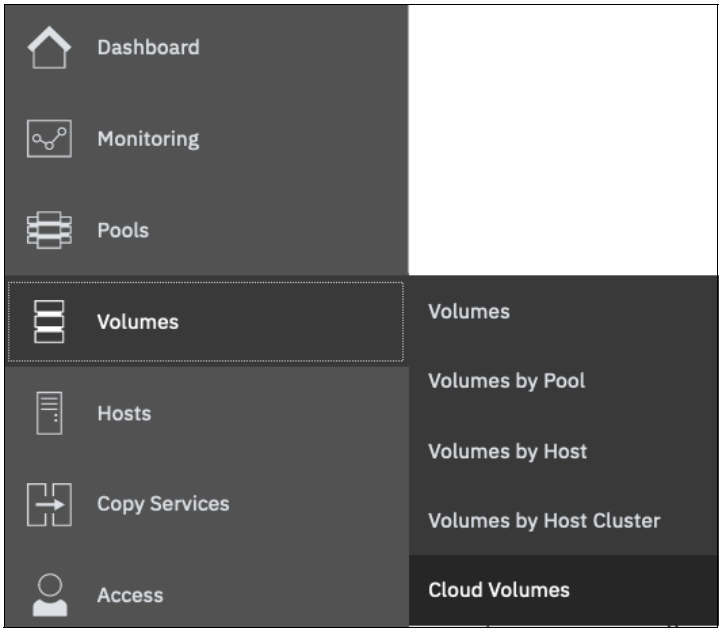


Figure 10-70 Cloud volumes menu

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 10-71.



Figure 10-71 Cloud volumes window

3. Click **Add Volumes** to enable cloud-snapshot on volumes. A new window opens, as shown in Figure 10-72. Select the volumes that you want to enable Cloud Tiering for and click **Next**.

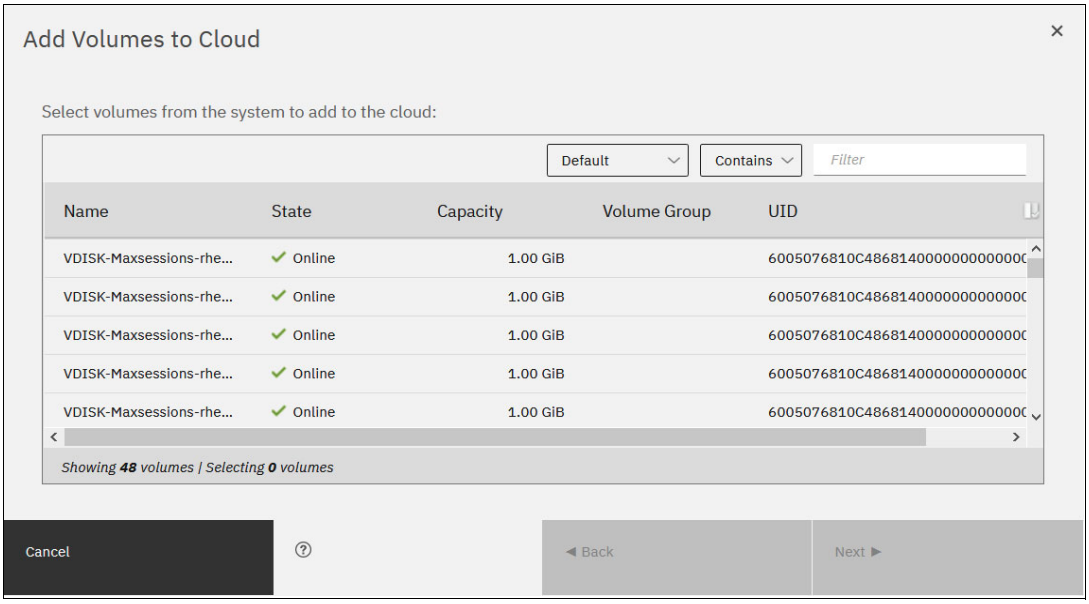


Figure 10-72 Adding volumes to Cloud Tiering

4. IBM Spectrum 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 10-73. You can also select the second option, even if another full copy of the volume exists.

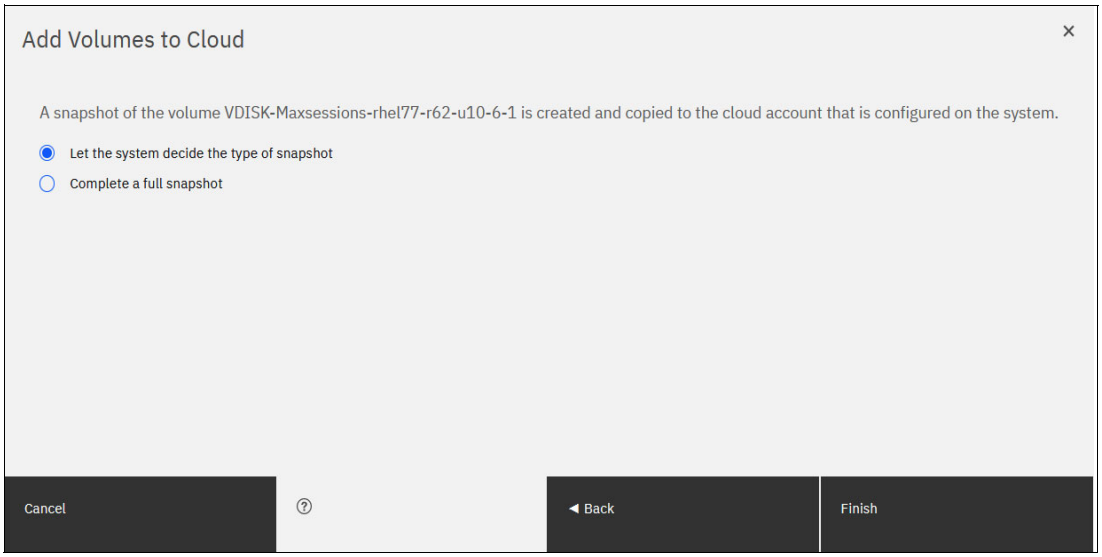


Figure 10-73 Selecting if a full copy is made or if the system decides

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 assigned by the IBM Spectrum 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 10-74 shows an example of a Cloud Volumes list.

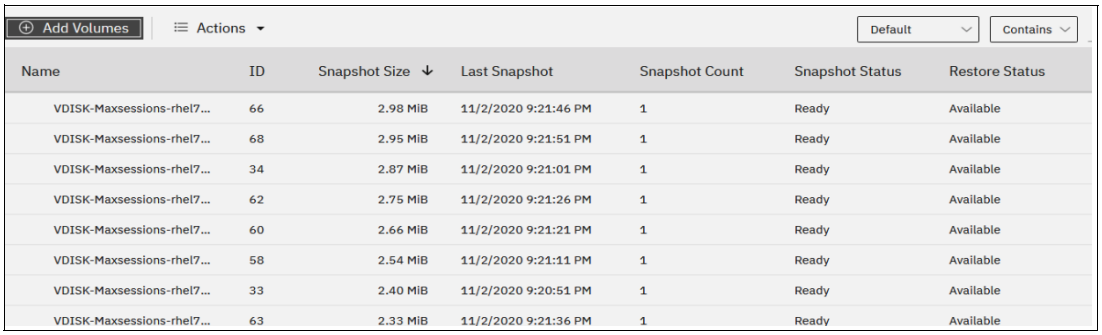


Figure 10-74 Cloud Volumes list example

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 10-75.

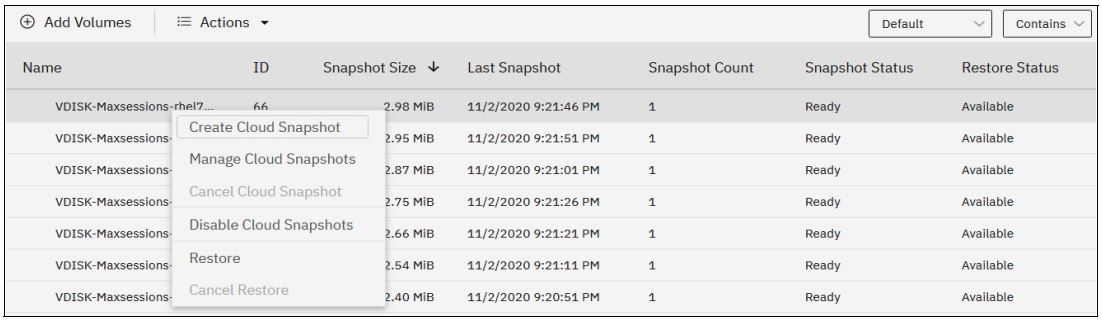


Figure 10-75 Available actions in Cloud Volumes window

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 list of PiT copies appear and provide details about their status, type, and snapshot date, as shown in Figure 10-76.

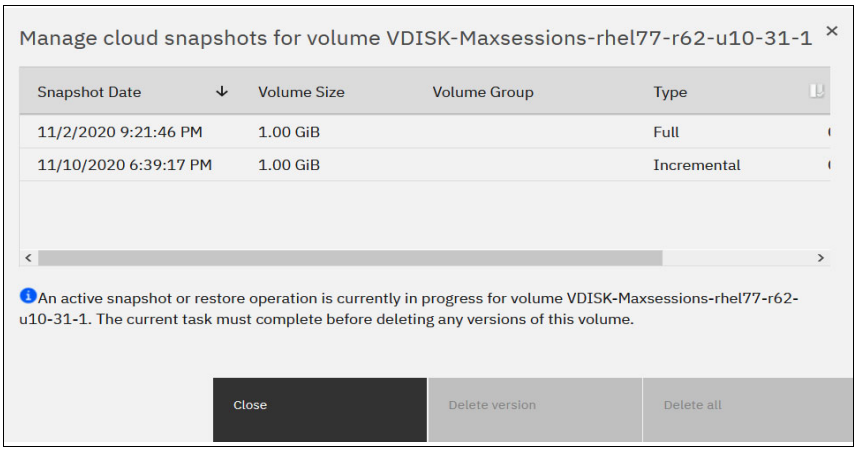


Figure 10-76 Deleting versions of a volume's snapshots

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 Spectrum 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 Spectrum 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 10-77.

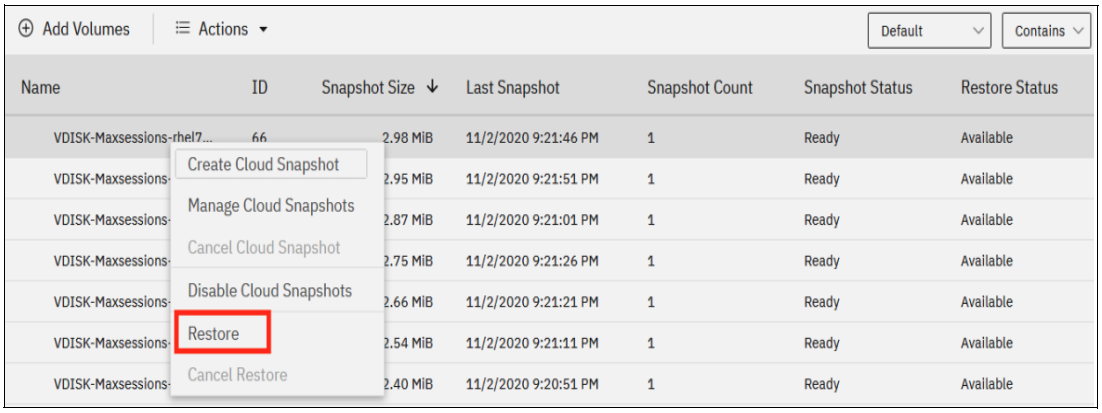


Figure 10-77 Selecting a volume to restore a snapshot from

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 10-78). Select the version that you want to restore and then, click **Next**.

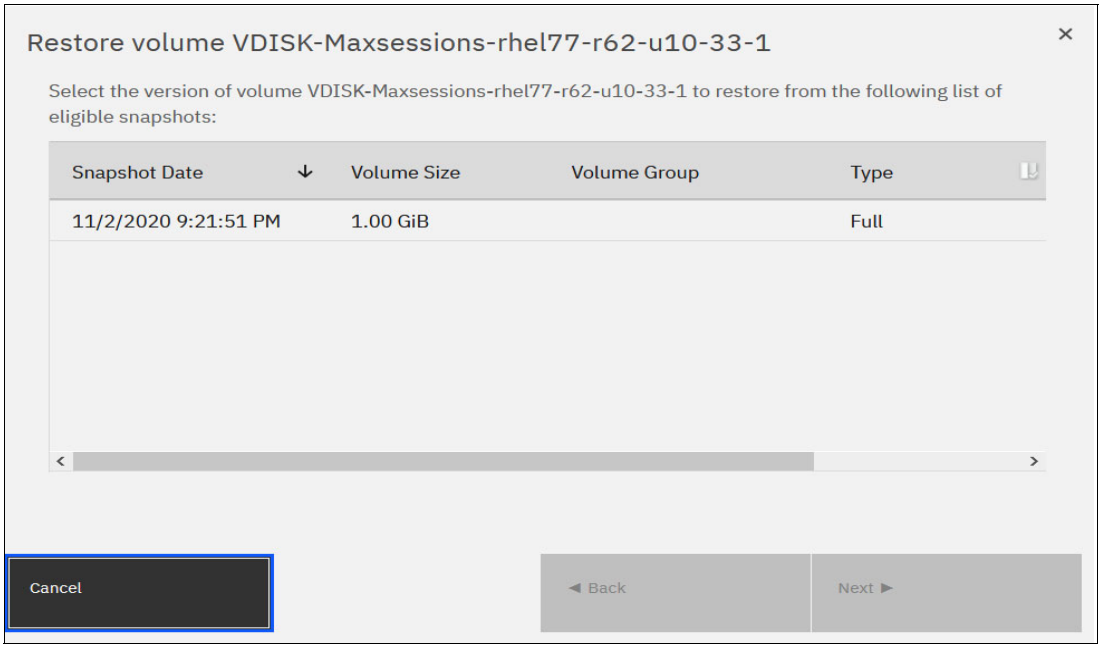


Figure 10-78 Selecting a snapshot version to restore

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 Spectrum 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 10-79. Make a selection and click **Next**.

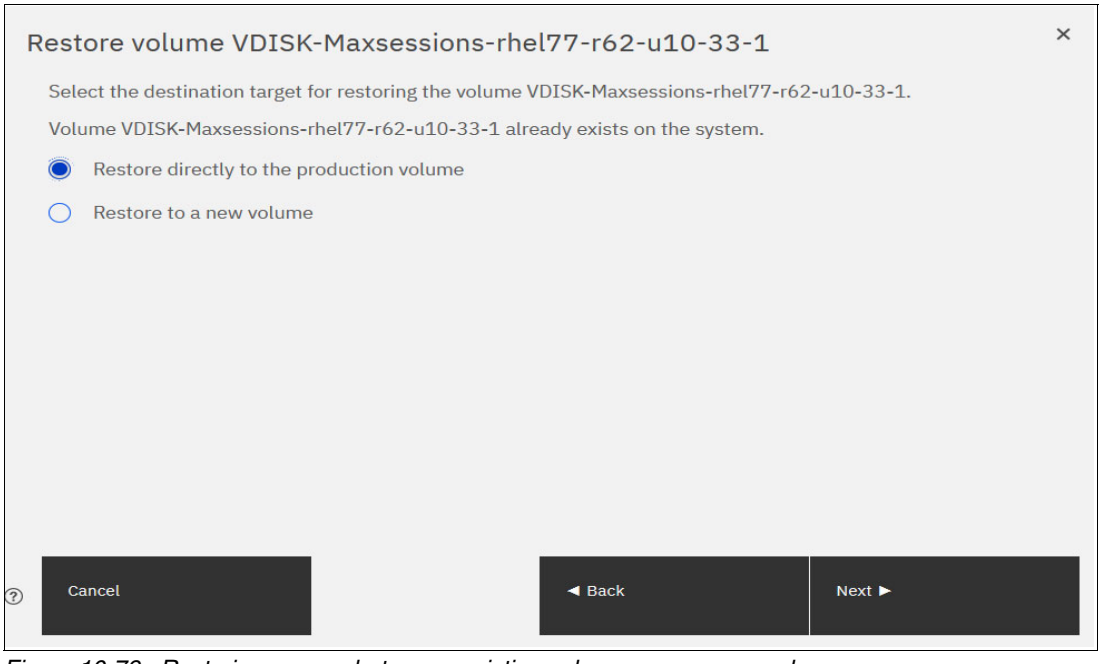


Figure 10-79 Restoring a snapshot on an existing volume or on a new volume

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 are 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 10-80:

   – 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**.

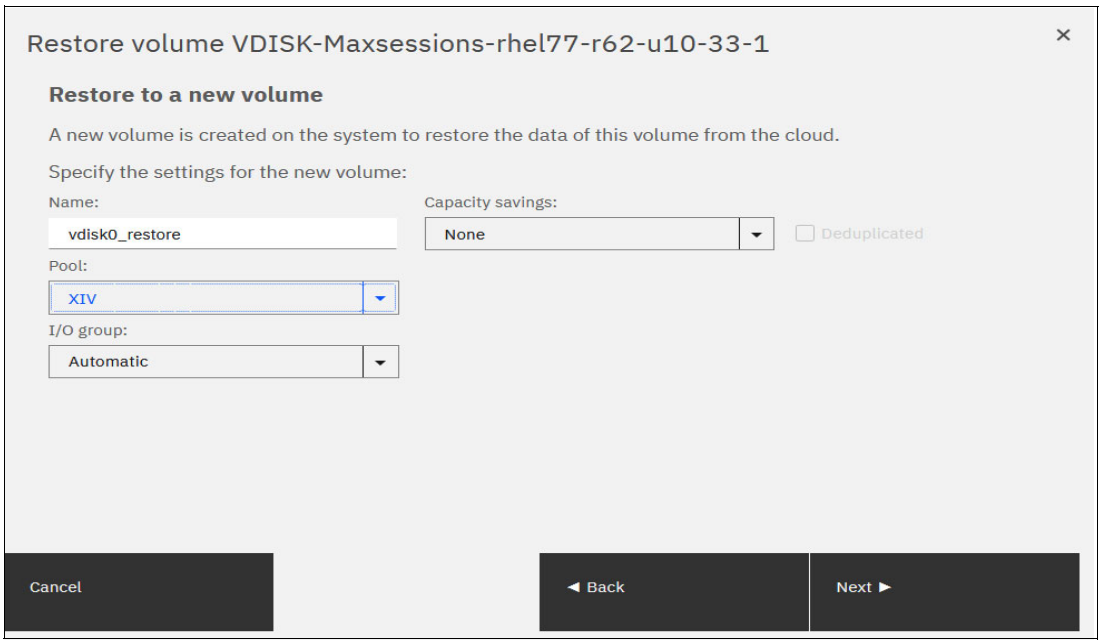


Figure 10-80 Restoring a snapshot to a new volume

6. A Summary window is displayed so you can review the restoration settings, as shown in Figure 10-81 on page 691. 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.

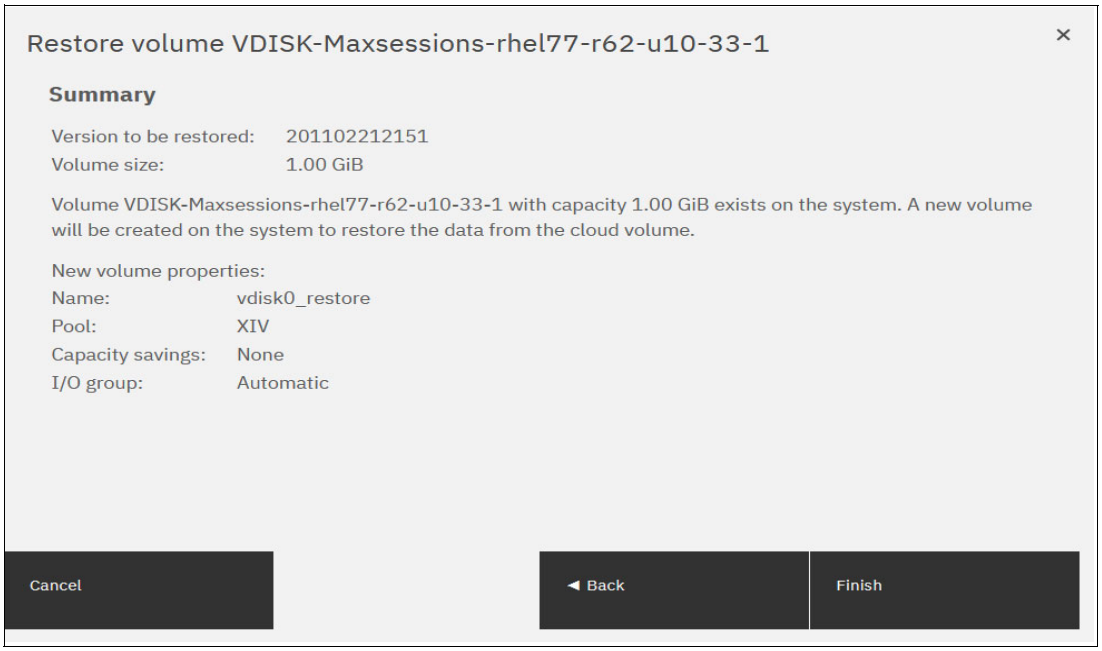


Figure 10-81 Restoring a snapshot summary

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 also is 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 Spectrum 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 Spectrum 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 Spectrum 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, “Creating and provisioning volumes” on page 331.

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 Spectrum Virtualize system is similar to RC in the IBM System Storage DS8000 family at a functional level, but the implementation differs.

IBM Spectrum 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 Spectrum 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 Spectrum 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 are used, a certain 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 Spectrum Virtualize system 1	IBM Spectrum 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 FlashSystems system layers

An IBM Spectrum 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 Spectrum Virtualize based systems. The IBM SAN Volume Controller is always set to replication layer. This parameter *cannot* be changed.

In the storage layer, an IBM FlashSystems 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 FlashSystems 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 FlashSystems 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 FlashSystems 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 Spectrum Virtualized based system can exist as a backend or host entity.
- No system partnerships can exist.
- No other IBM Spectrum 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 10-2.

Example 10-2 Output from the lssystem command showing the system layer

```
IBM_SVC1:superuser>lssystem
id 0000020421400428
name FS9110
...
code_level 8.4.2.0 (build 154.20.2109031944000)
...
layer storage
...
```

Note: Consider the following rules for creating remote partnerships between the IBM SAN Volume Controller and IBM FlashSystems Family systems:

- An IBM SAN Volume Controller is always in the replication layer.
- By default, the IBM FlashSystems 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 Spectrum Virtualize based system in the replication layer can virtualize an IBM FlashSystems system in the storage layer.

10.8.2 Multiple IBM Spectrum Virtualize systems replication

Each IBM Spectrum 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 10.10.3, “IP partnership limitations” on page 732.

Figure 10-82 shows an example of a multiple systems mirroring configuration.

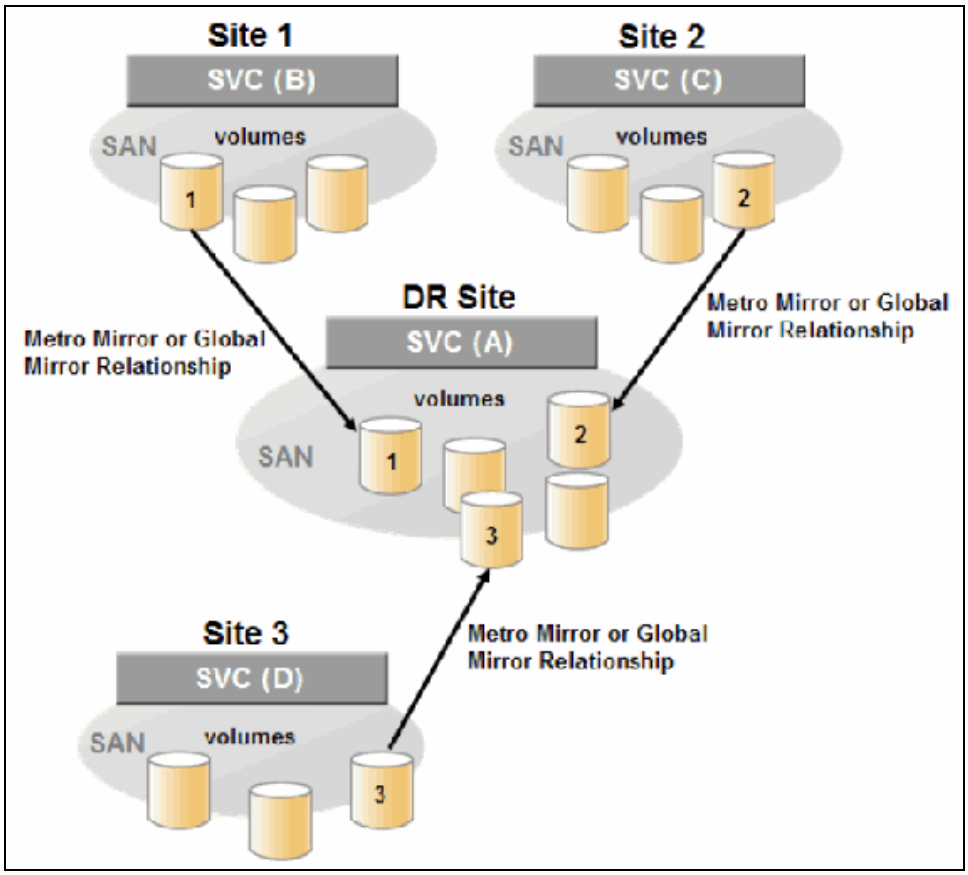


Figure 10-82 Multiple systems mirroring configuration example

Supported multiple systems mirroring topologies

Multiple systems mirroring supports various partnership topologies, as shown in the example in Figure 10-83. This example is a star topology (A → B, A → C, and A → D).

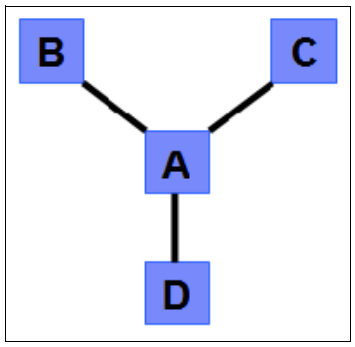


Figure 10-83 Star topology

Figure 10-83 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 10-84 shows an example of a triangle topology (A → B, A → C, and B → C).

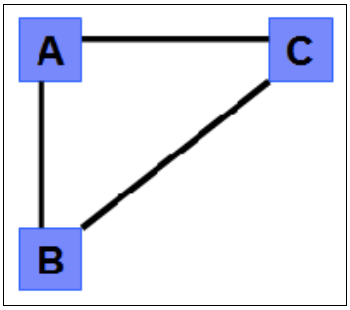


Figure 10-84 Triangle topology

Figure 10-85 shows an example of an IBM Spectrum Virtualize system fully connected topology (A → B, A → C, A → D, B → D, and C → D).

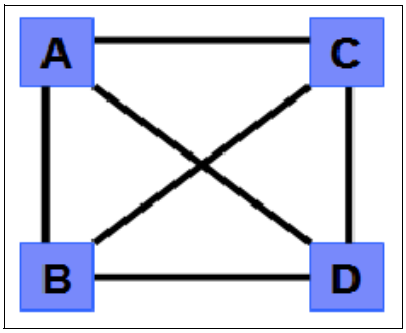


Figure 10-85 Fully connected topology

Figure 10-85 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 10-86 shows a daisy-chain topology.

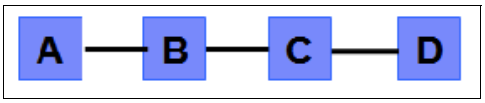


Figure 10-86 Daisy-chain topology

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 FlashSystems if the IBM FlashSystems system is set to the replication layer.

With the IBM Spectrum Virtualize code version 8.3.1, a three-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 1.6.4, “Business continuity with three-site replication” on page 51, or *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597.

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 Spectrum Virtualize MM and GM implementation operates 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 10.3.13, “FlashCopy and image mode volumes” on page 632.

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 Spectrum 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 10-87 shows the concept of RC consistency groups.

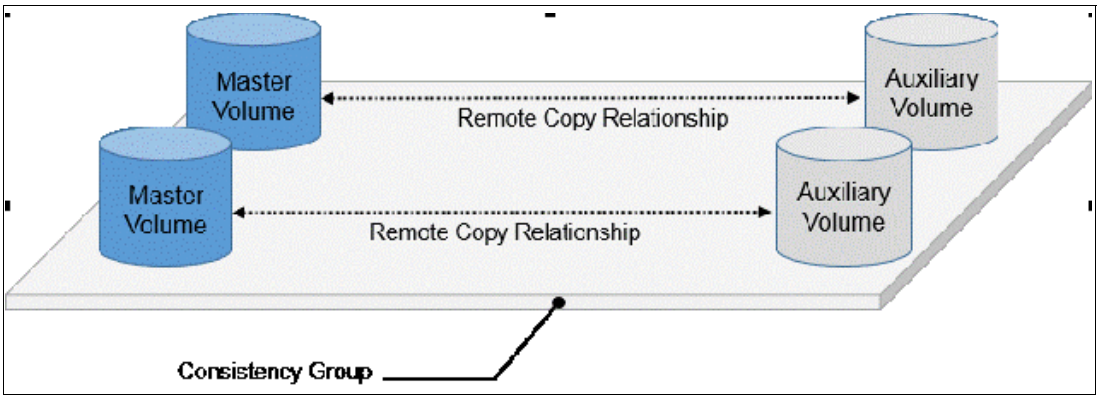


Figure 10-87 Remote Copy consistency group

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 10.10.7, “States of IP partnership” on page 736.

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 Spectrum 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 Spectrum 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 FlashSystems 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 Spectrum 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 only to 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.

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 Spectrum 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 Spectrum 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 10-88 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.

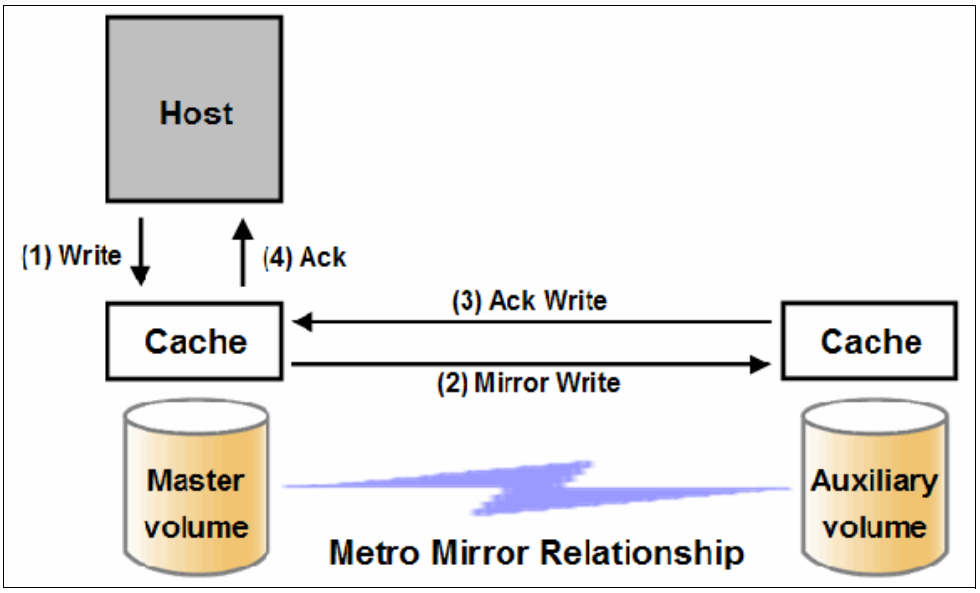


Figure 10-88 Write on volume in Metro Mirror relationship

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 Spectrum Virtualize does not support more than 300 km (186.4 miles).

10.8.7 Metro Mirror features

The IBM Spectrum 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 Spectrum 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 Spectrum 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 Spectrum Virtualize systems that are configured in a partnership.

- Intercluster and intracluster MM can be used concurrently.
- IBM Spectrum 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 Spectrum 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 Spectrum 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 Spectrum Virtualize features the following attributes:

- A partnership is created between two IBM Spectrum 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 Spectrum 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 Spectrum 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

Although GM is available for intracluster, it 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 Spectrum Virtualize system.

Intercluster Global Mirror

Intercluster GM operations require a pair of IBM Spectrum 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 10-89 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.

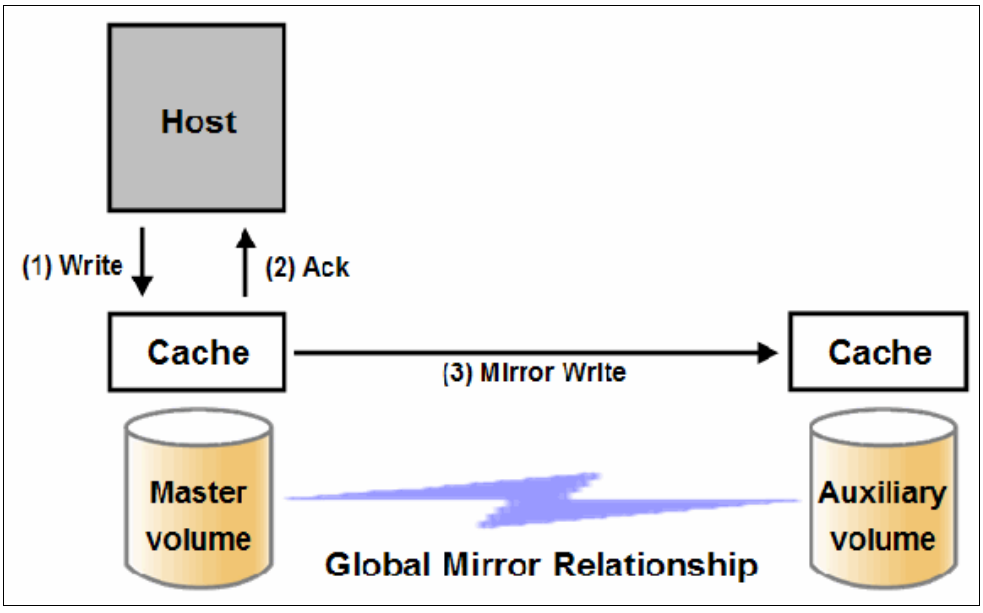


Figure 10-89 Global Mirror write sequence

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 write 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 Spectrum 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 Spectrum 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**
SV1	250 ms	80 ms	10 ms
SV2	250 ms	80 ms	10 ms
SA2	250 ms	80 ms	10 ms

10.8.12 Global Mirror features

IBM Spectrum Virtualize GM supports the following features:

- Asynchronous RC of volumes that are dispersed over metropolitan-scale distances.
- IBM Spectrum Virtualize implements the GM relationship between a volume pair, with each volume in the pair being managed by an IBM Spectrum Virtualize system.
- IBM Spectrum Virtualize supports intracluster GM where both volumes belong to the same system (and I/O group).

- An IBM Spectrum Virtualize system can be configured for partnership with 1 - 3 other systems. For more information about IP partnership restrictions, see 10.10.3, “IP partnership limitations” on page 732.
- Intercluster and intracluster GM can be used concurrently, but not for the same volume.
- IBM Spectrum 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 Spectrum 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 Spectrum 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 10-90 shows a colliding write sequence example.

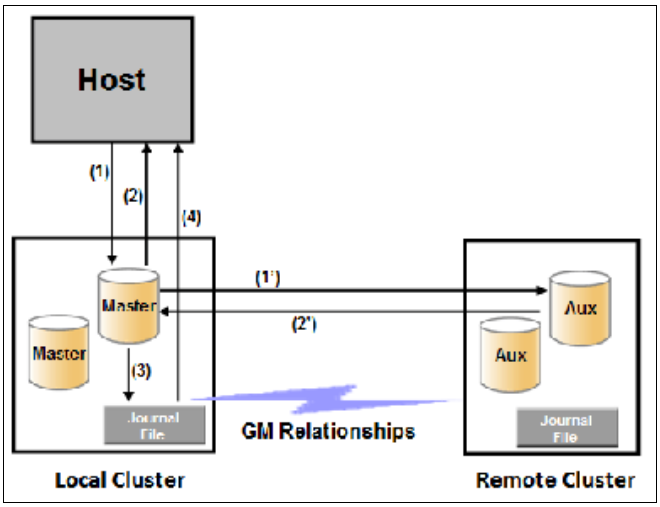


Figure 10-90 Colliding writes example

The following numbers correspond to the numbers that are shown in Figure 10-90:

- (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 has functionality that is designed to 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 10-91 shows a simple Global Mirror relationship without change volumes.

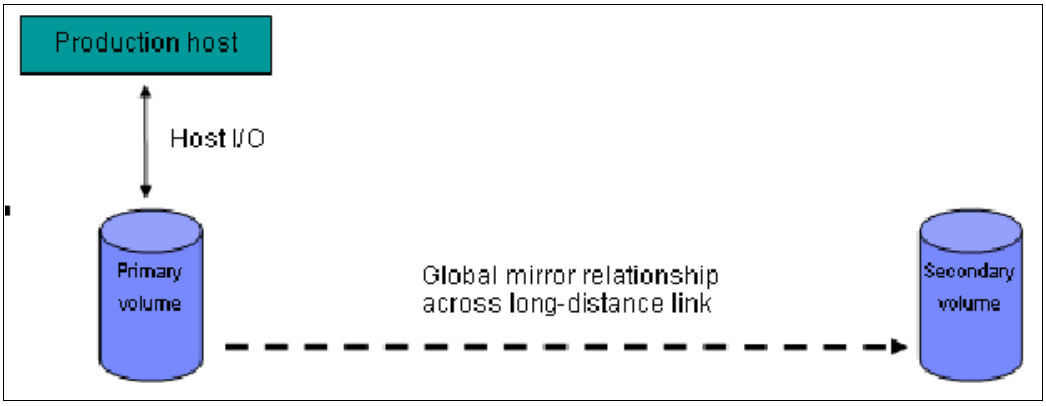


Figure 10-91 Global Mirror without change volumes

With change volumes, this environment looks as it is shown in Figure 10-92.

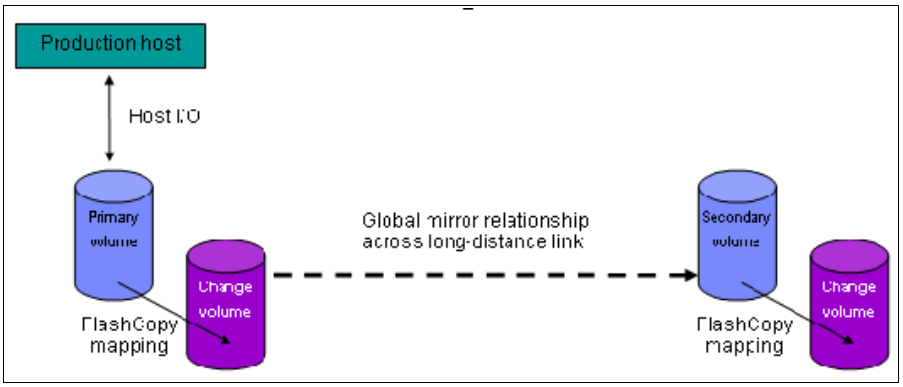


Figure 10-92 Global Mirror with Change Volumes

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, “IBM Safeguarded Copy” on page 607.

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 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 Spectrum 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 703. 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 also can be a reason to prefer asynchronous GM over synchronous MM. Otherwise, the use cases for MM/GM are the same.

10.8.19 IBM Spectrum Virtualize HyperSwap topology

The IBM HyperSwap topology is based on IBM Spectrum 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 Volume size per I/O group	A per I/O group limit of 1024 TB 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 512 MiB of bitmap space for the I/O group and allows 10 MiB of space for all remaining copy services features.

10.8.23 Remote Copy states and events

This section describes the various states of a MM/GM relationship and the conditions that cause them to change. In Figure 10-93 shows an overview of the status that can apply to a MM/GM relationship in a connected state.

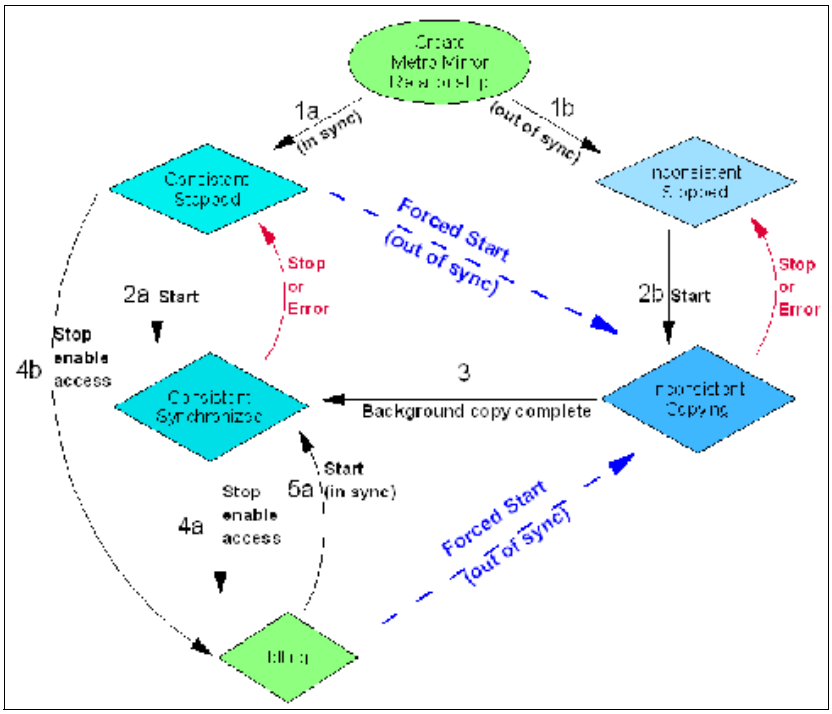


Figure 10-93 Metro Mirror or Global Mirror mapping state diagram

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 10-93:

- 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 716.

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 spans 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 Spectrum 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 10.3.13, “FlashCopy and image mode volumes” on page 632.

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 713.

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.

10.9 Remote Copy commands

This section presents commands that must be issued to create and operate RC services.

10.9.1 Remote Copy process

The MM/GM process includes the following steps:

1. A system partnership is created between two IBM Spectrum 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:

- Commands to create, delete, and manipulate relationships and consistency group
- Commands to 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.

10.9.2 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.

Note: This command is not supported on IP partnerships. Use **mkippartnership** for IP connections.

10.9.3 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 the IBM Spectrum Control, to help identify and resolve the problem.

10.9.4 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 an 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/GM relationships between volumes on the IBM Spectrum 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 200 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.

10.9.5 Creating a Metro Mirror/Global Mirror consistency group

Run the **`mkrcconsistgrp`** command to create an empty MM/GM consistency group.

The MM/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.

10.9.6 Creating a Metro Mirror/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.

10.9.7 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.

10.9.8 Changing Metro Mirror/Global Mirror consistency group

Run the **chrcconsistgrp** command to change the name of an MM/GM consistency group.

10.9.9 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.

10.9.10 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.

10.9.11 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.

10.9.12 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.

10.9.13 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.

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.

10.9.14 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.

10.9.15 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.

10.9.16 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.

10.10 Native IP replication

IBM Spectrum Virtualize can implement RC services by using FC connections or IP connections. This section describes the IBM Spectrum Virtualize IP replication technology and implementation.

Demonstration: The IBM Client Demonstration Center shows how data is replicated by using GMCV (cycling mode set to `multiple`). This configuration perfectly fits the new IP replication functionality because it is well-designed for links with high latency, low bandwidth, or both.

For more information, see this web page (log in required).

10.10.1 Enhancements

IBM Spectrum Virtualize 8.4.2 includes 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

Some of the new features are described next. For more information about the new features, see this IBM Documentation web page.

Portsets

Portsets group 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 10-94 shows the new administration model for creating portsets and configuring IP replication.

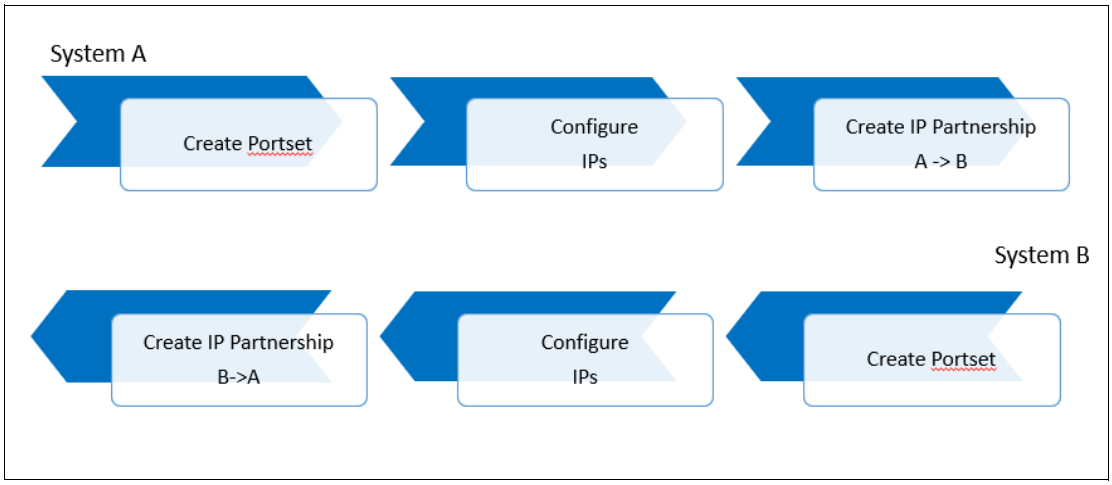


Figure 10-94 New administration model for creating portsets and configuring IP replication

Spectrum Virtualize 8.4.2 now allows up to three IP Replication partnerships per system (the previous limit was one in code versions earlier than 8.4.2). Use cases include three-site replication and having DR both at a remote site or in the cloud.

Example 10-3 shows the new IP replication commands.

Example 10-3 New IP replication 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
```

10.10.2 Native IP replication technology

Remote Mirroring over IP communication is supported on the IBM SAN Volume Controller and IBM FlashSystems Family systems by using Ethernet communication links. The IBM Spectrum 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 Spectrum 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 10-95.

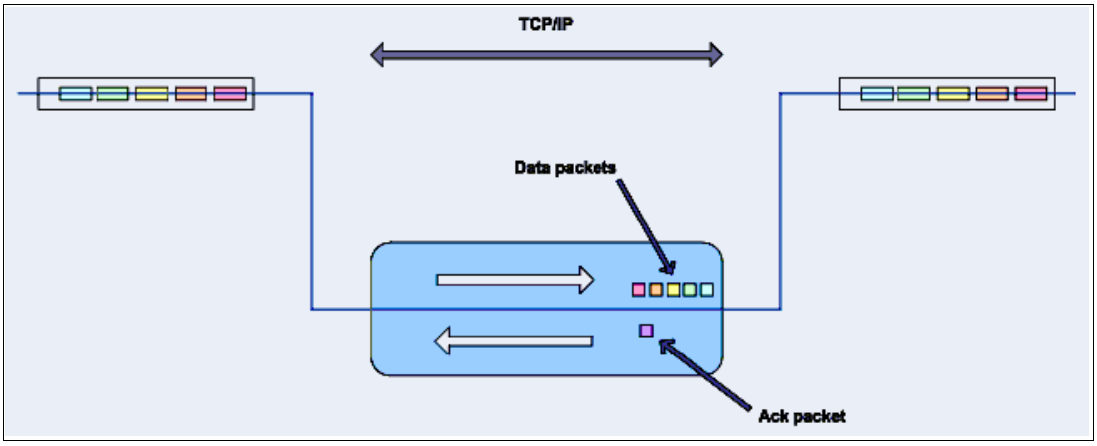


Figure 10-95 Typical Ethernet network data flow

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 10-96.

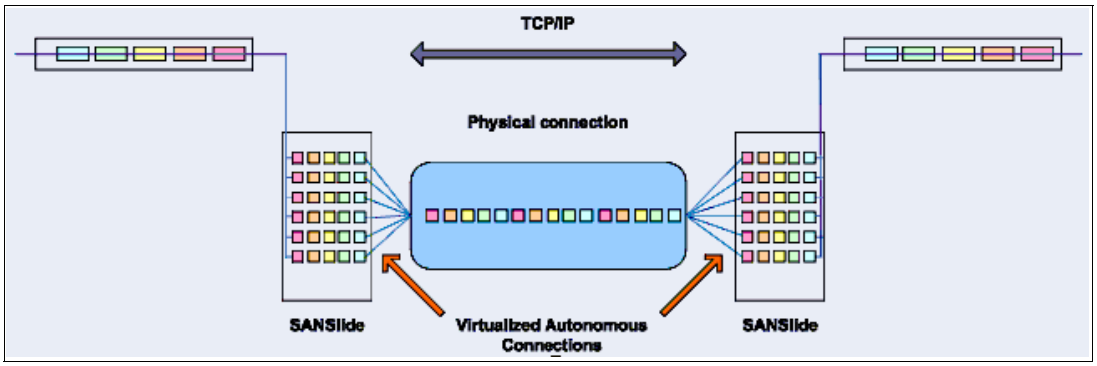


Figure 10-96 Optimized network data flow by using Bridgeworks SANSlide technology

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.

10.10.3 IP partnership limitations

The following prerequisites and assumptions must be considered before IP partnership between two IBM Spectrum Virtualize systems can be established:

- The IBM Spectrum 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 Spectrum 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 (four systems total).
- Only a single partnership 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 10-97 shows scaling host I/O.

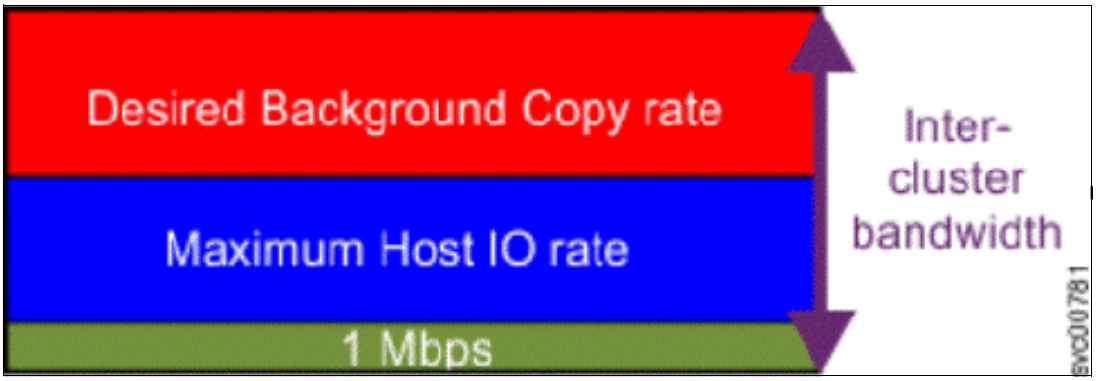


Figure 10-97 Scaling of host I/O

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.

10.10.4 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

10.10.5 VLAN support

Starting with 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 Spectrum 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.

10.10.6 IP partnership and terminology

The IP partnership terminology and abbreviations are listed in Table 10-15.

Table 10-15 Terminology for IP partnership

IP partnership terminology	Description
RC group or RC port group	The following numbers group a set of IP addresses that are connected to the same physical link. Therefore, only IP addresses that are part of the same RC group can form RC connections with the partner system: ► 0: Ports that are not configured for RC ► 1: Ports that belong to RC port group 1 ► 2: Ports that belong to RC port group 2 Each IP address can be shared for iSCSI host attach and RC functionality. Therefore, appropriate settings must be applied to each IP address.
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 Spectrum 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).

10.10.7 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.

State	Systems connected	Support for active RC I/O	Comments
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 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 **`cfgportip`** CLI command), and completes the following steps:
   a. Configure the IP addresses for RC data.
   b. Add the IP addresses in the respective RC port group.
   c. Define 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.

10.10.8 Remote Copy groups

This section describes Remote Copy (RC) groups (or RC port groups) and different ways to configure the links between the two remote systems. The two IBM Spectrum Virtualize systems can be connected to each other over one link or at most, two links. To address the requirement to enable the systems to know about the physical links between the two sites, the concept of RC port groups was introduced.

RC port group ID is a numerical tag that is associated with an IP port of an IBM Spectrum Virtualize system to indicate to which physical IP link it is connected. Multiple nodes might be connected to the same physical long-distance link, and must therefore share RC port group ID.

In scenarios with two physical links between the local and remote clusters, two RC port group IDs must be used to designate which IP addresses are connected to which physical link. This configuration must be done by the system administrator by using the GUI or running the **`cfgportip`** CLI command.

Remember: IP ports on both partners must be configured with identical RC port group IDs for the partnership to be established correctly.

The IBM Spectrum Virtualize system IP addresses that are connected to the same physical link are designated with identical RC port groups. The system supports three RC groups: 0, 1, and 2.

The systems' IP addresses are, by default, in RC port group 0. Ports in port group 0 are not considered for creating RC data paths between two systems. For partnerships to be established over IP links directly, IP ports must be configured in RC group 1 if a single inter-site link exists, or in RC groups 1 and 2 if two inter-site links exist.

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 a RC port group.

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 Spectrum Virtualize controller node must have only a single RC port group that is configured: 1 or 2. The remaining IP addresses must be in RC port group 0.

10.10.9 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 Spectrum 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 10-98 (configuration 1).

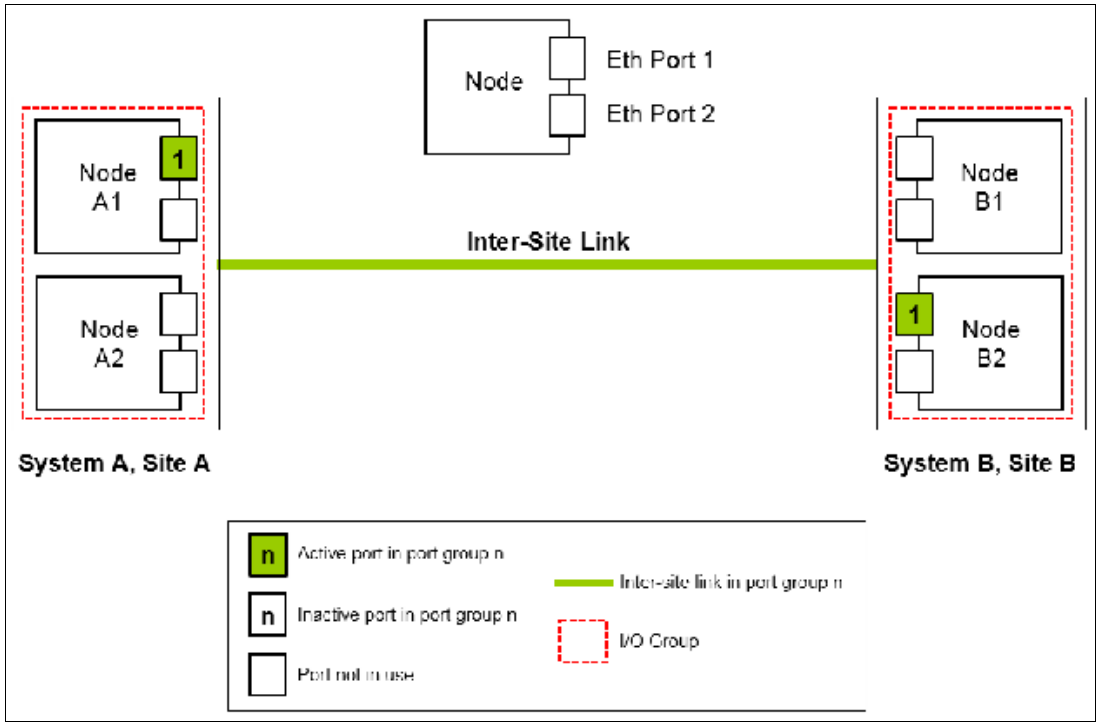


Figure 10-98 Single link with only one Remote Copy port group configured in each system

As shown in Figure 10-98, two systems are available:

- System A
- System B

A single RC port group 1 is created on Node A1 on System A and on Node B2 on System B because only a single inter-site link is used to facilitate the IP partnership traffic. An administrator might choose to configure the RC port group on Node B1 on System B rather than Node B2.

At any time, only the IP addresses that are configured in RC port group 1 on the nodes in System A and System B 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 a RC port group that is configured, and no failover ports are configured.
- If the Node A1 in System A or the Node B2 in System B 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 configured), as shown in Figure 10-99 (configuration 2).

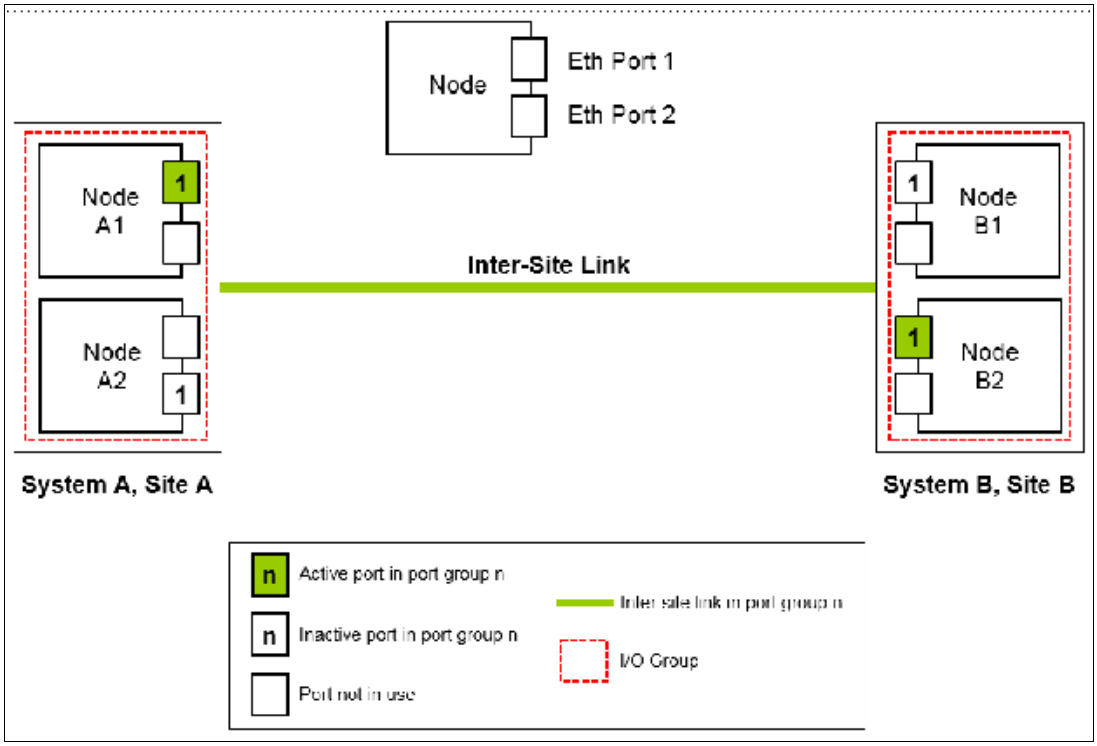


Figure 10-99 One Remote Copy group on each system and nodes with failover ports configured

As shown in Figure 10-99, two systems are available:

- System A
- System B

A single RC port group 1 is configured on two Ethernet ports, one each on Node A1 and Node A2 on System A. Similarly, a single RC port group is configured on two Ethernet ports on Node B1 and Node B2 on System B.

Although two ports on each system are configured for RC port group 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 A1 fails in System A, IP partnership continues servicing replication I/O from Ethernet Port 2 because a failover port is configured on Node A2 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 **`lsportip`** output (reported as `used`).

This configuration has the following characteristics:

- Each node in the I/O group has the same RC port group that is configured. However, only one port in that RC port group is active at any time at each system.
- If the Node A1 in System A or the Node B2 in System B 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 10-100 on page 741 (configuration 3).

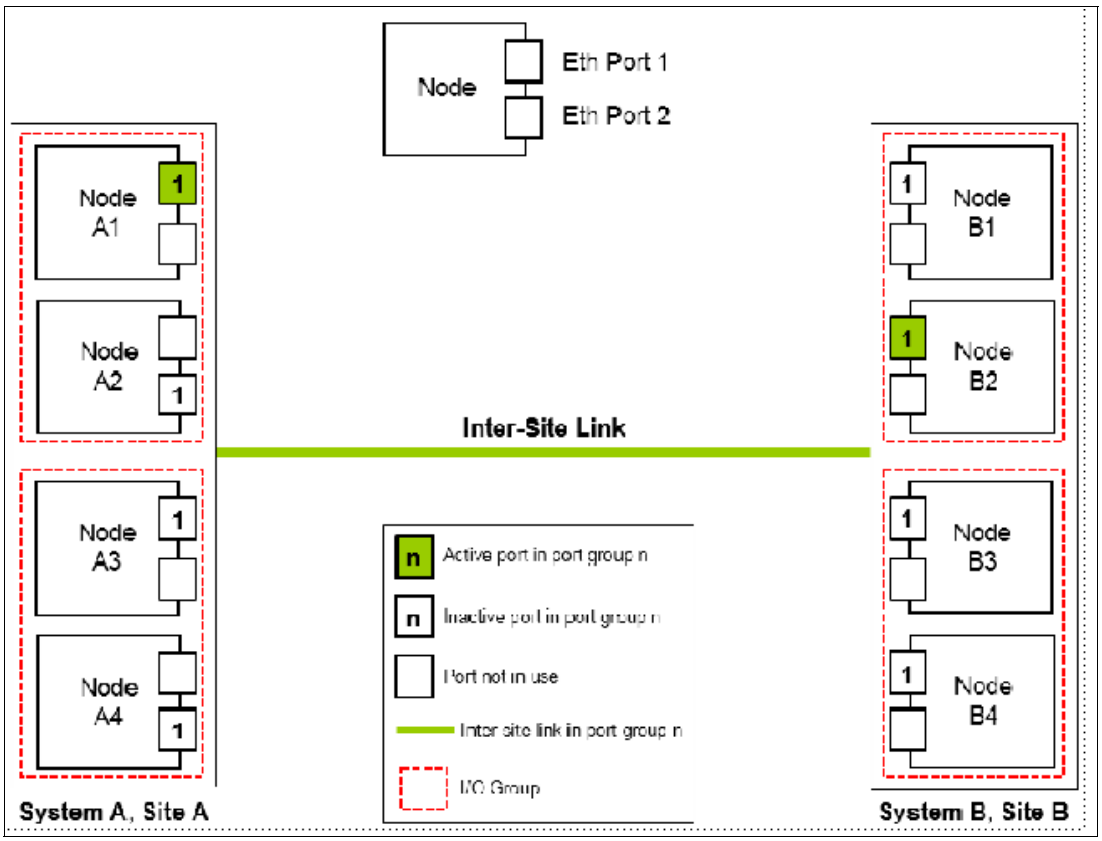


Figure 10-100 Multinode systems single inter-site link with only one RC port group

As shown in Figure 10-100, two 4-node systems are available:

- System A
- System B

A single RC port group 1 is configured on nodes A1, A2, A3, and A4 on System A, Site A; and on nodes B1, B2, B3, and B4 on System B, Site B. Although four ports are configured for RC group 1, only one Ethernet port in each RC port group 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 A1 fails in System A, the IP partnership selects one of the remaining ports that is configured with RC port group 1 from any of the nodes from either of the two I/O groups in System A. 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 **`lsportip`** view (reported as `used`).

This configuration has the following characteristics:

- Each node has the RC port group that is configured in both I/O groups. However, only one port in that RC port group remains active and participates in IP partnership on each system.
- If the Node A1 in System A or the Node B2 in System B 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.

- Eight-node system in IP partnership with four-node system over single inter-site link, as shown in Figure 10-101 (configuration 4).

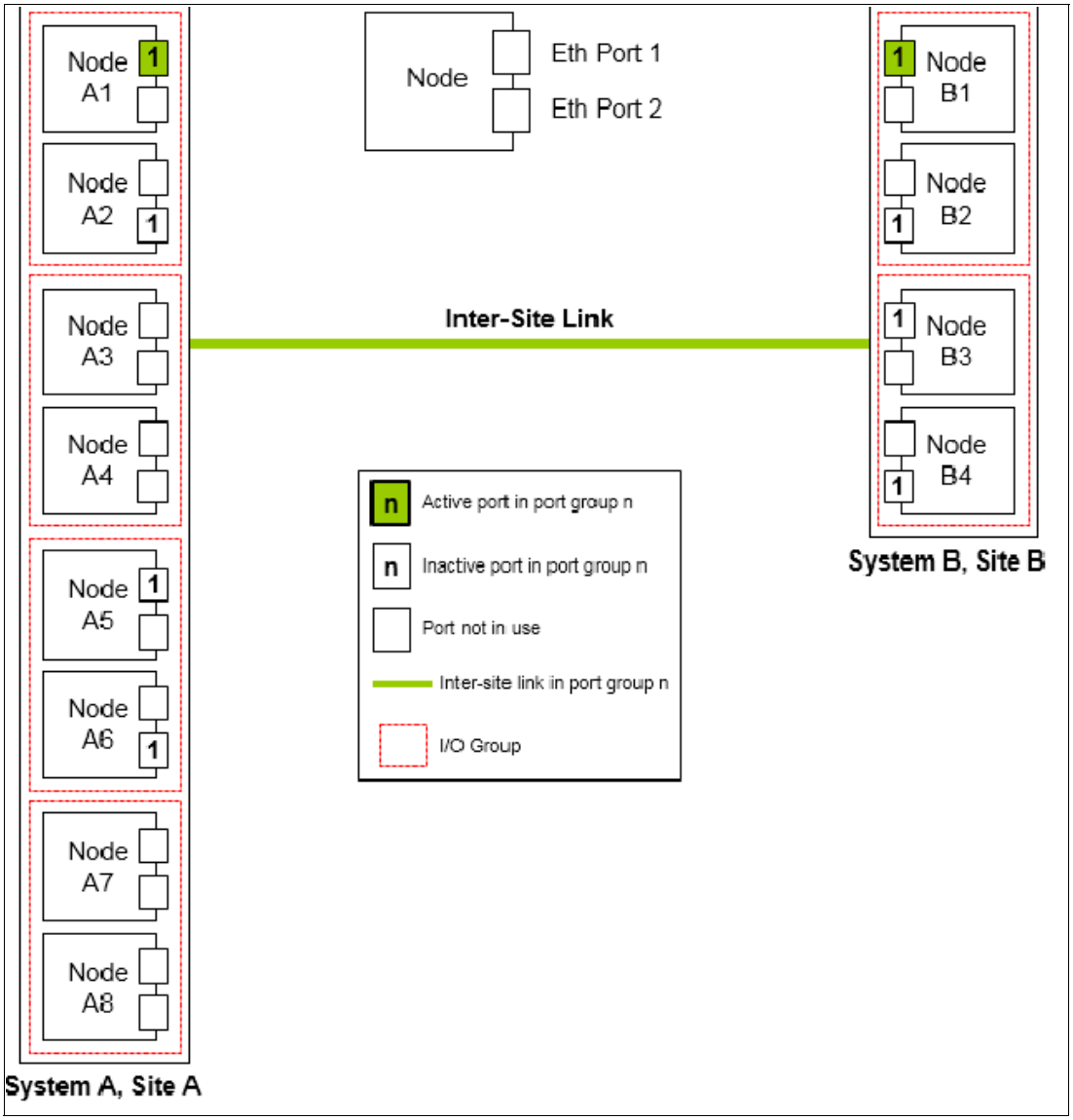


Figure 10-101 Multinode systems single inter-site link with only one Remote Copy port group

As shown in Figure 10-101, an eight-node system (System A in Site A) and a four-node system (System B in Site B) are used. A single RC port group 1 is configured on nodes A1, A2, A5, and A6 on System A at Site A. Similarly, a single RC port group 1 is configured on nodes B1, B2, B3, and B4 on System B.

Although four I/O groups (eight nodes) are in System A, any two I/O groups at maximum are supported to be configured for IP partnerships. If Node A1 fails in System A, IP partnership continues by using one of the ports that is configured in RC port group from any of the nodes from either of the two I/O groups in System A.

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 particular IP port is actively participating in IP partnership process is provided in **`lsportip`** output (reported as `used`).

This configuration features the following characteristics:

- Each node has the RC port group 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 A1 in System A or the Node B2 in System B 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 10-102 on page 744 (configuration 5).

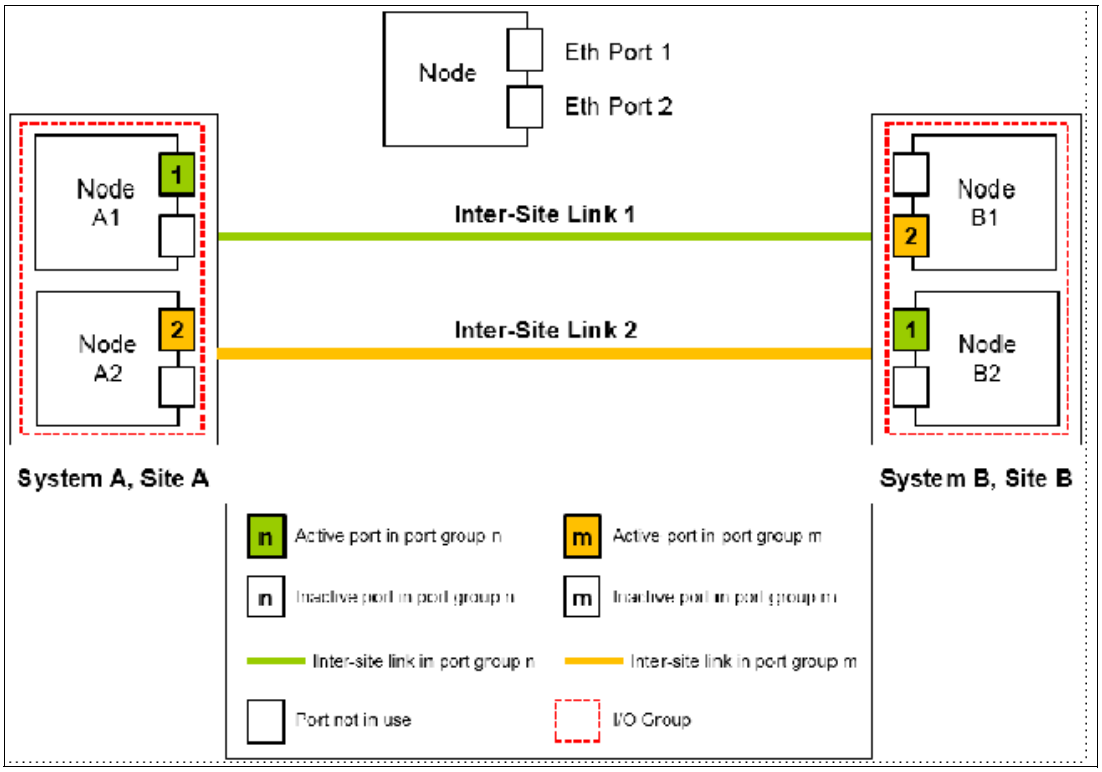


Figure 10-102 Dual links with two Remote Copy groups on each system configured

As shown in Figure 10-102, RC port groups 1 and 2 are configured on the nodes in System A and System B 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 A1 on System A fails), the IP partnership continues only from the available IP port that is configured in RC port group 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 port groups are configured.
- Each node has only one IP port in RC port group 1 or 2.
- Both the IP ports in the two RC port groups 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 port group. Therefore, if two links of 10 Mbps each are available and you have 20 Mbps of effective link bandwidth, 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 10-103 (configuration 6).

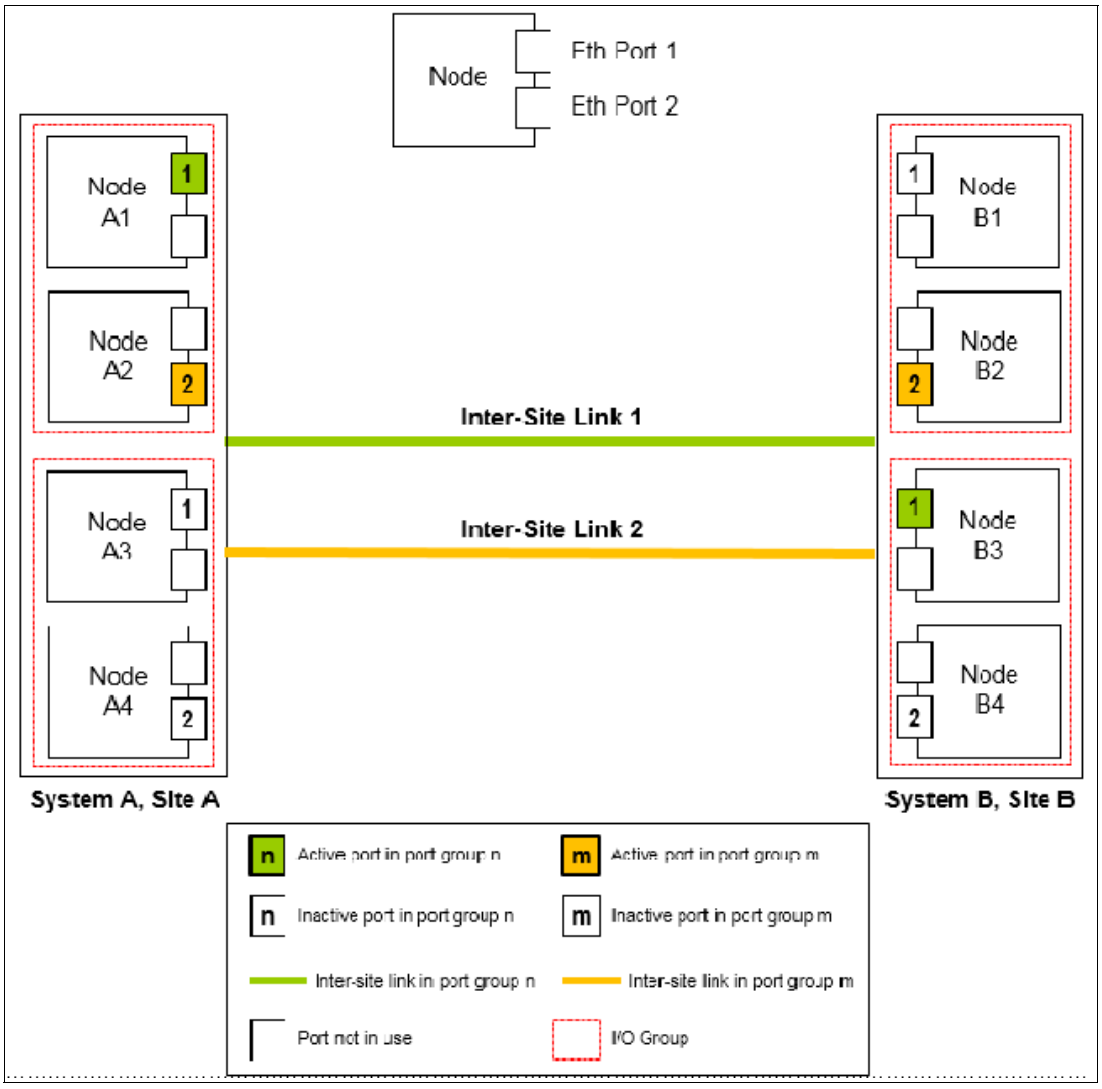


Figure 10-103 Multinode systems with dual inter-site links between the two systems

As shown in Figure 10-103, two 4-node systems are used:

- System A
- System B

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 port groups 1 or 2.

Although two ports are configured in RC port groups 1 and 2 on each system, only one IP port in each RC port group on each system actively participates in IP partnership. The other ports that are configured in the same RC port group act as standby ports in the event of failure. Which port in a configured RC port group participates in IP partnership at any moment is determined by a path configuration algorithm.

In this configuration, if Node A1 fails in System A, IP partnership traffic continues from Node A2 (that is, RC port group 2) and at the same time the failover also causes discovery in RC port group 1.

Therefore, the IP partnership traffic continues from Node A3 on which RC port group 1 is configured. The details of the particular IP port that is actively participating in IP partnership process is provided in the **`lsportip`** output (reported as `used`).

This configuration has the following characteristics:

- Each node has the RC port group that is configured in the I/O groups 1 or 2. However, only one port per system in both RC port groups remains active and participates in IP partnership.
- Only a single port per system from each configured RC port group participates simultaneously in 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 port group 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.

- Eight-node system in IP partnership with a four-node system over dual inter-site links, as shown in Figure 10-104 on page 747 (configuration 7).

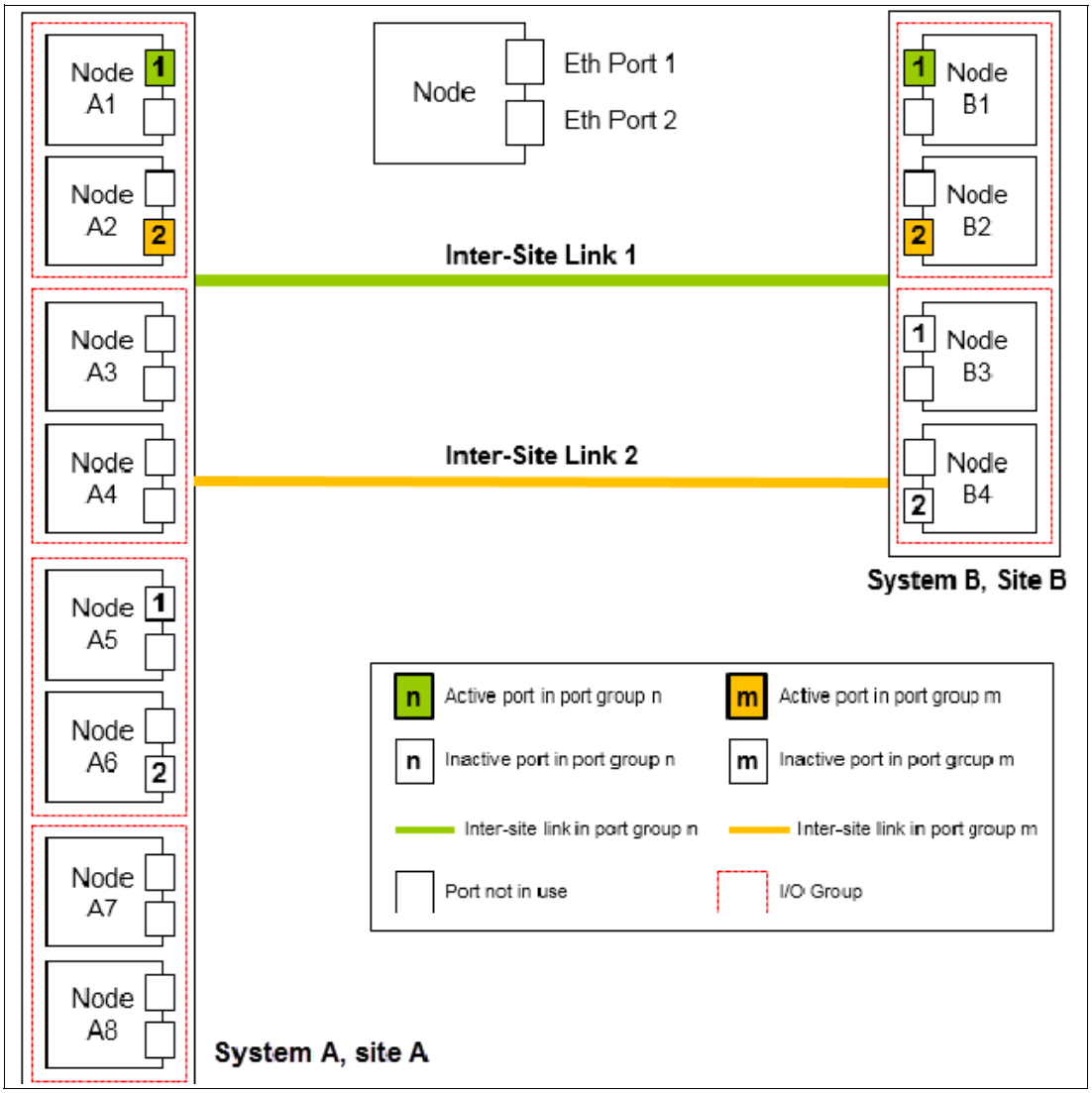


Figure 10-104 Multinode systems with dual inter-site links between the two systems

As shown in Figure 10-104, an eight-node System A in Site A and a four-node System B in Site B is used. Because a maximum of two I/O groups in IP partnership is supported in a system, although four I/O groups (eight nodes) exist, nodes from only two I/O groups are configured with RC port groups in System A. 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 port groups 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 A5 and Node A6 are configured with RC port groups properly, active IP partnership traffic on both of the links might be driven from Node A1 and Node A2 only.

If Node A1 fails in System A, IP partnership traffic continues from Node A2 (that is, RC port group 2). The failover also causes IP partnership traffic to continue from Node A5 on which RC port group 1 is configured. The details of the particular IP port actively participating in IP partnership process is provided in the **`lsportip`** output (reported as `used`).

This configuration has the following characteristics:

- Two I/O groups with nodes in those I/O groups are configured in two RC port groups because two inter-site links are used for participating in IP partnership. However, only one port per system in a particular RC port group remains active and participates in IP partnership.
- One port per system from each RC port group participates in 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 port group.
- The path selection algorithm starts discovery of available ports in the affected RC port group 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 10-105 (configuration 8).

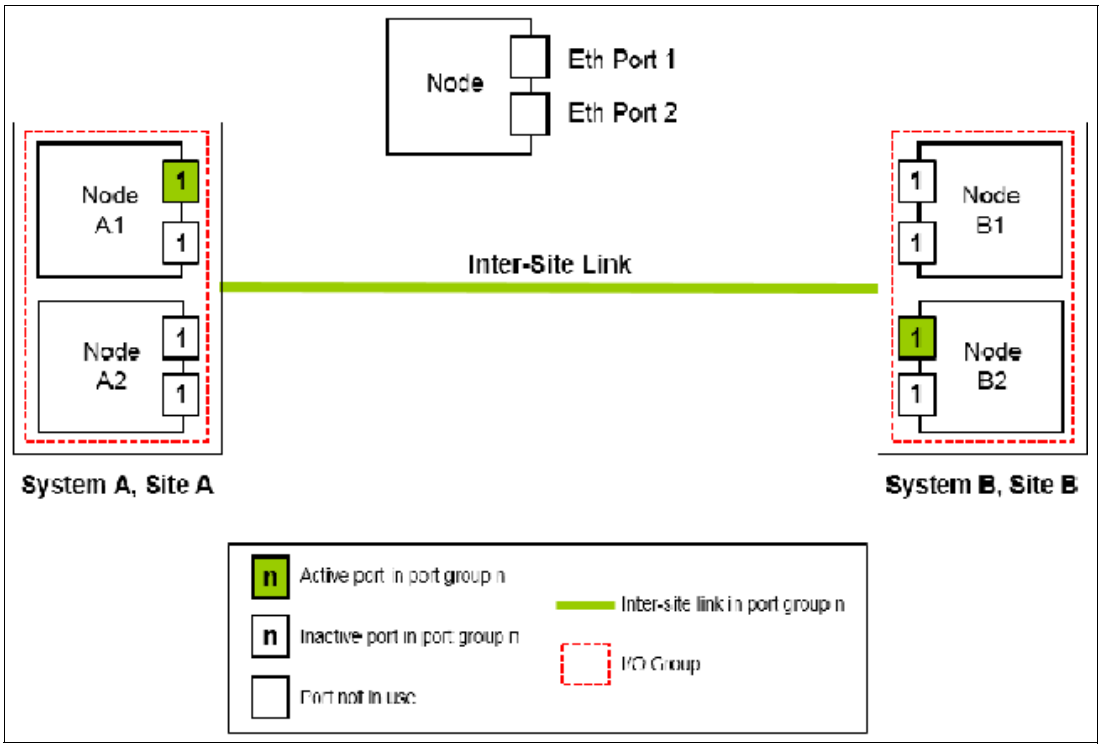


Figure 10-105 Two node systems with single inter-site link and Remote Copy port groups configured

As shown in Figure 10-105, this configuration is similar to Configuration 2, but differs because each node now has the same RC port group 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 group on the same node is *not supported*.

- An example of an *unsupported* configuration for a dual inter-site link is shown in Figure 10-106 (configuration 9).

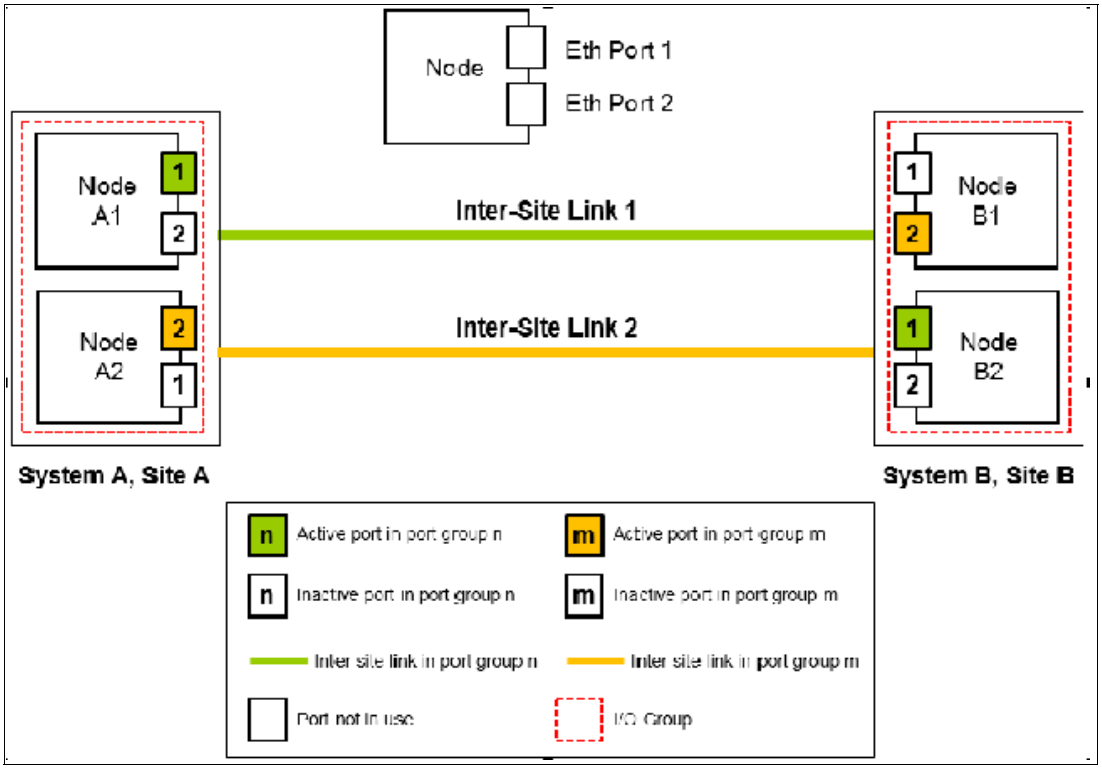


Figure 10-106 Dual Links with two Remote Copy Port Groups with failover Port Groups configured

As shown in Figure 10-106, this configuration is similar to Configuration 5, but differs because each node now also has two ports that are configured with RC port groups. 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 10-107 (configuration 10).

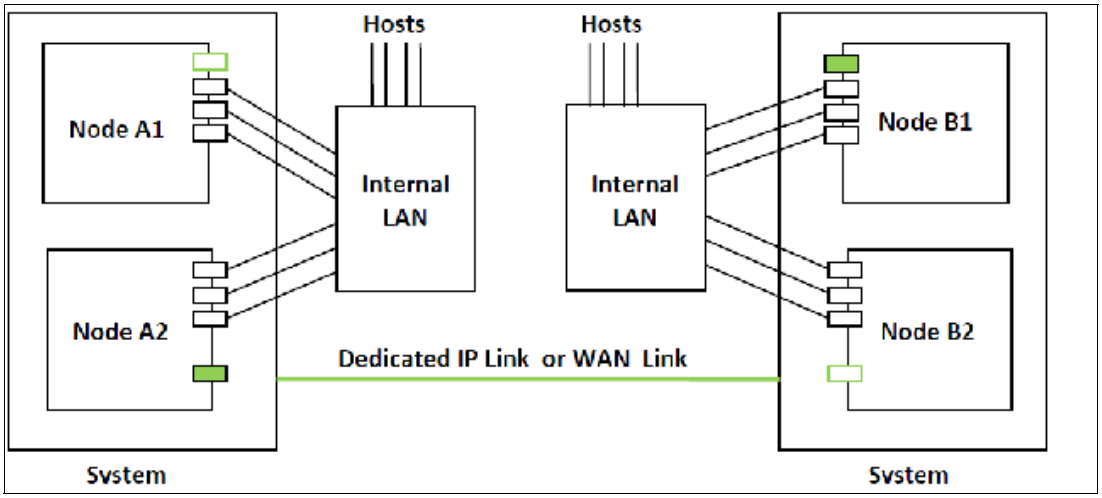


Figure 10-107 Deployment example

In this configuration, one port on each node in System A and System B is configured in RC group 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 group: 1
   - Host: No
   - 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 10-107 on page 749 is an example deployment for the configuration that is shown in Figure 10-101 on page 743. Ports that are shared with host access are shown in Figure 10-108 (configuration 11).

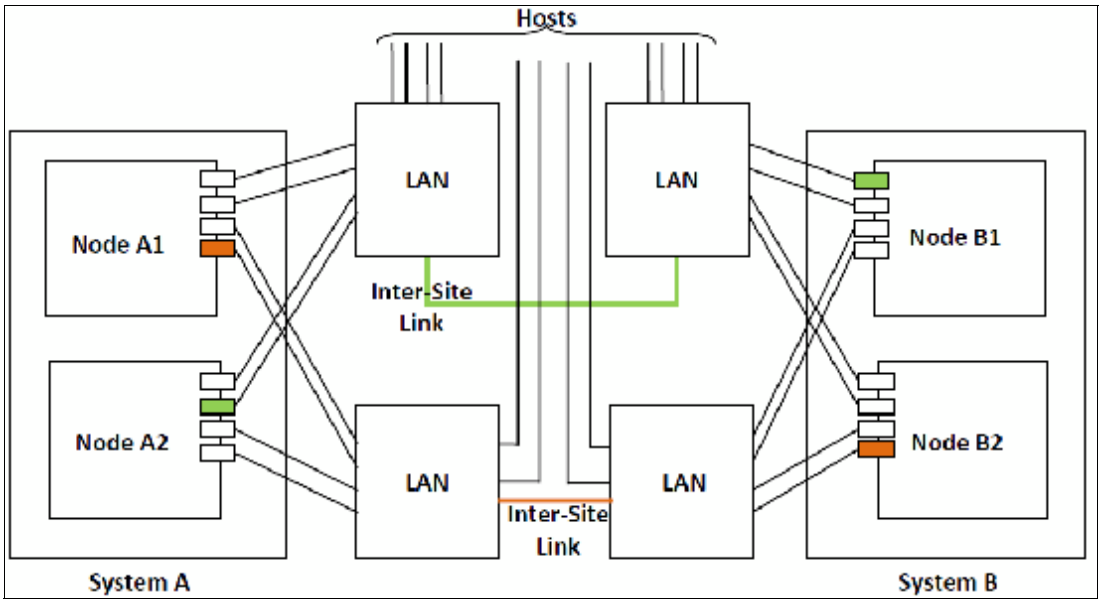


Figure 10-108 Deployment example

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 port group 1
     - Host: Yes
     - Assign IP address

   - Node 2:

     - Port 4, remote copy port group 2
     - Host: Yes
     - Assign IP address

d. Configure IP ports for RC on System B1 by using the following settings:

   - Node 1:

     - Port 1, remote copy port group 1
     - Host: Yes
     - Assign IP address

   - Node 2:

     - Port 4, remote copy port group 2
     - Host: Yes
     - Assign IP address

e. Check the MTU levels across the network as set (default MTU is 1500 on SAN Volume Controller and Spectrum 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.

10.11 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 a 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 10-109.

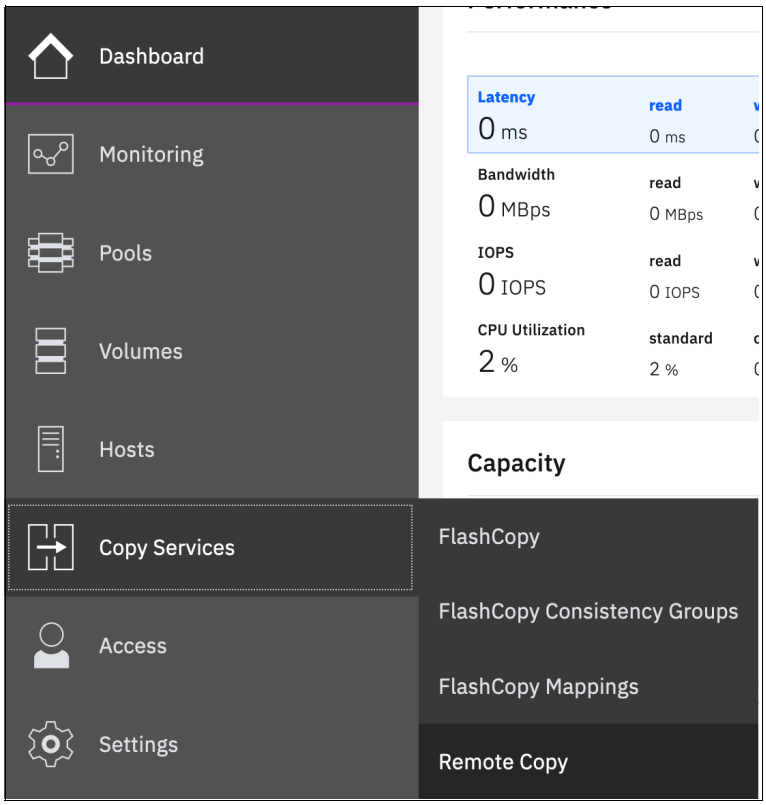


Figure 10-109 Remote Copy menu

The Remote Copy panel is displayed, as shown in Figure 10-110.

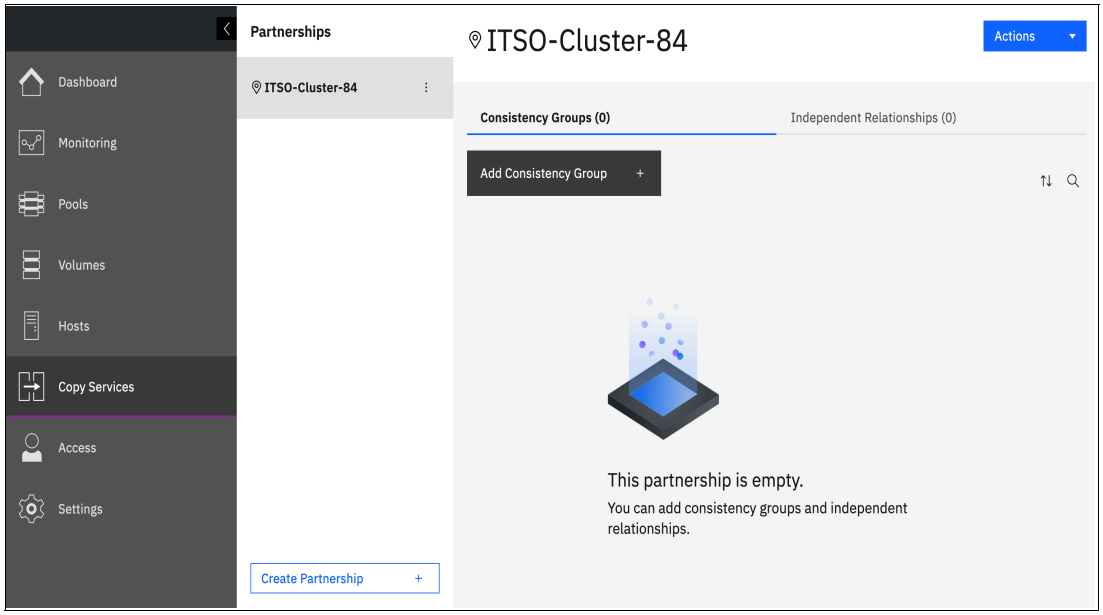


Figure 10-110 Remote Copy panel

10.11.1 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 10.11.2, “Creating Remote Copy relationships” on page 755.

To create an FC partnership between IBM Spectrum Virtualize systems by using the GUI, open the Remote Copy panel (see Figure 10-110 on page 752) 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 10-111. In this example, we use a two-site partnership.

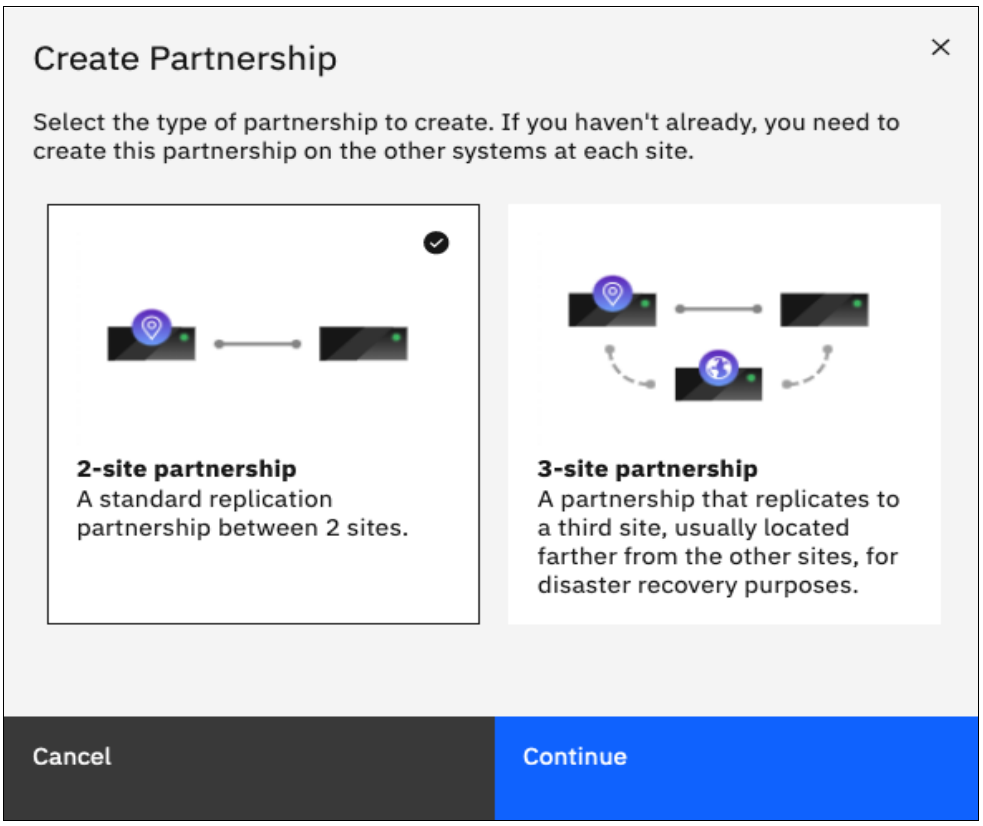


Figure 10-111 Select Remote Copy topology

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 10-112):

   `This system does not have any candidates.`

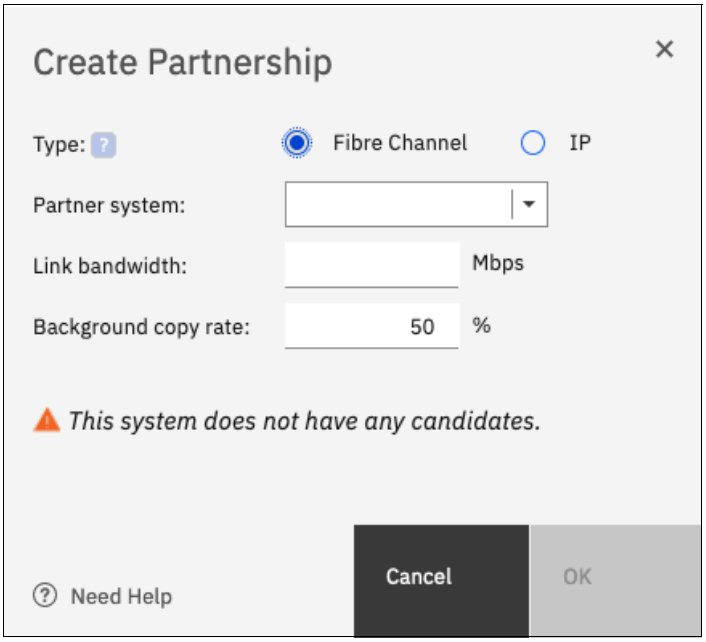


Figure 10-112 Creating a Partnership details

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 10-113.

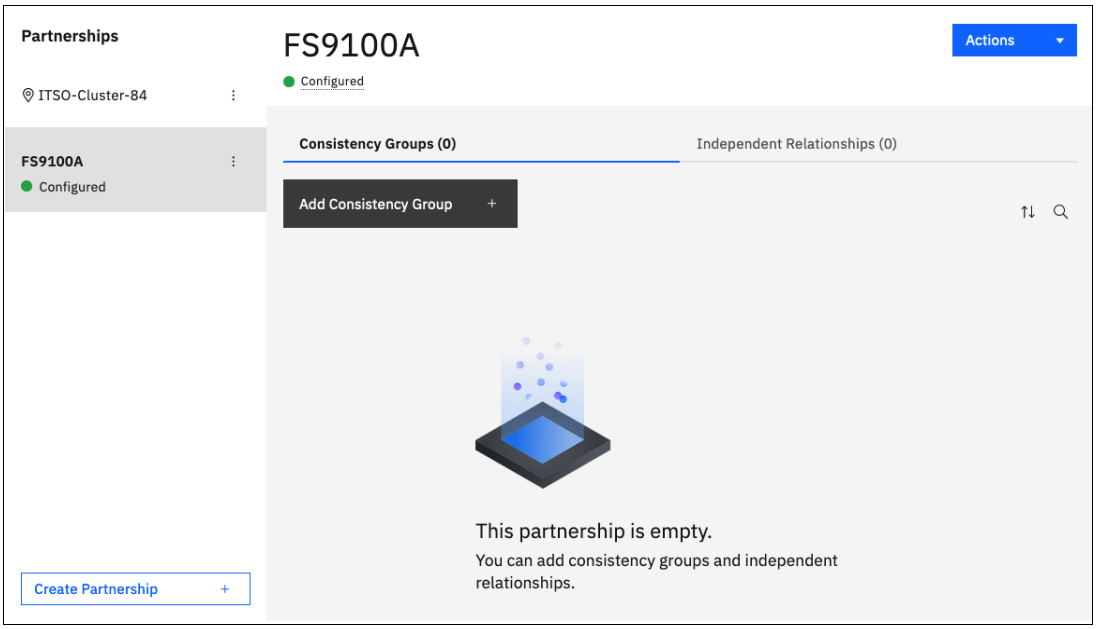


Figure 10-113 Fully configured FC partnership

10.11.2 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 a 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 a 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 10-114.

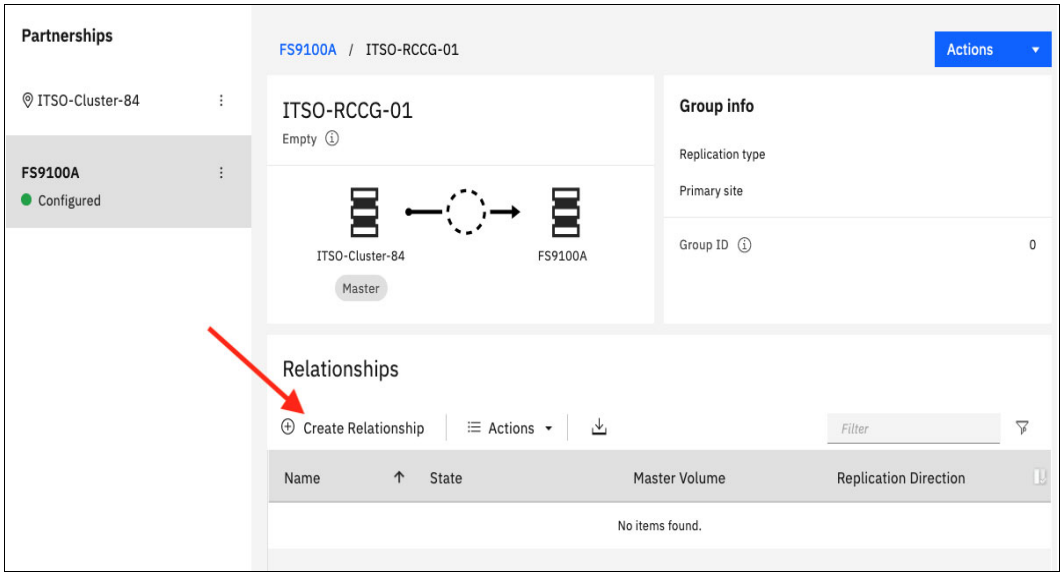


Figure 10-114 Creating a Remote Copy Relationship in an existing consistency group

4. If you want to add a standalone relationship, select the **Independent Relationships** tab and click **Create Relationship**, as shown in Figure 10-115.

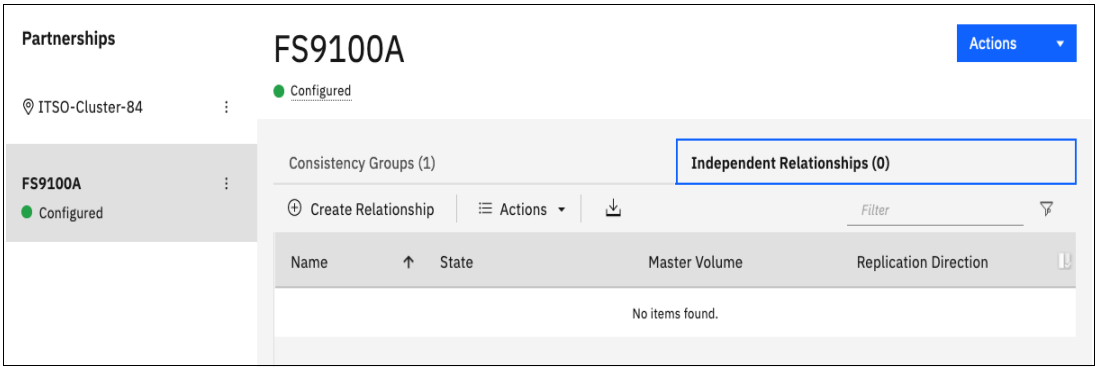


Figure 10-115 Creating Standalone Remote Copy relationships

5. In the Create Relationship window, select one of the following types of relationships that you want to create, as shown in Figure 10-116:
   - MM
   - GM (with or without Consistency Protection)
   - GMCV

   Click **Next**.

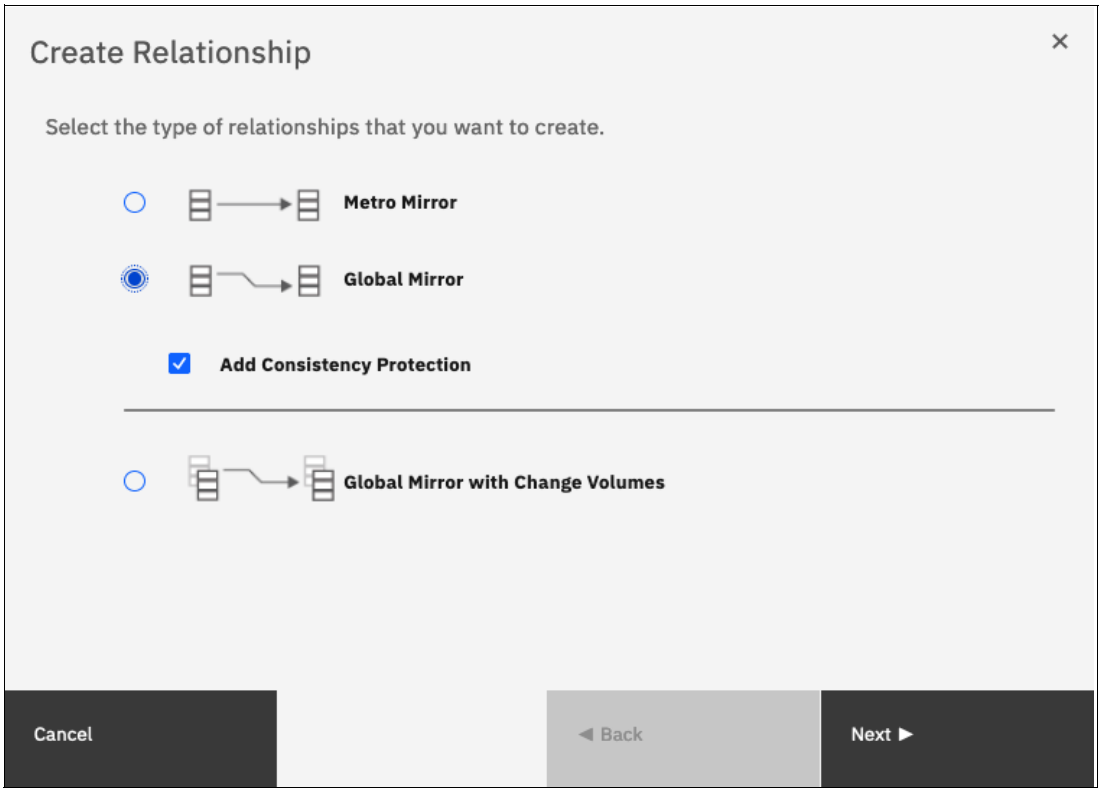


Figure 10-116 Creating a Remote Copy relationship

6. In the next window, select the target system for this RC relationship and click **Next**, as shown in Figure 10-117.

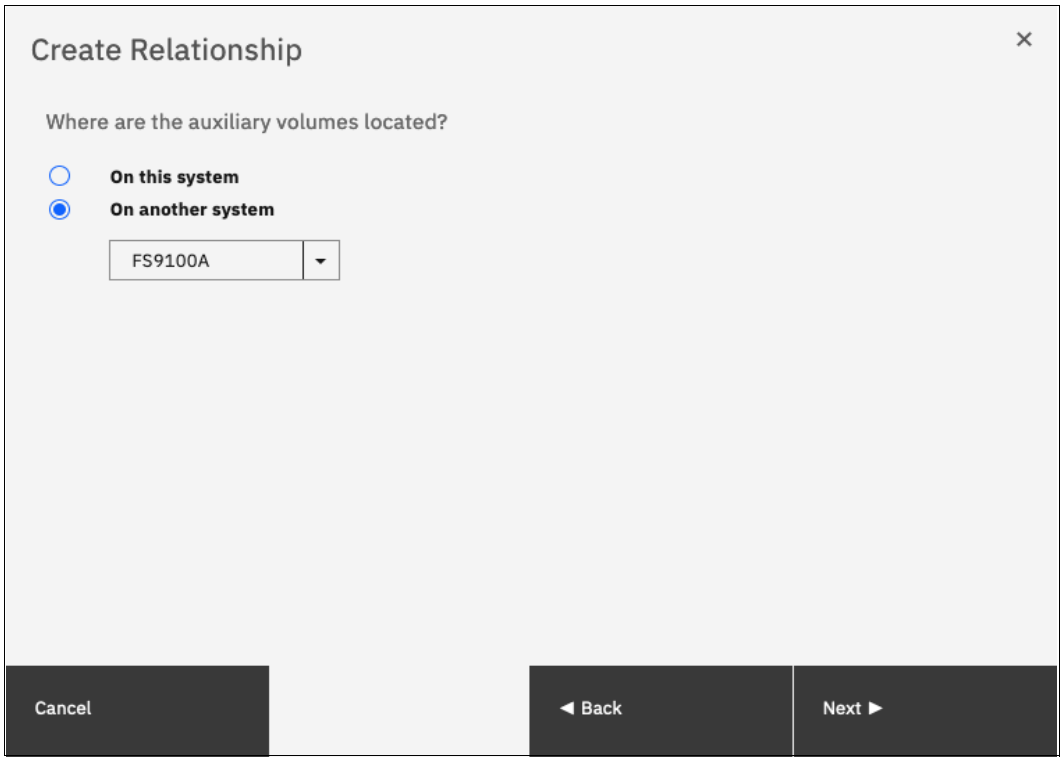


Figure 10-117 Selecting the target system for the RC relationship

7. Select the master and auxiliary volumes, as shown in Figure 10-118. Click **Add**.

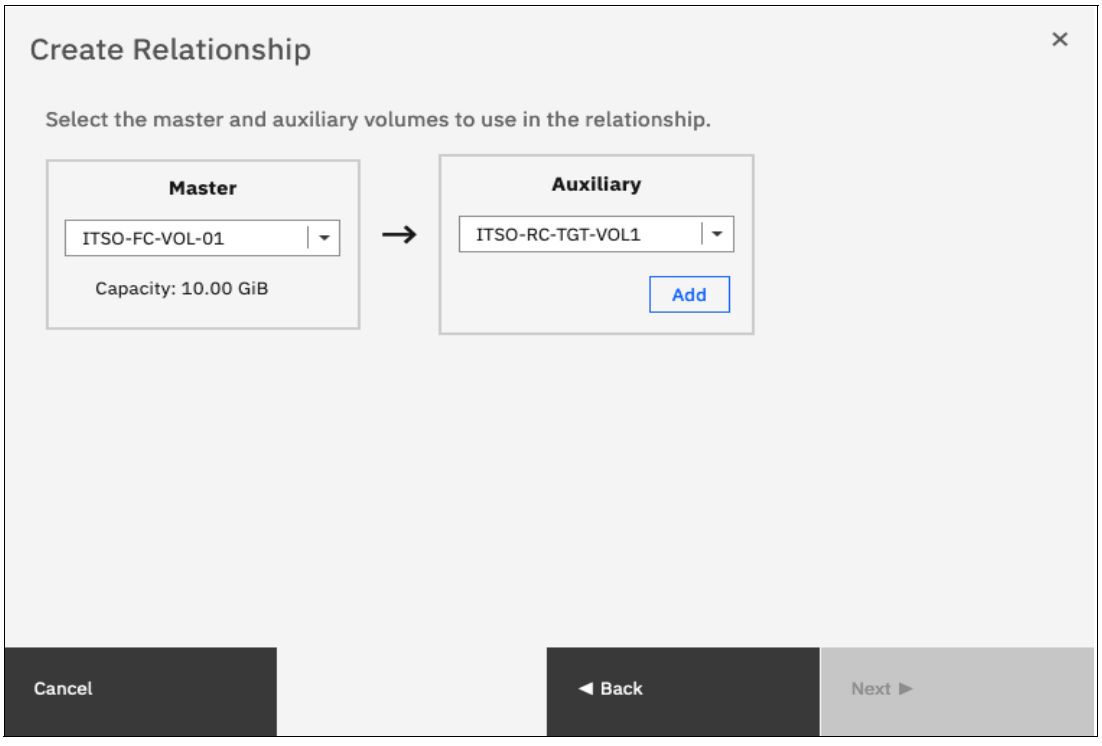


Figure 10-118 Selecting the master and auxiliary volumes

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.

8. In the next window, you can add change volumes if needed, as shown in Figure 10-119. Click **Finish**.

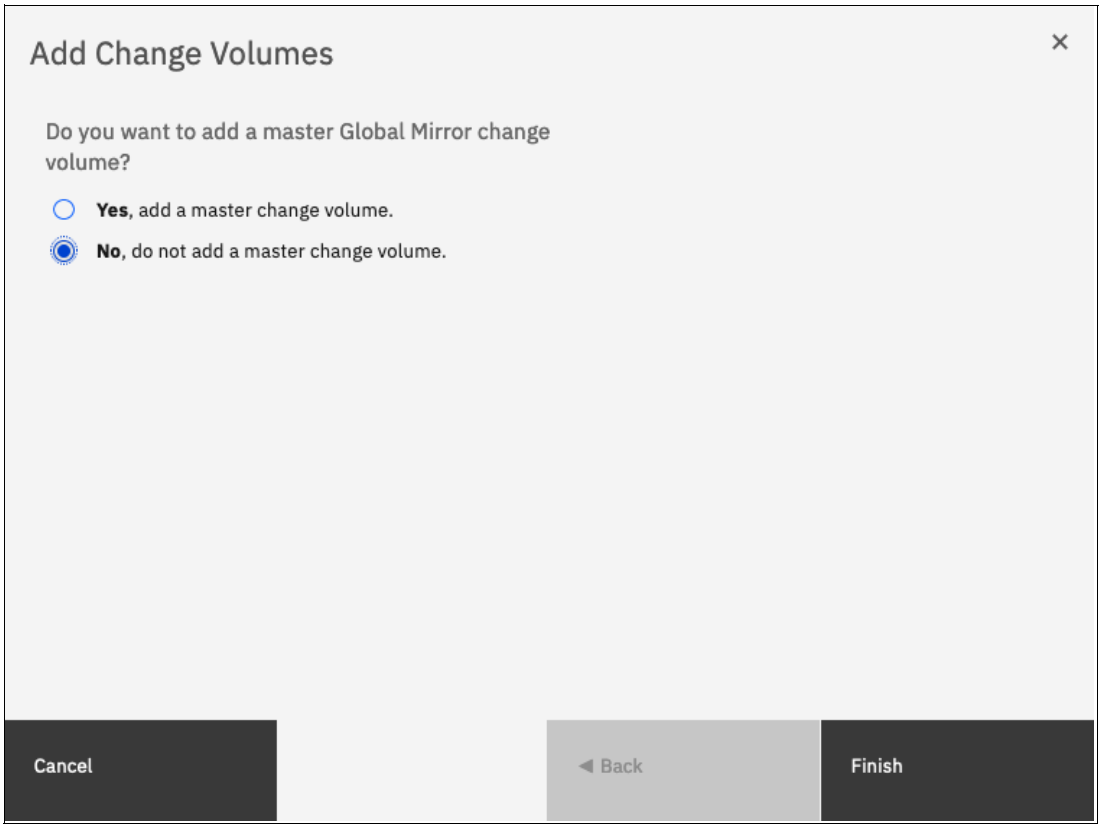


Figure 10-119 Add Change Volumes panel

9. Repeat the steps 7 and 8 to add relationships if needed, (see Figure 10-120). When you are finished creating relationships, click **Next**.

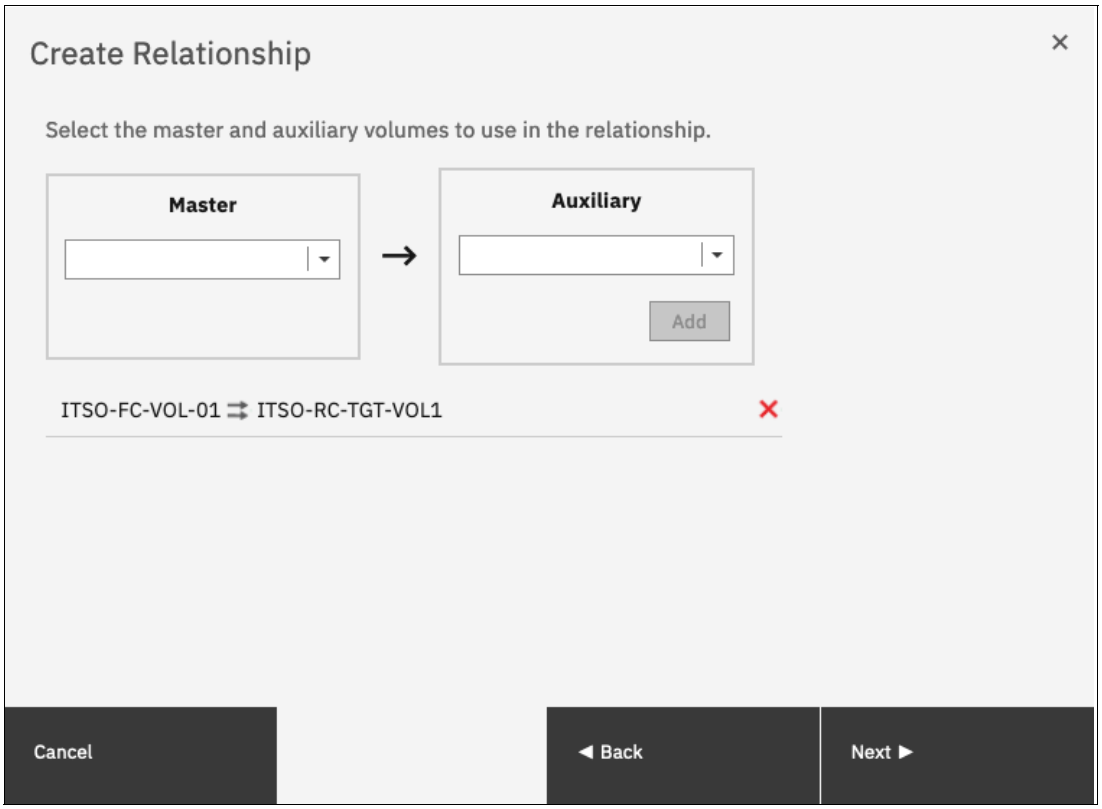


Figure 10-120 Checking and adding the relationship

10. In the next window, select whether the volumes are synchronized so that the relationship is created, as shown in Figure 10-121. Click **Next**.

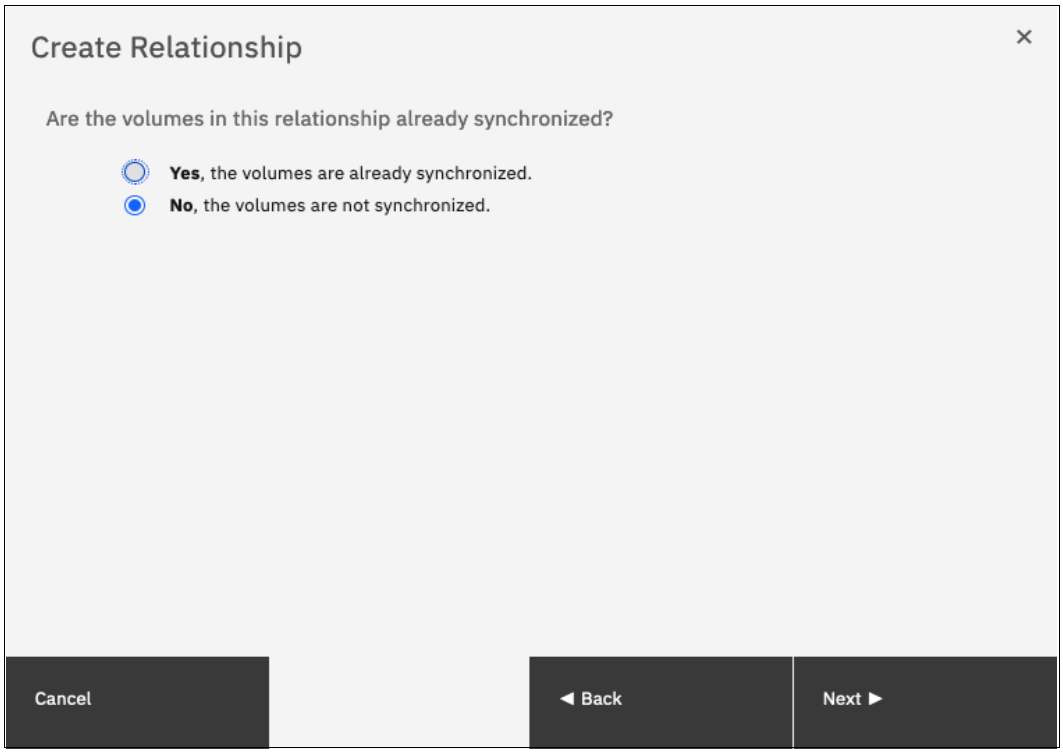


Figure 10-121 Selecting if volumes are synchronized

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 10-122. Then, click **Finish**.

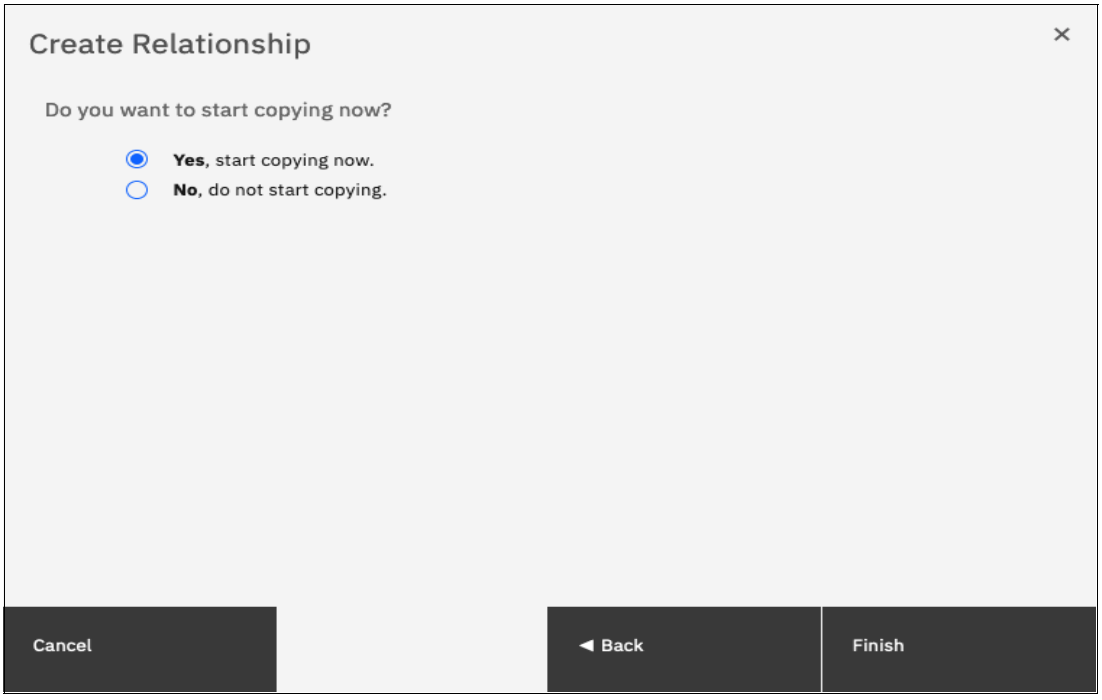


Figure 10-122 Start copying Remote Copy relationship

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.

10.11.3 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 10-123.

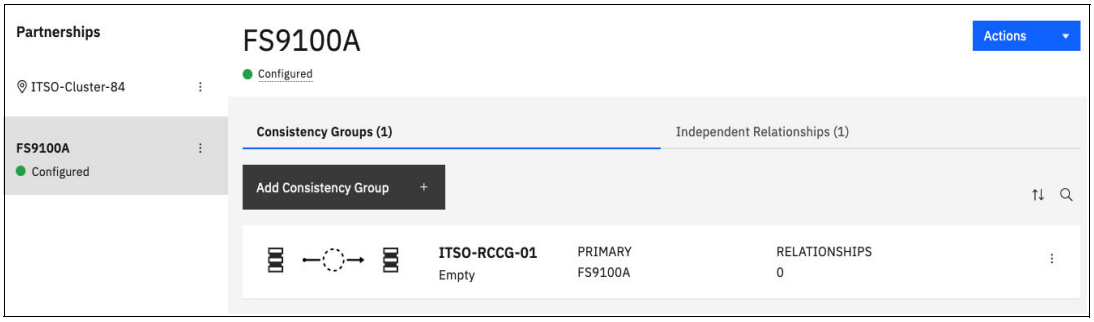


Figure 10-123 Creating a Remote Copy consistency group

2. Enter a name for the consistency group, select the target system and click **Add**, as shown in Figure 10-124. It is then added to the configuration with no relationships.

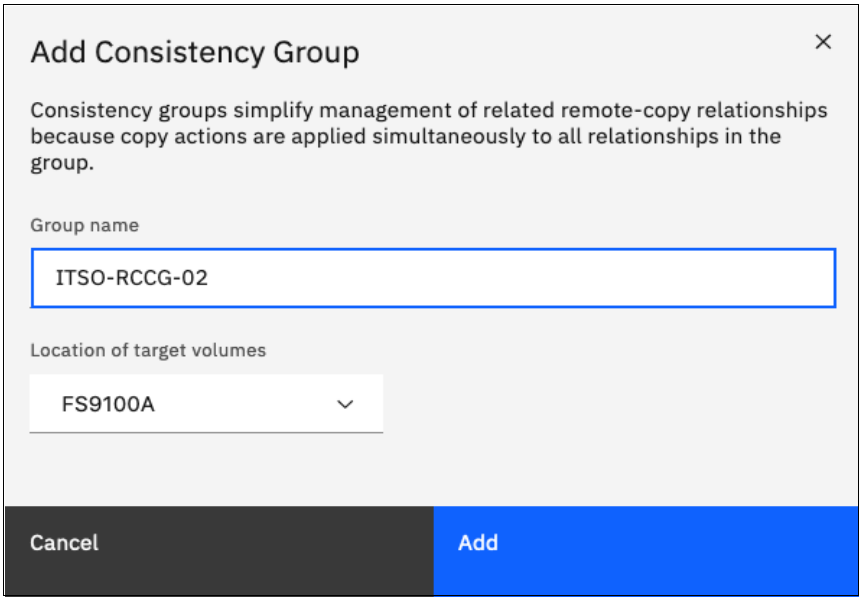


Figure 10-124 Entering a name for the new consistency group

3. You can add standalone relationships to the recently added consistency group by selecting the **Independent Relationships** tab, right-clicking the wanted relationship, and then, clicking **Add to Consistency Group**. You also can create relationships directly for this consistency group by selecting it in the **Consistency Group** tab, as shown in Figure 10-125, and then, clicking **Create Relationship**.

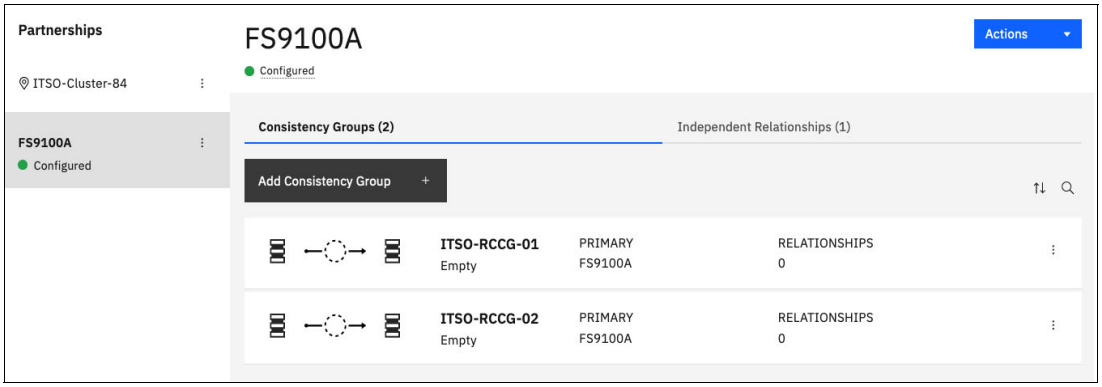


Figure 10-125 Remote Copy Consistency Groups tab

For more information about creating a Remote Copy relationship, see 10.11.2, “Creating Remote Copy relationships” on page 755.

10.11.4 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, click the consistency group‘s name to view its relationships.
3. Right-click the relationships to be renamed and select **Rename**, as shown in Figure 10-126.

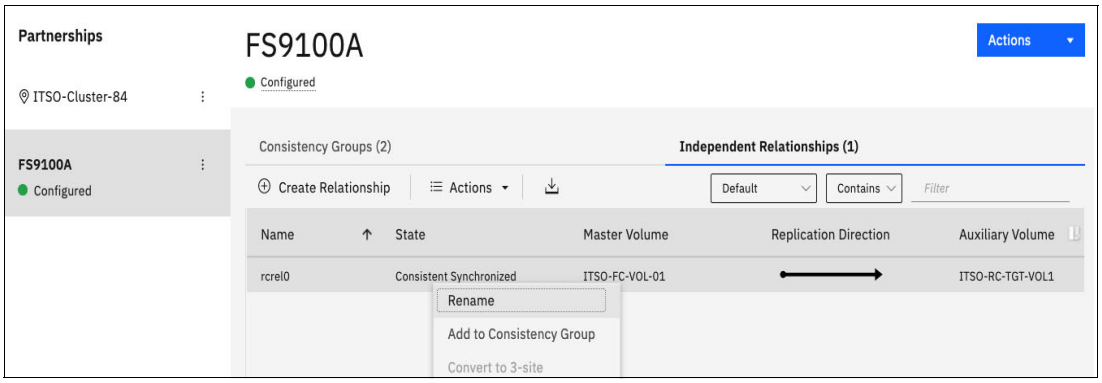


Figure 10-126 Renaming Remote Copy relationships

4. Enter the new name that you want to assign to the relationships and click **Rename**, as shown in Figure 10-127.

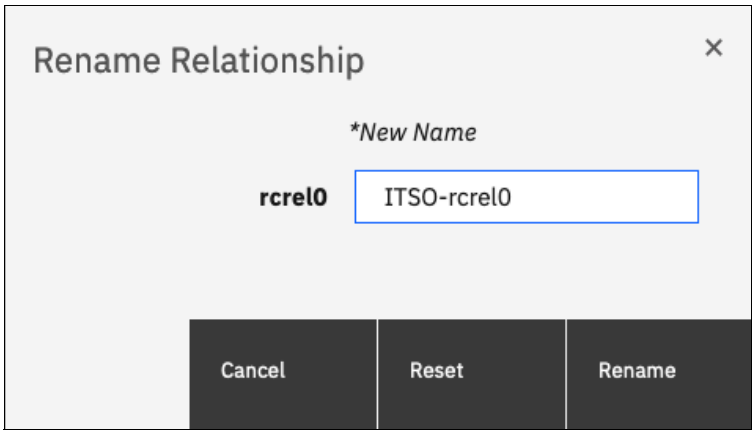


Figure 10-127 Renaming Remote Copy relationships

RC relationship name: Names can include 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.

10.11.5 Renaming a Remote Copy consistency group

To rename a 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 10-128.

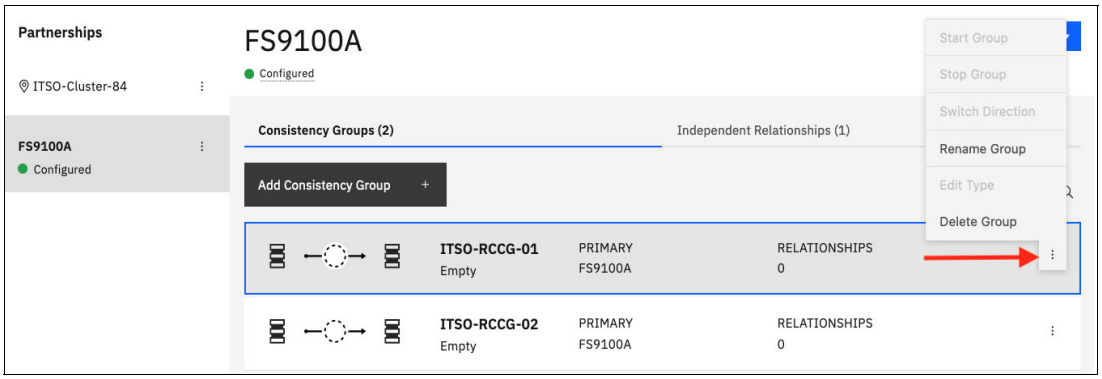


Figure 10-128 Renaming a Remote Copy consistency group

3. Enter the new name that you want to assign to the consistency group and click **Rename**, as shown in Figure 10-129.

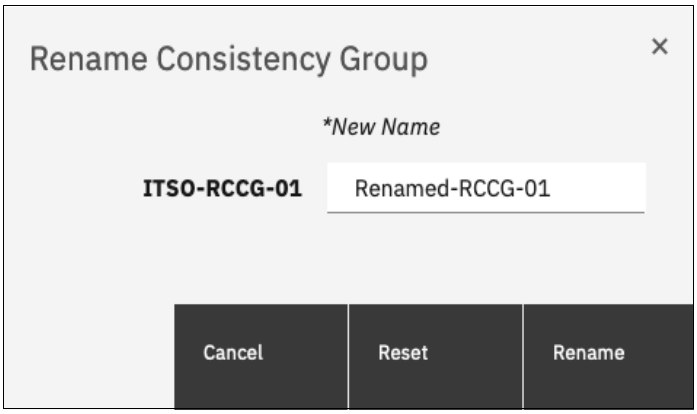


Figure 10-129 Entering new name for consistency group

RC consistency group name: Names can include 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.

10.11.6 Moving standalone Remote Copy relationships to consistency group

To add one or multiple stand-alone relationships to a 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 10-130.



Figure 10-130 Moving relationships to a consistency group

3. Select the consistency group for this RC relationship by using the menu, as shown in Figure 10-131. Click **Add to Consistency Group** to confirm your changes.

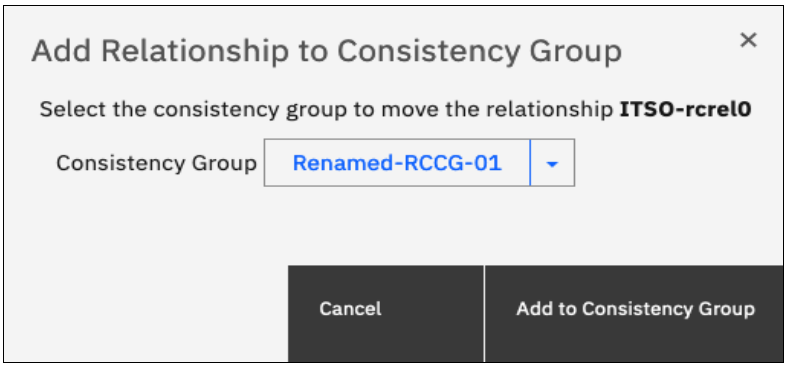


Figure 10-131 Selecting the consistency group to add the relationships to

10.11.7 Removing Remote Copy relationships from consistency group

To remove one or multiple relationships from a 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 10-132.

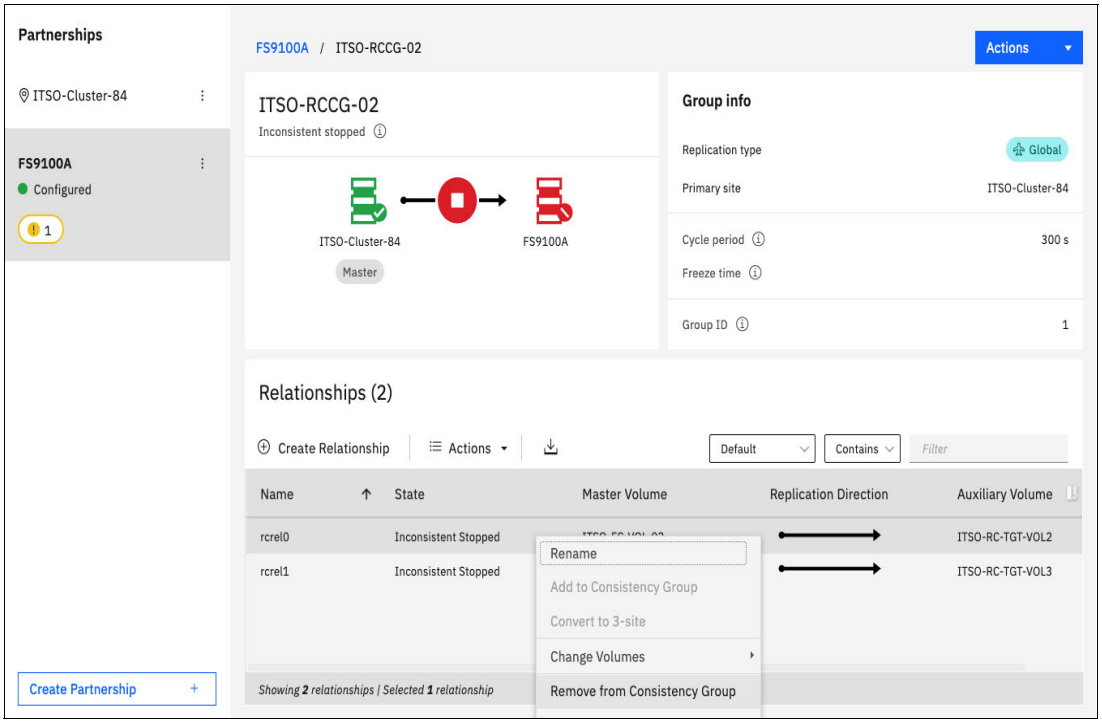


Figure 10-132 Removing relationships from a consistency group

4. Confirm your selection and click **Remove**, as shown in Figure 10-133.

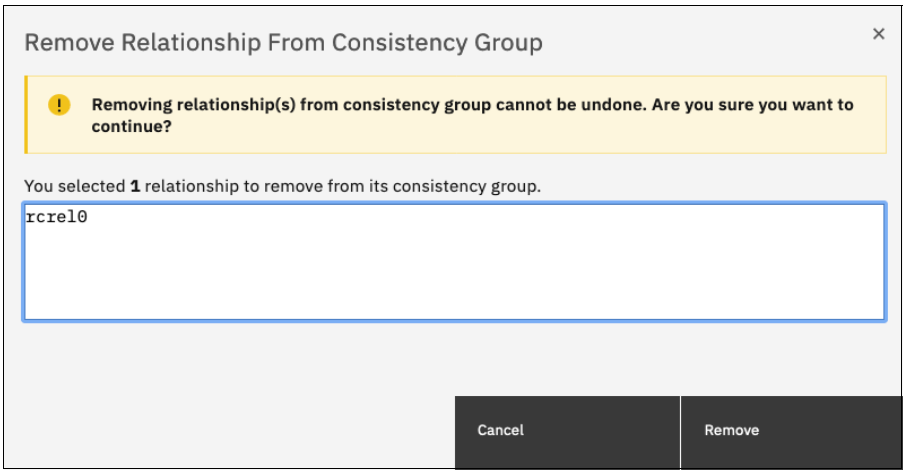


Figure 10-133 Confirm the removal of relationships from a consistency group

10.11.8 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 10-134.

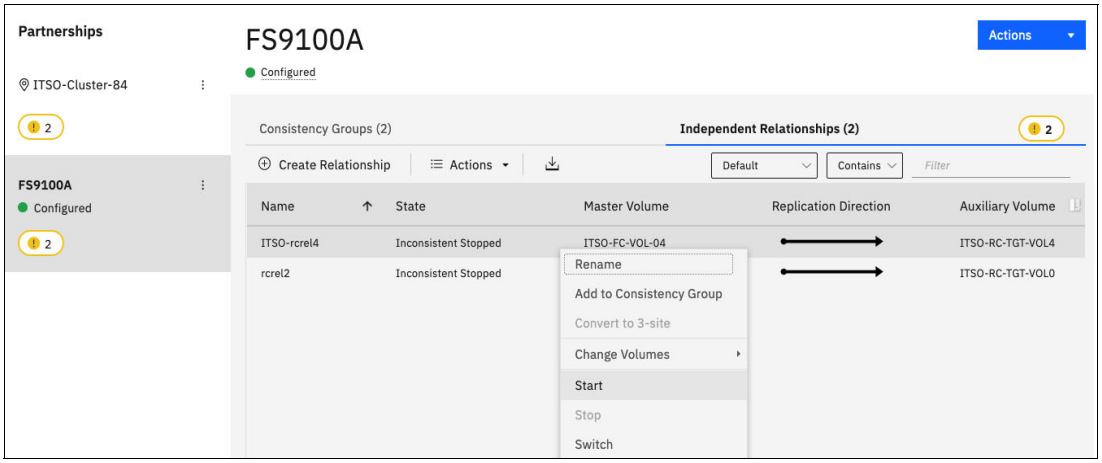


Figure 10-134 Starting Remote Copy relationships

10.11.9 Starting a Remote Copy consistency group

When a 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 10-135.

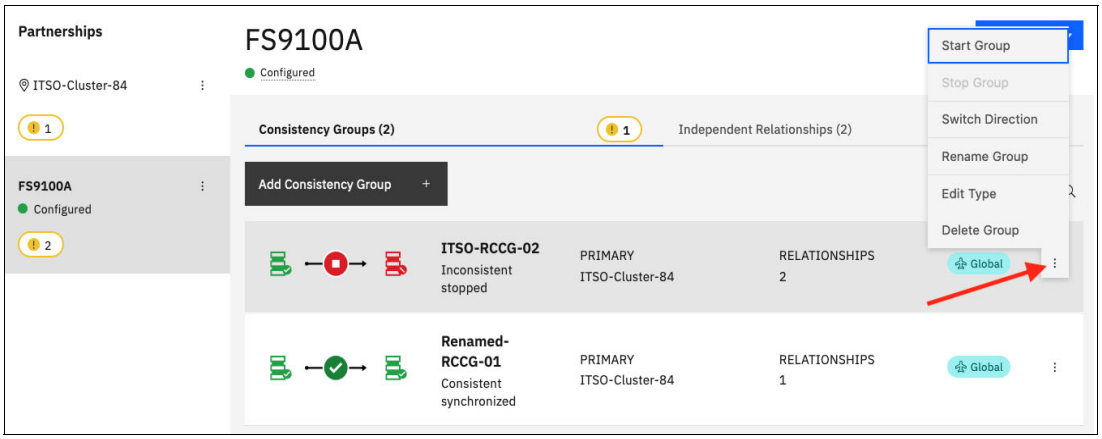


Figure 10-135 Starting a Remote Copy consistency group

10.11.10 Switching a relationship copy direction

When a 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 10-136.

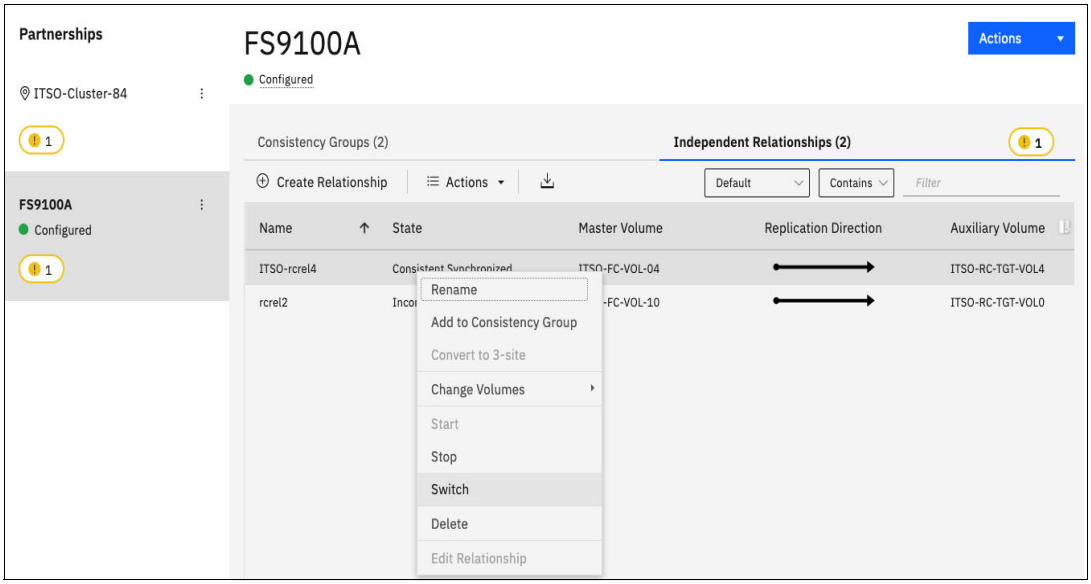


Figure 10-136 Switching Remote Copy relationship direction

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 10-137. Click **Yes**.

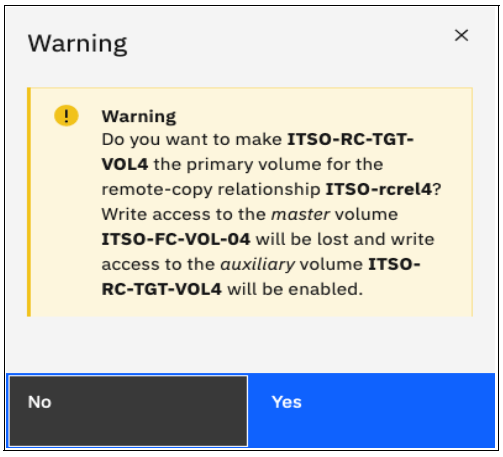


Figure 10-137 Switching master-auxiliary direction of relationships changes the write access

When a 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 10-138.

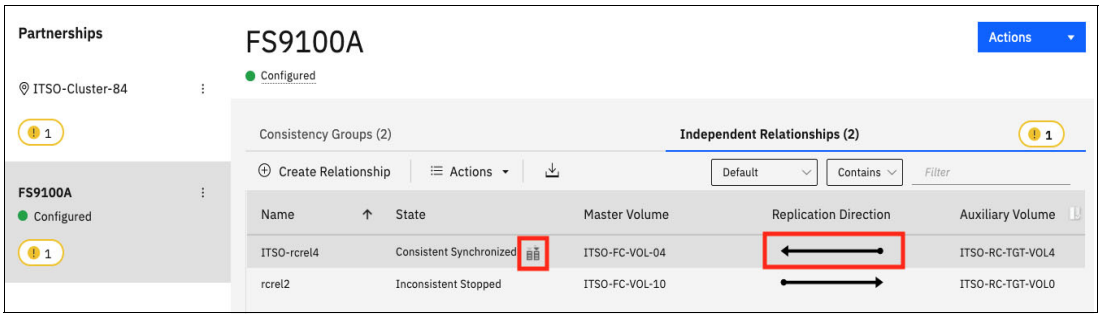


Figure 10-138 Switched Remote Copy Relationship

10.11.11 Switching a consistency group direction

When a 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 a 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 wanted consistency group to be switched and select **Switch Direction**, as shown in Figure 10-139.

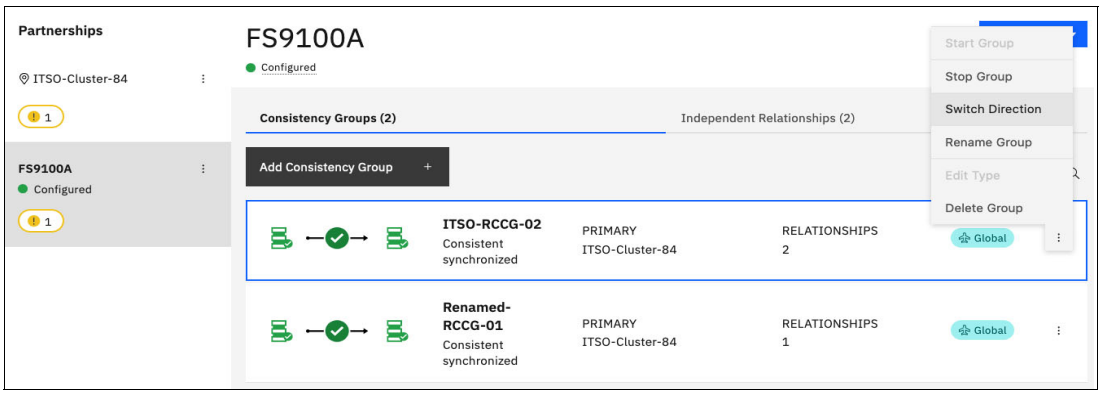


Figure 10-139 Switching a consistency group direction

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 10-140. Click **Yes**.

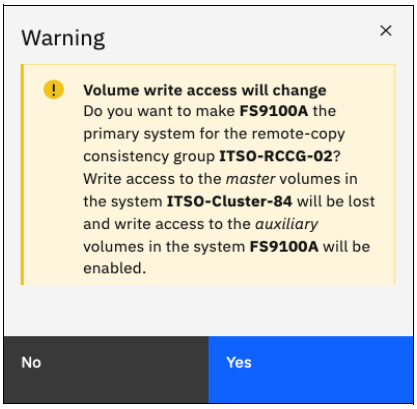


Figure 10-140 Switching direction of consistency groups changes the write access

10.11.12 Stopping Remote Copy relationships

When a 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 10-141.

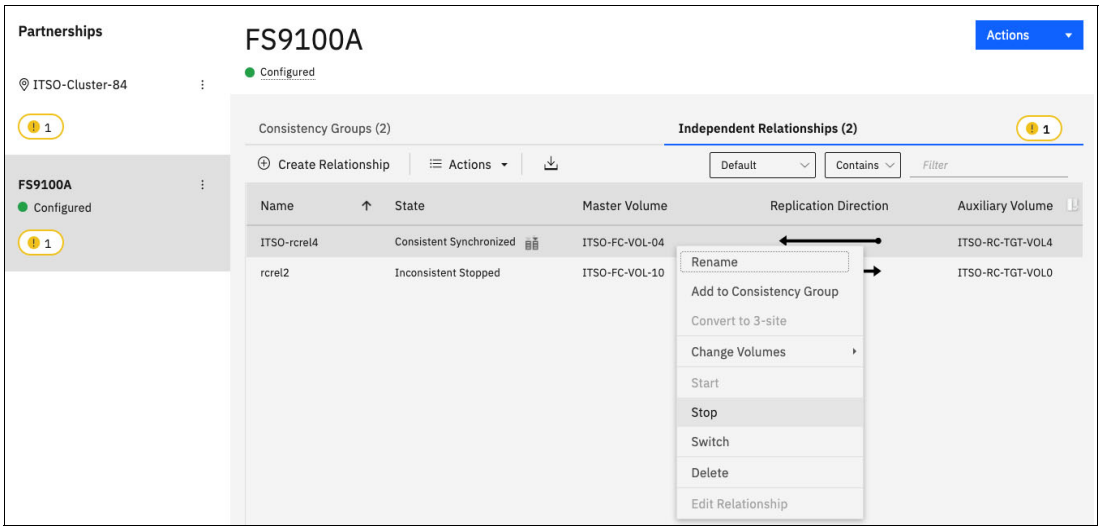


Figure 10-141 Stopping a Remote Copy relationship

3. When a 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 10-142.

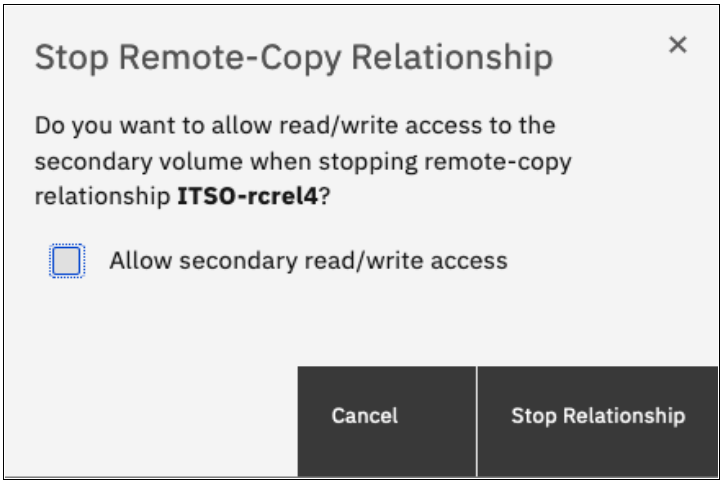


Figure 10-142 Grant access in read and write to the auxiliary volume

10.11.13 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 10-143.

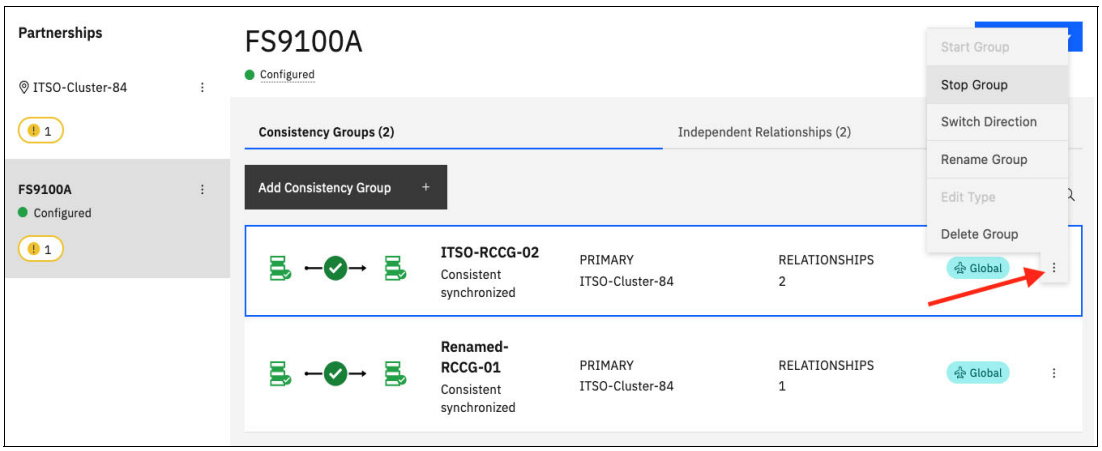


Figure 10-143 Stopping a consistency group

3. When a 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 10-144.

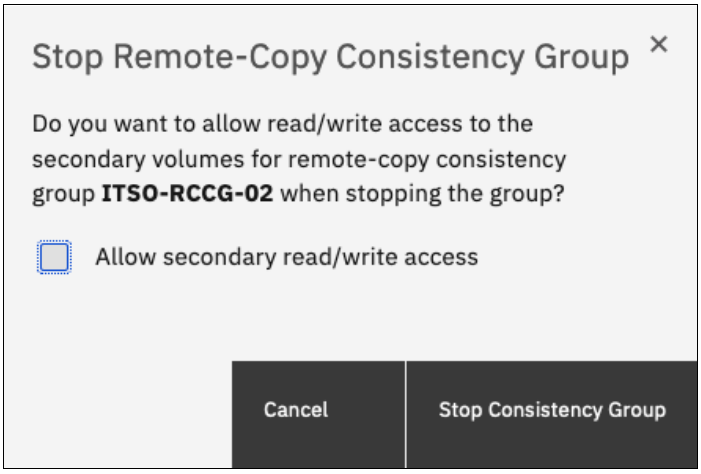


Figure 10-144 Grant access in read and write to the auxiliary volumes

10.11.14 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 10-145.

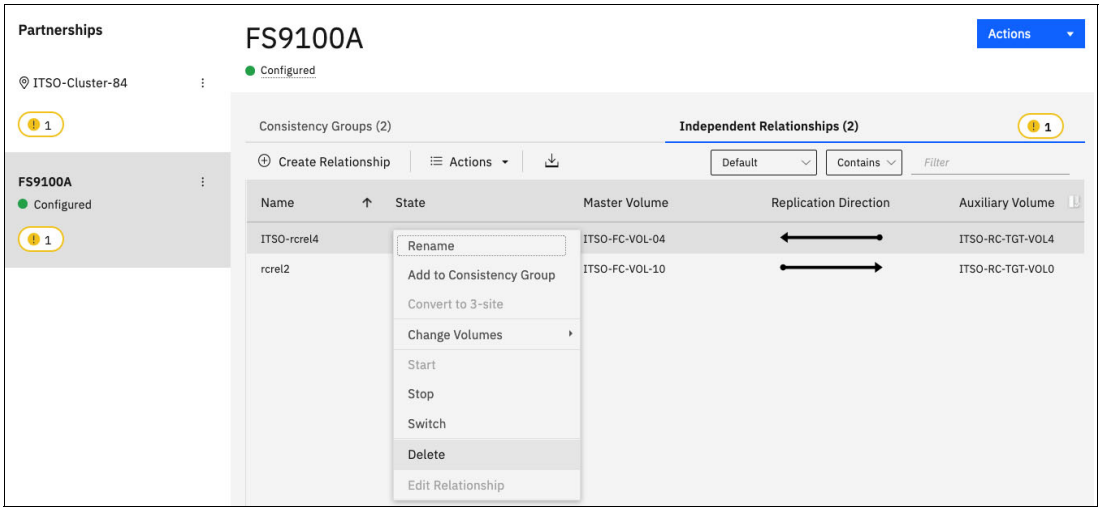


Figure 10-145 Deleting Remote Copy Relationships

3. A confirmation message is displayed that requests that the user enter the number of relationships to be deleted, as shown in Figure 10-146.

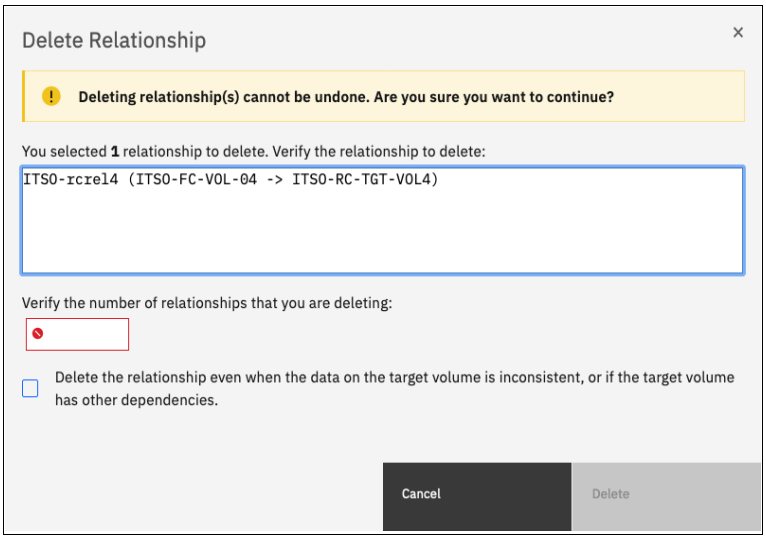


Figure 10-146 Confirmation of relationships deletion

10.11.15 Deleting a consistency group

To delete a 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 10-147.

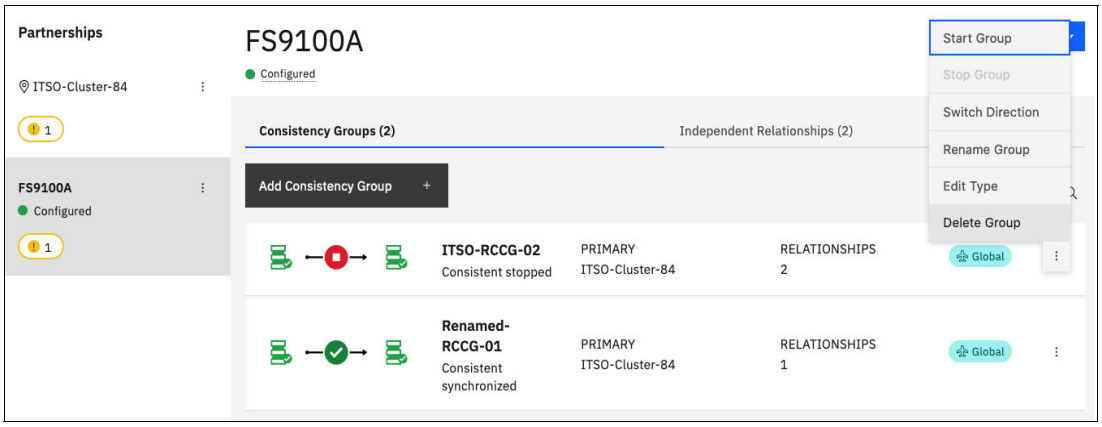


Figure 10-147 Deleting a consistency group

3. A confirmation message is displayed, as shown in Figure 10-148. Click **Yes**.

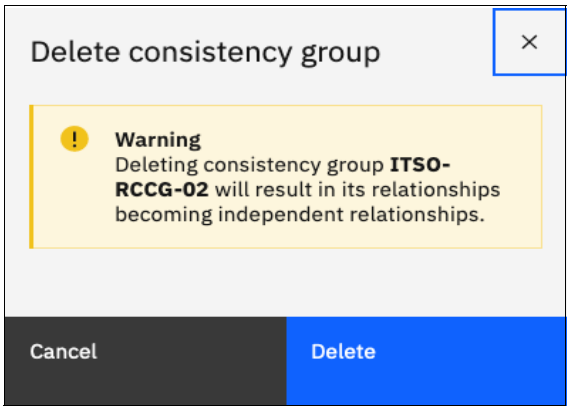


Figure 10-148 Confirmation of a consistency group deletion

Important: Deleting a consistency group does *not* delete its RC mappings.

10.12 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 Spectrum 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 function when using the default values are used[a]
0	20 MiB	1024 MiB	40 TiB of remote mirroring volume capacity

a. 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 10-149.

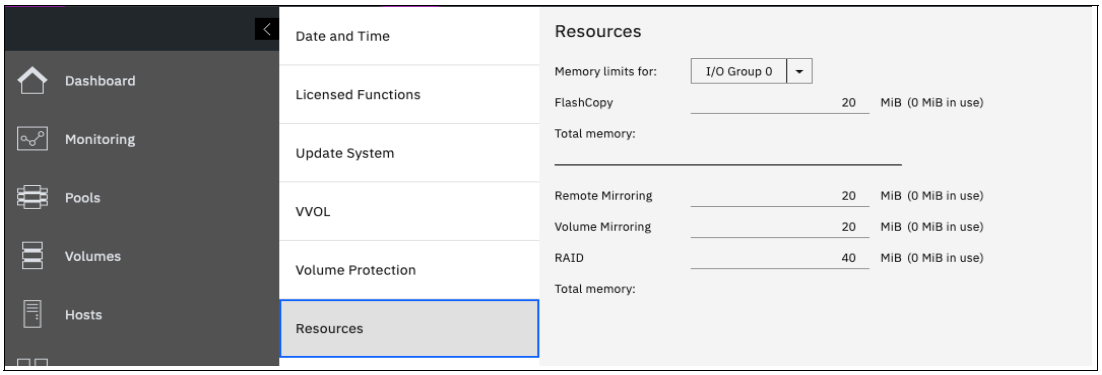


Figure 10-149 Modifying resources allocation

At the time of this writing, this GUI option is not available for other IBM Spectrum Virtualize based systems, so it can be adjusted by using the **chiogrp** CLI command. For more information about the syntax, see this IBM Documentation web page.

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 often is 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. The 985003 and 985004 errors are the IDs of entries in the Spectrum 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 a 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 Spectrum 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 Spectrum Virtualize node ports where each node port comes from a different IBM Spectrum 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 could 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 Spectrum 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.

11

Ownership groups

The ownership groups feature, or object-based access control (OBAC), provides a method of implementing a multi-tenant solution on the IBM SAN Volume Controller systems. Its principles of operations and implementation steps are provided in this chapter.

This chapter includes the following topics:

- 11.1, “Ownership groups principles of operations” on page 782
- 11.2, “Implementing ownership groups on a new system” on page 783
- 11.3, “Migrating objects to ownership groups” on page 788

© Copyright IBM Corp. 2021. All rights reserved.

11.1 Ownership groups principles of operations

Ownership groups allow 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. 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 (for example, superuser) 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

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, it important to understand the following guidelines:

- Users can be in only one ownership group at a time (that applies to both local and remotely authenticated users).
- Objects can be within at most one ownership group.
- Global resources, such as drives, enclosures, arrays, cannot be assigned to ownership groups.
- Global users that do not belong to any ownership group can view and manage (depending on their user role) all resources on the system, including those that belong to an ownership group, and 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 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.

Child pools 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 it can be defined explicitly or inherited from these explicitly 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 IBM Documentation.

When the user logs on to the management GUI or command-line interface (CLI), only resources that can access 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.

11.2 Implementing ownership groups on a new system

This section describes ownership group implementation process for a new system, which has no volumes and users that must be migrated to ownership groups.

11.2.1 Creating an ownership group

Complete the following steps to create an ownership group:

1. To create the first ownership group, select **Access** → **Ownership Groups**, as shown in Figure 11-1. Enter a name for the group and click **Create Ownership Group**.

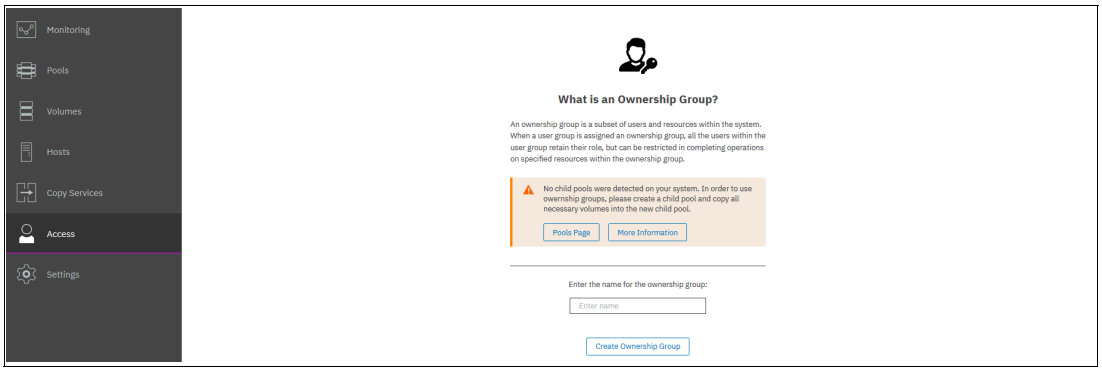


Figure 11-1 Creating a first ownership group

After the first group is created, the pane changes to ownership group mode, as shown in Figure 11-2. The new ownership group has no user groups and no resources assigned to it.

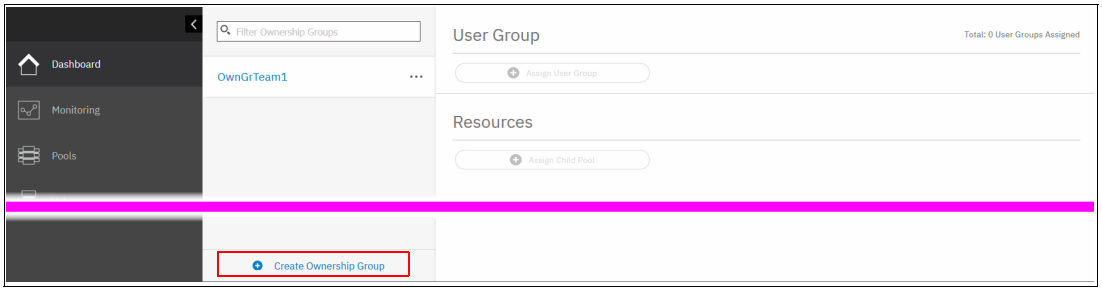


Figure 11-2 Ownership groups management pane

2. To create more ownership groups, click **Create Ownership Group**.

11.2.2 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 click **Create User Group**. The Create User Group window opens, as shown in Figure 11-3. Within it, enter the User Group name and select an ownership group to tie this user group to and select a role for the users in this group.

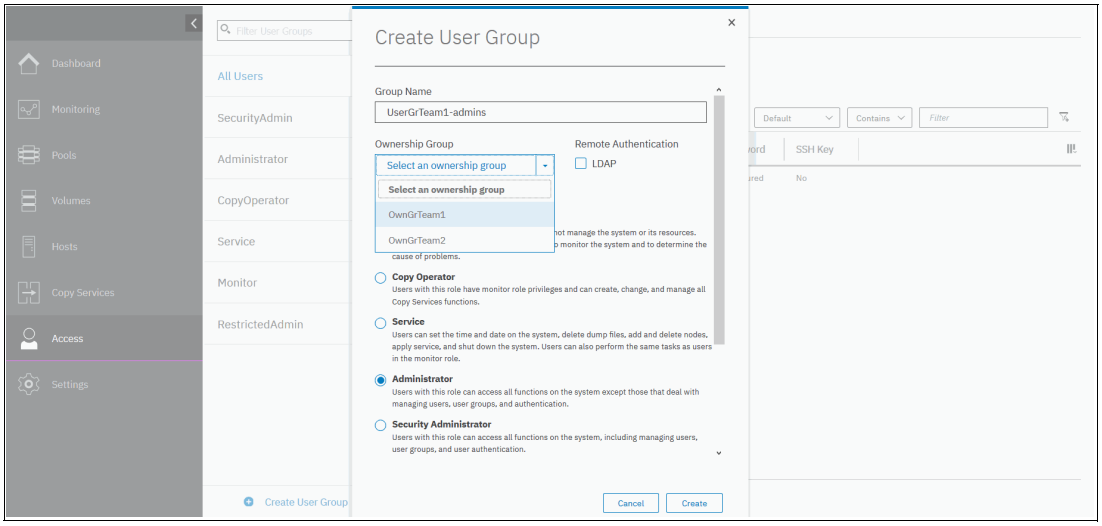


Figure 11-3 Creating and assigning a user group

For more information about user roles, see IBM Documentation.

To create a volume, host, and other objects in an ownership group, users must have an Administrator or Restricted Administrator role. Users with Security Administrator role cannot be assigned to an ownership group.

You can also set up a user group to use remote authentication; that is, Lightweight Directory Access Protocol (LDAP), if it is enabled. To set up remote authentication, select the **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 also must create a user (or users) and assign it to a created user group, as shown in Figure 11-4.

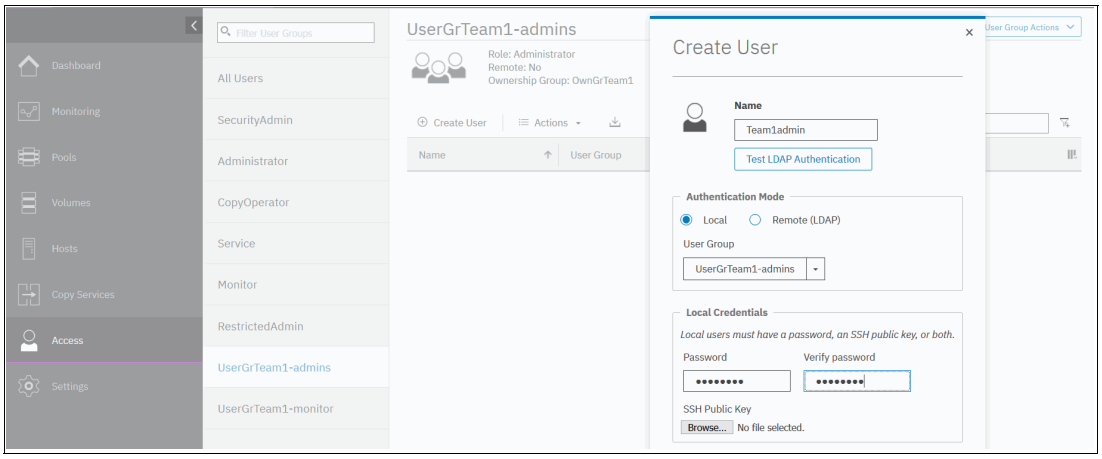


Figure 11-4 Creating a user

You can manage user groups that are assigned to an ownership group by selecting **Access** → **Ownership Groups**, as shown in Figure 11-5. To assign a user group that exists but is not yet assigned to any ownership group, click **Assign User Group**. To unassign, click the "**...**" icon that is next to the assigned user group name.

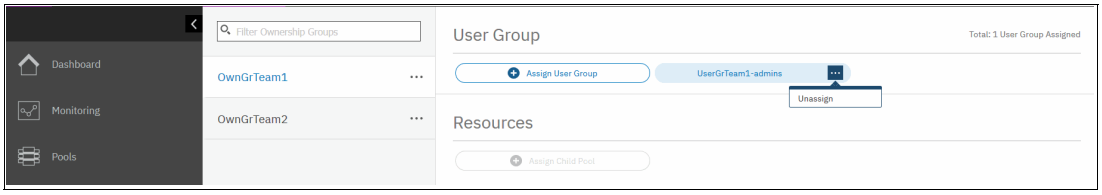


Figure 11-5 Unassigning user groups

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 groups resources.

11.2.3 Creating ownership group resources

To create ownership group volumes and other resources, a child pool must be created and assigned to the ownership group.

Select **Pools** → **Pools**, right-click a parent pool that is designated to be a container for child pools, and click **Create Child Pool**, as shown in Figure 11-6 on page 786.

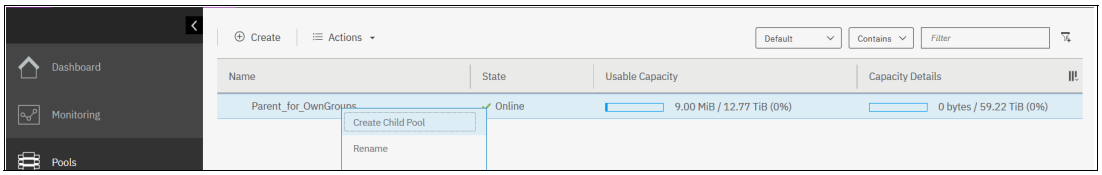


Figure 11-6 Creating a child pool

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 11-7. Ownership group objects can use only capacity that is provisioned for them with the child pool.

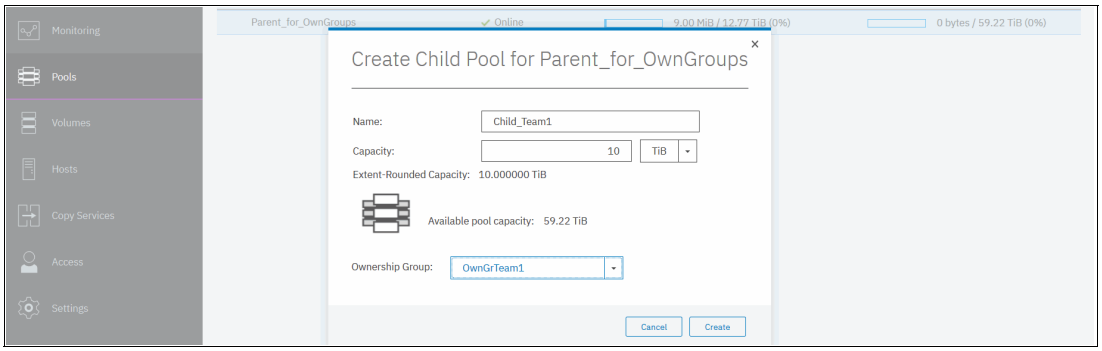


Figure 11-7 Creating a child pool and assigning it to an ownership group

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 pane at **Access** → **Ownership Groups** changes to show the assigned and available resources, as shown in Figure 11-8.

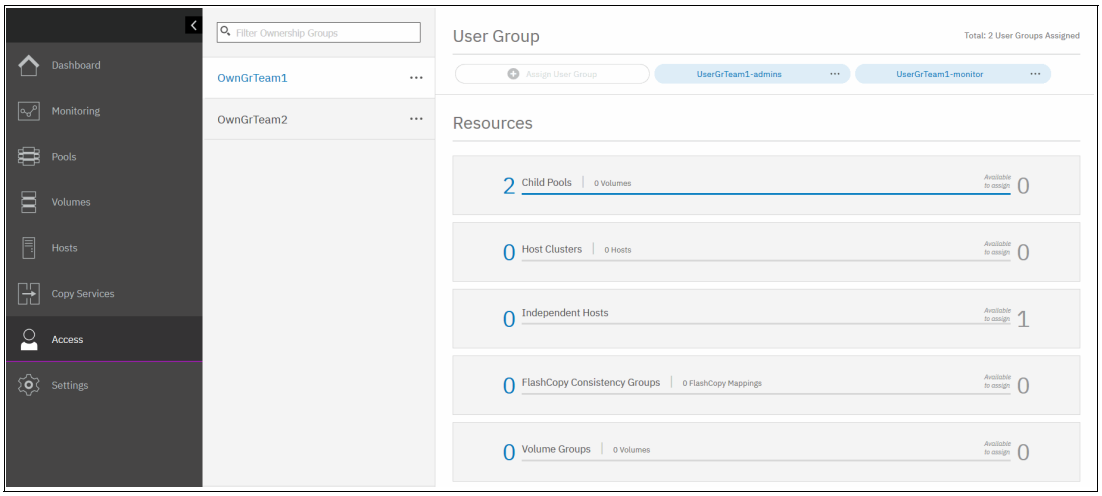


Figure 11-8 Ownership group management pane

Any volumes that are created on a child pool that are assigned to an ownership group inherit 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 these objects, see Chapter 6, “Creating and provisioning volumes” on page 331, Chapter 7, “Hosts” on page 443, and Chapter 10, “Advanced Copy Services” on page 605.

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.

11.2.4 Listing ownership group resources

By default, the Ownership Group attribute is not enabled in the GUI panes that list volumes and other objects that can be owned. For convenience, the system administrator can enable this attribute, as shown in Figure 11-9.

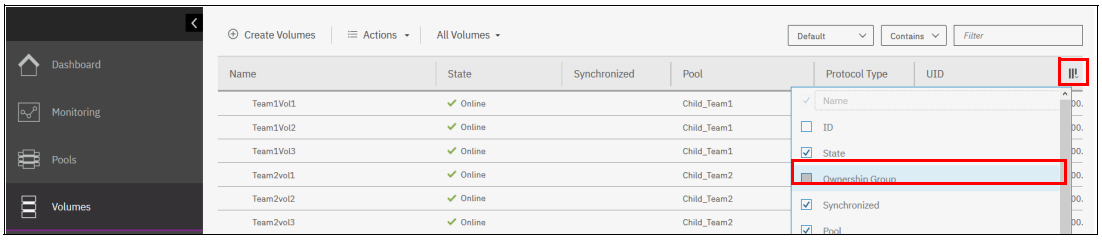


Figure 11-9 Enabling ownership group attribute display

As an example, the volume listing for a system administrator looks as shown in Figure 11-10.

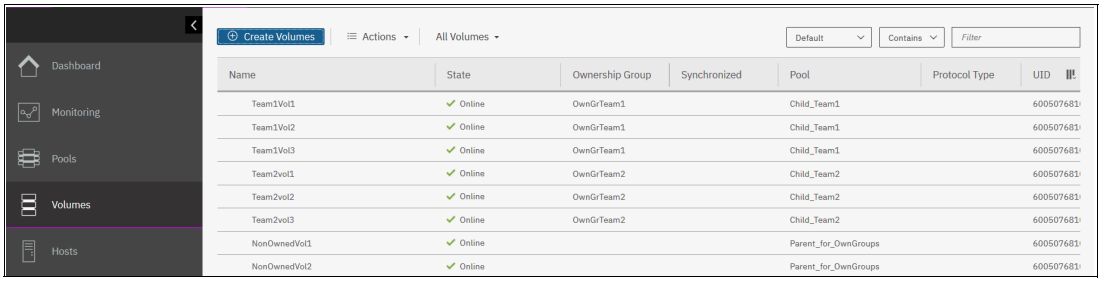


Figure 11-10 Listing volumes for all ownership groups

The system administrator can see and manage resources of all ownership groups, and resources that are not assigned to any of the groups.

When the ownership group user logs in, they can see and manage only resources that are assigned to their group. The example that is shown in Figure 11-11 shows the dashboard pane for an ownership group user with the Administrator role.

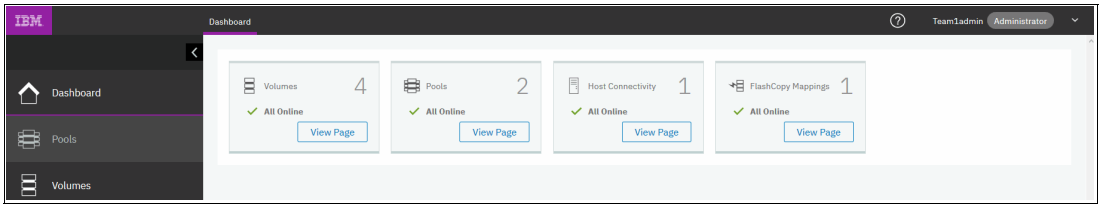


Figure 11-11 Ownership group administrator view

It is different from a dashboard visible to system administrator. Global system performance and capacity parameters are not shown. Only tiles for resources that are assigned to this ownership group are displayed. Eight volumes are configured on a system, as shown in Figure 11-10 on page 787. These volumes 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 only view them.

11.2.5 Actions on ownership groups

The global system administrator can rename or remove an ownership group. Select **Access** → **Ownership Groups**, click the **"..."** icon that is next to the group name, and select the required task, as shown in Figure 11-12.

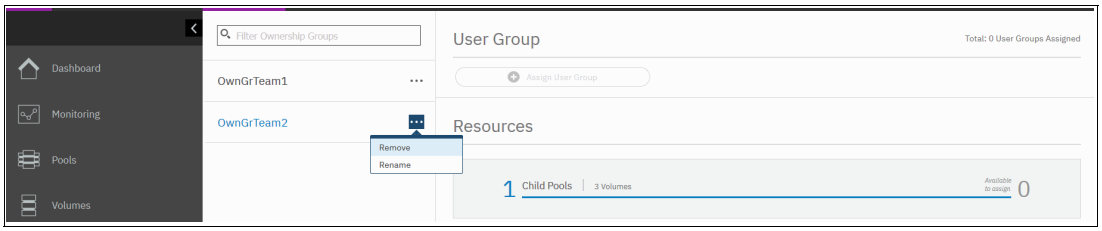


Figure 11-12 Renaming or removing an ownership group

When an ownership group is removed with the GUI, all ownership assignment information for all the objects of the ownership group is removed, but the objects remain configured. Only the system administrator can manage those resources later.

11.3 Migrating objects to ownership groups

If you want to use ownership groups for objects 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 pool is 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.

Consider the following example scenario in which resources are non-disruptively migrated to an ownership group:

1. Create a child pool. Do not assign it to an ownership group yet.

   Migrate volumes that must be assigned to an ownership group to that child pool by using the volume migration or volume mirroring function. The example that is shown in Figure 11-13 shows volume `NonOwnedVol1`, 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.

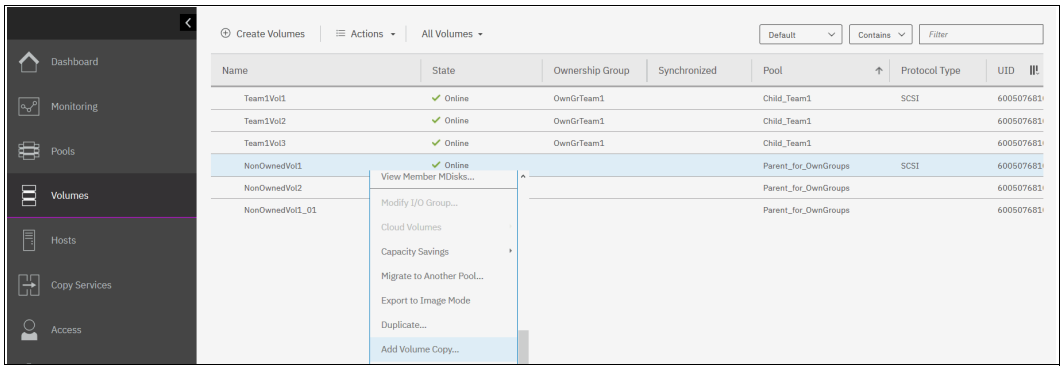


Figure 11-13 Adding volume copy for migration

2. To start migration, right-click the volume and select **Add Volume Copy**. You can also select **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.

   In the **Volume Copy** window, select the child pool that is to be assigned later to an ownership group, as shown in Figure 11-14.

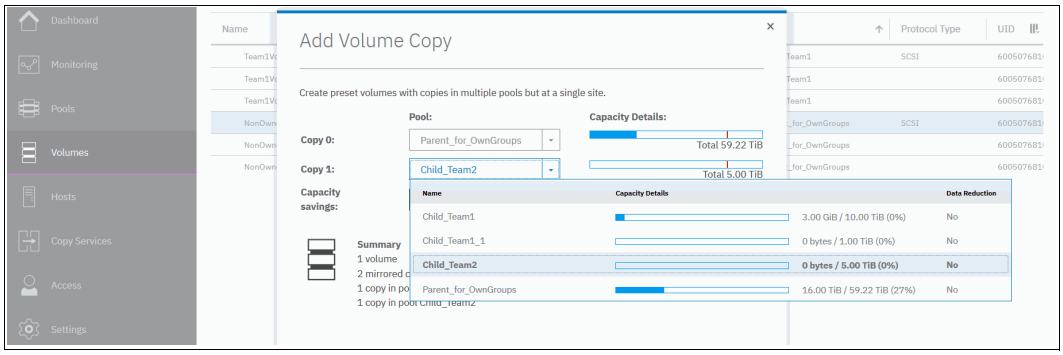


Figure 11-14 Migrating to a child pool

3. Repeat the previous step for all volumes that must belong to an ownership group. Then, remove the source copies.
4. Create an ownership group as described in 11.2.1, “Creating an ownership group” on page 783. Assign a user group to it, as described in 11.2.2, “Assigning users to an ownership group” on page 784.

5. As shown in Figure 11-15, select **Access** → **Ownership Groups**. Then, select the wanted ownership group and click **Assign Child Pool**.



Figure 11-15 Assign child pool to an ownership group

6. Select a child pool to assign, as shown in Figure 11-16.

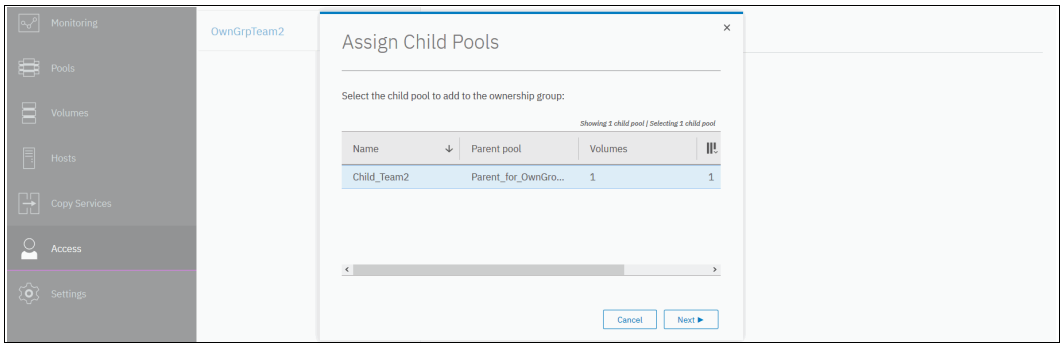


Figure 11-16 Selecting child pool to assign

After you click **Next**, the system notifies you that more resources inherit ownership from a volume and as the volume is mapped to a host, the host becomes an ownership group object, as shown in Figure 11-17.

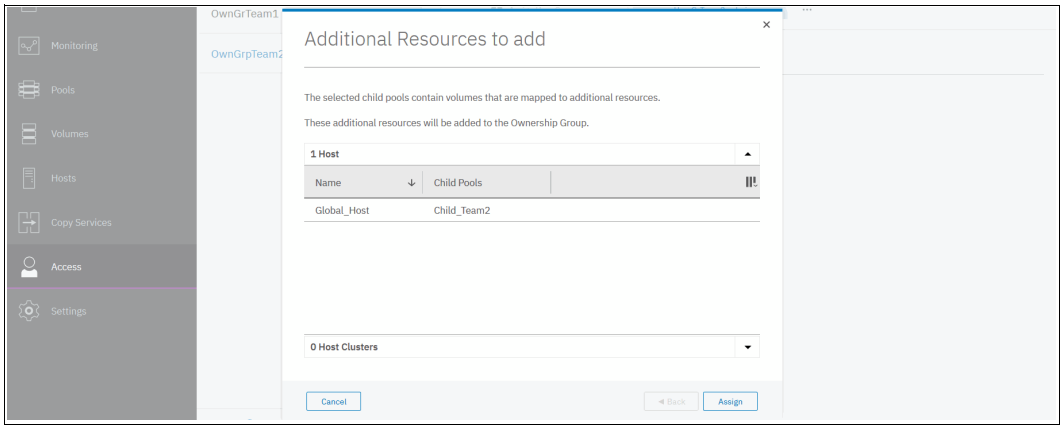


Figure 11-17 More resources to add

As shown in Figure 11-18, a volume and a host both belong to an ownership group. As a host and a volume are in a group, host mapping inherits ownership and becomes a part of an ownership group, too.

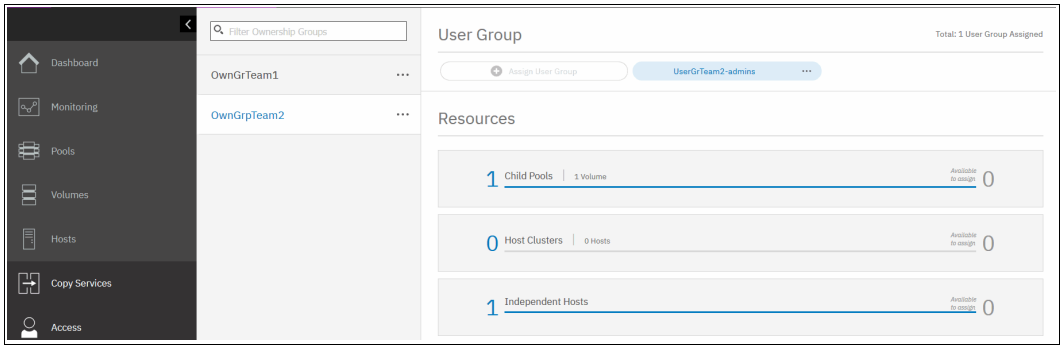


Figure 11-18 Resources of an ownership group

After these steps are completed, a child pool is assigned to an ownership group. If you must migrate volumes to it later, the same process can be used. However, one volume copy is in an owned child pool, and the original copy remains in an unowned parent pool during migration. Such a condition causes Inconsistent ownership, as shown in Figure 11-19.

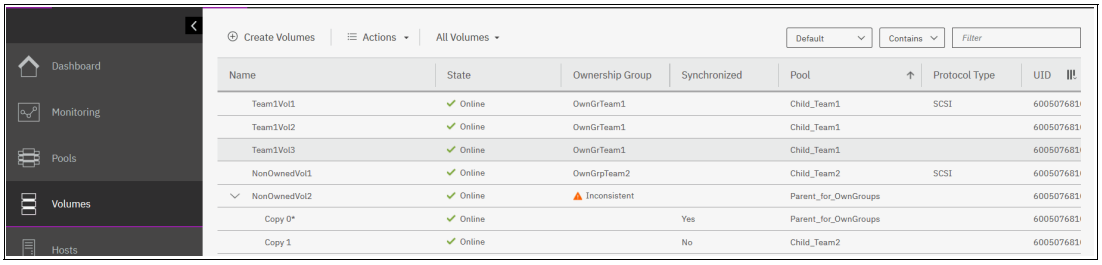


Figure 11-19 Example of inconsistent volume ownership

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

Encryption

Encryption protects against the potential exposure of 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 Spectrum Virtualize support optional encryption of data at-rest.

This chapter includes the following topics:

- 12.2, “Planning for encryption” on page 795
- 12.3, “Defining encryption of data at-rest” on page 796
- 12.4, “Activating encryption” on page 800
- 12.5, “Enabling encryption” on page 809
- 12.6, “Configuring more providers” on page 832
- 12.7, “Migrating between providers” on page 835
- 12.8, “Recovering from a provider loss” on page 838
- 12.9, “Using encryption” on page 839
- 12.10, “Rekeying an encryption-enabled system” on page 848
- 12.11, “Disabling encryption” on page 851

© Copyright IBM Corp. 2021. All rights reserved.

12.1 General types of encryption across IBM Spectrum Virtualize

Within IBM Spectrum Virtualize, IBM SAN Volume Controller, and IBM FlashSystems, 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 next.

12.1.1 Externally virtualized storage

Data is decrypted and encrypted as we issue read/write I/Os to the external storage. You can have an encryption key per storage pool or per child storage pool. Migrating a volume between pools (by using volume mirroring) 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, then 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 managed disks (MDisks) within the storage pool. If an external storage controller supports encryption, you can choose to allow it to encrypt the data instead.

12.1.2 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. Multi-tenancy is possible only if you have more than one array and storage pool, which usually is not practical.

You can migrate volumes from a non-encrypted 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.

12.1.3 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 uses 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 and a cryptographic erase is possible only by deleting the entire storage pool and arrays.

12.2 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 configuring encryption 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 particular 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 will data-at-rest encryption be implemented?
- How can it be demonstrated that the proposed solution realizes the set of goals?

12.3 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 Spectrum Virtualize is defined by the following characteristics:

- *Data-at-rest* means that the data is encrypted on the end device (drives).
- 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.
- XTS-AES-256 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.

12.3.1 Encryption methods

Two types of encryption are available on devices that are running IBM Spectrum Virtualize: hardware encryption and software encryption. Both methods of encryption protect against the potential exposure of sensitive user data that is stored on discarded, lost, or stolen media. Both can also 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 presented in the system, 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 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 (for example, pools) are created as encrypted by default.

12.3.2 Encrypted data

IBM Spectrum Virtualize performs data-at-rest encryption, which is the process of encrypting data that is stored on the end devices, such as physical drives.

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 Spectrum 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 12-1 shows an encryption example. Encrypted disks and encrypted data paths are marked in blue. Unencrypted disks and data paths are marked in red. The server sends unencrypted data to an IBM SAN Volume Controller 2145-DH8 system, which stores hardware-encrypted data on internal disks. The data is mirrored to a remote Storwize V7000 Gen1 system by using RC. The data that is flowing through the RC link is not encrypted. Because the Storwize V7000 Gen1 (2076-324) cannot perform any encryption activities, data on the Storwize V7000 Gen1 is not encrypted.

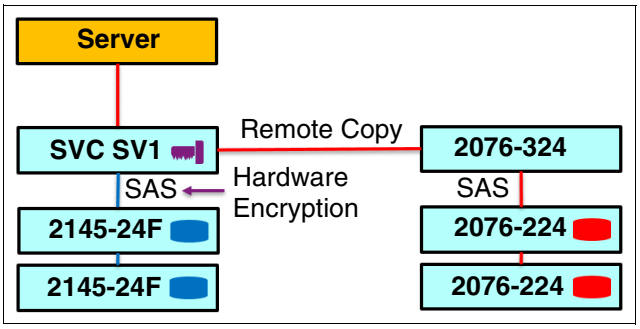


Figure 12-1 Encryption on single site

To enable encryption of both data copies, the Storwize V7000 Gen1 must be replaced by an encryption capable (with optional encryption enabled), IBM Spectrum Virtualize system as shown in Figure 12-2. After such replacement, both copies of data are encrypted, but the RC communication between both sites remains unencrypted.

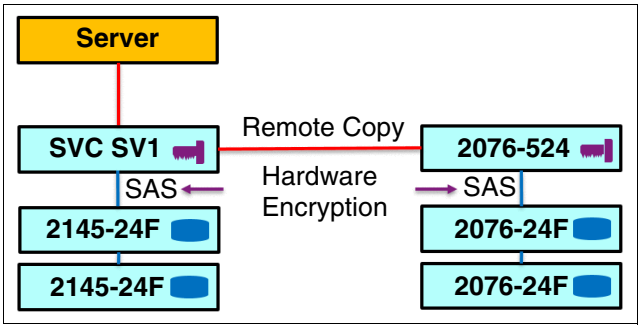


Figure 12-2 Encryption on both sites

Figure 12-3 shows an example configuration that uses software and hardware encryption. Software encryption is used to encrypt an external virtualized storage system (2076-324 in Figure 12-3). Hardware encryption is used for internal, SAS-attached disk drives.

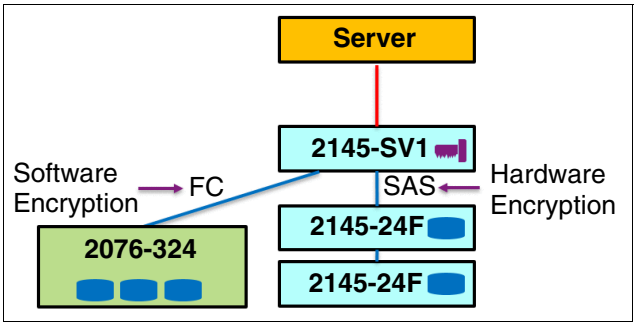


Figure 12-3 Example of software encryption and hardware encryption

Placement of hardware encryption and software encryption in the IBM Spectrum Virtualize code stack is shown in Figure 12-4. Because compression is performed before encryption, it is possible to realize the benefits of compression for the encrypted data.

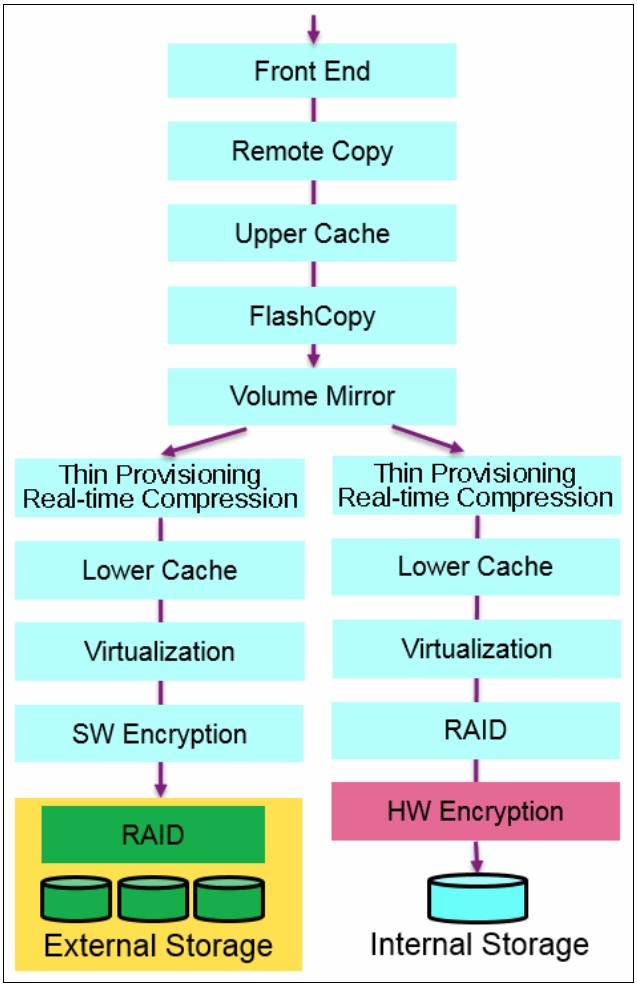


Figure 12-4 Encryption placement in the IBM Spectrum Virtualize software stack

Each volume copy can use different encryption methods (hardware and software). It also can have volume copies with different encryption status (encrypted versus unencrypted). The encryption method depends only on the pool that is used for the specific copy. You can migrate data between different encryption methods by using volume migration or volume mirroring.

12.3.3 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.
- Two 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.

12.3.4 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 runs out, 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.

12.4 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. It is mandatory when purchasing your Spectrum Virtualize storage system with an encryption license that you add the three USB keys to your configuration. See the Note at the end of this section with example output from the EConfig tool when Encryption is enabled on a Spectrum Virtualize Storage System.
- You must activate the license that you purchased.
- Encryption must be enabled.

Activation of the license can be performed in one of two ways:

- Automatic activation: Used when you have the authorization code and the workstation that is being used to activate the license has access to 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 Spectrum 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 Spectrum 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.

12.4.1 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 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/search for IBM Data Storage Feature Activation to retrieve your license keys.

Ensure that you have the following information:

- 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 12.4.5, “Manual license activation” on page 808.

12.4.2 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**, as shown in Figure 12-5.

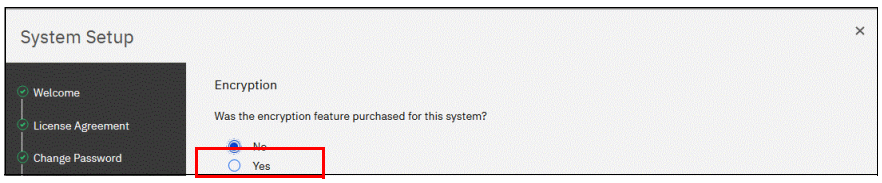


Figure 12-5 Encryption activation during initial system setup

The Encryption window displays information about your storage system, as shown in Figure 12-6.

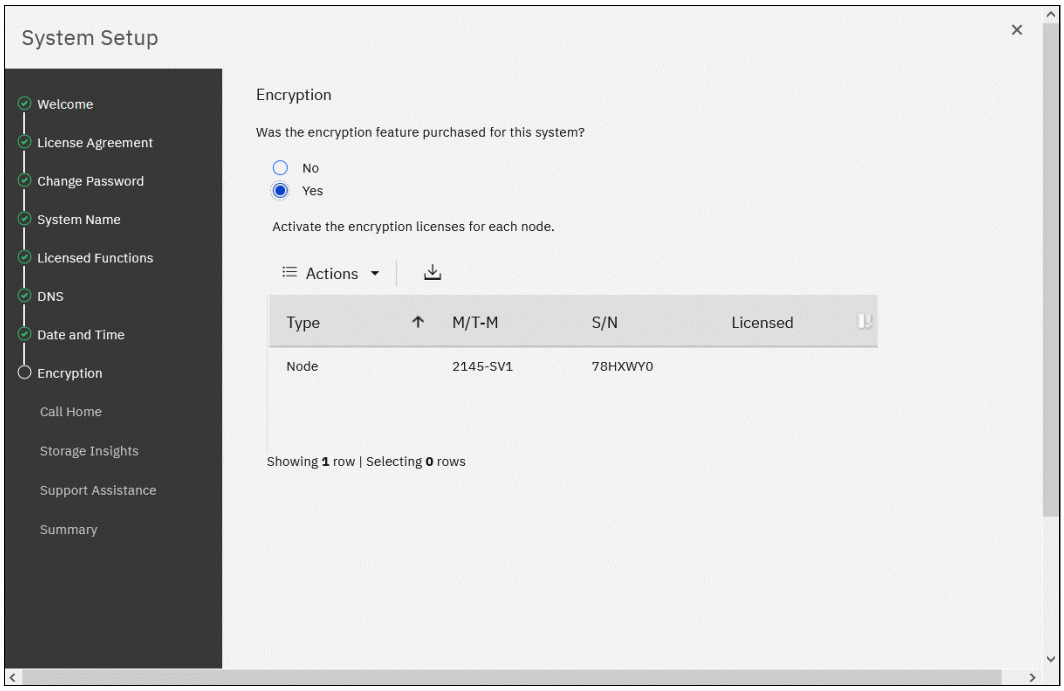


Figure 12-6 Information storage system during initial system setup

Note: If you configured cluster on one of the nodes, and have not added the node to the cluster, 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 with two license activation options (**Activate License Automatically** and **Activate License Manually**), as shown in Figure 12-7. Use either option to activate encryption.

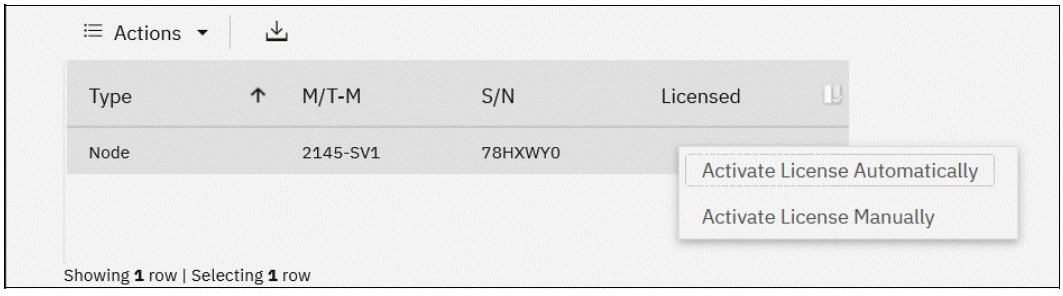


Figure 12-7 Selecting license activation method

For more information about how to complete the automatic activation process, see 12.4.4, “Activating the license automatically” on page 805. For more information about how to complete a manual activation process, see 12.4.5, “Manual license activation” on page 808.

3. After either activation process is complete, a green check mark is shown in the column that is labeled “Licensed” next to a node for which the license was enabled. You can proceed with the initial system setup by clicking **Next** (see Figure 12-8).

Note: Every enclosure needs an active encryption license before you can enable encryption on the system. Attempting to add a non-licensed enclosure to an encryption-enabled system fails.

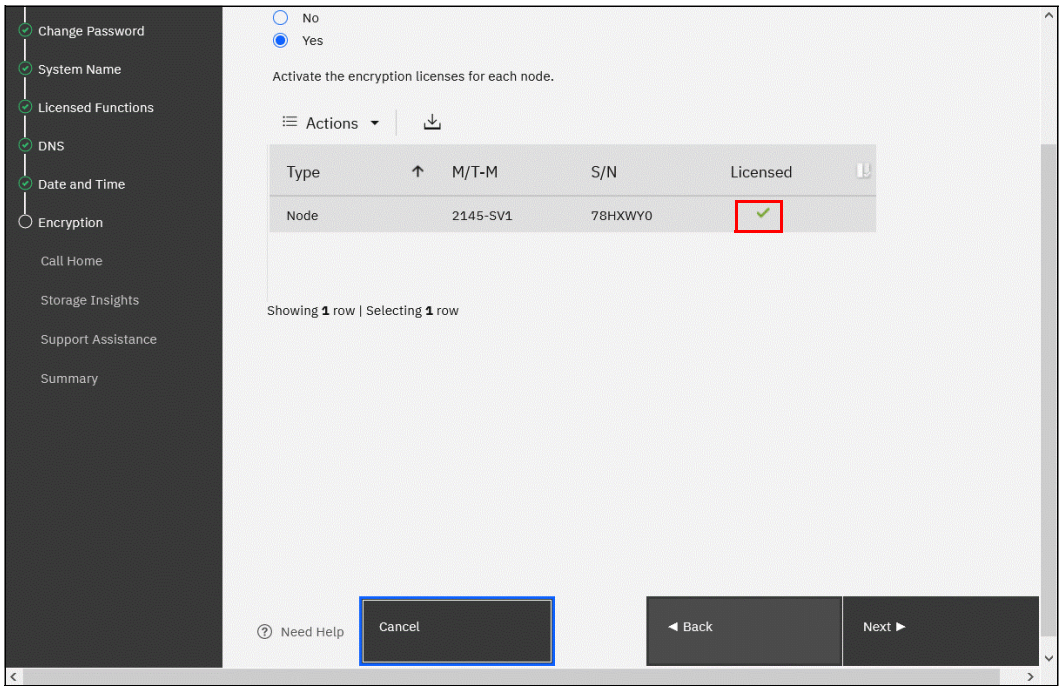


Figure 12-8 Successful encryption license activation during initial system setup

12.4.3 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 12-9.

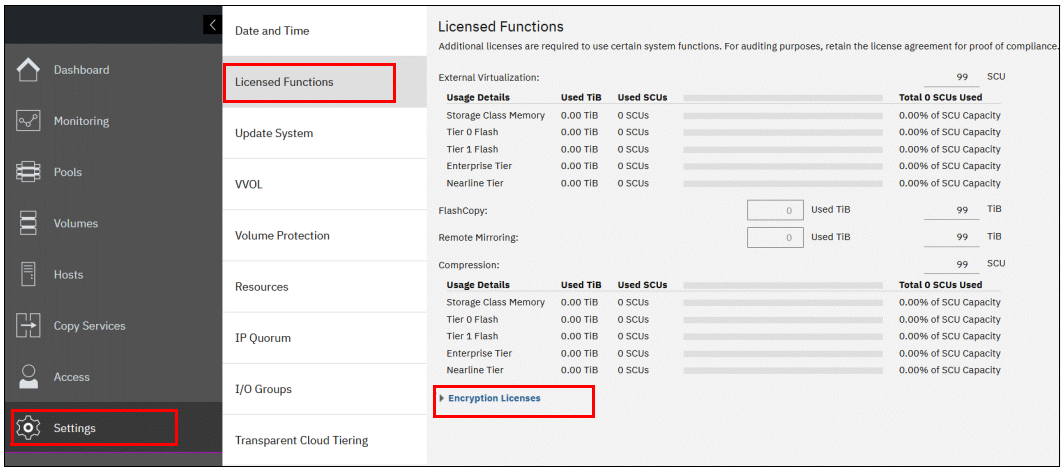


Figure 12-9 Expanding Encryption Licenses section on the Licensed Functions window

3. The Encryption Licenses window displays information about your nodes. Right-click the enclosure on which you want to install an encryption license. This action opens a menu with two license activation options (**Activate License Automatically** and **Activate License Manually**), as shown in Figure 12-10. Use either option to activate encryption.

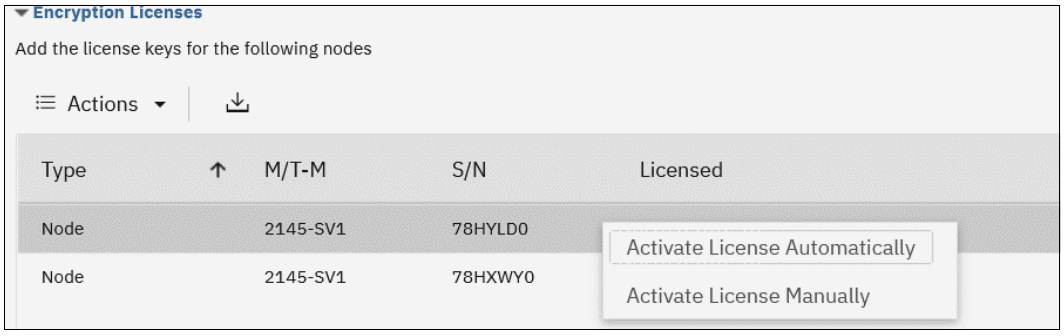


Figure 12-10 Select the node on which you want to enable the encryption

For more information about how to complete an automatic activation process, see 12.4.4, “Activating the license automatically” on page 805. For more information about how to complete a manual activation process, see 12.4.5, “Manual license activation” on page 808.

After either activation process is complete, a green check mark is shown in the column that is labeled “Licensed” for the node, as shown in Figure 12-11.

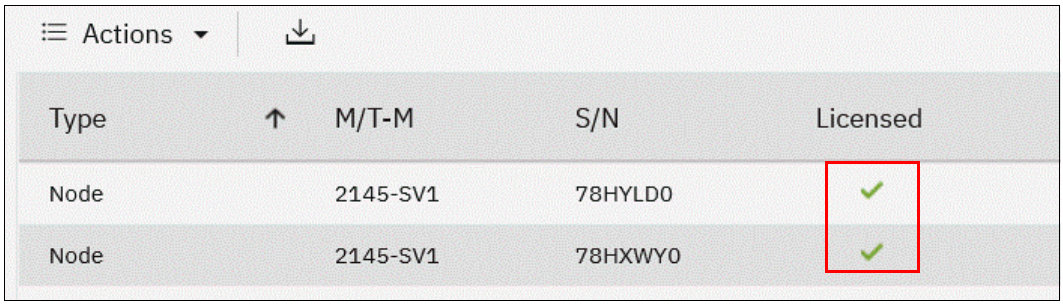


Figure 12-11 Successful encryption license activation on a running system

12.4.4 Activating the license automatically

The automatic license activation is the faster method to activate the encryption license for IBM Spectrum 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 12-12.

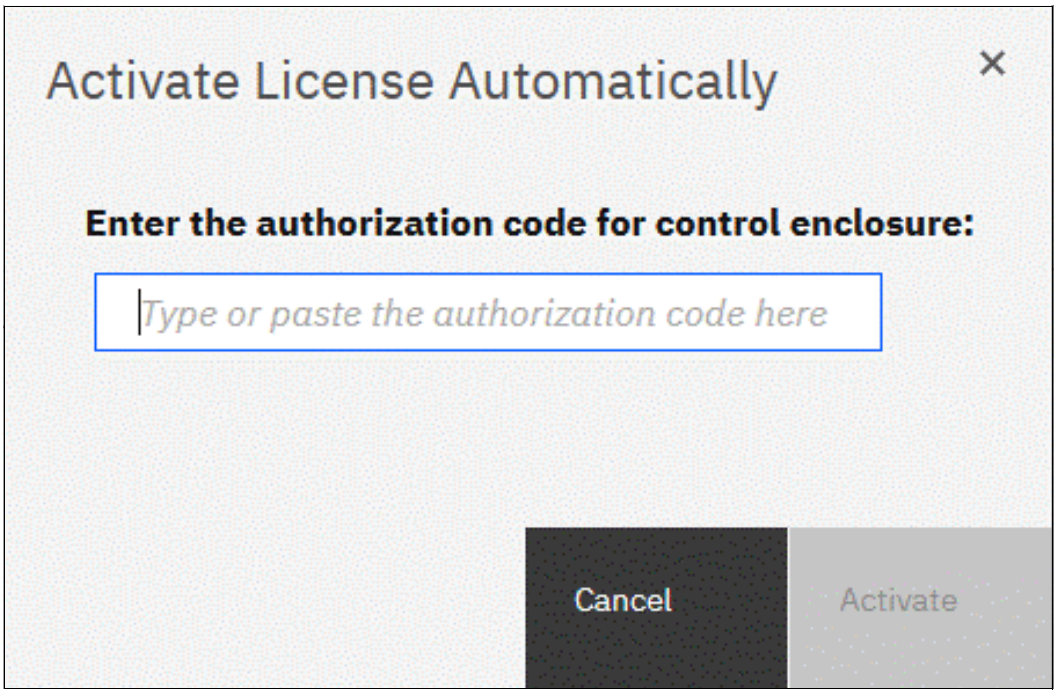


Figure 12-12 Encryption license Activate License Automatically window

2. Enter the authorization code that is specific to the node that you selected, as shown in Figure 12-13. Click **Activate**.

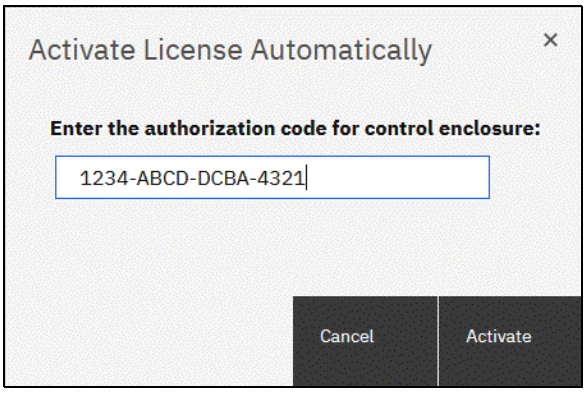


Figure 12-13 Entering an authorization code

The system connects to IBM to verify the authorization code and retrieves the license key. Figure 12-14 shows a window that is displayed during this connection process. If everything works correctly, the procedure takes less than a minute.

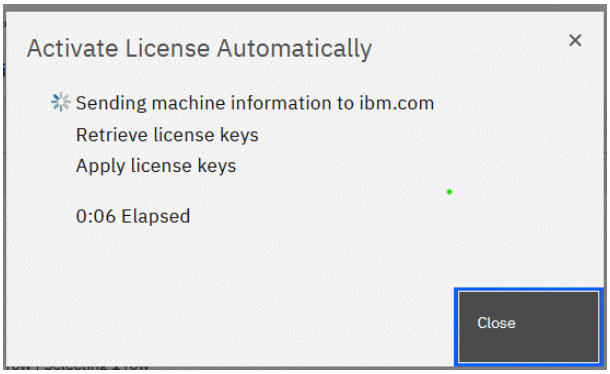


Figure 12-14 Activating encryption

After the license key is retrieved, it is automatically applied, as shown in Figure 12-15.

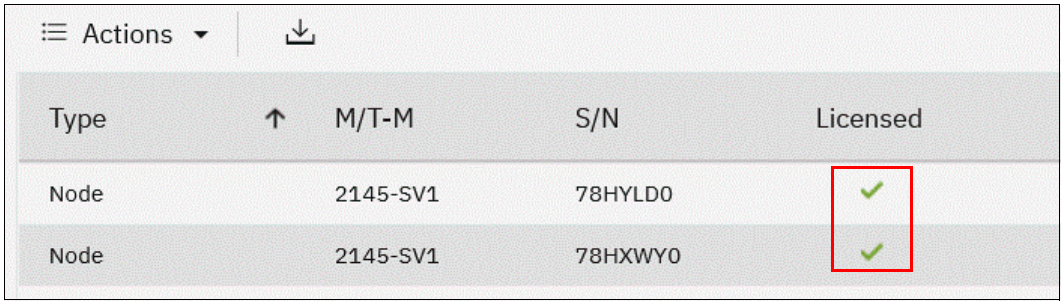


Figure 12-15 Successful encryption license activation

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 12-16).

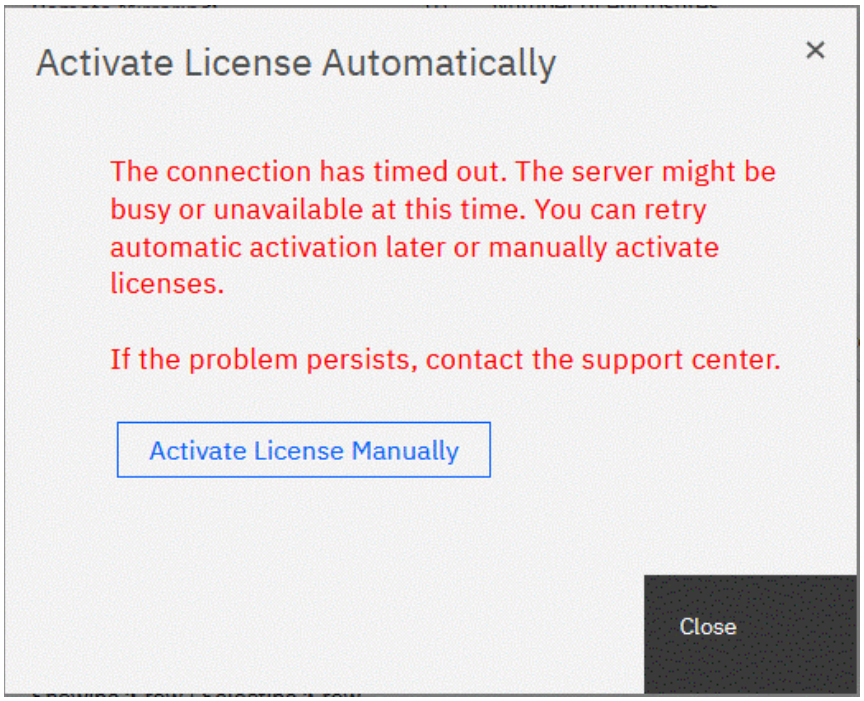


Figure 12-16 Authorization code failure

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 12.4.5, “Manual license activation” on page 808.

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 appropriate activation process. If you use a license key when the system expects an authorization code, the system displays an error message.

12.4.5 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 12-17.

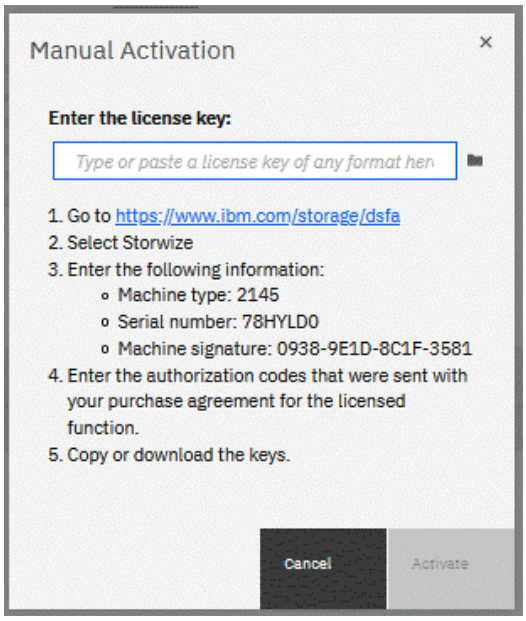


Figure 12-17 Manual encryption license activation window

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 in 12.4.1, “Obtaining an encryption license” on page 801.
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 12-18.

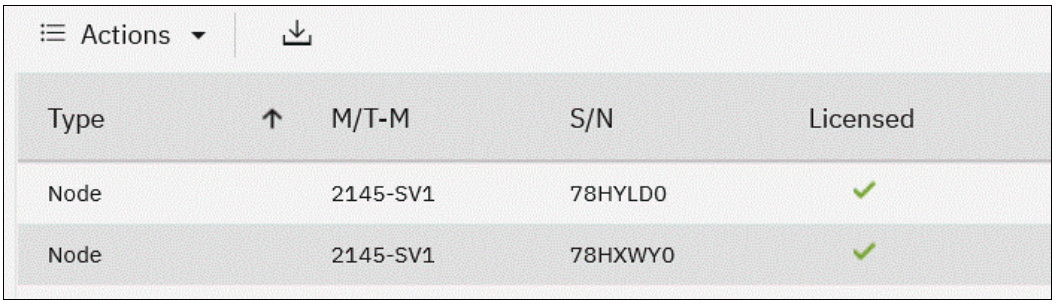


Figure 12-18 Successful encryption license activation

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 12-19.

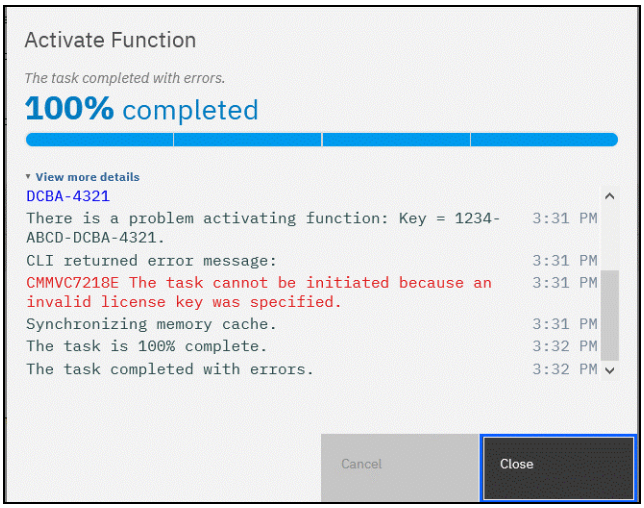


Figure 12-19 License key failure

12.5 Enabling encryption

This section describes the process to create and store system master access key copies, 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.

Two types of key servers are supported by IBM Spectrum Virtualize:

- IBM Security Guardium Key Lifecycle Manager (SGKLM), which was introduced in IBM Spectrum Virtualize V7.8.
- Gemalto SafeNet KeySecure, introduced in IBM Spectrum Virtualize V8.2.

For more information about supported key servers, see Supported Key Servers - IBM Spectrum Virtualize.

IBM Spectrum Virtualize code V8.1 introduced the ability to define up to four encryption key servers, which is a preferred configuration because it increases key provider availability. In this version, support for simultaneous use of both USB flash drives and 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 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.
- Possibility to use hardware security modules (HSMs) for encryption key generation.
- Ability to replicate keys between servers and perform automatic backups.
- Implementations follow an open standard (Key Management Interoperability Protocol [KMIP]) that aids in interoperability.
- Ability to audit operations related to key management.
- Ability to separately manage encryption keys and physical access to storage systems.

USB flash drives have the following characteristics:

- Physical access to the system might be required to process a rekey operation.
- No moving parts with almost no read or write operations to the USB flash drive.
- 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 security of the encryption keys. Pay special attention to ensure secure creation, management, and storage of the encryption keys.

12.5.1 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 12-20.

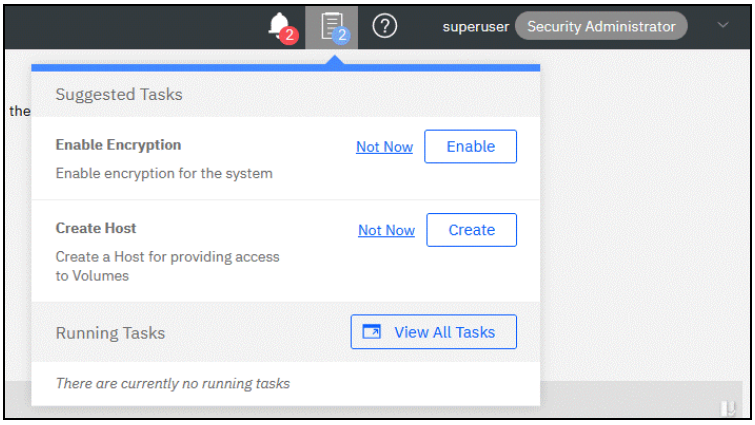


Figure 12-20 Enable Encryption from the Suggested Tasks *window*

You can also click **Settings** → **Security** → **Encryption** and then, click **Enable Encryption**, as shown in Figure 12-21.

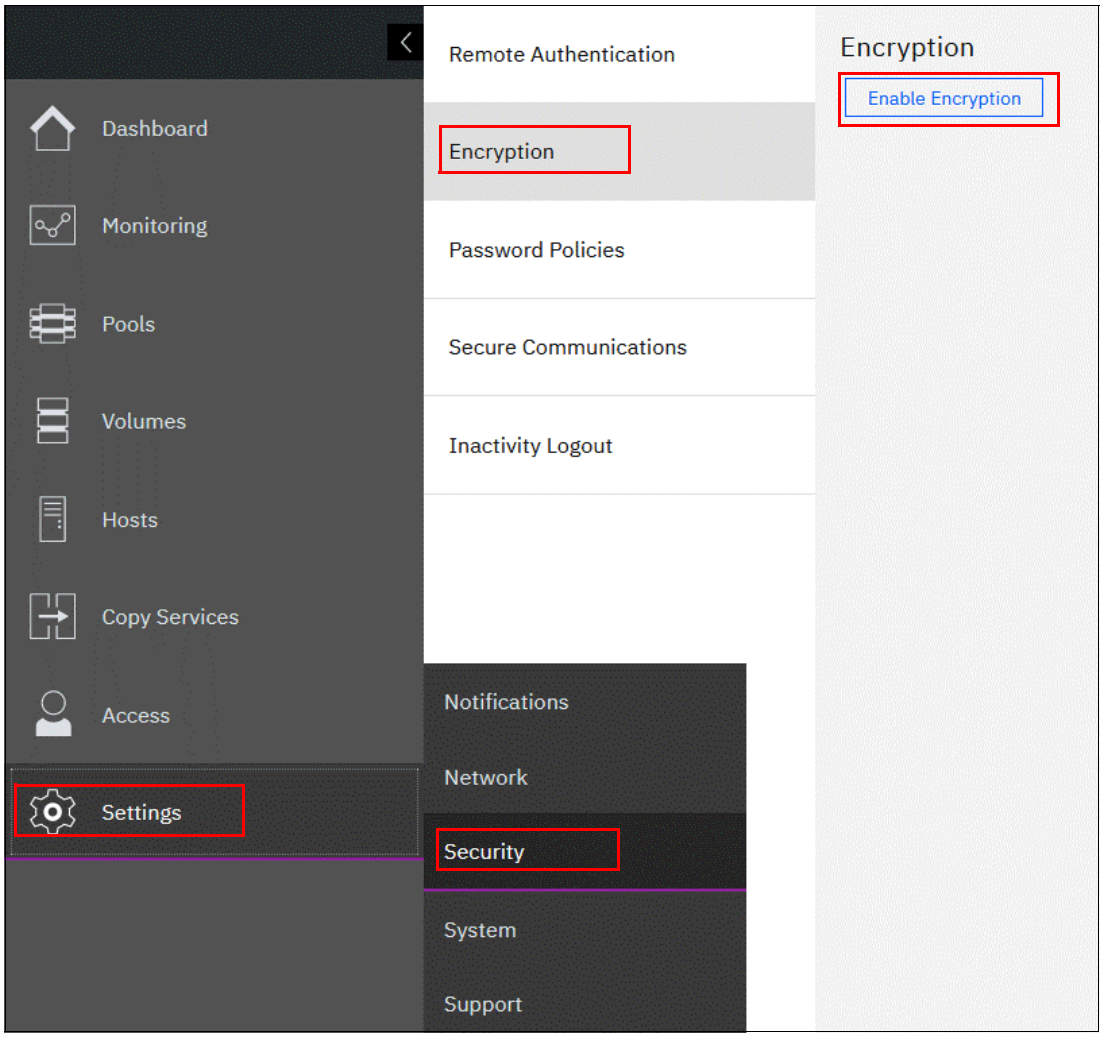


Figure 12-21 Enable Encryption from the Security pane

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 12-22. You can enable one or both providers.

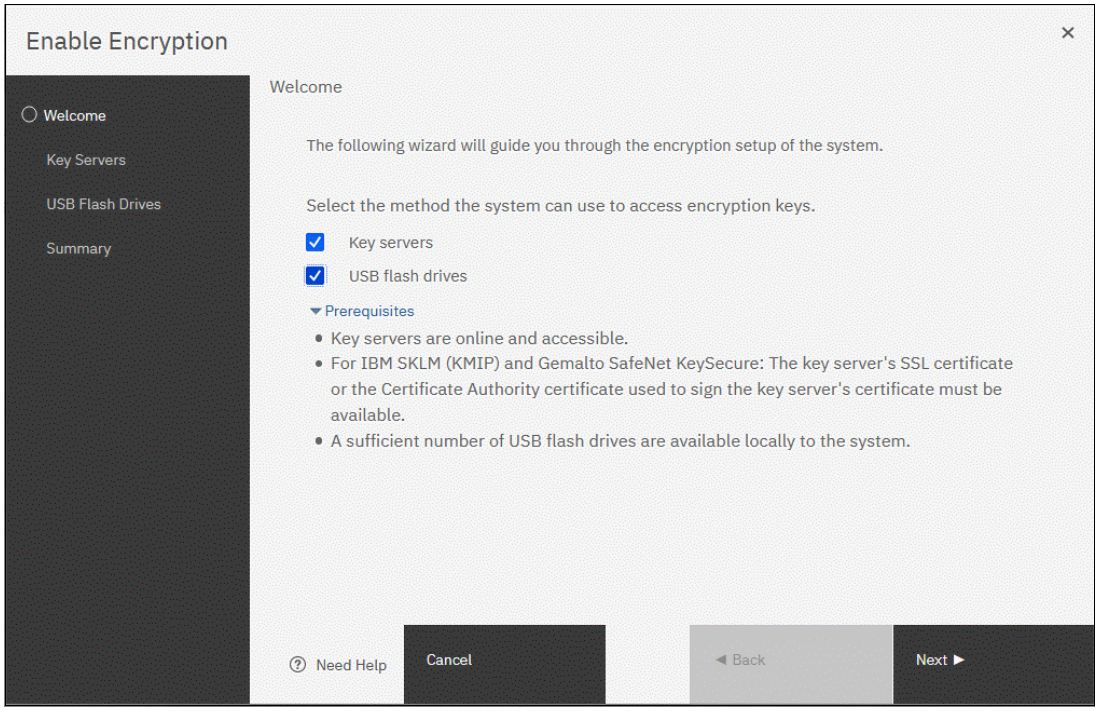


Figure 12-22 Enable Encryption wizard Welcome window

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 12.5.2, “Enabling encryption by using USB flash drives” on page 812.

For more information about how to enable encryption by using key servers as the sole encryption key provider, see 12.5.3, “Enabling encryption by using key servers” on page 816.

12.5.2 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 key 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 12-24 on page 815 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 are 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 used to store the encryption key come from a faulty batch.

Every encryption key copy must be stored securely to maintain 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 click **Next**, as shown in Figure 12-23.

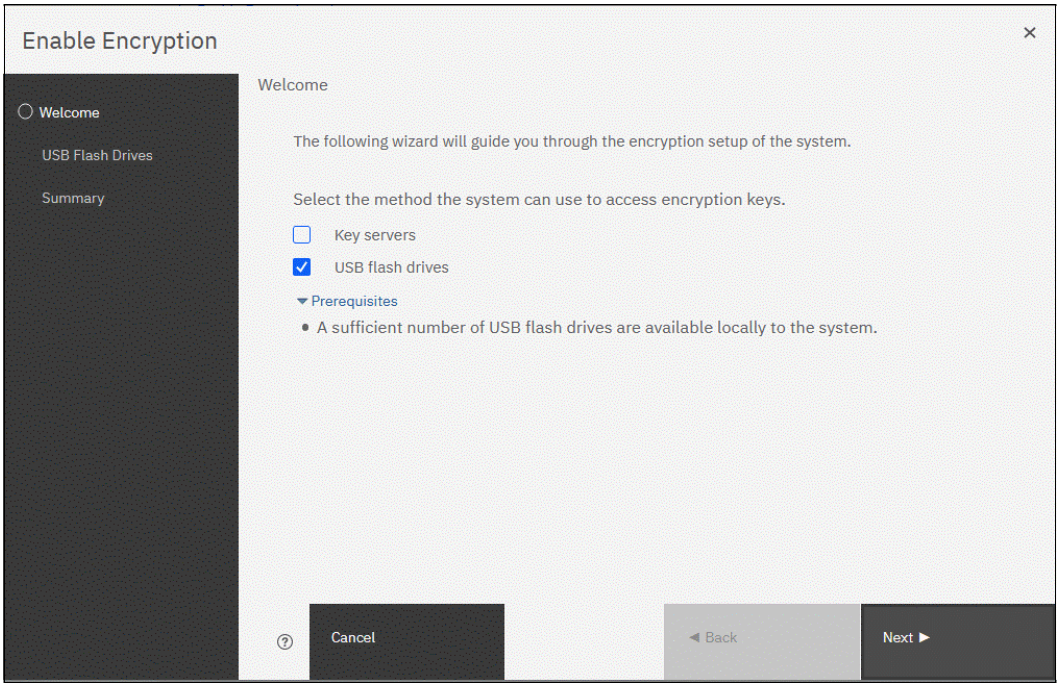


Figure 12-23 Selecting USB flash drives in the Enable Encryption wizard

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.

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 12-24.

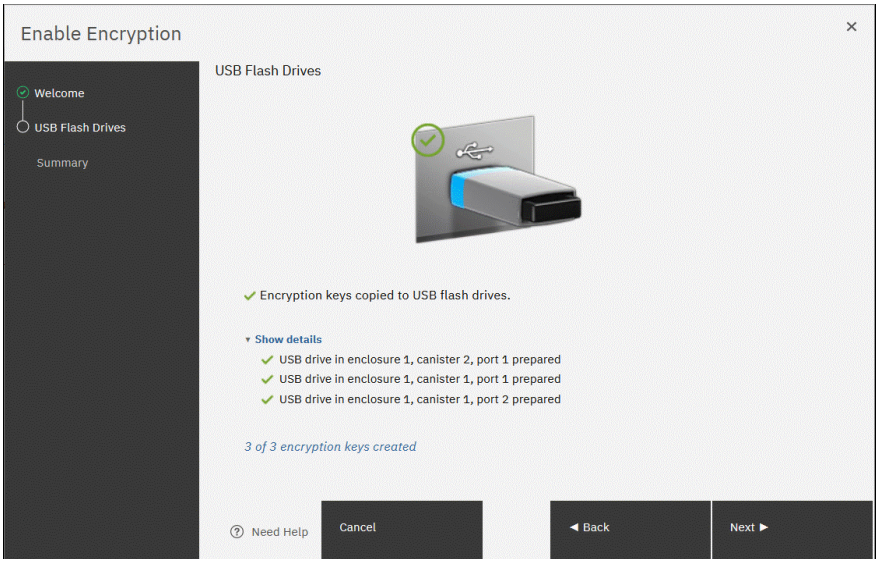


Figure 12-24 Writing the master access key to USB flash drives

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 12-25. Click **Finish** to finalize the encryption enablement.



Figure 12-25 Commit the encryption enablement

You receive a message confirming that the encryption is now enabled on the system, as shown in Figure 12-26.

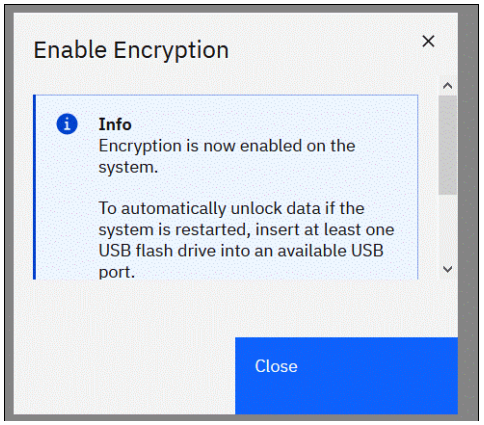


Figure 12-26 Encryption enabled message that uses USB flash drives

4. Confirm that encryption is enabled and verify which key providers are in use by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-27.

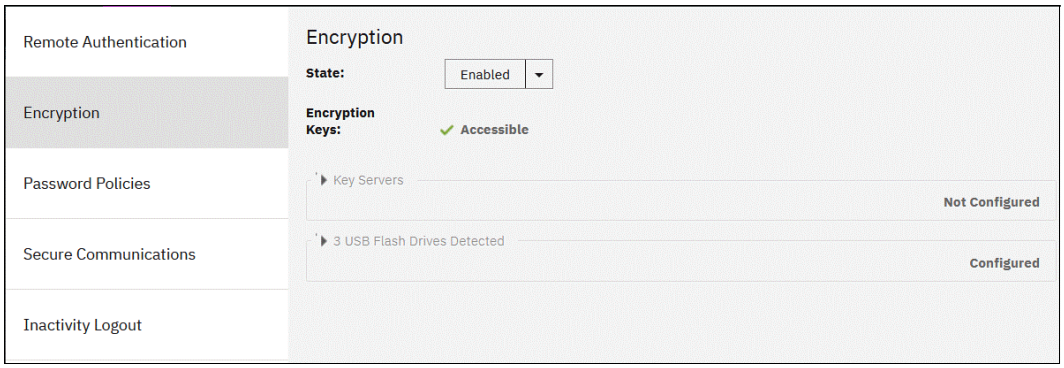


Figure 12-27 Encryption view showing using USB flash drives as the enabled provider

12.5.3 Enabling encryption by using key servers

A key server is a centralized system that receives and then distributes encryption keys to its clients, including IBM Spectrum Virtualize systems.

IBM Spectrum Virtualize supports use of the following key servers as encryption key providers:

- Security Guardium® Key Lifecycle Manager (SGKLM)
- Gemalto SafeNet KeySecure

Note: Support for SGKLM was introduced in IBM Spectrum Virtualize V7.8. Support for Gemalto SafeNet KeySecure was introduced in IBM Spectrum Virtualize V8.2.1.

SGKLMSGKLM and SafeNet KeySecure support 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 features encryption 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 Spectrum Virtualize code V8.1 and later supports up to four key server objects that are defined in parallel. However, only one key server type (SGKLM or KeySecure) 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.

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 Spectrum 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 an `IBM Spectrum_VIRT` device group for IBM Spectrum Virtualize systems.

For more information about completing these tasks, see the SGKLM 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 click **Next**, as shown in Figure 12-28 on page 818.

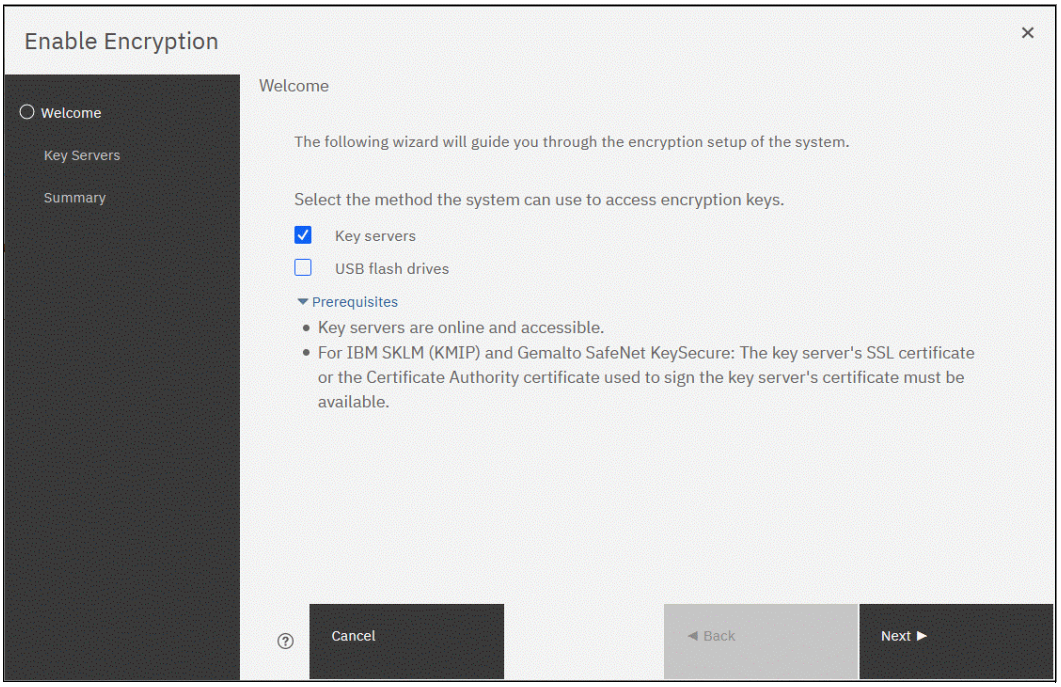


Figure 12-28 Selecting Key server as the only provider in the Enable Encryption wizard

3. Select **IBM SKLM (with KMIP)** as the key server type, as shown in Figure 12-29.

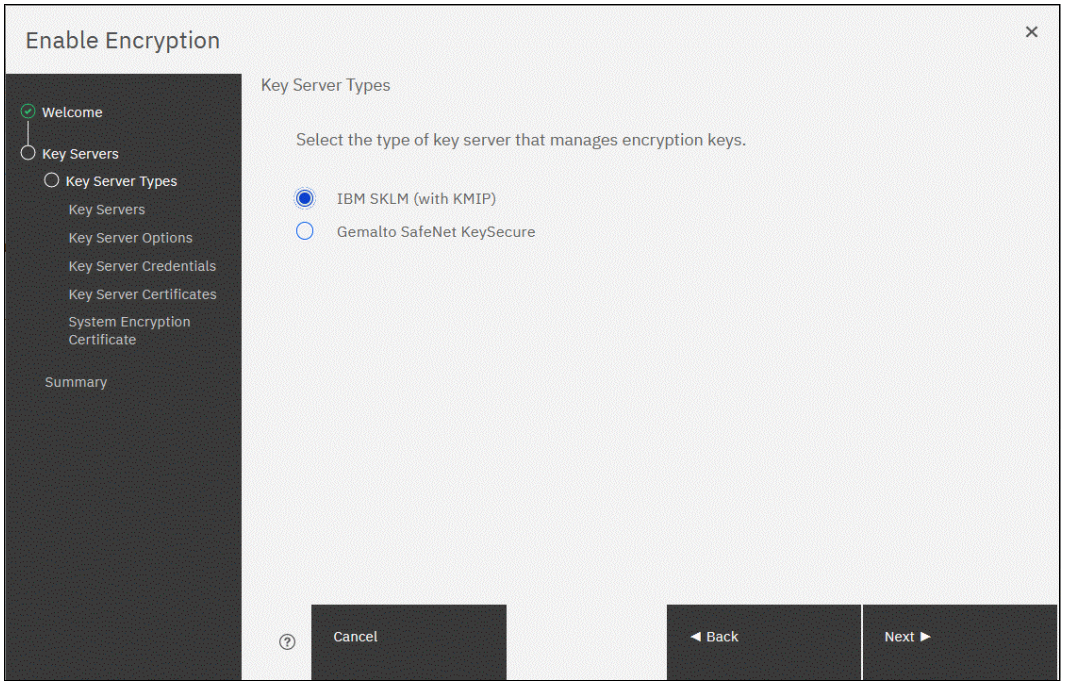


Figure 12-29 Selecting SGKLM as key server type

4. The wizard moves to the Key Servers tab, as shown in Figure 12-30. 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 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 12-30.

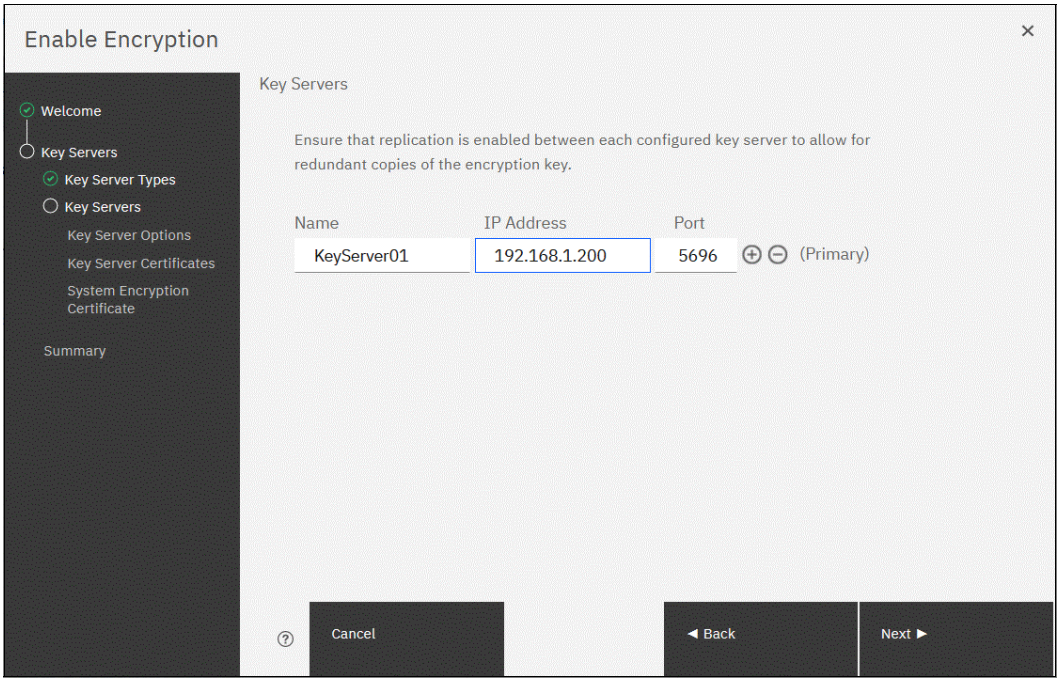


Figure 12-30 Configuration of the primary SGKLM server

5. If you want to add secondary SGKLM servers, click the “**+**” symbol and enter the data for secondary SGKLM servers (see Figure 12-31). You can define up to four SGKLM servers. Click **Next** when you are done.

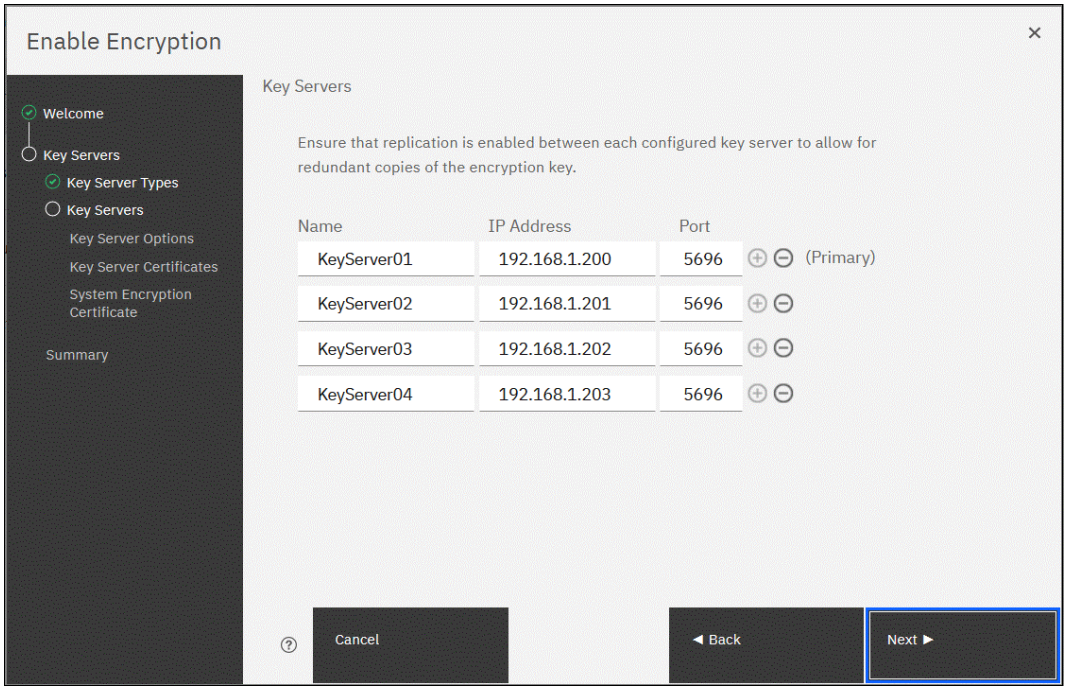


Figure 12-31 Configuring multiple SGKLM servers

6. The next window in the wizard is a reminder that IBM Spectrum_VIRT device group that is dedicated for IBM Spectrum 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 12-32.

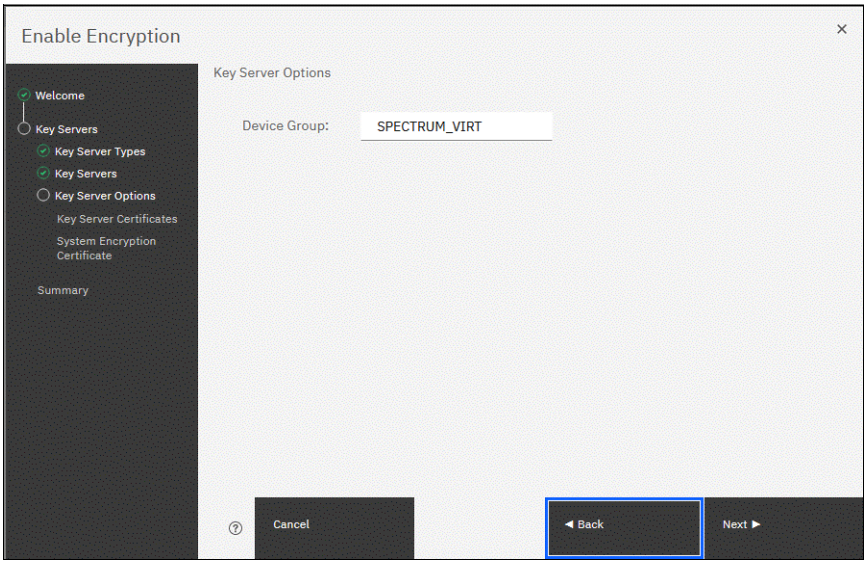


Figure 12-32 Checking key server device group

7. Enable secure communication between the IBM Spectrum 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 from each of key servers directly.

   After uploading any of the certificates in the window that is shown in Figure 12-33, click **Next**.

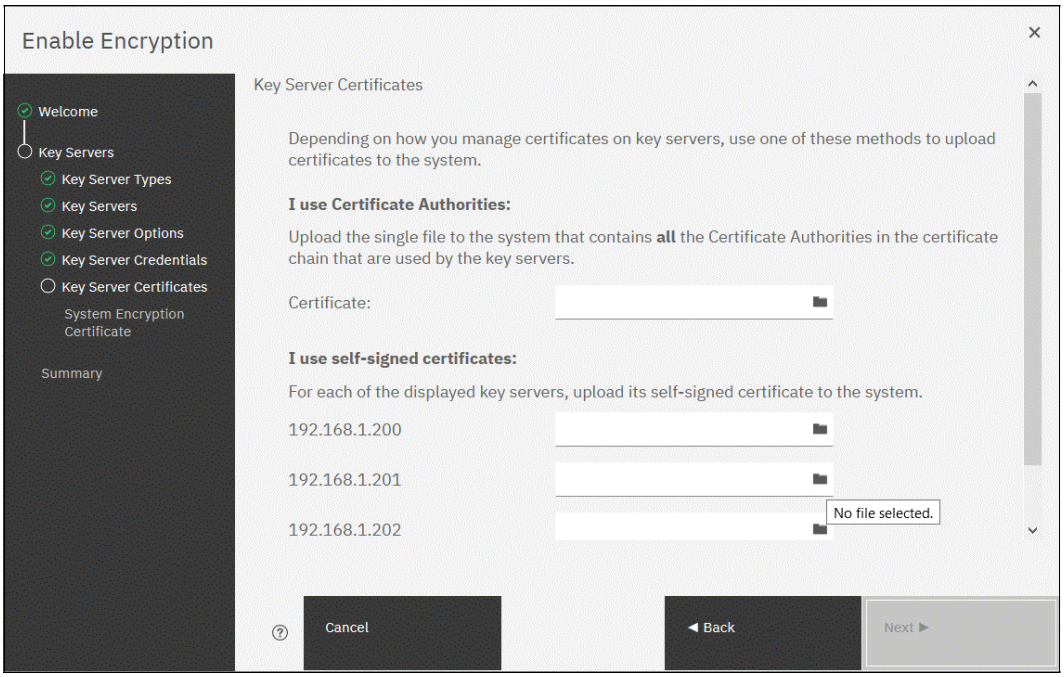


Figure 12-33 Uploading key servers or certificate authority SSL certificate

8. Configure the SGKLM key server to trust the public key certificate of the IBM Spectrum Virtualize system. You can download the IBM Spectrum Virtualize system public SSL certificate by clicking **Export Public Key**, as shown in Figure 12-34. Install this certificate in the SGKLM key server in the IBM Spectrum_VIRT device group.

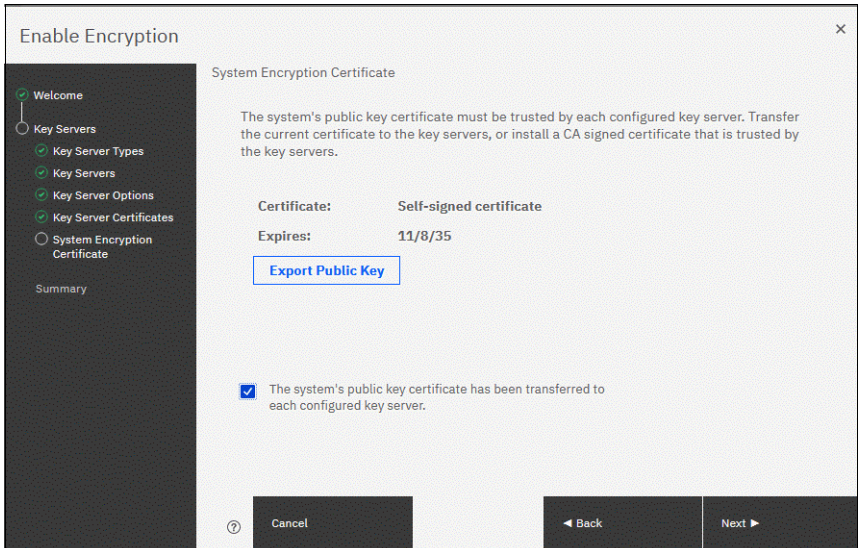


Figure 12-34 Downloading the IBM Spectrum Virtualize SSL certificate

9. When the IBM Spectrum Virtualize system public key certificate is installed on the SGKLM key servers, acknowledge this installation by clicking the check box that is below the Export Public Key button and click **Next**.
10. The key server configuration is shown in the Summary tab, as shown in Figure 12-35. Click **Finish** to create the key server object and finalize the encryption enablement.



Figure 12-35 Finish enabling encryption 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 12-36. The `Online` state indicates which SGKLM servers are detected as available by the system.

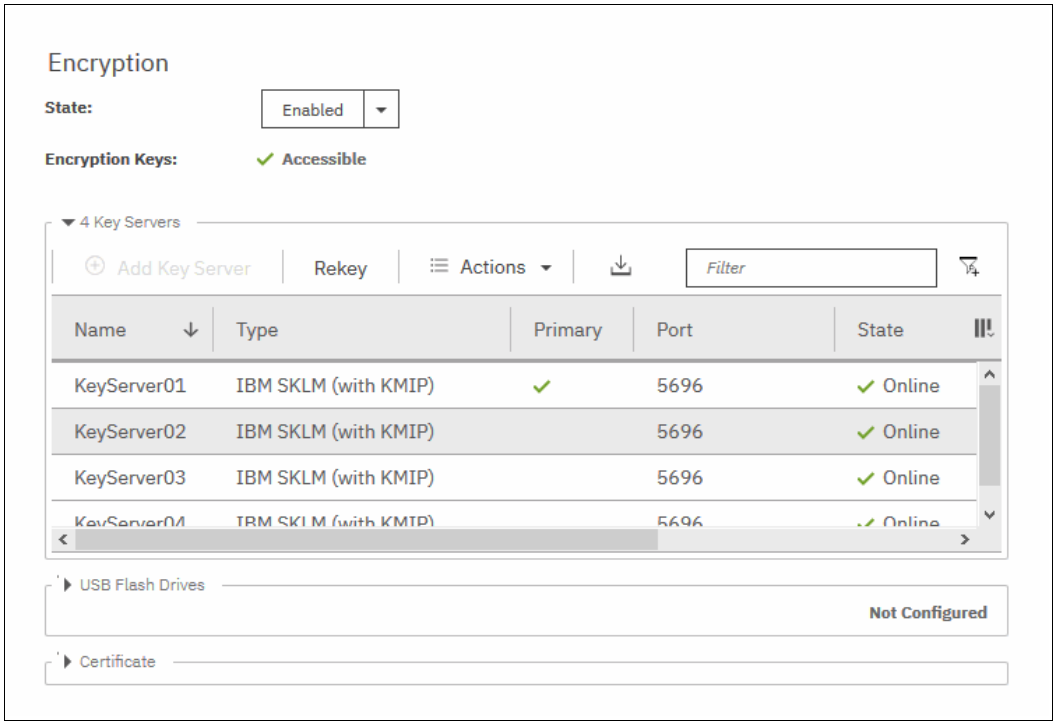


Figure 12-36 Encryption enabled with only SGKLM servers as encryption key providers

Enabling encryption by using SafeNet KeySecure

IBM Spectrum 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.

IBM Spectrum 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 Spectrum Virtualize for redundancy (they can coexist with USB flash drive encryption).

It is not possible to have both SafeNet KeySecure and SGKLM key servers configured at the same time in IBM Spectrum 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, 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 user name and password for IBM Spectrum 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 Spectrum 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 click **Next**, as shown in Figure 12-37.

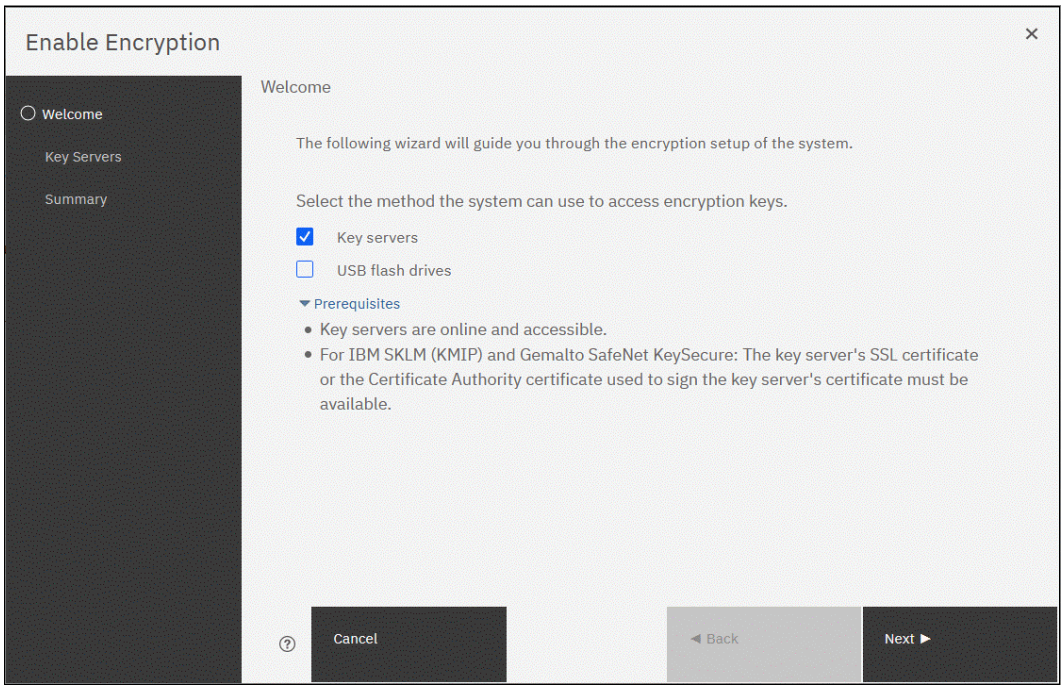


Figure 12-37 Selecting Key servers as the only provider in the Enable Encryption wizard

3. In the next window, you can choose between IBM SGKLM or Gemalto SefeNet KeySecure server types, as shown in Figure 12-38. Select **Gemalto SafeNet KeySecure** and click **Next**.

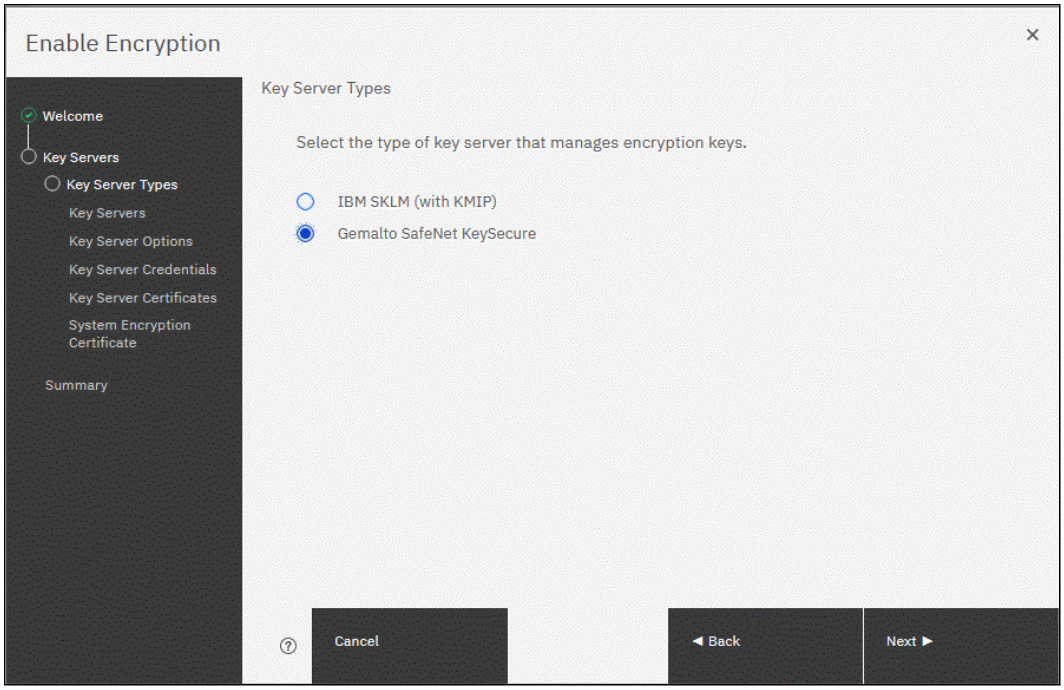


Figure 12-38 Selecting Gemalto SafeNet KeySecure as key server type

4. Add up to four SafeNet KeySecure servers in the next wizard window, as shown in Figure 12-39. For each key server, enter the name, IP address, and TCP port for KMIP protocol (default value is 5696). Because the server name is only a label, it does not need to be the real host name of the server.

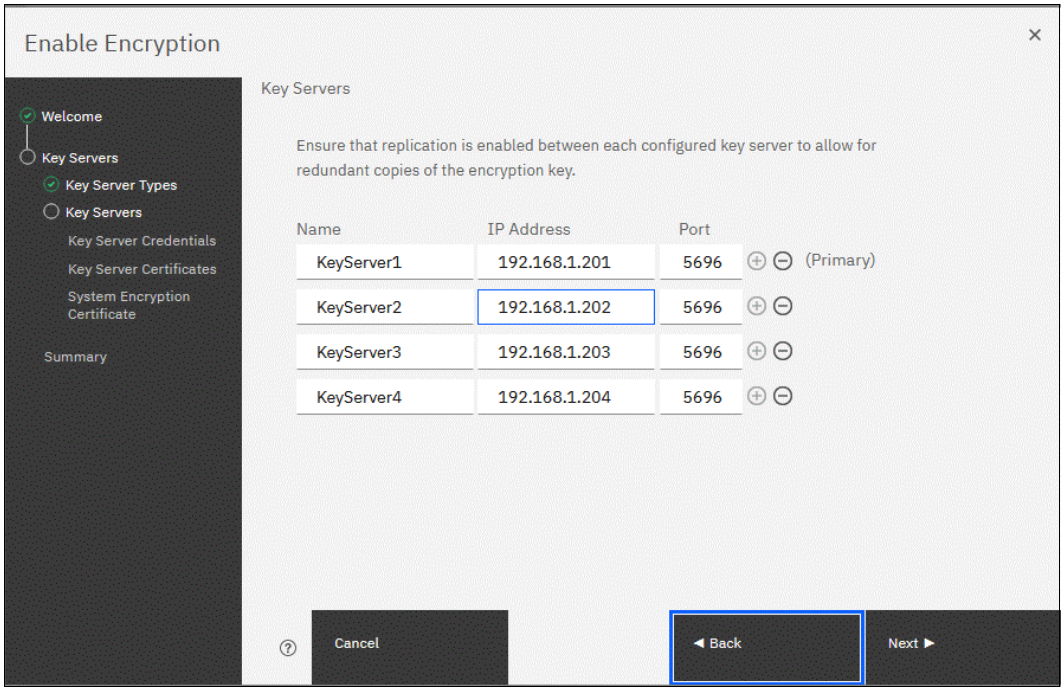


Figure 12-39 Configuring multiple SafeNet KeySecure servers

Although Gemalto SafeNet KeySecure uses an active-active clustered model, IBM Spectrum 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 key servers credentials (user name and password), as shown in Figure 12-40. This setting is optional because it depends on how SafeNet KeySecure servers are configured.

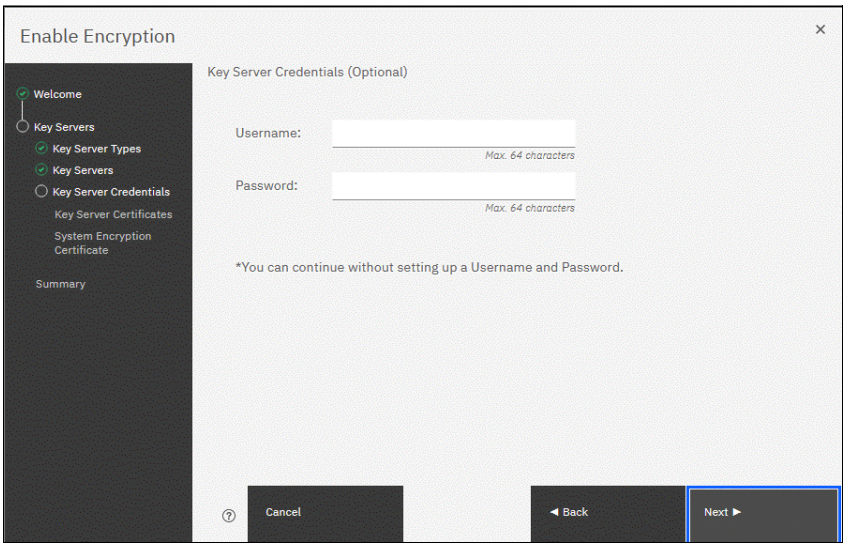


Figure 12-40 Key server credentials input (optional)

5. Enable secure communication between the IBM Spectrum 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 key servers directly. After uploading any of the certificates in the window that is shown in Figure 12-41, click **Next**.

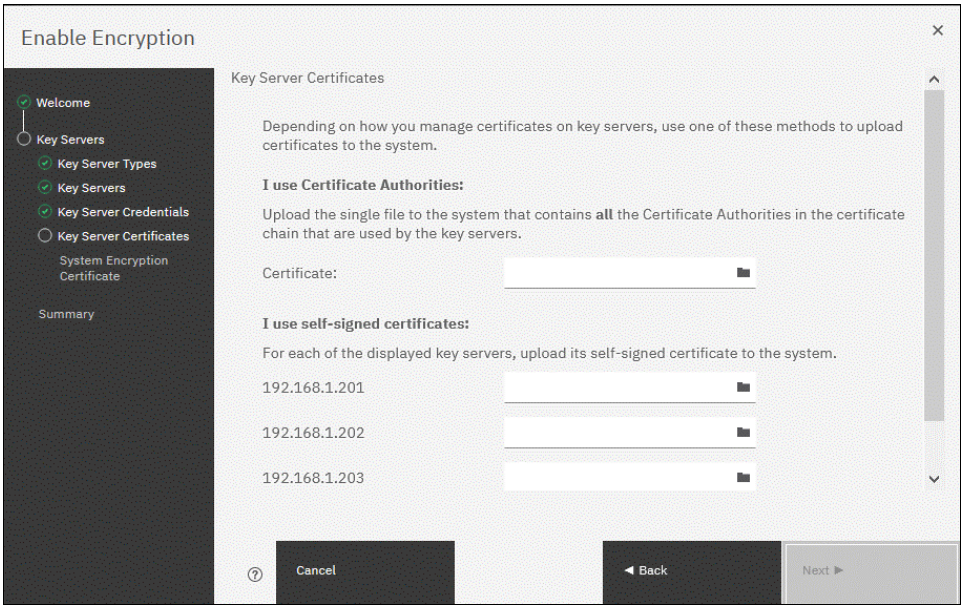


Figure 12-41 Uploading SafeNet KeySecure key servers certificate

6. Configure the SafeNet KeySecure key servers to trust the public key certificate of the IBM Spectrum Virtualize system. You can download the IBM Spectrum Virtualize system public SSL certificate by clicking **Export Public Key**, as shown in Figure 12-41. After adding the public key certificate to the key servers, select the **The system's public key certificate has been transferred to each configured key server** check box and click **Next**.

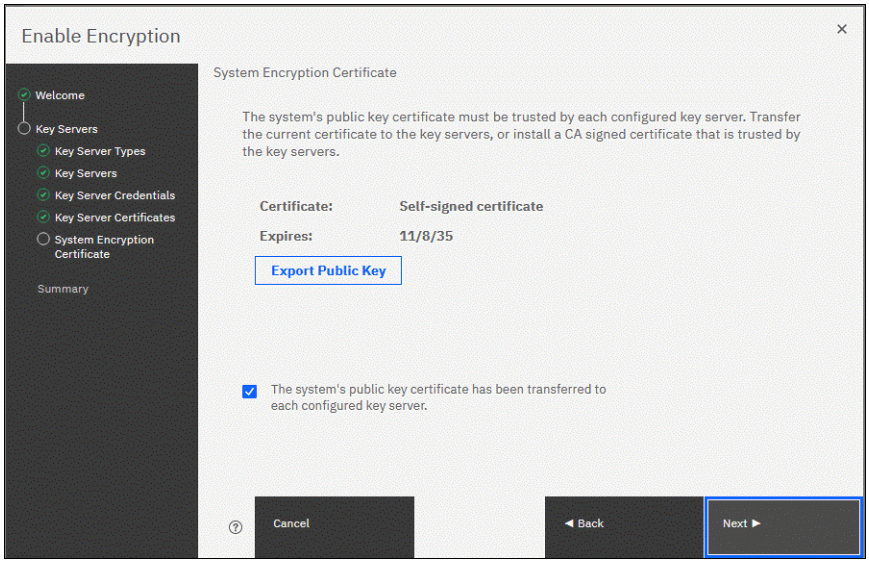


Figure 12-42 Downloading the IBM Spectrum Virtualize SSL certificate

7. The key server configuration is shown in the Summary tab, as shown in Figure 12-43. Click **Finish** to create the key server object and finalize the encryption enablement.

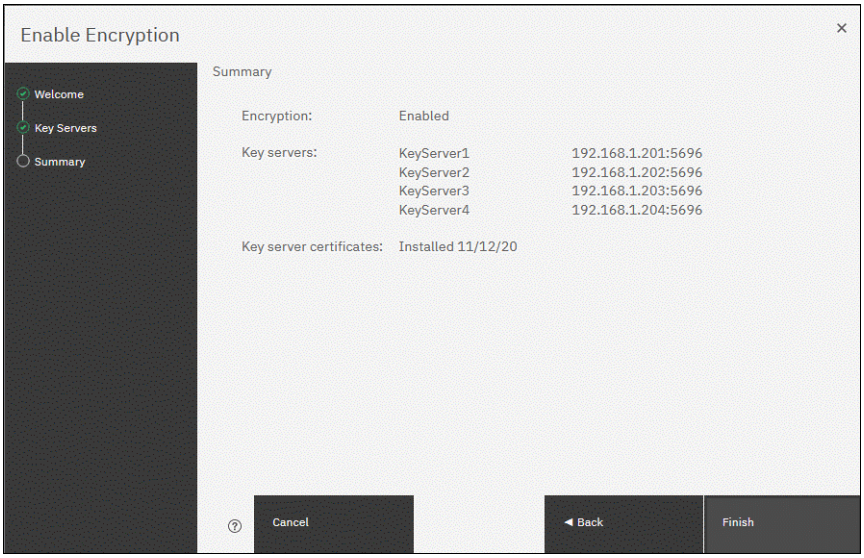


Figure 12-43 Finish enabling encryption using SafeNet KeySecure key servers

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 12-44. Check whether the four servers are shown as online, which indicate that all four SafeNet KeySecure servers are detected as available by the system.

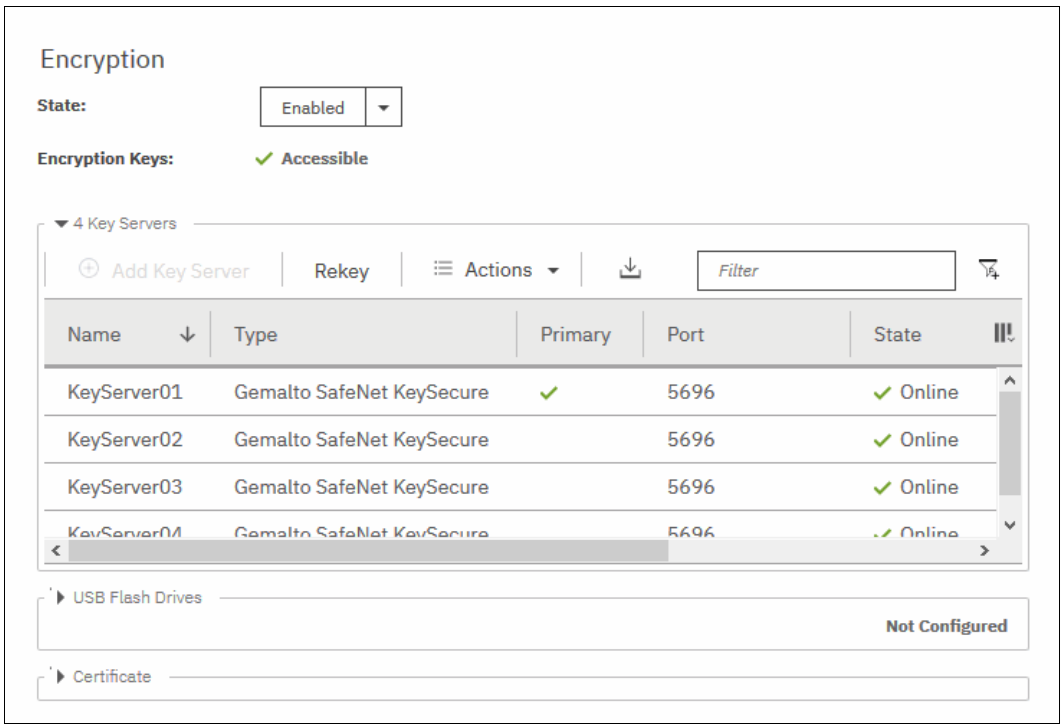


Figure 12-44 Encryption enabled with 4 SafeNet KeySecure key servers

12.5.4 Enabling encryption by using both providers

IBM Spectrum Virtualize allows parallel use of both USB flash drive and one type of key server (SGKLM or SafeNet KeySecure) as encryption key providers. It is possible to configure both providers in a single run of encryption enable wizard. To perform this configuration process, the system must meet requirements of both key server (SGKLM of SafeNet KeySecure) and USB flash drive encryption key providers.

Note: Make sure that the key management server functionality is fully independent from an encrypted storage that has encryption 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 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 Spectrum 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 both USB flash drives and a key servers, confirm the requirements that are described in 12.5.2, “Enabling encryption by using USB flash drives” on page 812, and 12.5.3, “Enabling encryption by using key servers” on page 816.

To enable encryption by using a key server and USB flash drive, complete the following steps:

1. Ensure that you have service IPs 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 12-45).

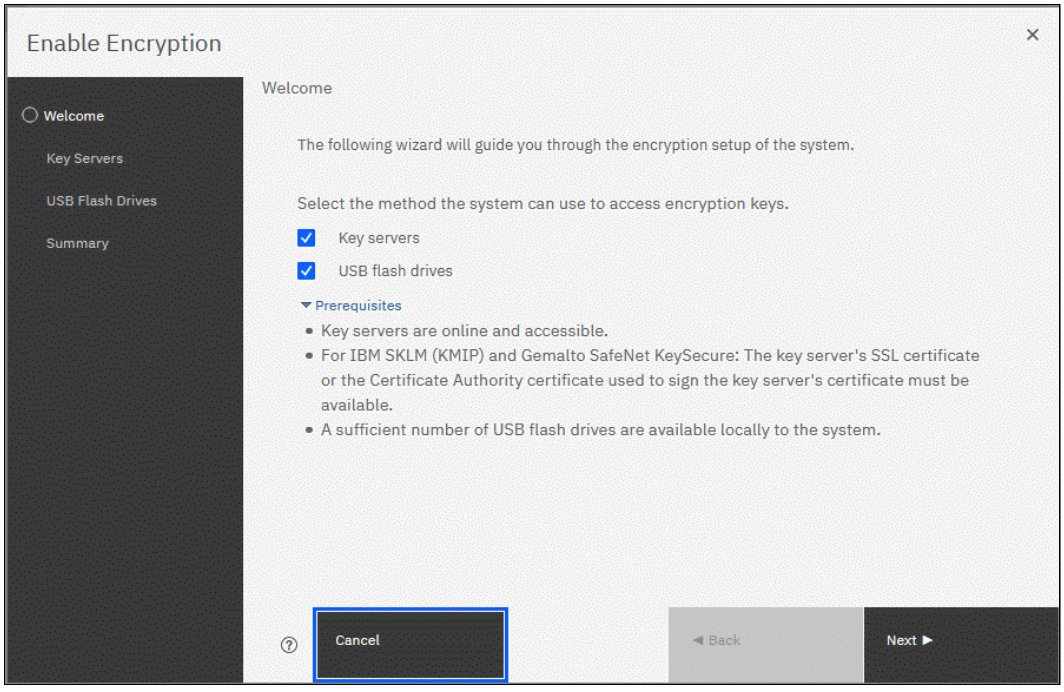


Figure 12-45 Selecting Key servers and USB flash drives in the Enable Encryption wizard

3. The wizard moves to the Key Server Types window, as shown in Figure 12-46. Then, select the key server type that manages the encryption keys.

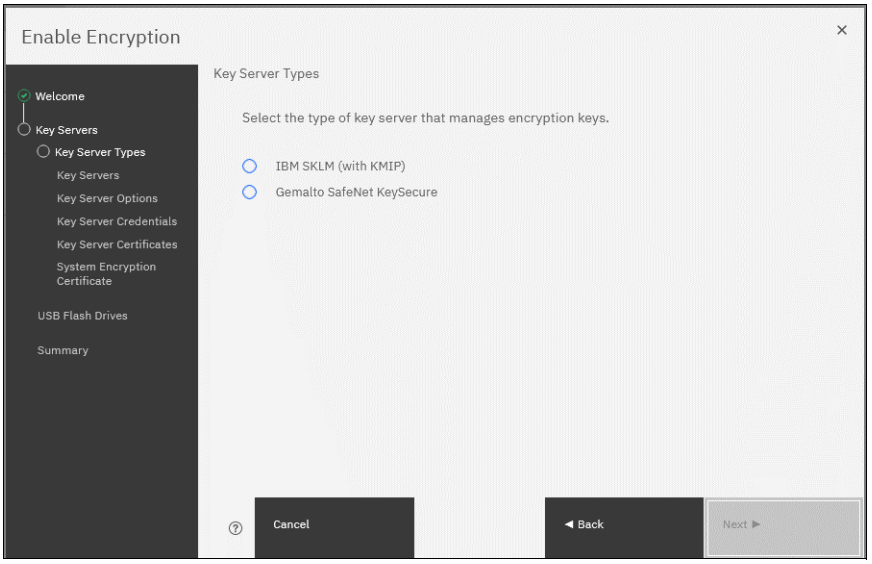


Figure 12-46 Selecting the key server type

The next windows that are displayed are the same as described in 12.5.3, “Enabling encryption by using key servers” on page 816, 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 requests plugging in more USB flash drives. You cannot proceed until the required minimum number of USB flash drives is detected by the system.

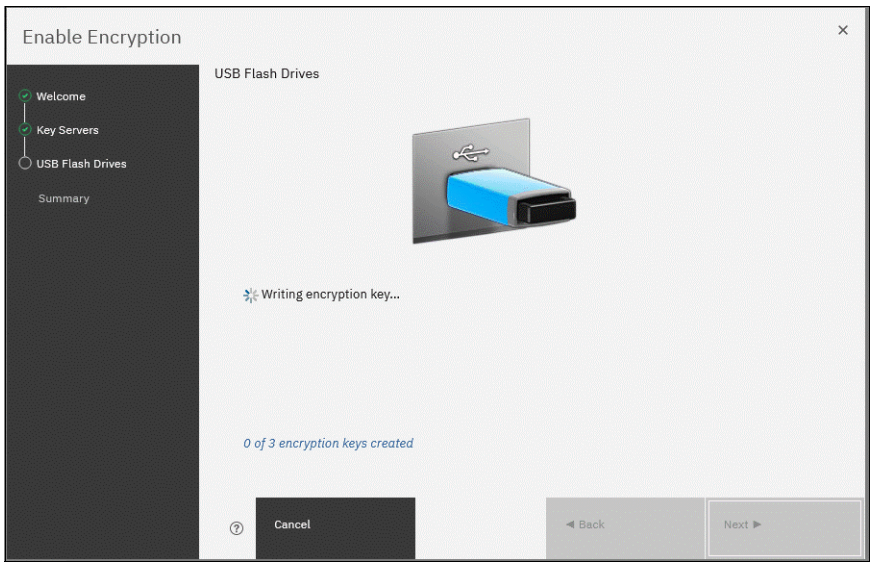


Figure 12-47 Prompt to insert USB flash drives

After at least three USB flash drives are detected, the system writes master access key to each of the drives, see Figure 12-47 on page 830. Notice that the system attempts to write the encryption key to any flash drive it detects. Therefore, it is crucial to maintain physical security of the system during this procedure.

4. After copying encryption keys to USB flash drives, a window is shown with 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 12-48. 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.

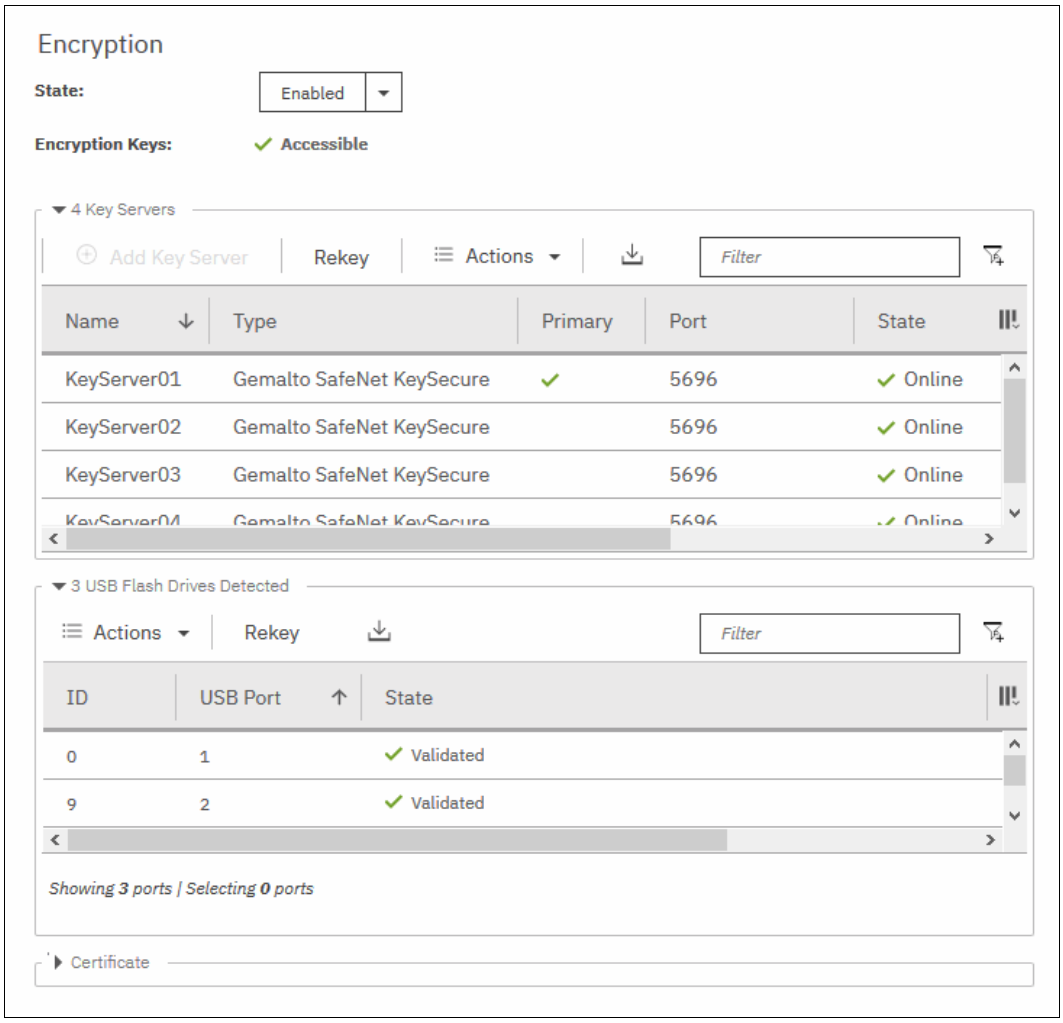


Figure 12-48 Encryption enabled with both USB flash drives and key servers

12.6 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 Spectrum 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.

12.6.1 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 12-49. To enable 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.

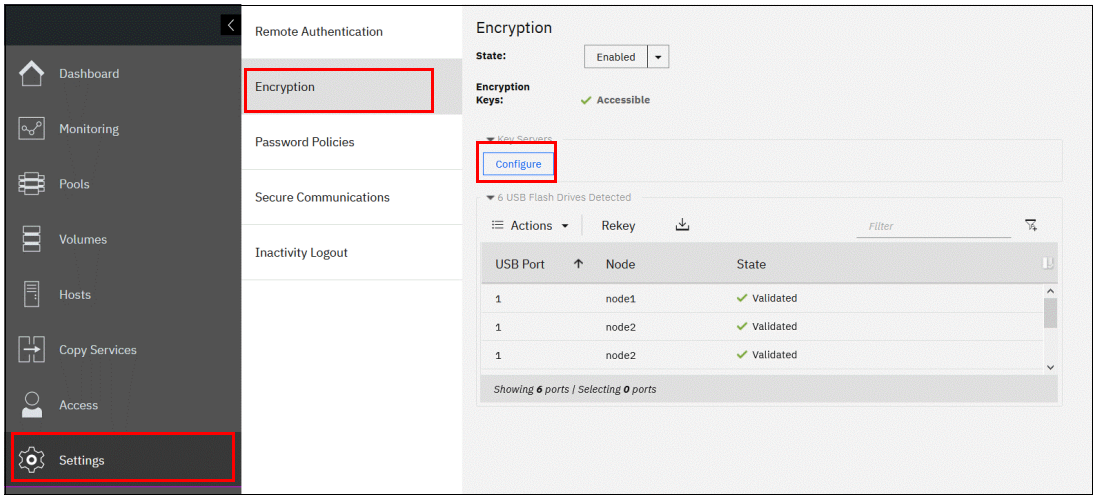


Figure 12-49 Enable key servers as a second provider

2. Complete the steps that are required to configure the key server provider, as described in 12.5.3, “Enabling encryption by using key servers” on page 816. 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 12-50.

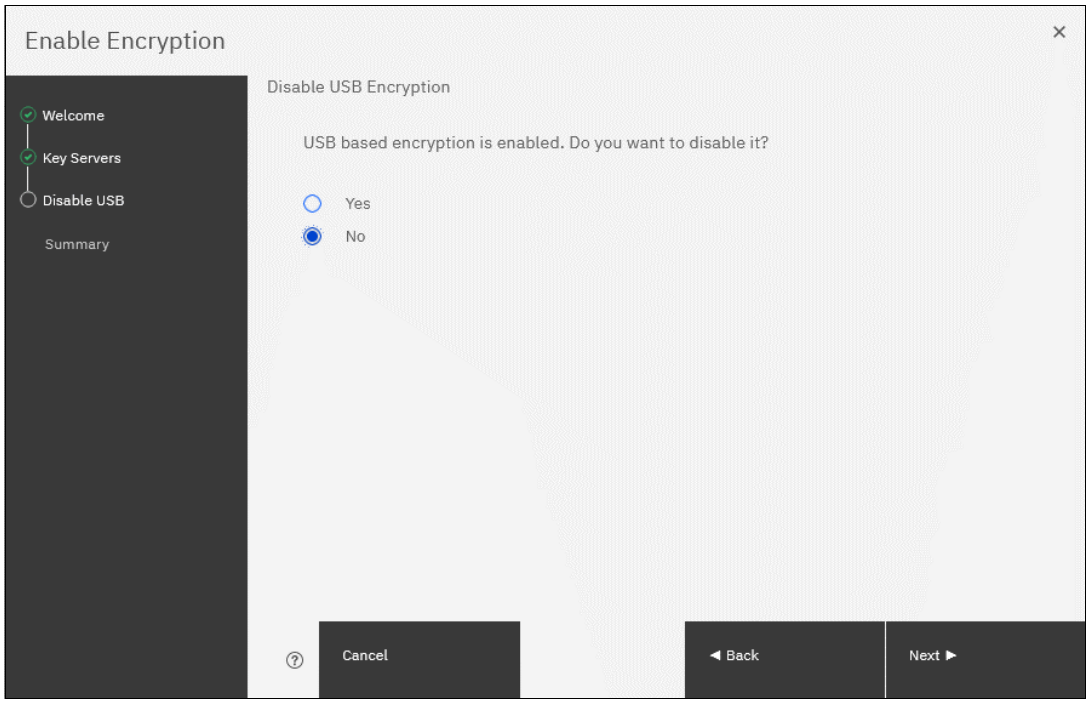


Figure 12-50 Do not disable USB flash drive encryption key provider

This choice is confirmed on the summary window before the configuration is committed, as shown in Figure 12-51.



Figure 12-51 Configuration summary before committing

3. After you click **Finish**, the system configures keys 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 12-52. 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.

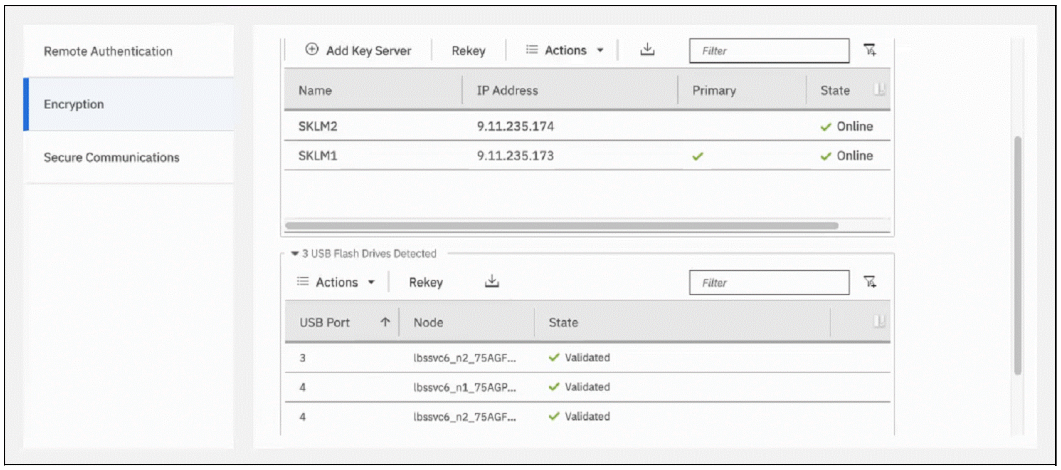


Figure 12-52 Encryption enabled with two key providers available

12.6.2 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 12.5.2, “Enabling encryption by using USB flash drives” on page 812. You cannot disable key server providers during this process.

   After successful completion of the process, you are presented with a message confirming that both encryption key providers are enabled.

3. You can confirm that encryption is enabled and verify which key providers are in use by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-53. 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.

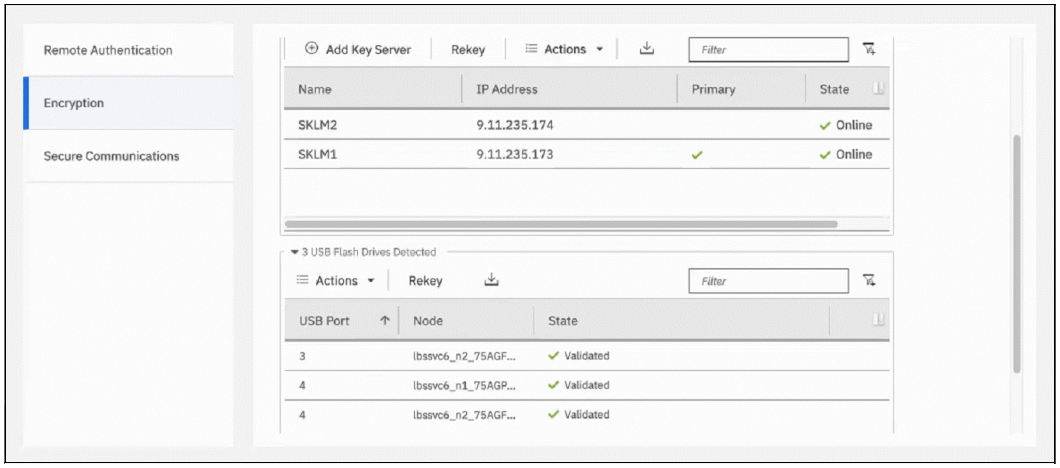


Figure 12-53 Example of encryption enabled with two key providers available

12.7 Migrating between providers

IBM Spectrum Virtualize V8.1 introduced support for simultaneous use of both 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.

12.7.1 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 12.6.1, “Adding key servers as a second provider” on page 832, but when completing step 2 on page 832, select **Yes** instead of **No** (see Figure 12-54 on page 836). This action deactivates the USB flash drives provider, and the procedure completes with only key servers configured as key provider.

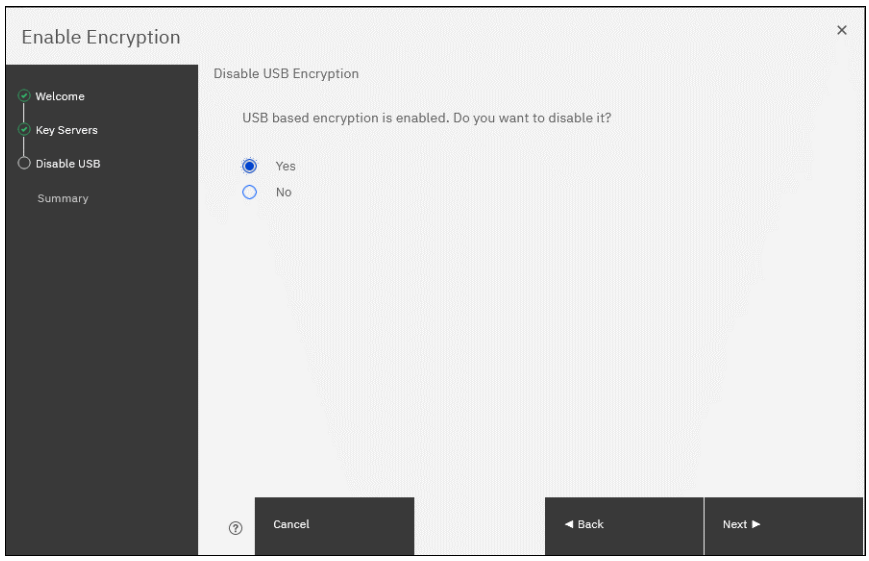


Figure 12-54 Disable USB flash drive provider while changing to SGKLM provider

12.7.2 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 12.6.2, “Adding USB flash drives as a second provider” on page 834.

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.

12.7.3 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 Spectrum Virtualize system that is configured with IBM SGKLM key servers (as shown in Figure 12-55) to SafeNet KeySecure servers.

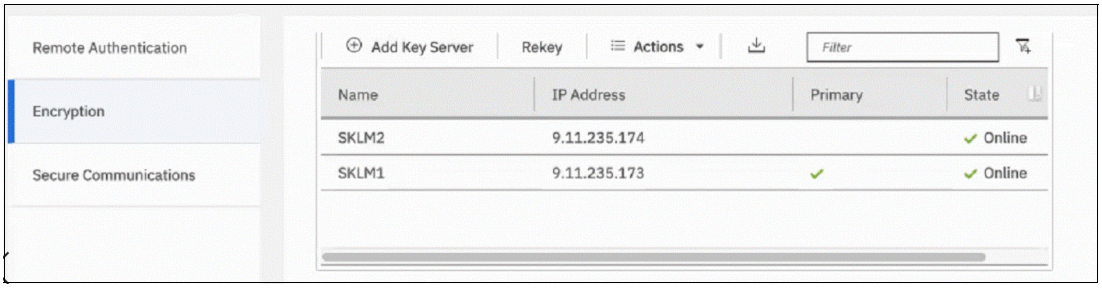


Figure 12-55 IBM Spectrum Virtualize encryption configured with IBM SGKLM servers

Complete the following steps to migrate to Gemalto SafeNet KeySecure:

1. Migrate from key server encryption to USB flash drives encryption, as described in 12.7.2, “Changing from encryption key server to USB flash drive provider” on page 836. After this step, only USB flash drives encryption are configured, as shown in Figure 12-56.

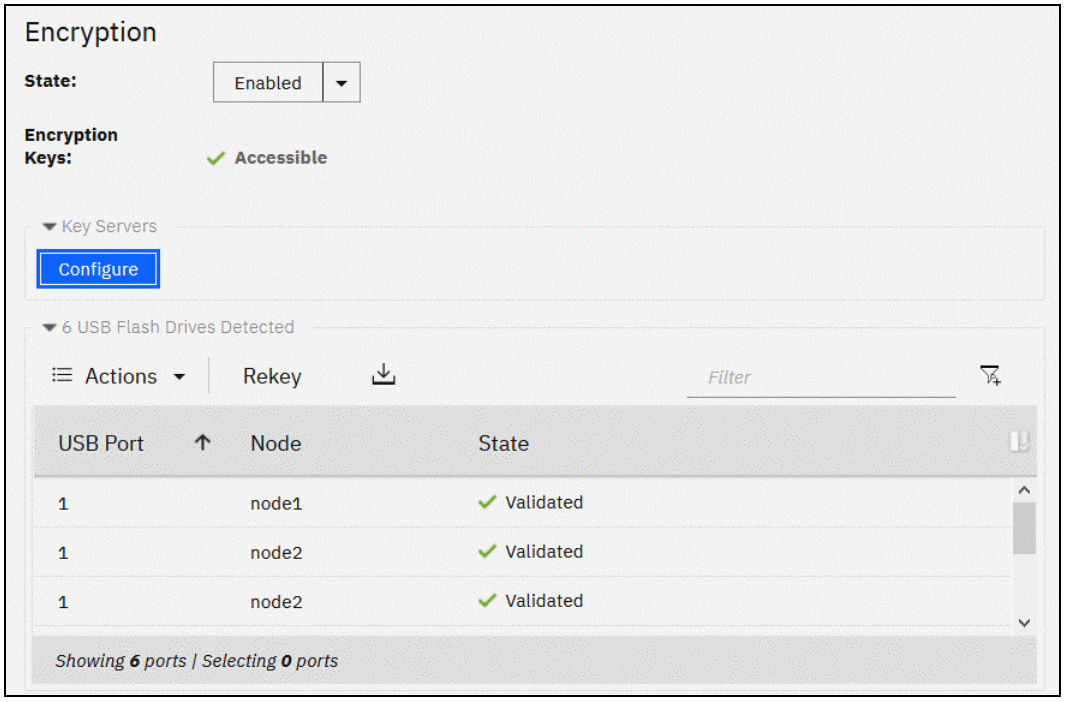


Figure 12-56 IBM SAN Volume Controller encryption configured with USB Flash Drives

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 12.7.1, “Changing from USB flash drive provider to encryption key server” on page 835. After completing this step, the other key server type are configured as encryption provider in IBM Spectrum Virtualize, as shown in Figure 12-57.

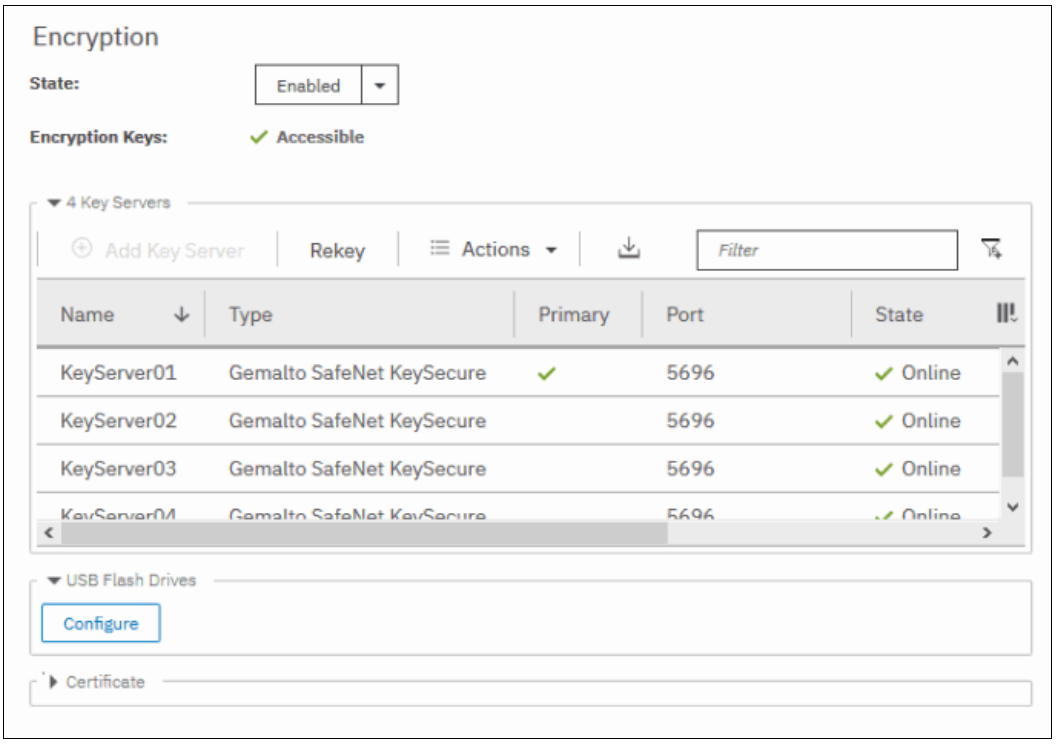


Figure 12-57 IBM SAN Volume Controller encryption configured with SafeNet KeySecure

12.8 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 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 12.6, “Configuring more providers” on page 832.

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 by default created as encrypted.

Some unsupported configurations are actively policed in code. For example, no support exists for creating unencrypted child pools from encrypted parent pools. However, the following exceptions exist:

- 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 Spectrum 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, 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 271. After encryption is enabled, any new pool is created by default as encrypted, as shown in Figure 12-58.

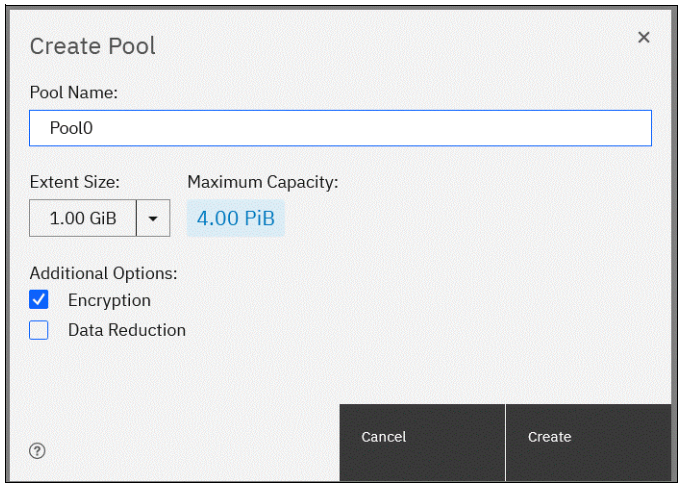


Figure 12-58 Create Pool window basic

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 12-59.

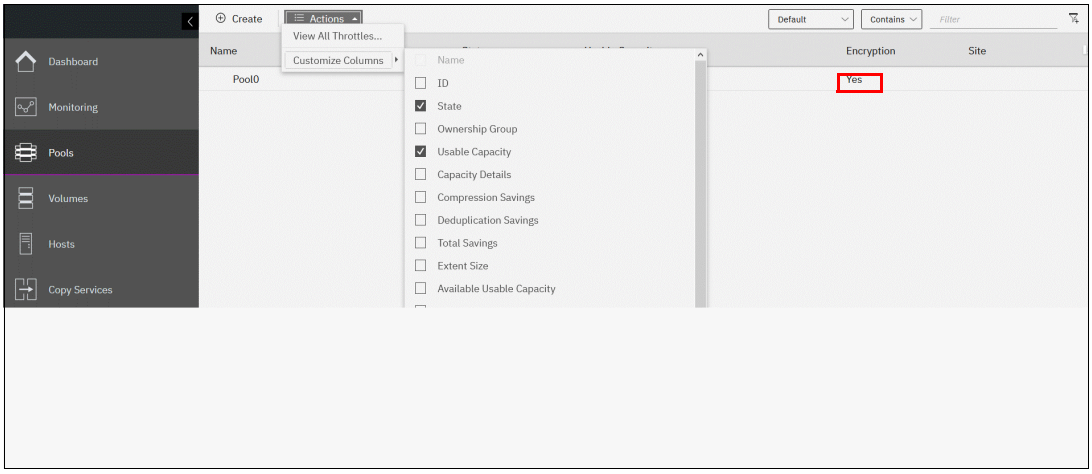


Figure 12-59 Pool encryption state

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 there are any unencrypted MDisks 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 as MDisks are encrypted, but the pool still allows unencrypted MDisks to be added, which makes it possible to mix the configuration and if unencrypted MDisks are added, the pool reverts to unencrypted state. In this case, data partially ends up on an unencrypted MDisk.

Caution is necessary if encrypting previously unencrypted pools by adding encrypted MDisks and removing unencrypted and also expanding it later with new MDisks.

How to add encrypted storage to encrypted pools is described next. You can mix and match storage encryption types in a pool. Figure 12-60 shows an example of an encrypted pool that contains storage by using different encryption methods.

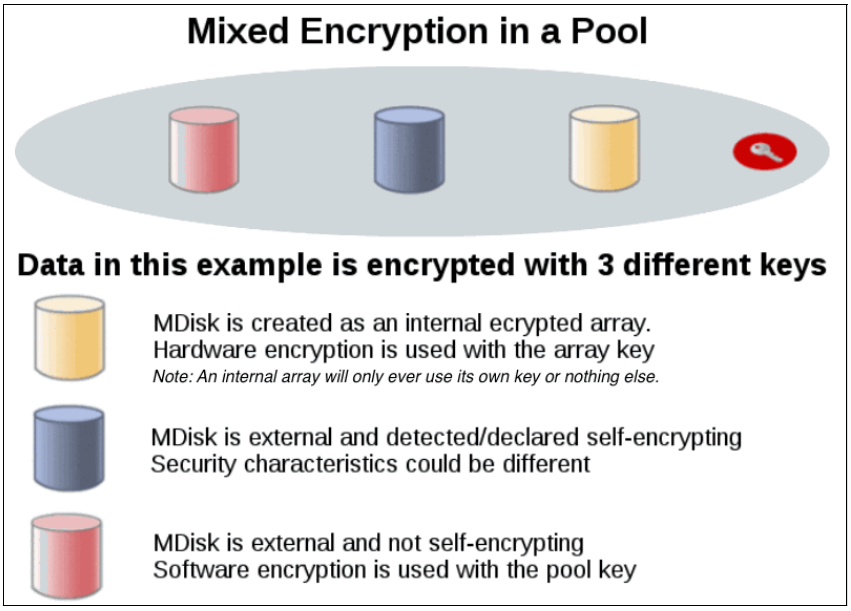


Figure 12-60 Mix and match encryption in a pool

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 271.

If the parent pool is encrypted, every child pool also must 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 12-61.

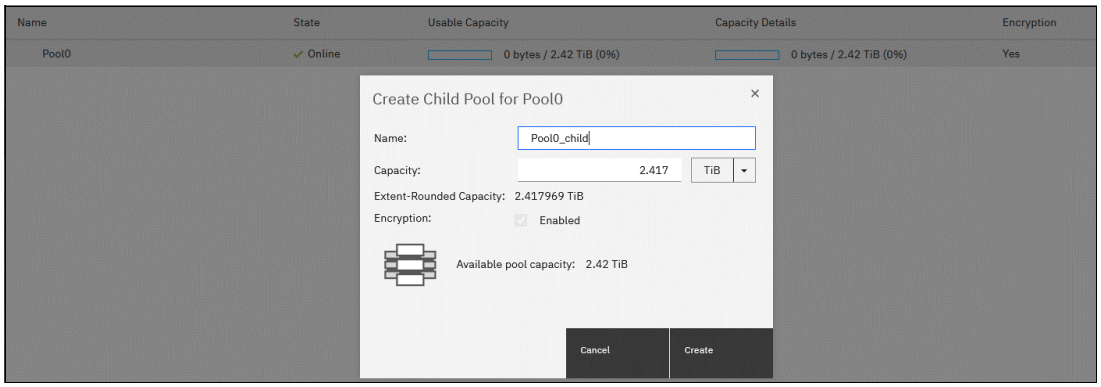


Figure 12-61 Create a child pool of an encrypted parent pool

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 must not contain any unencrypted internal arrays. If any unencrypted internal array is in the unencrypted pool, when you try to create a child pool and select the option to set as encrypted, it is created as unencrypted.
- 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 12-62. The example shows an encrypted child pool with non-encrypted parent pool.

.

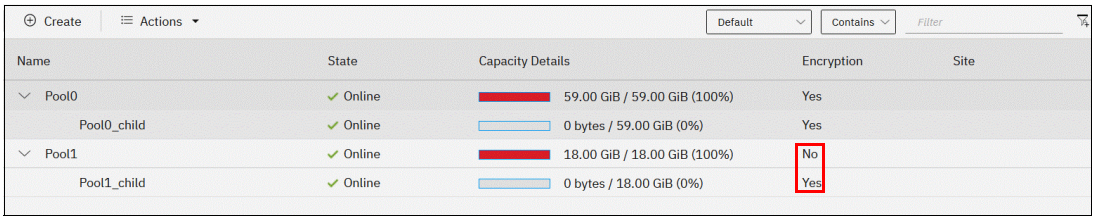


Figure 12-62 Child pool encryption state

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 271.

After encryption is enabled, all newly built arrays are hardware encrypted by default. In this case, the GUI does not allow you to create an unencrypted array. To create an unencrypted array, the command-line interface (CLI) must be used.

Example 12-1 shows how to create an unencrypted array by using the CLI.

Example 12-1 Creating an unencrypted array 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 also can right-click the table header to customize columns and select **Encryption**, as shown in Figure 12-63.

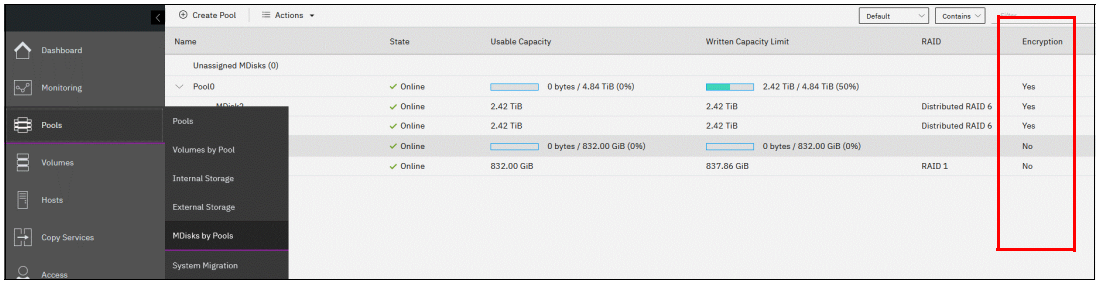


Figure 12-63 Array encryption state

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 12-64.

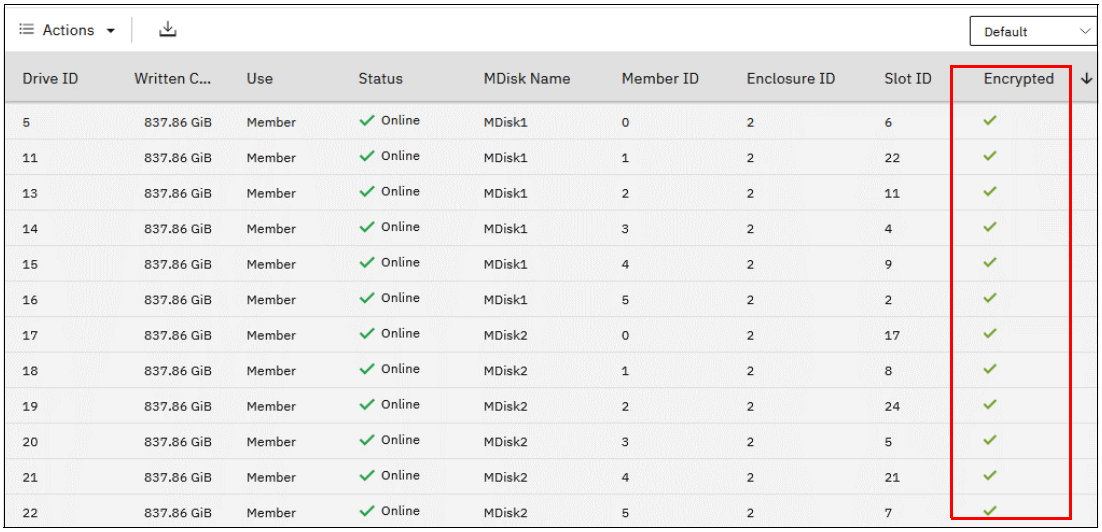


Figure 12-64 Drive encryption state

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 271. 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 12-65 shows an example in which self-encrypting MDisk is in an unencrypted pool.

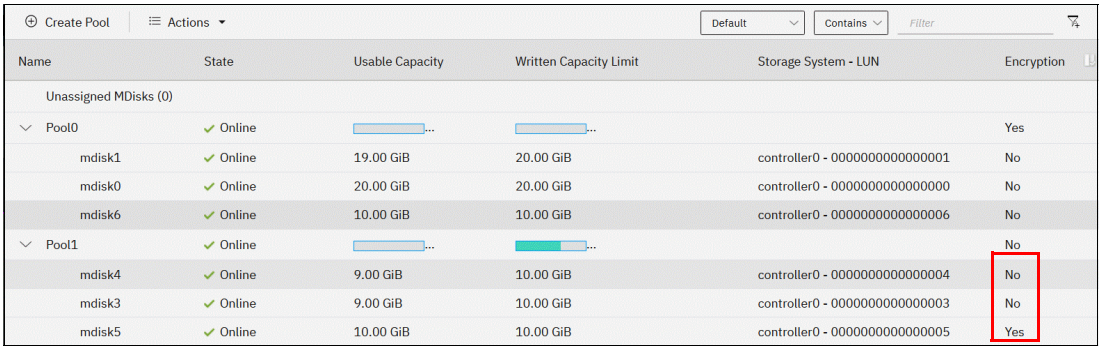


Figure 12-65 MDisk encryption state

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 marks 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 misconfigure 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 ends up with unencrypted data on an extent in an encrypted storage pool.

However, all data that is written to any MDisk that is a part of 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 waste of 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 12-66.

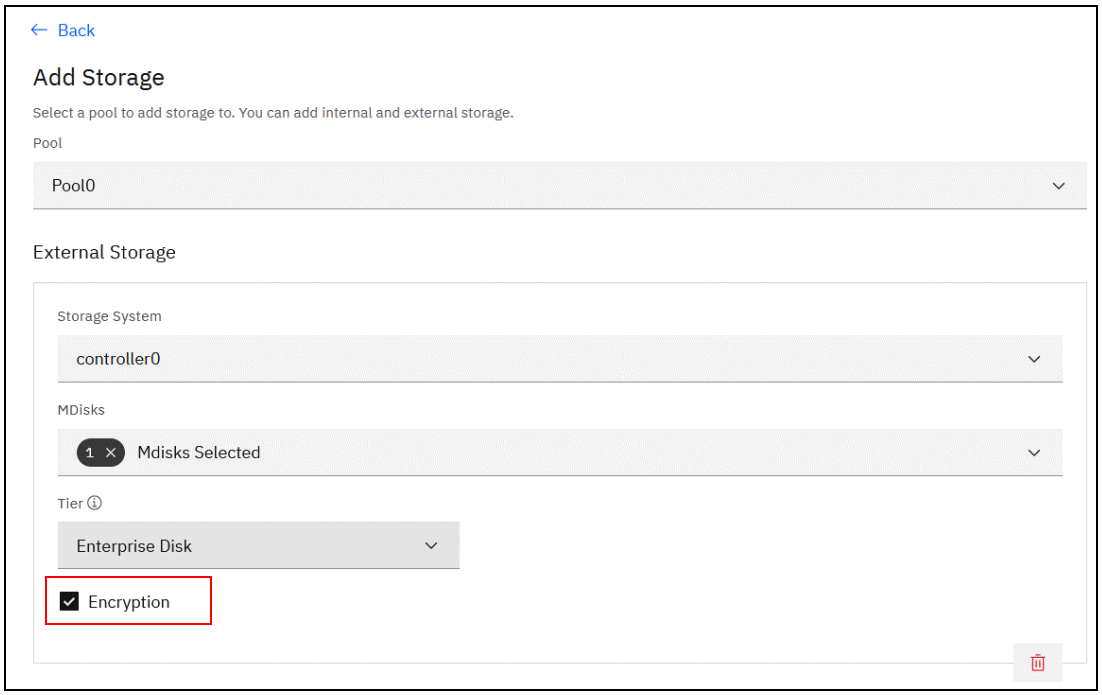


Figure 12-66 Declaring MDisk as externally encrypted

IBM Spectrum Virtualize products can detect that an MDisk is self-encrypting by using the SCSI Inquiry page C2. MDisks that are provided by other IBM Spectrum Virtualize products report this page correctly. The Encryption check box 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 CLI command **`chmdisk -encrypt no`**. 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 in Cluster GUI **Pools** → **External Storage** (see Figure 12-67).

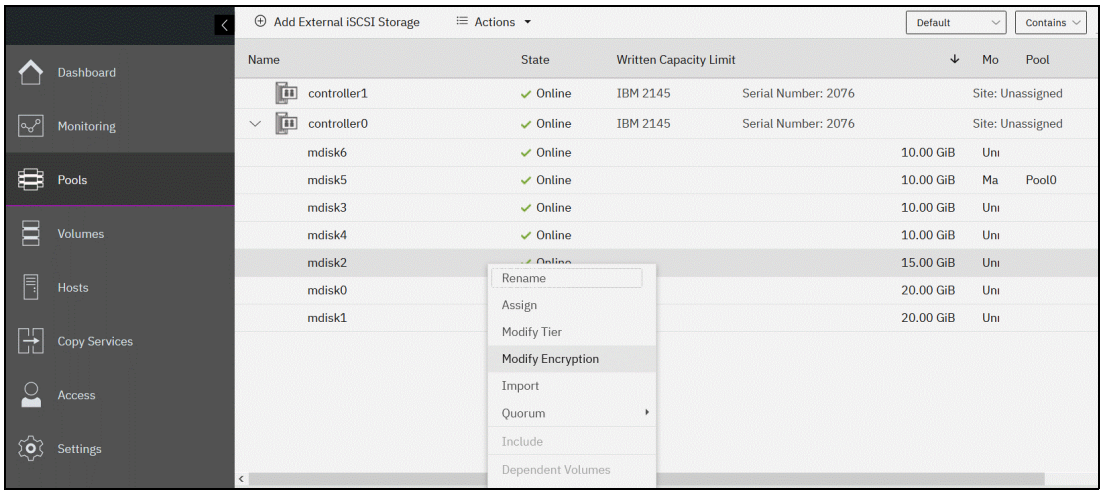


Figure 12-67 Overriding external encryption setting for external MDisk

If external encryption setting is overridden for self-encrypting MDisk, the Encryption check box is not automatically set in the Add Storage dialog when adding this MDisk to the pool (see Figure 12-66 on page 845).

To check whether an MDisk was declared as self-encrypting, select **Pools** → **MDisk by Pools** and verify the information in the Encryption column, as shown in Figure 12-68.

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 12-68 MDisk self-encryption state

The value that is shown in the Encryption column shows the property of objects in respective rows. Therefore, in the configuration that is shown in Figure 12-68, `Pool0` is encrypted, so every volume 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. 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, “Creating and provisioning volumes” on page 331. 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 12-69.

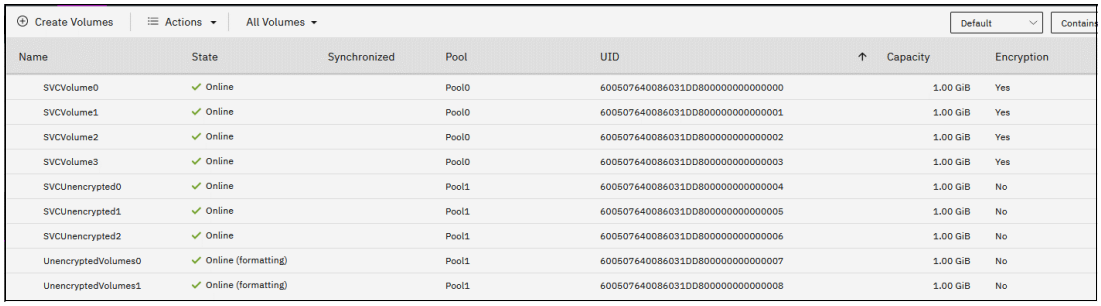


Figure 12-69 Volume view customization

A volume is reported as encrypted only if all the volume copies are encrypted, as shown in Figure 12-70.

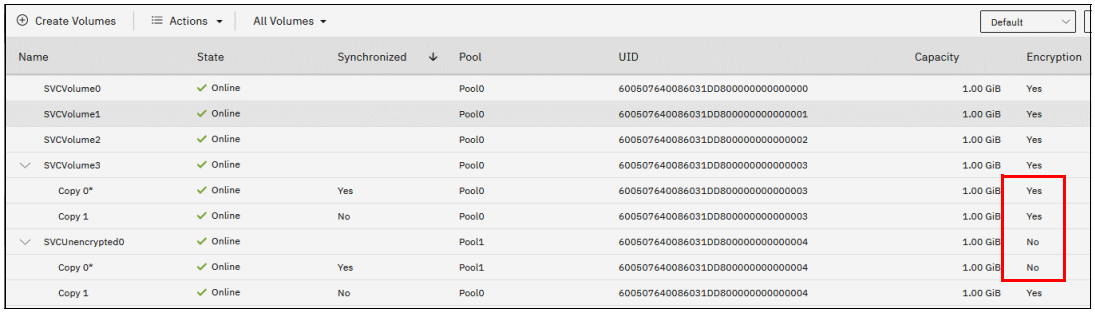


Figure 12-70 Volume encryption status depending on volume copies encryption

When creating volumes, make sure to select encrypted pools to create encrypted volumes, as shown in Figure 12-71.

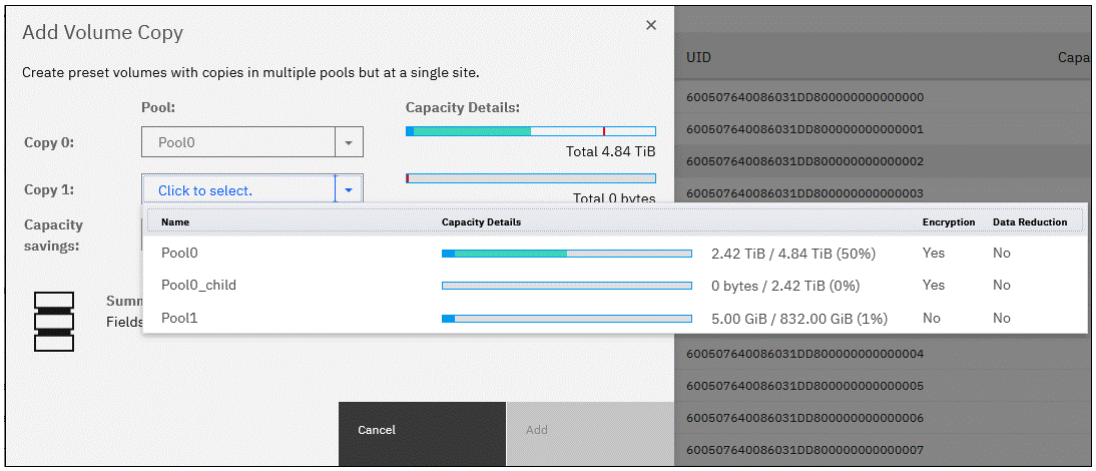


Figure 12-71 Create an encrypted volume by selecting an encrypted pool

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 Chapter 12, “Encryption” on page 793.

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.

12.10 Rekeying an encryption-enabled system

Changing the master access key is a security requirement. *Rekeying* is the process of replacing 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 defined in your organization’s security policy and whenever you suspect that the key might have been 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.

12.10.1 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 kept on the key server provider, complete the following steps:

1. Select **Settings** → **Security** → **Encryption**. Ensure that Encryption Keys shows that all configured SGKLM servers are reported as `Accessible`. Click **Key Servers** to expand the section.
2. Click **Rekey**, as shown in Figure 12-72.

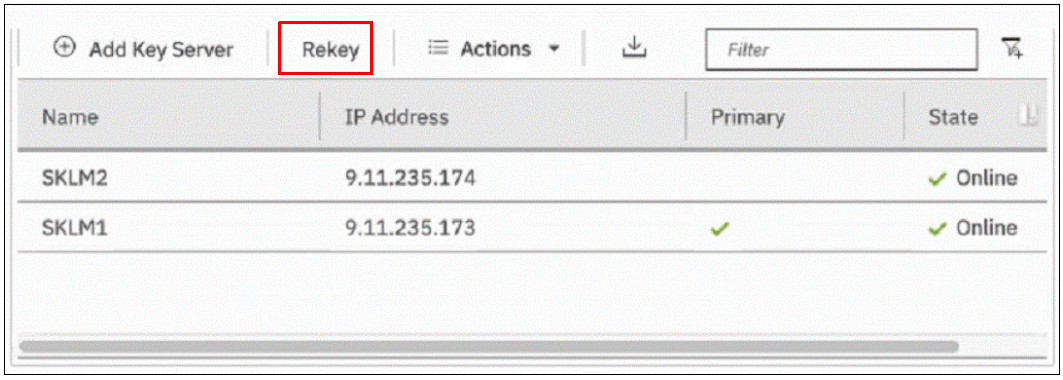


Figure 12-72 Start rekey on SGKLM key server

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.

12.10.2 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 12-73. Click **Rekey**. You need at least three USB flash drives, with at least one reported as `Validated` to process with rekey.

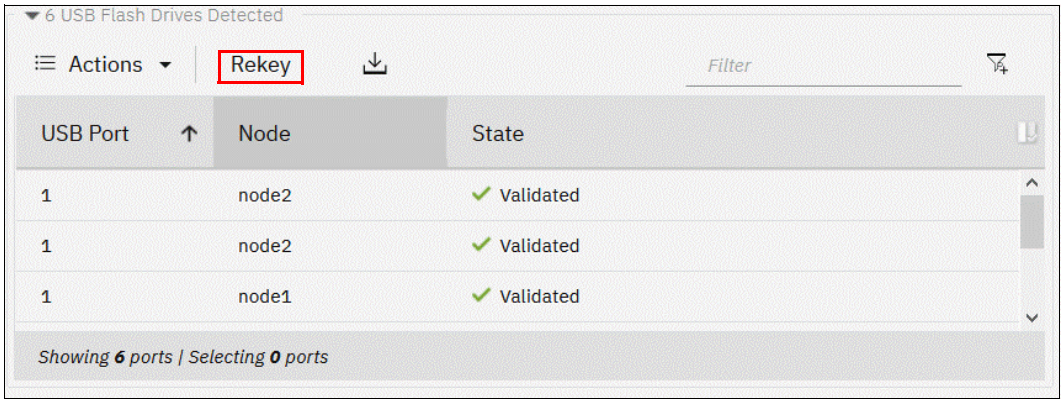


Figure 12-73 Start rekey on USB flash drives provider

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 12-74. Click **Commit** to finalize the rekey operation.

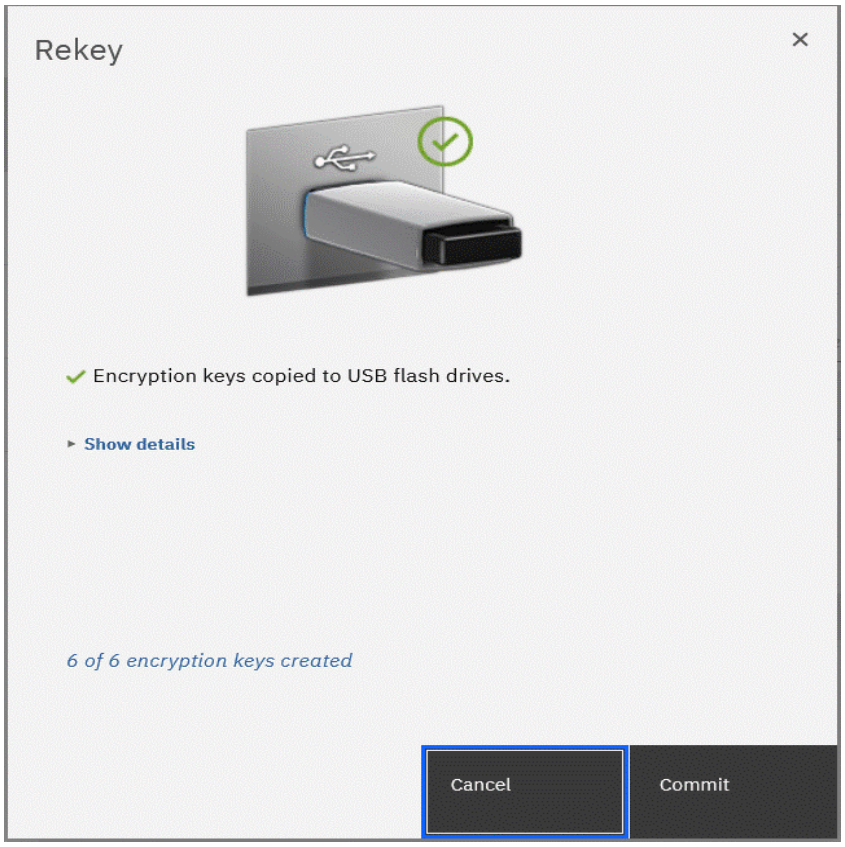


Figure 12-74 Writing new keys to USB flash drives

4. You receive a message confirming the rekey operation was successful. Click **Close**.

12.11 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 12-75 shows an example for a system with both encryption key providers configured.

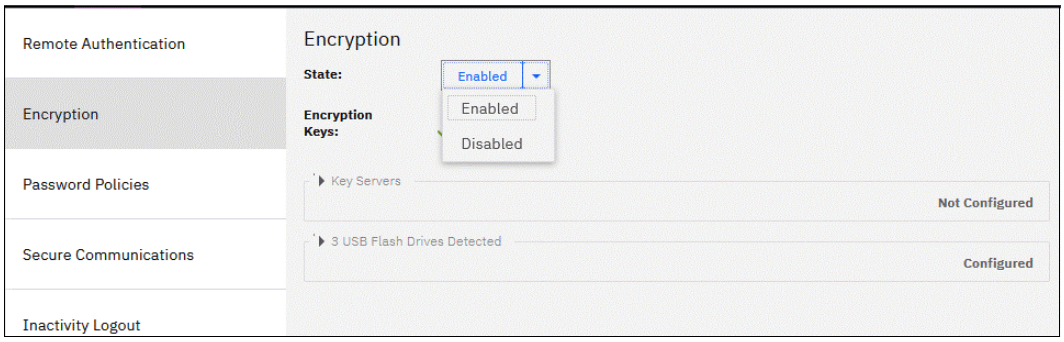


Figure 12-75 Disabling encryption on a system with both providers

2. You receive a message confirming that encryption was disabled. Figure 12-76 shows the message when a key server is used.

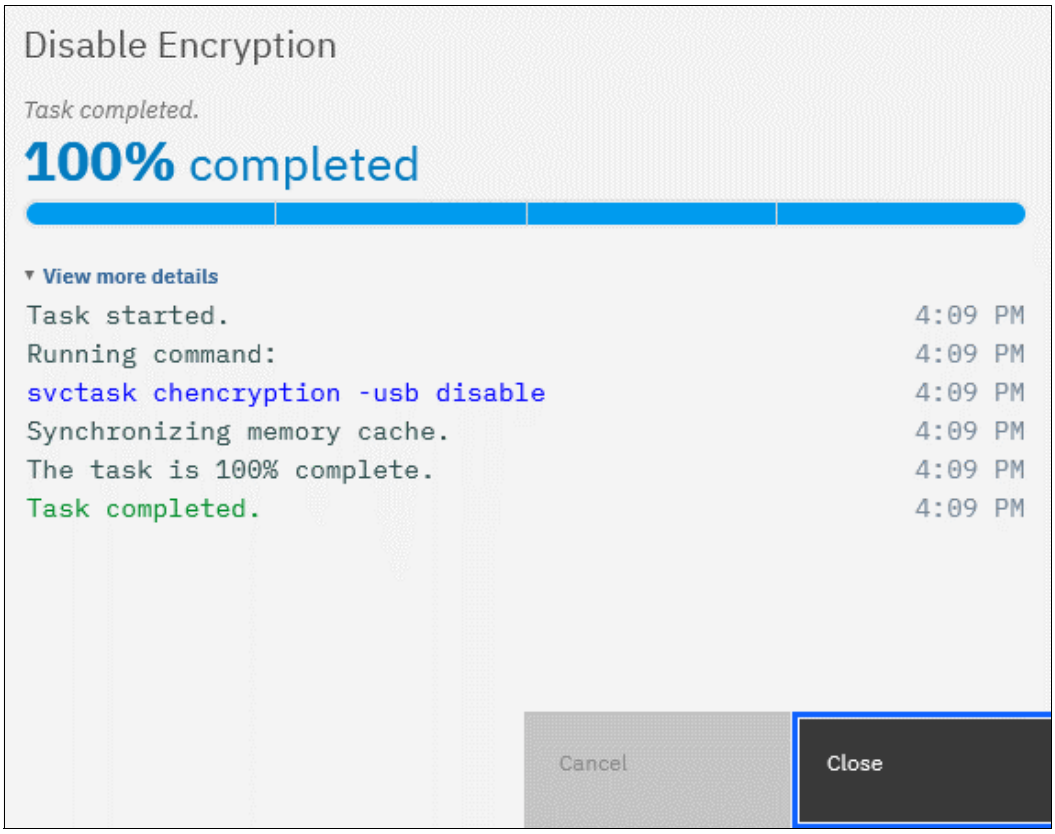


Figure 12-76 Encryption disabled

13

Reliability, availability, serviceability, monitoring and troubleshooting

Many methods are available to manage, monitor, and troubleshoot IBM Spectrum Virtualize.

This chapter introduces useful, common procedures to maintain IBM Spectrum Virtualize. It includes the following topics:

- 13.1, “Reliability, availability, and serviceability” on page 854
- 13.2, “Removing or adding a node to or from the system” on page 862
- 13.3, “Configuration backup” on page 865
- 13.4, “IBM Spectrum Virtualize Code Upgrade” on page 869
- 13.5, “Health checker feature” on page 880
- 13.6, “Troubleshooting and fix procedures” on page 881
- 13.7, “Monitoring and Event Notification” on page 889
- 13.8, “Audit log” on page 905
- 13.9, “Collecting support information by using the GUI, CLI, and USB” on page 908
- 13.10, “Service Assistant Tool” on page 916
- 13.11, “IBM Storage Insights Monitoring” on page 920

© Copyright IBM Corp. 2021. All rights reserved.

13.1 Reliability, availability, and serviceability

Reliability, availability, and serviceability (RAS) are important concepts in the design of the IBM Spectrum Virtualize systems family. Hardware features, software features, design considerations and operational guidelines all contribute to make the system reliable.

High levels of availability are achieved by using the following methods:

- The Redundant Array of Independent Disks (RAID) capabilities of the underlying disk arrays or virtualized back-end storage controllers.
- Auto-restart of hung nodes.
- Integrated battery backup units (BBUs) to provide write-cache protection if a site power failure occurs.
- Volume copies in another pool backed by another storage controller to prevent from back-end failures.
- Host system failover capabilities by using N_Port ID Virtualization (NPIV).
- Deploying advanced multi-site configurations such as HyperSwap and Stretched Cluster.
- Hot Spare Node option to provide complete node redundancy and failover.

High levels of serviceability are provided by using following methods:

- Cluster error logging
- Asynchronous error notification
- Automatic dump capabilities to capture software detected issues
- Concurrent diagnostic procedures
- Directed Maintenance Procedures (DMPs) with guided online replacement procedures
- Concurrent log analysis and memory dump data recovery tools
- Concurrent maintenance of all IBM SAN Volume Controller components
- Concurrent upgrade of IBM Spectrum Virtualize software and firmware
- Concurrent addition or deletion of nodes in the clustered system
- Automatic software version leveling when replacing or adding a node
- Detailed status and error conditions that are displayed by light-emitting diode (LED) indicators
- Error and event notification through Simple Network Management Protocol (SNMP) traps, syslog, and email
- Optional Remote Support Assistant
- IBM Storage Insights

The heart of an IBM Spectrum Virtualize system is one or more pairs of *nodes*. The nodes share the read and write data workload from the attached hosts and to the disk arrays. This section examines the RAS features of an IBM SAN Volume Controller system, monitoring, and troubleshooting.

13.1.1 IBM SAN Volume Controller nodes

IBM SAN Volume Controller nodes work as a redundant clustered system. Each IBM SAN Volume Controller node is an individual server within the clustered system on which the IBM Spectrum Virtualize software runs.

IBM SAN Volume Controller nodes are always installed in pairs, forming an I/O group. A minimum of one I/O group and a maximum of four I/O groups constitute an IBM SAN Volume Controller clustered system (the terms *system* and *cluster* are used synonymously). Many of the components that make up IBM SAN Volume Controller nodes include LEDs that indicate its status and or its activity.

Figure 13-1 shows the ports and indicator lights on the IBM SAN Volume Controller node model 2145-SV1.

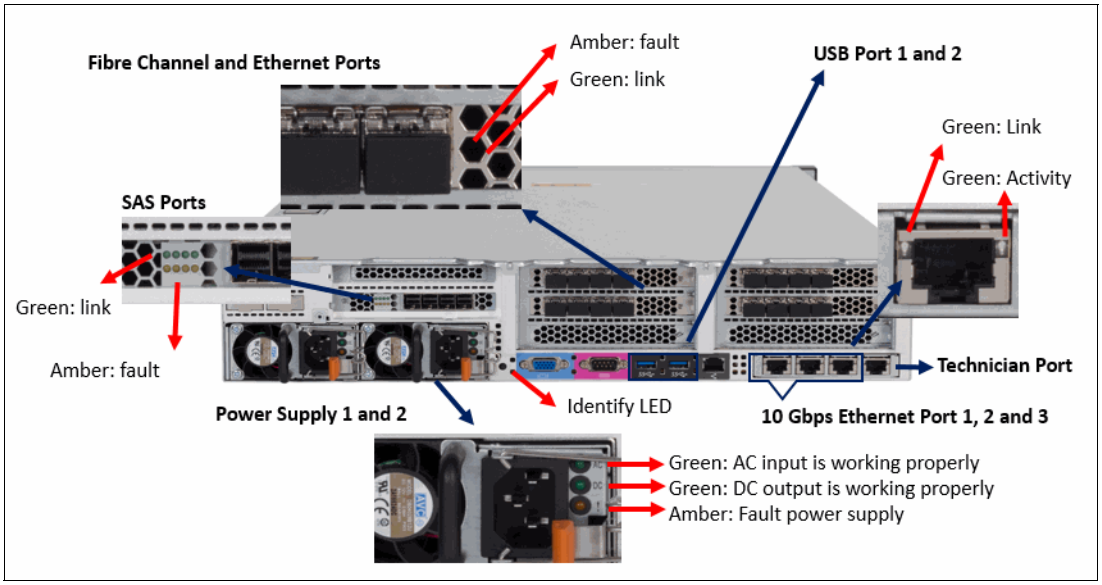


Figure 13-1 2145-SV1 Ports and Components

Figure 13-2 shows the ports and indicator lights on the IBM SAN Volume Controller node model 2145-SV2/SA2.

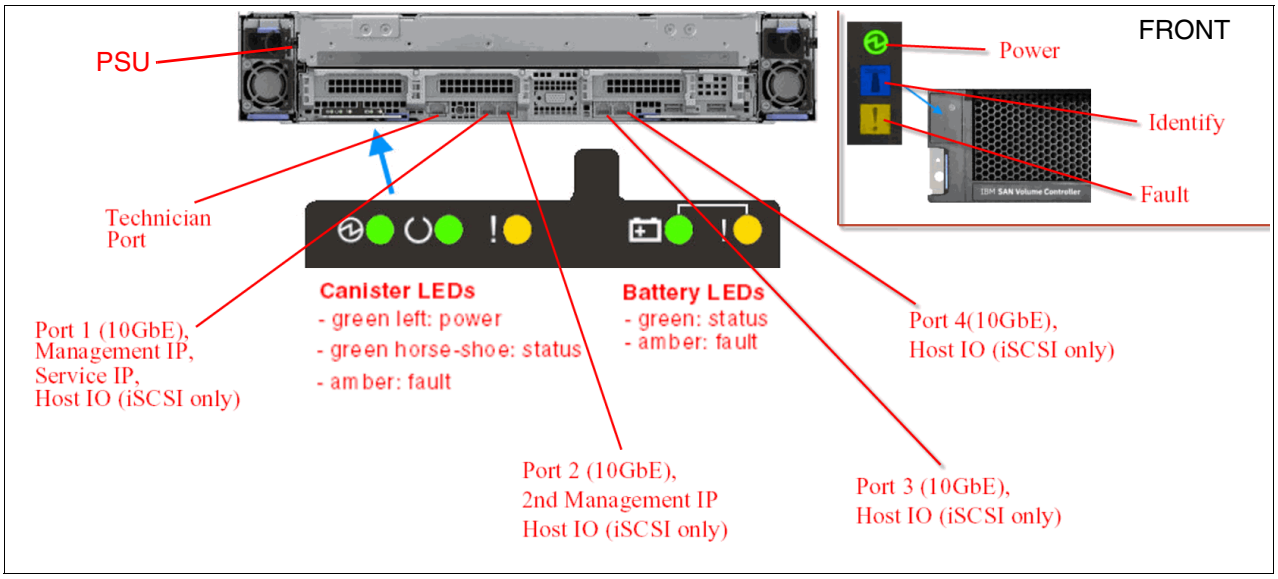


Figure 13-2 Rear ports and indicators of IBM SAN Volume Controller model 2145-SV2

Note: Unlike previous IBM SAN Volume Controller models, SV2/SA2 does not support direct serial-attached Small Computer System Interface (SCSI) (SAS) attached drive enclosures. Therefore, no optional SAS adapters are supported, nor are expansion enclosures.

Host interface cards

In this section, we focus on IBM SAN Volume Controller models that are marketed at the time of this writing: SV2 and SA2. Each IBM SAN Volume Controller node has three host interface card (HIC) slots. Nodes in the same I/O group must have the same HIC configuration. The available HICs are listed in Table 13-1.

Table 13-1 Available host interface cards

Protocol/HIC Type	Feature Code	Supported number of cards	Ports	Allowed slots
16 Gb Fibre Channel (FC)/Fibre Channel over NVMe	AH14	0-3	4	1, 2, 3
32 Gb FC/Fibre Channel over NVMe	AH1D	0-3	4	1, 2, 3
25 Gb iSCSI/RDMA over Converged Ethernet (RoCE)/Non-Volatile Memory Express (NVMe) over Ethernet	AH16	0-3	2	1, 2, 3
25 Gb iSCSI/internet Wide-area RDMA Protocol (iWARP)/NVMe over Ethernet	AH17	0-3	2	1, 2, 3

Note: To achieve the best possible performance, it is recommended to configure the system with a minimum of two host bus adapters (HBAs) of any type. Also, Fibre Channel over Ethernet (FCoE) is *not* supported. For more information, see 2.6, “Fibre Channel SAN configuration planning” on page 65.

Table 13-2 lists the meaning of port LEDs for an FC configuration.

Table 13-2 Fibre Channel link LED status

Port LED	Color	Meaning
Link status	Green	Link is up, connection established.
Speed	Amber	Link is not up or a speed fault exists.

USB ports

Two USB ports are available in the rear of each node. They are not numbered and do not have any indicators that are associated. These USB ports can be used for initial cluster setup, storing encryption keys, and node status or log collection.

Ethernet Ports and their LED status

Four 10-Gigabit Ethernet (GbE) ports and one 1-Gigabit Ethernet port are available on each node. However, not all ports are equal. Their functions are listed in Table 13-3. The location of the technician port on a node is shown in Figure 13-2 on page 856.

Table 13-3 On-board Ethernet Ports and their function

On-board Ethernet port	Speed	Function
1	10 GbE	Management IP, Service IP, Host I/O (iSCSI only), Replication I/O
2	10 GbE	Secondary Management IP, and Host I/O (iSCSI only), Replication I/O
3	10 GbE	Host I/O (iSCSI only), Replication I/O
4	10 GbE	Host I/O (iSCSI only), Replication I/O
T	1 GbE	Technician Port: service activities and initial setup by way of direct attached PC with automatic (DHCP) IP address assignment

Ethernet LED Statuses are listed in Table 13-4.

Table 13-4 Ethernet LED status

LED	Color	Meaning
Link state	Green	On when an Ethernet link exists.
Activity	Amber	Flashing when activity occurs on the link.

Node front status LEDs

Three LEDs are in a row on the front of the node that indicates the status and the functions of the node (see Table 13-5).

Table 13-5 Node Front LEDs

Position	Color	Name	State	Meaning
Top	Green	Power	On	The CPU is active (not in standby) in the node (canister).
			Slow flash	Node is in standby mode.
			Off	No power to the canister or it is running on battery.
Middle	Blue	Identify	On	Node is being identified. Fault LED on rear of canister will also be flashing.
			Off	Default.
Bottom	Amber	Fault	On	Error. Node is in service state or there is a problem with the node starting. Also illuminates when battery is charging and node is waiting to come up.
			Off	No Faults. Candidate/Active State.

Node rear status LEDs

Three status LEDs in a row on the rear of the node indicate the status and the functions of the node (see Table 13-6 on page 858).

Table 13-6 Rear Node LEDs

Position	Color	Name	State	Meaning
Left	Green	Power	On	The node is started and active. If the fault LED is off, the node is an active member of a cluster or candidate. If the fault LED is also on, node is in service state or in error, preventing the software to start.
			Flashing (2 Hz)	Node is started and in standby mode.
			Flashing (4 Hz)	Node is running power-on self-test (POST).
			Off	No power to the node 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 for or in a service state.
			Flashing (4 Hz)	The node is performing a fire hose dump. Never unplug the node now.
			Off	No power to the node or node is in standby mode.

Position	Color	Name	State	Meaning
Right	Amber	Fault	On	The node is in a service state, or in error, for example, a POST error that is preventing the software from starting.
			Flashing (2 Hz)	Node is being identified.
			Off	Node is either in candidate or active state.

Battery LEDs

Immediately to the right of the canister LEDs, with a short gap in between, are the Battery LEDs that provide the status of the battery (see Table 13-7).

Table 13-7 Battery LEDs

Position	Color	Name	State	Meaning
Left	Green	Status	On	Indicates that the battery is fully charged and thus has sufficient charge to complete two fire hose dumps.
			Flashing (2 Hz)	Indicates that the battery has sufficient charge to complete a single fire hose dump only.
			Flashing (4 Hz)	Indicates that the battery is charging and has insufficient charge to complete a single fire hose dump.
			Off	Indicates that the battery is not available for use (for example, missing or containing a fault).
Right	Amber	Fault	On	Indicates that a battery has a fault or condition occurred. The node enters service state.
			Off	Indicates that there are no known battery faults or conditions. An exception to this status is when a battery has insufficient charge to complete a single fire hose dump.

13.1.2 Power

IBM SAN Volume Controller nodes accommodate two power supply units (PSUs) for normal operation. For this reason, it is highly advised to supply AC power to each PSU from different power distribution units (PDUs) and, if possible, from different power circuits. For more information, see 2.2, “Planning for availability” on page 61.

IBM SAN Volume Controller SV2/SA2 nodes are equipped with one Backup Battery Unit (BBU) to maintain data integrity in case of an input power failure. If the AC input power fails, the node continues to operate on battery power for five seconds before it is being shut down gracefully after cache content and the system data are saved to the node’s internal drive, which is also called a *fire hose dump*.

A fully charged battery is designed to sustain two fire hose dumps. 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.

Figure 13-3 shows two PSUs that are present in an IBM SAN Volume Controller SV2/SA2 node. The controller PSU on a SV2/SA2 model 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 supplied to this PSU.

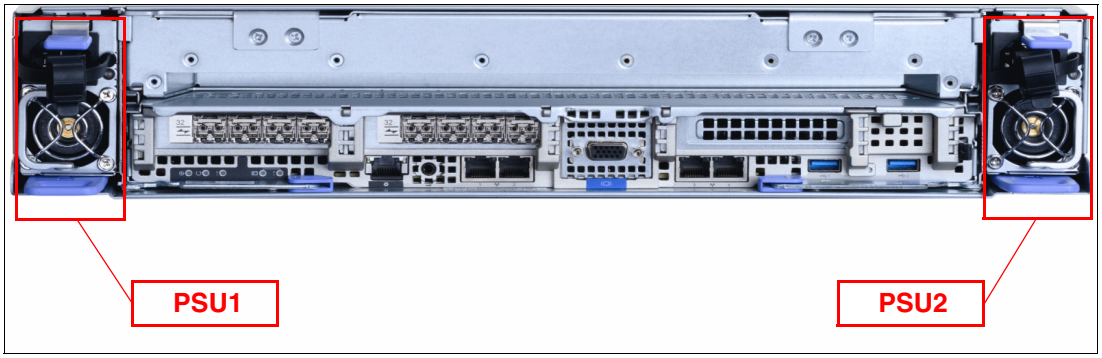


Figure 13-3 IBM SAN Volume Controller power supplies

Power supplies in IBM SAN Volume Controller nodes are hot-swappable and can be replaced without the need to shut down a node.

If the power is interrupted in one node for less than 5 seconds, the node does not perform a fire hose dump, and continues operation from the battery.

13.1.3 Shutting down an IBM SAN Volume Controller cluster

You can safely shut down an IBM SAN Volume Controller cluster by using the GUI or CLI.

Important: *Never* shut down an IBM SAN Volume Controller clustered system by removing PSUs, unplugging both power cables from the node PSUs, or switching off AC input power source. Doing so can lead to inconsistency or loss of data that is staged in the cache.

Before shutting down the cluster, stop all hosts that feature volumes that are allocated from the system. This step can be skipped for hosts that have volumes that also are provisioned with mirroring (host-based mirror) from another storage device. However, doing so incurs errors that are related to lost storage paths or disks in the hosts' error log.

You can shut down a single node or the entire clustered system. When you shut down only one node, host I/O can continue through the partner node in the same I/O group. The system's overall redundancy is reduced because this I/O group then depends on the remaining node.

Also, the write cache is deactivated for this I/O group because host write I/Os cannot be mirrored to the partner node any longer. When you shut down the entire cluster, IBM SAN Volume Controller nodes before SV2/SA2 must be powered on locally to restart the system.

If input power to the IBM SAN Volume Controller clustered system is removed for more than a few minutes (for example, if the machine room power is shut down for maintenance), it is important that you shut down the IBM SAN Volume Controller system immediately.

Shutting down the system while it is still connected to the main power ensures that the internal node batteries are still fully charged when the power is restored.

If main power is removed while the system is still running, the loss of power is detected and node shutdown process starts. This shutdown can take several minutes to complete.

Although the internal batteries are sufficiently powered to perform the shutdown, the node batteries are drained.

The nodes start when power is restored. However, if the node batteries have insufficient charge to survive another power failure so that the node can perform another clean shutdown, the node enters service mode. You do *not* want the batteries to run out of power in the middle of the node's shutdown.

It can take up to approximately three hours to charge the batteries sufficiently for a node to come online.

Important: When a node shuts down upon losing input AC power, the node dumps the content of the write cache to its internal boot drive so that the cached data can be retrieved and de-staged to the back-end storage when the system restarts.

The IBM SAN Volume Controller internal batteries are designed to survive at least two power failures in a short period. After that period, the nodes do not come online until the batteries have sufficient power to survive another immediate power failure.

During maintenance activities, if the batteries detect power and then detect a loss of power multiple times (the nodes start and shut down more than once in a short time), you might discover that you unknowingly drained the batteries. In this case, you must wait until they are charged sufficiently before the nodes start again.

To shut down your IBM SAN Volume Controller system, complete the following steps:

1. From the **Monitoring** → **System Hardware** window, click **System Actions**, as shown in Figure 13-4 on page 861. Click **Power Off System**.

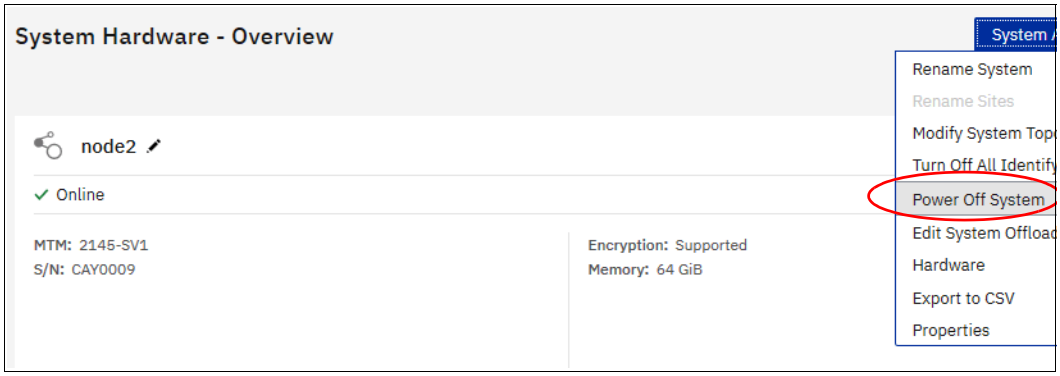


Figure 13-4 Action window to power off the system

A confirmation window opens, as shown in Figure 13-5.

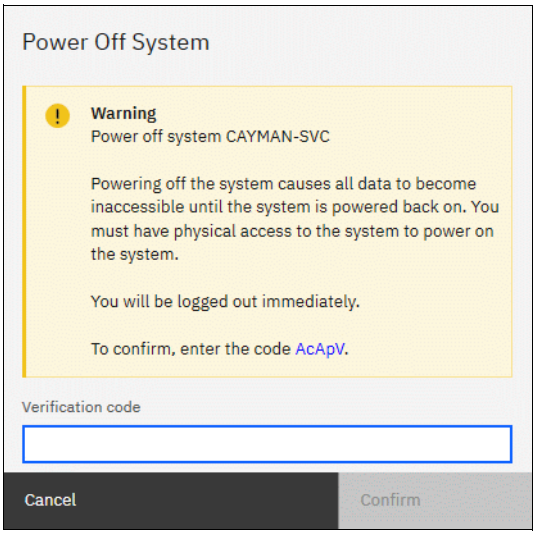


Figure 13-5 Confirmation window to confirm the shutdown of the system

2. Before you continue, ensure that all FlashCopy mappings, Remote Copy (RC) relationships, data migration operations, and forced deletions are stopped.
3. Enter the generated confirmation code and click **OK** to begin the shutdown process.

> **Attention:** Pay special attention when encryption is enabled on one or more storage pools. You must insert a USB drive with the stored encryption keys, or you must ensure that your IBM SAN Volume Controller can communicate with the configured key server (for instance IBM Security Guardium Key Lifecycle Manager) to retrieve the encryptions keys. Otherwise, the data is unreadable after restart.

13.2 Removing or adding a node to or from the system

Situations exist in which IBM Support staff might advise you to temporarily remove a node from the system configuration. Typical use cases for this situation might be a concurrent node replacement, a stuck code upgrade, or other maintenance activities.

The easiest way to remove a node from the cluster is to use a CLI command. Run the **`lsnode`** command to display all nodes' IDs, names status, and other details, as shown in Example 13-1. Make sure that the partner node of the node to be removed from the I/O group and from the cluster is online, which is done to ensure host I/O does not get interrupted.

Example 13-1 lsnode output (redacted for better readability)

```
IBM_2145:ITSO-SVC:superuser>lsnode
id name  WWNN      status IO_group_id IO_group_name config_node hardware iscsi_name
panel_name
1  node1  500507680100A288 online 0           io_grp0       yes         SV1
iqn.1986-03.com.ibm:2145.itso-svc.node1 78GZYT0
2  node2  50050768OC009B73 online 0           io_grp0       no          SV1
iqn.1986-03.com.ibm:2145.itso-svc.node2 78GZYL0
```

Also, before removing a node, ensure that no volumes depend on the node to be removed. While the **rmnode** CLI command (when issued without the *-force* parameter) conducts a background check for any dependant volumes, it is a best practice to manually conduct this check up front by using the **lsdependentvdisks** command, as shown in Example 13-2. The wanted result is that no output is produced because no VDisks depend on the node to be removed.

Example 13-2 lsdependentvdisks output

```
IBM_2145:ITSO-SVC:superuser>lsdependentvdisks -node node2
```

In this example, we remove node2 from the cluster. Run the **rmnode 2** command, as shown in Example 13-3.

Example 13-3 removing a node with rmnode command

```
IBM_2145:ITSO-SVC:superuser>rmnode node2
```

A node can also be removed by using the management GUI. Select **Monitoring** → **System Hardware**. Select the node to be removed, which brings up the **System Hardware - Node Details** window. At the top of the window, right-click and select **Node Actions** → **Remove**, as shown in Figure 13-6, which brings up a confirmation window before removing the node from the cluster.

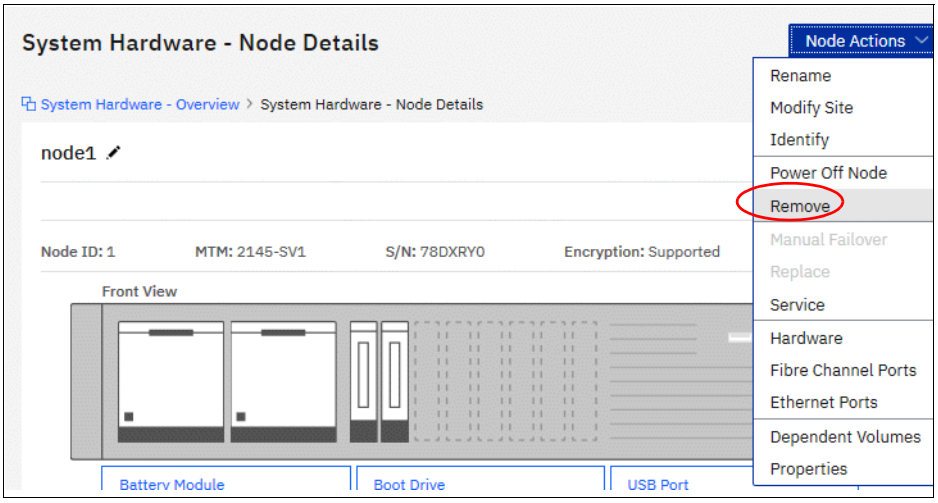


Figure 13-6 Remove Node GUI command

After the node was removed from the cluster, it no longer appears in the output of the **lsnode** command nor in the list of nodes in the GUI view of **Monitoring** → **System Hardware** → **System Hardware - Overview**.

Equally, the Service Assistant Tool (SAT) GUI does reflect the node is not longer an active member of the clustered system but a Candidate instead, as shown in Figure 13-7.

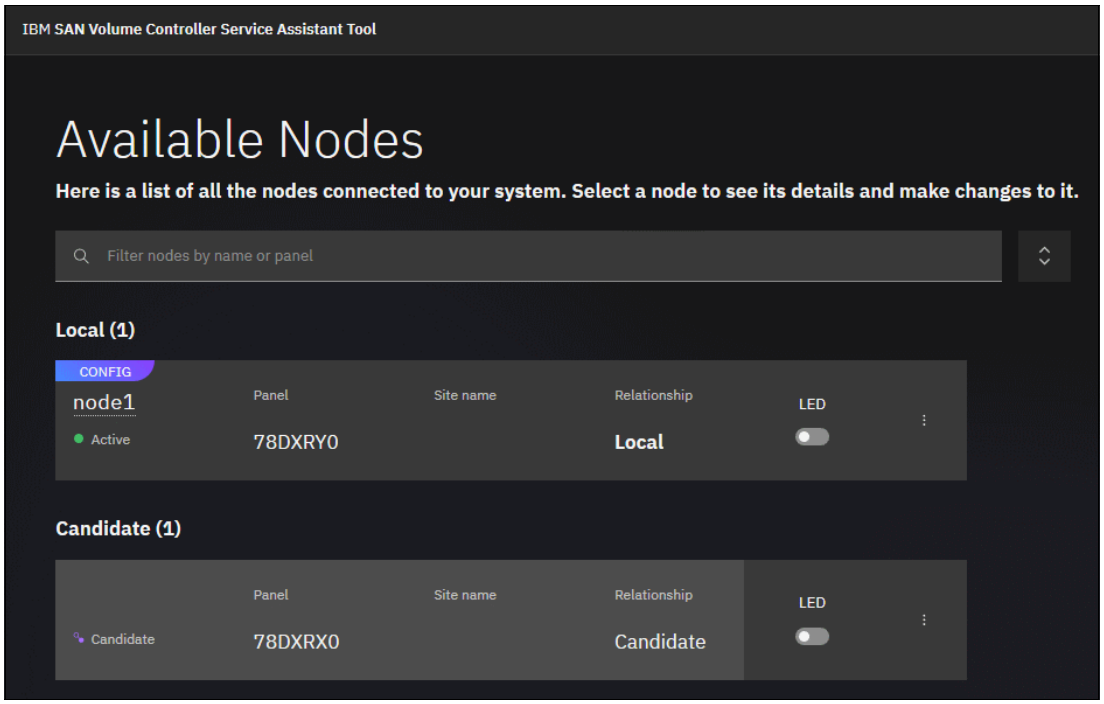


Figure 13-7 SAT GUI showing removed node as Candidate

Note: Write cache for that I/O group is flushed upon issuing the node removal command and remains disabled until both nodes in that I/O group are online again. This behavior is to is to prevent data loss because host write I/Os cannot be mirrored to the partner node.

A node with a Relationship of `Candidate` can be readded to the cluster by using the management GUI or CLI.

In the GUI, select **Add Node** in the **Monitoring** → **System Hardware Overview**. The Add Nodes view is displayed, as shown in Figure 13-8.

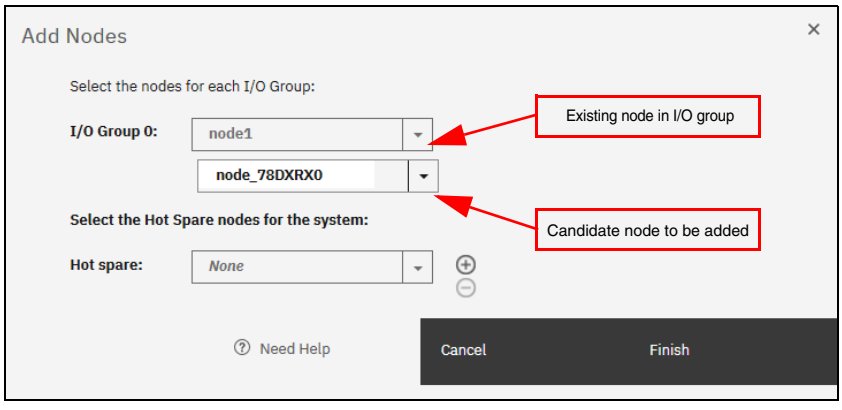


Figure 13-8 Adding a node by using the GUI

Alternatively, the node can be added back to the cluster by using **addnode** CLI command, as shown in Example 13-4.

Example 13-4 Adding a node by way of CLI

```
IBM_2145:ITSO-SVC:superuser>addnode -panelname 78DXRX0 -iogrp io_grp0 -name node2
Node, id [6], successfully added
```

Now, check again by using the **lsnode** CLI command to ensure that the node was successfully added back, as shown in Example 13-5.

Example 13-5 Node added by way of CLI addnode is back online (redacted for better readability)

```
IBM_2145:ITSO-SVC:superuser>lsnode
id name   WWNN             status IO_group_id IO_group_name config_node hardware
iscsi_name                              panel_name
1  node1  500507680100A288 online 0           io_grp0       yes         SV1
iqn.1986-03.com.ibm:2145.itso-svc.node1 78GZYT0
8  node2  500507680C009B73 online 0           io_grp0       no          SV1
iqn.1986-03.com.ibm:2145.itso-svc.node2 78GZYL0
```

Note: The `node id` features increased automatically upon adding the node to the cluster. It is *not* user configurable.

13.3 Configuration backup

A cluster's metadata consists (besides other components) of the clustered system's state and the configuration, including the association of host-facing volumes with the data that is stored in the back-end. This part of the hardened data is also known as *quorum data*. It is synchronized between all active cluster nodes and stored on the quorum MDisks, drives, or IP quorum devices. It can be used to restore the system after, for example, a loss of multiple cluster nodes during a system recovery, which is also known as *T3 recovery*.

Many user-configured settings are stored in a backup file on the configuration node's internal boot drive. This configuration backup occurs automatically as a scheduled task once a day at 1:00 AM cluster time by means of a cron job. The automatically generated backup file is named `svc.config.cron.xml_<config_node_panel_id>` and is in the `/dumps` directory.

Note: The configuration backup file alone (that is, without the quorum data) does not sufficiently restore the system status before a disaster. However, it can be used to restore the base configuration to prepare the system for restoring data from a backup. This procedure also is called a *T4 recovery*.

Because the configuration backup file can be downloaded from the system, it can be stored externally. The following options are available to download the configuration backup file from the system:

- Management GUI

  To download the configuration backup by using the GUI, click **Settings** → **Support** → **Support Package** → **Download Support Package** → **Download Existing Package**.

- Service Assistant Tool (SAT) GUI.

  Log on to the SAT GUI, select the configuration node and select **CONFIG node** → **Collect Logs** → **Download existing logs** → **svc.config.backup**.

- Secure Copy Protocol (SCP):

  - SCP is based on Secure Shell (SSH), which also is used to connect to the CLI.
  - SCP is the most suitable option for automating the process of downloading the configuration backup.
  - Because SCP uses an SSH session as a transport medium, SSH keys that are configured for the CLI login also can be used here; therefore, a password does not need to be entered.
- Support package alias snap

  A Spectrum Virtualize support package, also known as a *snap*, contains the configuration backup files from the system's `/dumps` folder.

You can download and save the configuration backup file by using the IBM Spectrum Virtualize GUI or CLI. On an ad hoc basis, manually perform this procedure because you 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 a specific Secure Copy Protocol (SCP). The CLI option is a best practice for an automated backup of the configuration.

Important: It is advisable to perform a daily backup of the IBM Spectrum Virtualize configuration backup file. A best practice is to automate this task. Before conducting any critical maintenance task, such as a code upgrade of the IBM Spectrum Virtualize software, a configuration backup must be triggered manually and copied off the system,

The backup file is updated by the cluster daily. Saving it after you make any changes to your system configuration also is important. It contains the configuration data of arrays, pools, volumes, and so on. The backup does *not* contain any volume data.

The following prerequisites must be met to successfully perform the configuration backup:

- 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: An ad hoc configuration backup can be done by using CLI only by running the **`svcconfig backup`** command. Then, the output of the command can be downloaded from the GUI.

13.3.1 Force configuration backup by using the CLI

In addition to the nightly scheduled config backup, you can use the CLI to trigger a configuration backup manually or by using 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 can choose to save the automated 1AM cron backup if no configuration changes were made,

Example 13-6 shows output of the **`svcconfig backup`** command.

Example 13-6 Saving the configuration by using the CLI

```
IBM_2145:ITSO-SVC:superuser>svcconfig backup
```

```
.....................................................................
.....................................................................
.....................
CMMVC6155I SVCCONFIG processing completed successfully
IBM_2145:ITSO-SVC:superuser>
```

The **svcconfig backup** command generates three files that provide information about the backup process and cluster configuration. These files are stored into the /dumps directory on the configuration node. Use the **lsdumps** command to list them (see Example 13-7).

Example 13-7 Listing the backup files by using the CLI

```
IBM_2145:ITSO-SVC:superuser>lsdumps | grep backup
26 svc.config.backup.bak_78CLRTO
65 svc.config.backup.log_78CLRTO
66 svc.config.backup.sh_78CLRTO
67 svc.config.backup.xml_78CLRTO
IBM_2145:ITSO-SVC:superuser>
```

Table 13-8 lists the files that are created by the backup process.

Table 13-8 Files that are created by the backup process

File name	Description
svc.config.backup.xml	Contains your cluster configuration data.
svc.config.backup.sh	Contains the names of the commands that were run to create the backup of the cluster.
svc.config.backup.log	Contains details about the backup, including any error information that might be reported.
svc.config.cron.xml	Written by the nightly automatic backup.

Save the current backup to a secure and safe location. The files can be downloaded by using **scp** (UNIX) or **pscp.exe** (Windows), as shown in Example 13-8. Replace the IP address with the cluster IP address of your IBM SAN Volume Controller and specify a local folder on your workstation. In this example, we are saving to C:\SVCbackups.

Example 13-8 Saving the config backup files to your workstation

```
C:\putty>
pscp.exe -unsafe superuser@10.41.160.201:/dumps/svc.config.backup.* c:\SVCbackups
Using keyboard-interactive authentication.
Password:
svc.config.backup.log_78C  |  33 kB |  33.6 kB/s | ETA: 00:00:00 | 100%
svc.config.backup.sh_78CL  |  13 kB |  13.9 kB/s | ETA: 00:00:00 | 100%
svc.config.backup.xml_78C  | 312 kB |  62.5 kB/s | ETA: 00:00:00 | 100%

C:\>dir SVCbackups

 Directory of C:\SVCbackups

17.10.2018  09:20    <DIR>          .
17.10.2018  09:20    <DIR>          ..
17.10.2018  09:20            34.415 svc.config.backup.log_78CLRTO
17.10.2018  09:20            14.219 svc.config.backup.sh_78CLRTO
```

By using the **-unsafe** option, you can use a wildcard for downloading all `svc.config.backup` files by using a single command.

Tip: If you encounter the following error when the **pscp.exe** command is run, consider upgrading your version of PuTTY:

```
Fatal: Received unexpected end-of-file from server
```

13.3.2 Saving the backup by using the GUI

Although it is not possible to generate an ad hoc backup, you can download the backup files by using the GUI. To do so, complete the following steps:

1. Select **Settings** → **Support** → **Support Package** → **Download Support Package** → **Download Existing Package**, as shown in Figure 13-9.

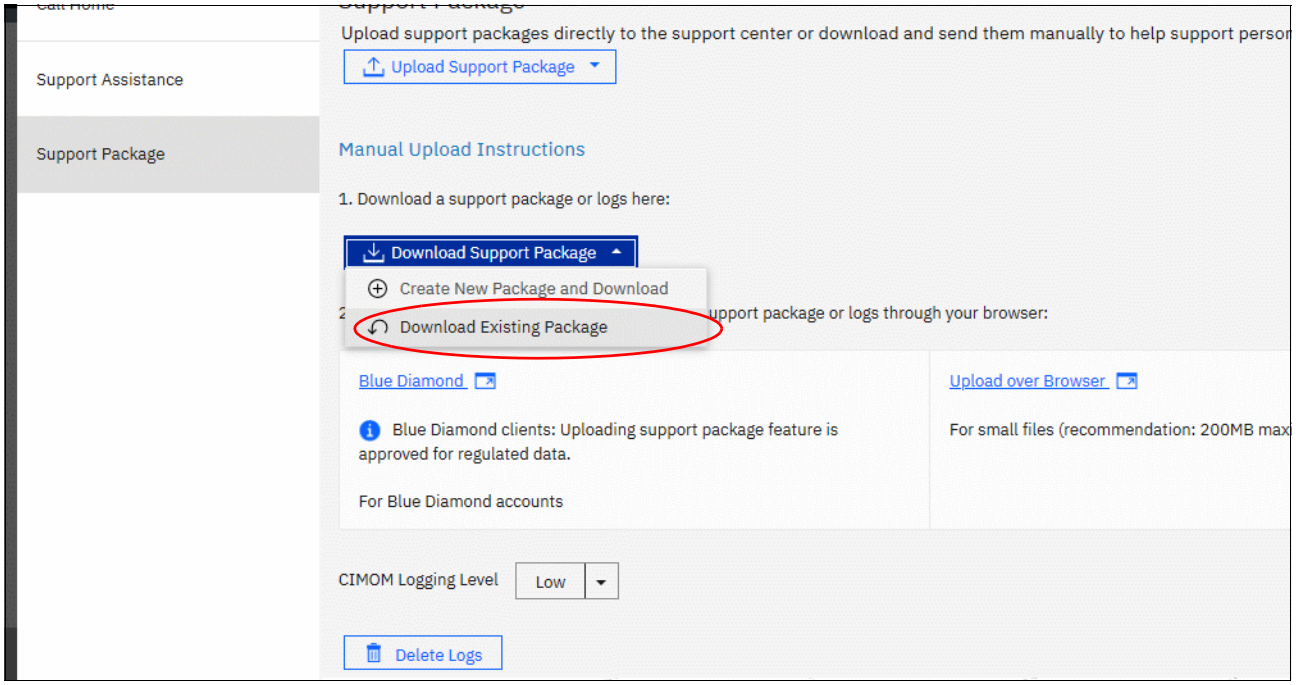


Figure 13-9 Downloading the support package

The Select Support Package or Logs to Download window opens, as shown in Figure 13-10. You can scroll through the list of files, or enter the file name or parts of the name it into the blue filter field.

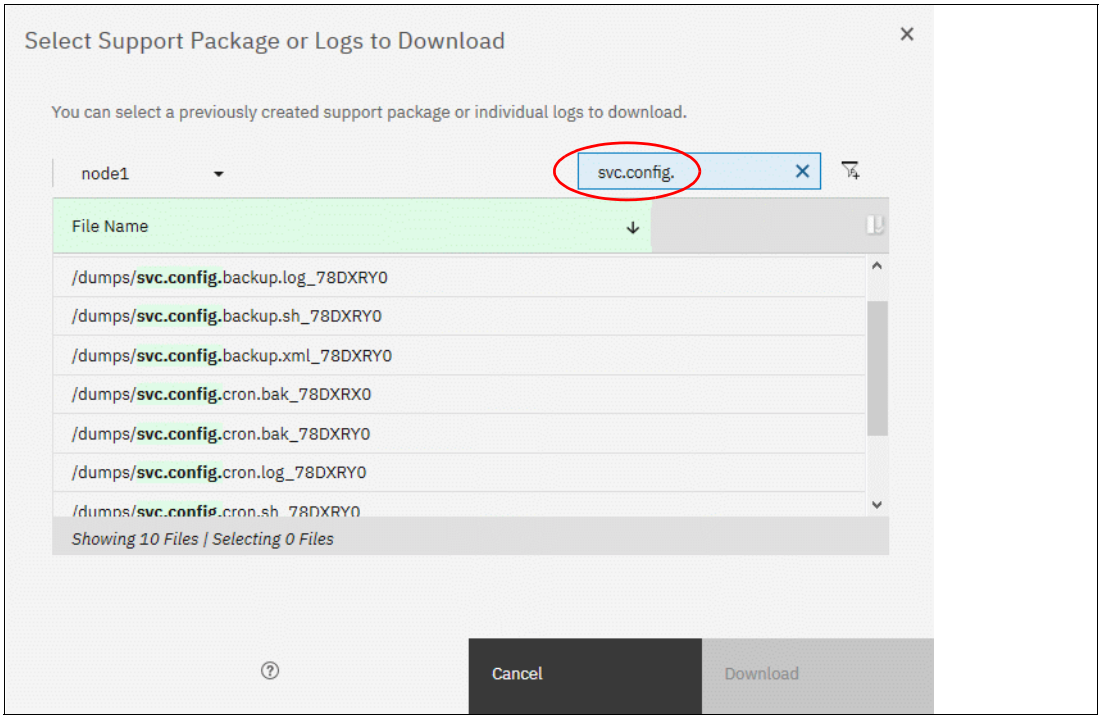


Figure 13-10 Select Support Package or Logs to Download window

2. Click the wanted file to select it or select multiple files by pressing the **Ctrl** key and then, click **Download**.

   Depending on your browser preferences, you might be prompted where to save the file or it downloads to your defined download directory.

   If you selected multiple files for download, they are downloaded as compressed archive.

Note: Make sure you selected the configuration node from the **node** drop-down menu in the upper left of the window. It is preselected normally when you browse to this menu. This selection is necessary because the backup files are stored on the configuration node.

13.4 IBM Spectrum Virtualize Code Upgrade

This section provides information about IBM Spectrum Virtualize software release numbering and describes the code upgrade procedure.

To stay updated about software updates, support bulletins, and further support information, log in to the IBM Support My Notifications web page.

Note: Regarding Spectrum Virtualize software, the terms *code update* and *code upgrade* are used interchangeably, as are the terms *software* and *code*.

13.4.1 Spectrum Virtualize Version Numbering

IBM Spectrum Virtualize software releases (for example, 8.4.2.0) are numbered following the VRMF naming scheme:

- V: Version
- R: Release
- M: Major
- F: Fix

The software update package file name ends in the target code level's VRMF value; that is, four positive integers that are separated by dots. For example, a software update package might use the following name:

```
IBM_SVC_INSTALL_8.4.2.0
```

13.4.2 Updating the Spectrum Virtualize system software

Updating the IBM Spectrum Virtualize software that is running on a clustered system is a concurrent task and host I/O is not interrupted. During the update process, most of the configuration tasks are restricted. Also, the system's redundancy and performance are degraded temporarily.

Therefore, a software update must be thoroughly planned and prepared. For example, it is not advisable to perform an update during periods of extraordinary high load on the system.

13.4.3 Precautions before the update

This section describes the precautions to take before a code upgrade is conducted.

Important: When you plan for an IBM Spectrum Virtualize code update, read and understand the Concurrent Compatibility and Code Cross Reference for Spectrum web page and the Release Notes of the target code level.

During the update, each individual node in your 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, host I/O can continue without interruption.

If set up and configured correctly, the hosts' multi-pathing software ensures continued access to the presented volumes. Therefore, we advise that you verify that all I/O paths between all hosts and the clustered system are fully functional.

In addition you might want to check for `degraded host` through cluster GUI view **Hosts** → **Hosts**,. Alternatively, use CLI command **`lshost -filtervalue status=degraded`**.

If you do not perform this check, hosts are at risk to lose connectivity to their volumes and experience I/O errors if they are not logged in to both nodes of an I/O group.

13.4.4 IBM Spectrum Virtualize upgrade test utility

The software upgrade test utility is an installable tool that checks for known issues that might cause problems during an IBM Spectrum Virtualize software update. Sometimes, it is also referred to as *CCU Checker*, whereby CCU stands for Concurrent Code Upgrade.

For more information about this utility, see the Software Upgrade Test Utility web page.

Always check for and download the latest version of the software upgrade test utility from IBM Fix Central. The utility is updated frequently to support new IBM Spectrum Virtualize code levels and new hardware, and updated checks to be conducted in preparation of a code update.

The software upgrade test utility can be downloaded in advance of the update process. Alternatively, it can be downloaded and run directly during the software update, as guided by the update wizard.

You can run the utility several times on a system to perform a readiness check-in preparation for a software update. Run this utility for a final time immediately before you apply the IBM Spectrum Virtualize code update.

The installation and use of this utility is nondisruptive. It does not require a restart of any nodes. Therefore, no interruption to host I/O occurs. The utility is installed and run on the current configuration node.

This utility is intended to supplement rather than duplicate the existing tests that are performed by the IBM Spectrum Virtualize update procedure (for example, checking for unfixed errors in the error log).

13.4.5 Performing the upgrade

To update the IBM Spectrum Virtualize software, complete the following steps:

1. In a supported web browser, enter your cluster IP address or host name.
   Log in to your system. Any account with at least `Administrator` privileges can perform the upgrade. Access to the `superuser` account is possible if a service action that incorporates the use of the Service Assistant Tool (SAT) GUI or CLI becomes necessary.
2. In the IBM SAN Volume Controller management home window, click **Settings** → **System** → **Update System**. The current software level and Test Only and Test & Update options are displayed, as shown in Figure 13-11.

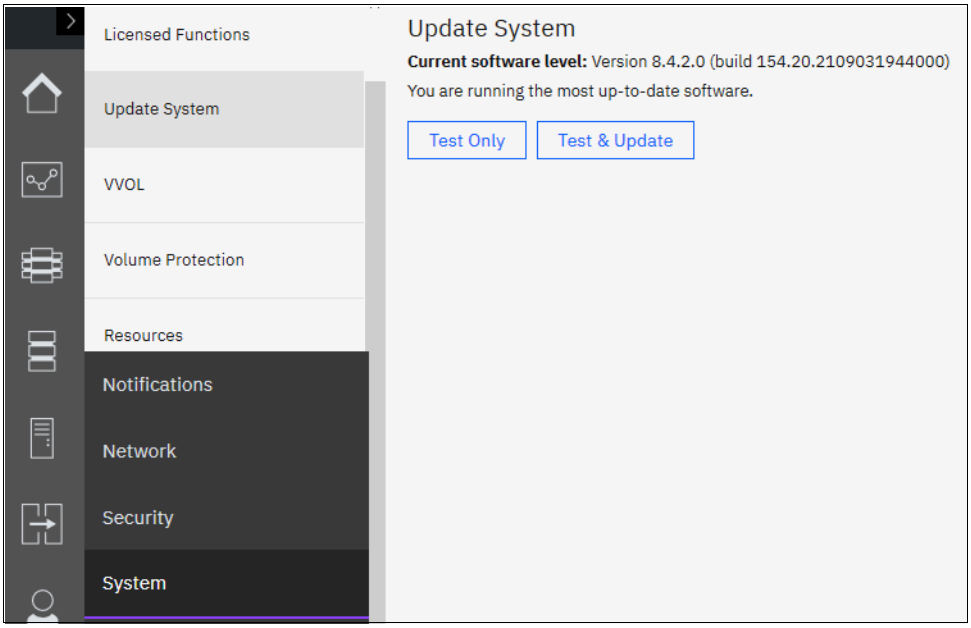


Figure 13-11 Update System window

3. From this window, you can select to run the update test utility and continue with the code update, or run only the test utility. For our example, we click **Test and Update**.
4. Because you downloaded both files from the SAN Volume Controller Support web page, you can click each folder, browse to the location where you saved the files, and upload them to the IBM SAN Volume Controller cluster. The uploaded code upgrade package is checked for integrity. The GUI detects and displays the target code level, as shown in Figure 13-12.

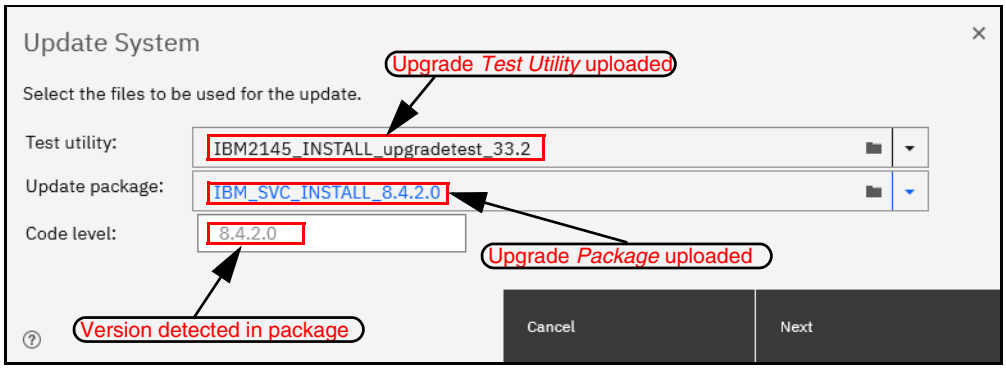


Figure 13-12 Test Utility and Update Package uploaded

5. Select the type of update that you want to perform, as shown in Figure 13-13. Select **Automatic update** unless IBM Support suggested a **Service Assistant Manual update**. The manual update might be preferable in cases where misbehaving host multipathing is known to cause loss of access.

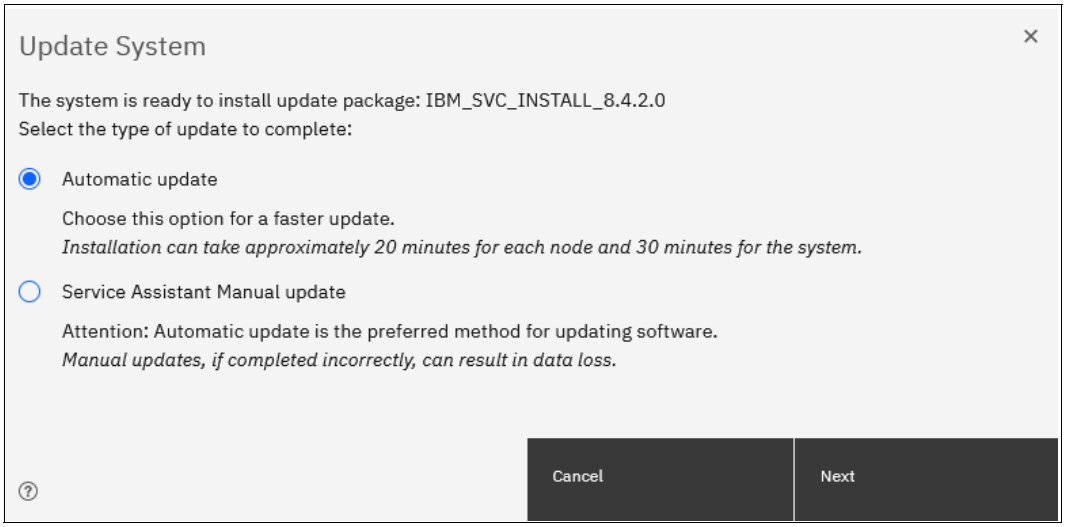


Figure 13-13 Software update type selection

Click **Next** to begin the update package upload process.

6. When updating from Version 8.1 or a later level, an extra window opens in which you can choose a fully automated update, such as one that pauses when *half the nodes* completed an update or *after each node* update, as shown in Figure 13-14. Click **Finish**.

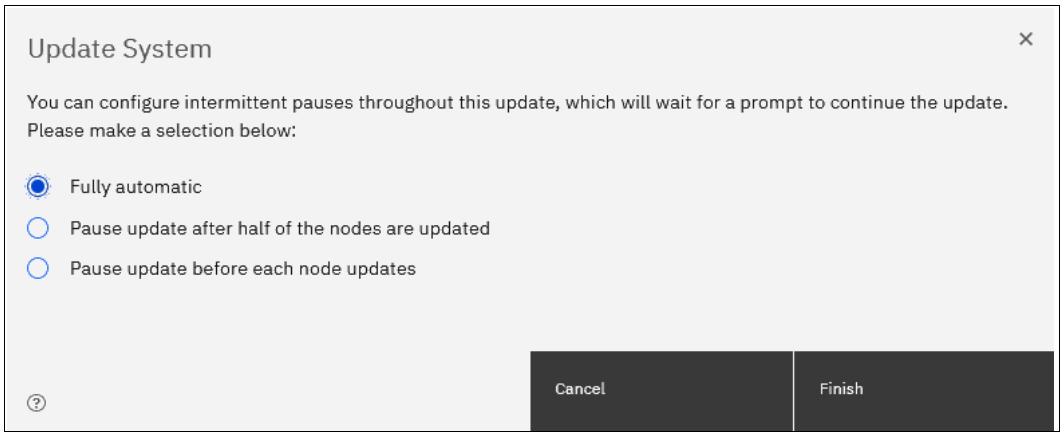


Figure 13-14 Update pause options

7. After the packages are uploaded, the update test utility conducts a check for any known issues that might affect a concurrent update of your system. The GUI helps identify any detected issues, as shown in Figure 13-15.

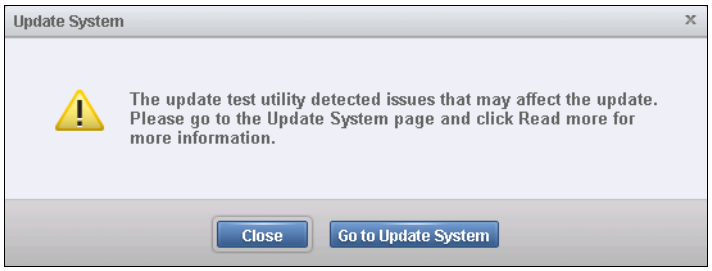


Figure 13-15 Issue detected

8. Click **Go to Update System** to return to the Update System window. Then, click **Read more**.
9. The results window opens and shows you the issues that were detected (see Figure 13-16). In our example, the warning is that email notification (Call Home) is not enabled. Although this condition is not recommended, it does not prevent the system update from running. Therefore, we can click **Close** and proceed with the update. However, you might need to contact IBM Support to resolve more serious issues before continuing.

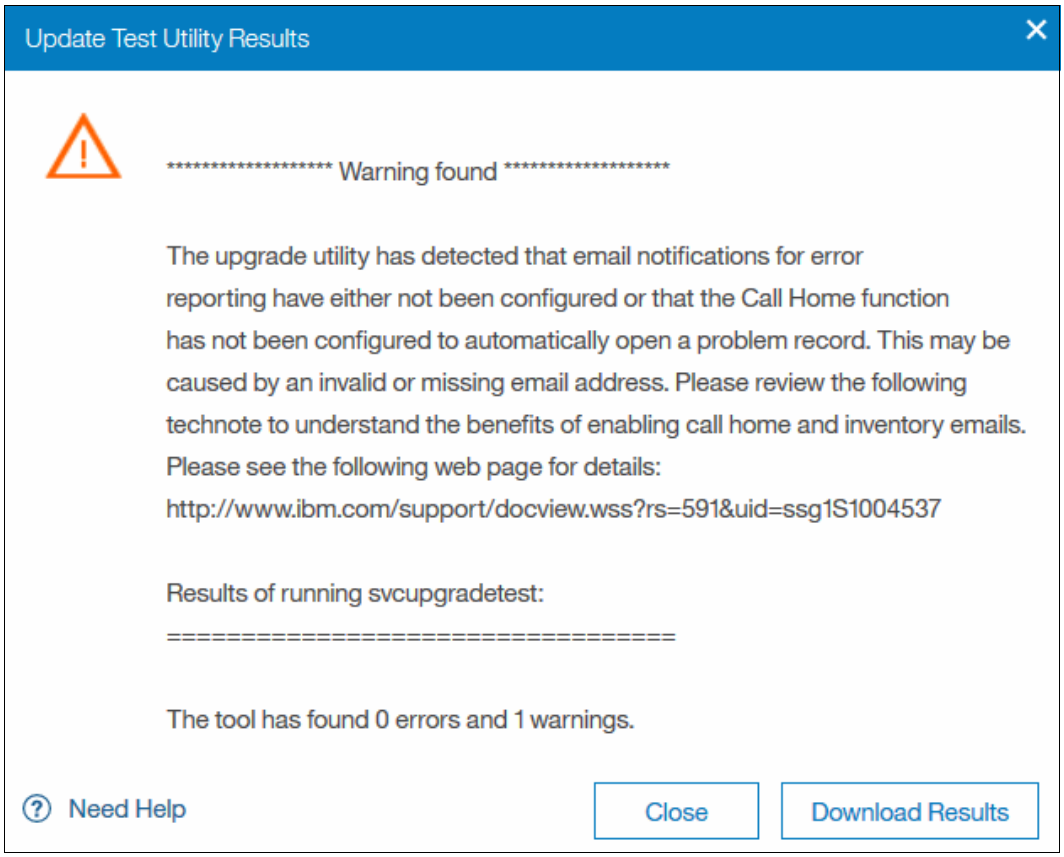


Figure 13-16 Description of the warning from the test utility

10. Click **Resume** in the Update System window and the update proceeds (see Figure 13-17).

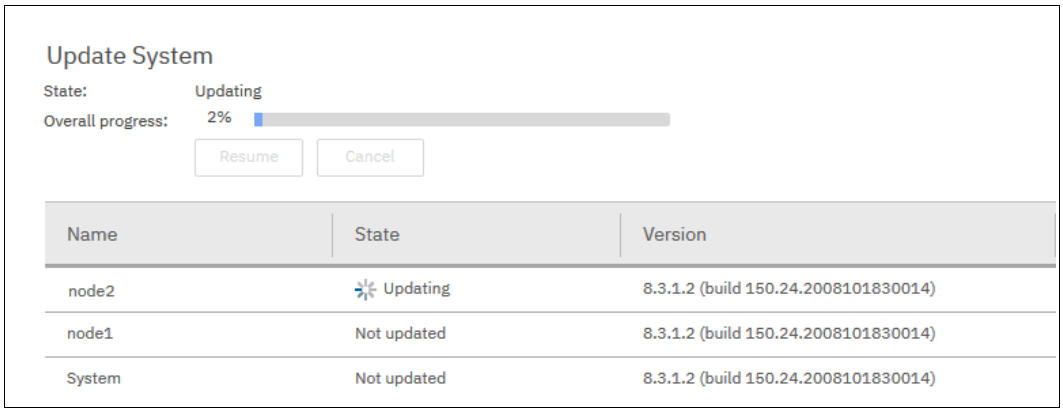


Figure 13-17 Resuming update

The system begins updating the IBM Spectrum Virtualize software by taking one node offline and installing the new code. This process takes approximately 20 minutes. After the node returns from the update, it is listed as `Complete`, as shown in Figure 13-18.

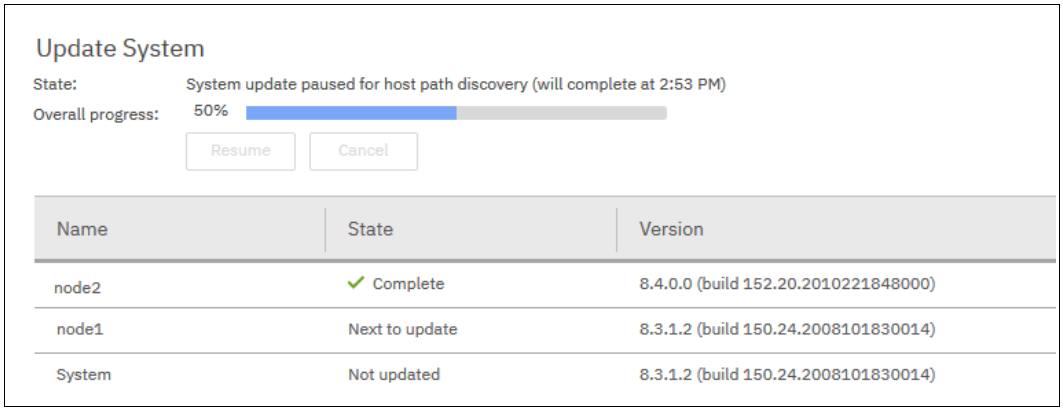


Figure 13-18 Update process pause

After a 30-minute pause to ensure that multi-pathing settled for the attached hosts, a node failover occurs and you temporarily lose connection to the GUI. A warning window opens, in which you are prompted to refresh the current session, as shown in Figure 13-19.

Tip: The 30-minute wait period can be adjusted by running the **`applysoftware`** command with the **`-delay (mins)`** parameter to begin the update instead of using the GUI.

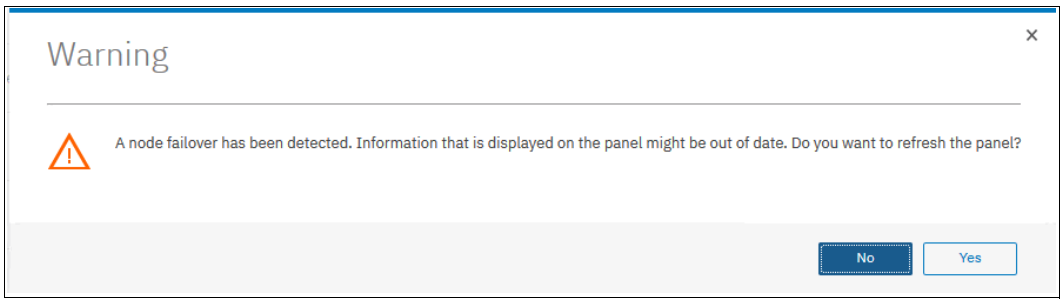


Figure 13-19 Node failover

You now see the Version 8.4 GUI and the status of the second node updating, as shown in Figure 13-20.

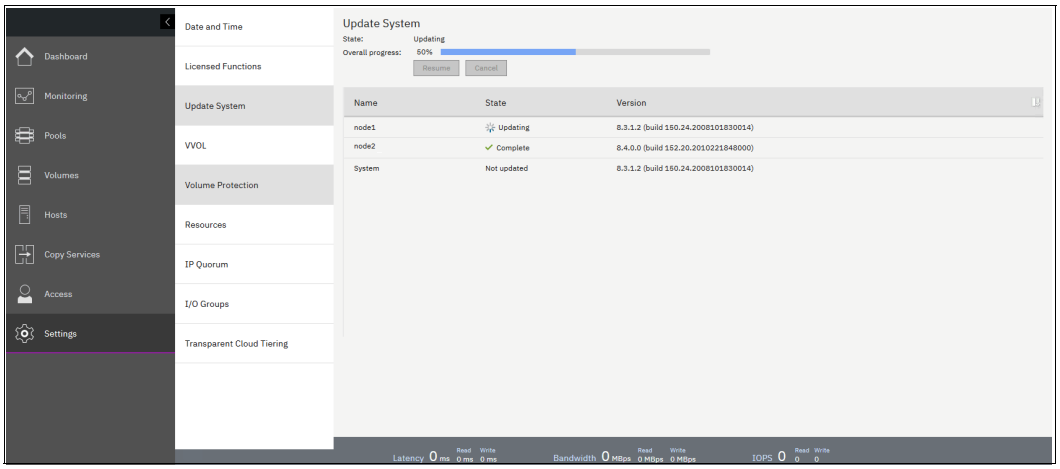


Figure 13-20 New GUI after node failover

After the second node has completed the update, the system is committing to the new code level as shown in Figure 13-21.

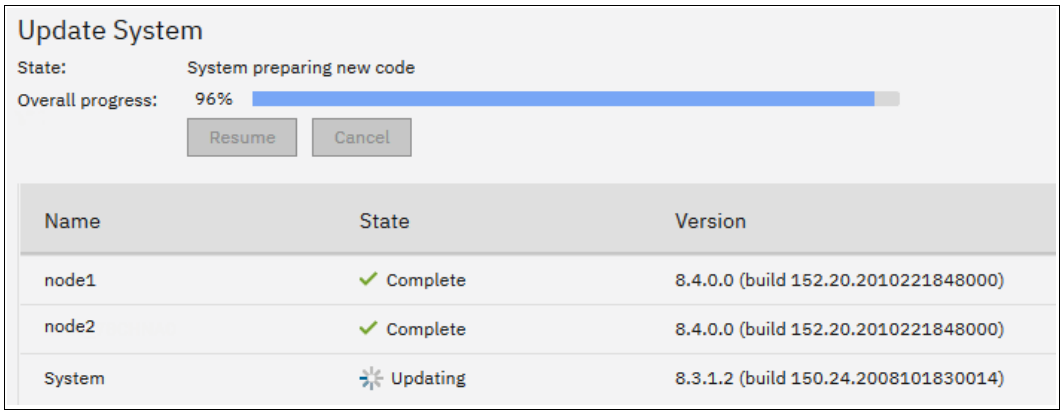


Figure 13-21 Updating the system level

The update process is complete when all nodes and the system are committed. The final status indicates the new level of code that is installed in the system.

13.4.6 Updating IBM Spectrum Virtualize with a hot spare node

IBM Spectrum Virtualize V8.1 introduced a new optional feature of Hot Spare Node. This feature is closely coupled with the N_Port ID Virtualization (NPIV) support (see 7.4, “N_Port ID Virtualization support” on page 447).

The hot spare node feature allows for IBM Spectrum Virtualize software updates to minimize performance impact and eliminates most of the risk concerning node redundancy in an I/O group during the update process. It does so by automatically swapping in a hot spare node after one minute to temporarily replace the node that is updating.

After the original node is updated, the hot spare node becomes a spare again, and is ready for the next node to update. The original node rejoins the cluster.

To use this feature, the spare node must have the same amount of memory and a matching FC port configuration as the other nodes in the cluster. Up to four hot spare nodes can be added to a cluster and must be zoned as part of the IBM SAN Volume Controller cluster.

Figure 13-22 shows how a spare node appears in the GUI.

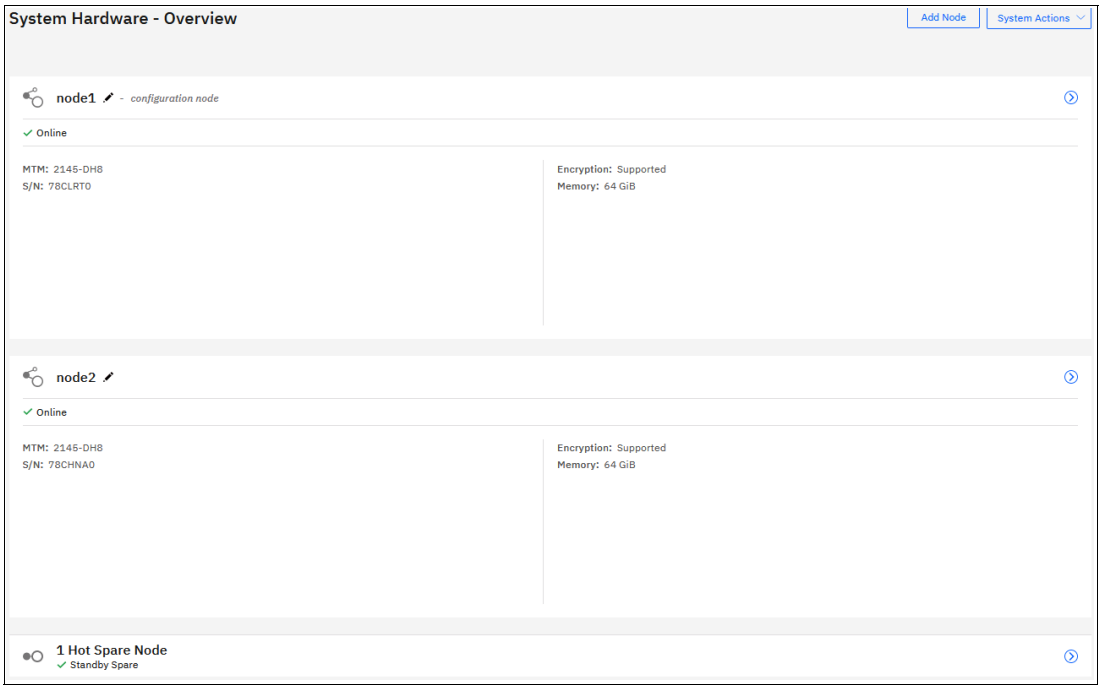


Figure 13-22 Hot Spare Node on the System Overview Pane

13.4.7 Updating the IBM SAN Volume Controller system manually

This example assumes that you have an 8-node cluster of the IBM SAN Volume Controller cluster, as listed in Table 13-9.

Table 13-9 The iogrp cluster

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

Note: It is recommended to perform manual update only if directed by IBM Support.

After uploading the upgrade utility test and software update package to the cluster by running the **pscp** command and running the utility test, complete the following steps:

1. Remove `node 2` from cluster, 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 window.
4. Browse and locate the code upgrade image that you downloaded and saved to your PC.
5. Upload the code and click **Update**.

   When the update is completed, a message caption opens that indicates that the software update completed. The node then restarts, and appears again in the service GUI after approximately 20 - 25 minutes in `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 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:** At this point, 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 IBM Spectrum Virtualize 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 `iogrp 0`.

> **Note:** A partner node becomes the configuration node because the original config 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 now complete.

13.5 Health checker feature

The IBM Spectrum Control health checker feature runs in IBM Cloud. Based on the weekly call home inventory reporting, it proactively creates recommendations. These recommendations are provided at the IBM Call Home Connect Cloud website, which is formerly known as *IBM Call Home Web*.

Note: For more information about Call Home, see the IBM Call Home Connect Cloud website (log in required).

Another feature is the *Critical Fix Notification* function, which enables IBM to warn IBM Spectrum Virtualize users that a critical issue exists in the level of code that they use. 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 points about this function:

- It warns users only about critical fixes; it does not warn about an outdated IBM Spectrum Virtualize software version.
- It works only if the browser also can access the internet. The IBM SAN Volume Controller systems do not need to be connected to the internet.
- The function cannot be disabled. Whenever a warning is displayed, it must be acknowledged (with the option to not warn the user again for that issue).

The decision about what a *critical* fix is subjective and requires judgment, which is exercised by the development team. As a result, clients might still encounter software issues that were not deemed critical. They should continue to review information about new code levels to determine whether update is to be done, even without a critical fix notification.

Important: Inventory notification must be enabled and operational for these features to work. It is preferred practice to enable Call Home and Inventory Reporting on your IBM Spectrum Virtualize systems.

13.6 Troubleshooting and fix procedures

The management GUI of IBM Spectrum Virtualize is a browser-based GUI for configuring and managing all aspects of your system. It provides extensive facilities to assist you in troubleshooting and correcting problems. This section explains how to use efficiently its features to avoid service disruption of your IBM SAN Volume Controller.

Figure 13-23 shows the Monitoring menu for System information, viewing Easy Tier reports, viewing events, or seeing real-time performance statistics, and running background tasks.

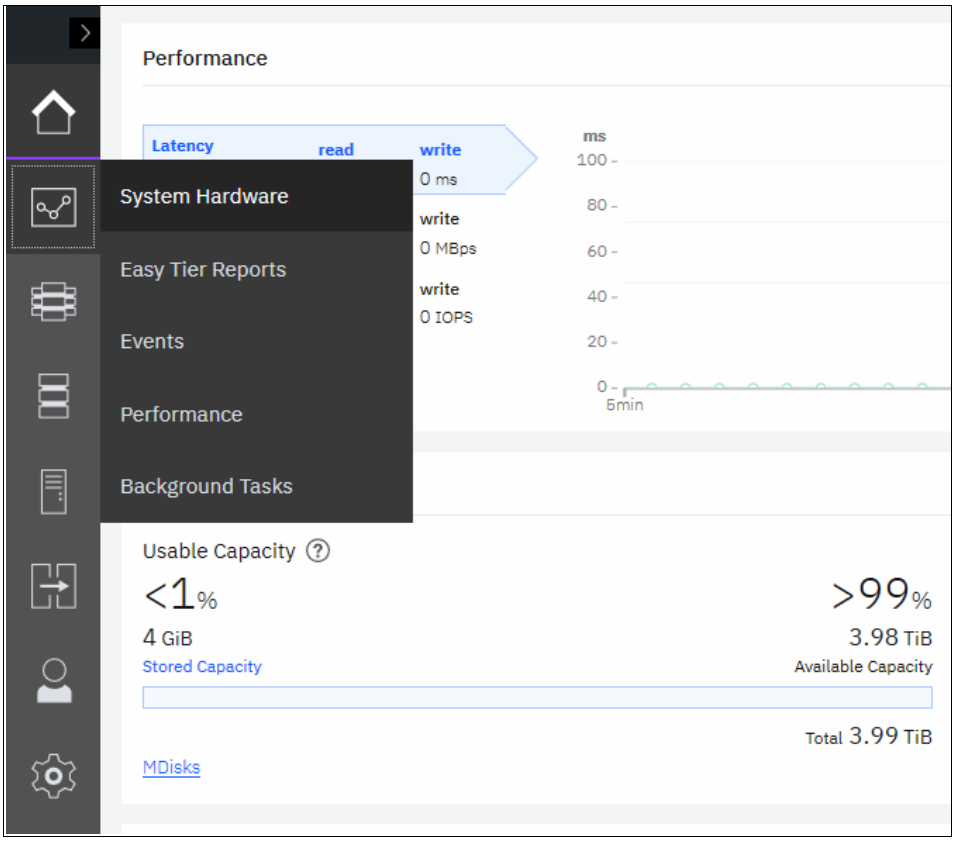


Figure 13-23 Monitoring options

Use the management GUI to manage and service your system. Click **Monitoring** → **Events** to list events that must be addressed, and start the associated Fix Procedures that guide you through the process of correcting problems.

The following presets can be used to select which events are displayed in the Events window (the default view shows Recommended Actions):

- Recommended Actions

  Shows unfixed alerts that require attention. Alerts are listed in priority order and must be resolved sequentially by using the available fix procedures.

- Unfixed Alerts

  Displays only the alerts that are not fixed.

- Unfixed Messages and Alerts

  Displays only the alerts and messages that are not fixed.

- Show All

  Displays all event types, whether they are fixed or unfixed.

Some events require a specific number of occurrences in 25 hours before they are shown 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 often are transient.

Important: The management GUI is the primary tool that is used to operate and service your system. Also, real-time monitoring must be established by using SNMP traps (outbound only), email notifications, or syslog messaging in an automatic manner.

13.6.1 Managing 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, physical storage, and available volumes by completing the following steps:

1. Click **Monitoring** → **Events** to see all problems that exist on the system (see Figure 13-24).

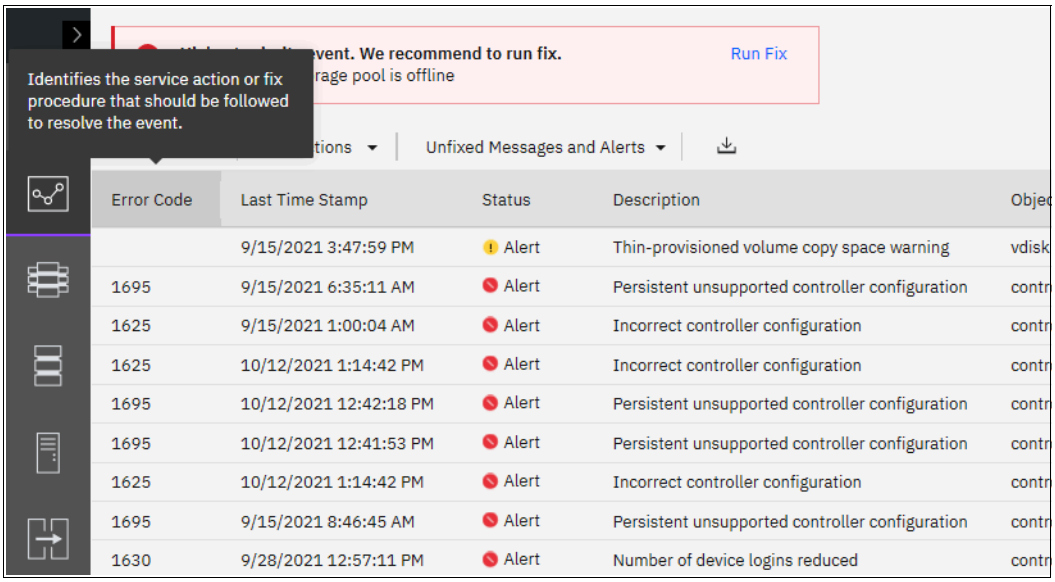


Figure 13-24 Messages in the Event log

2. Select **Recommended Actions** to show the most important unfixed events to be resolved (see Figure 13-25 on page 883). The Recommended Actions tab shows the highest priority maintenance procedure that must be run. Use the troubleshooting wizard so that IBM SAN Volume Controller can determine the correct order of maintenance procedures.

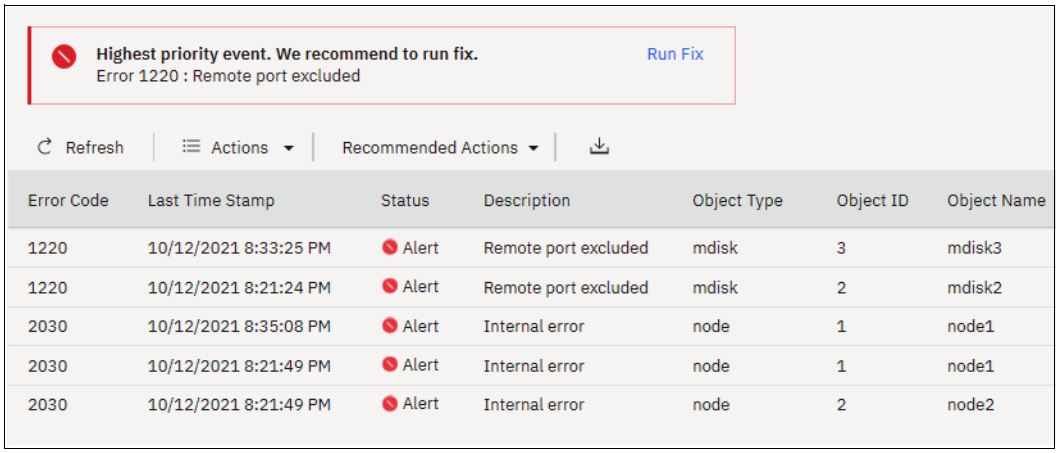


Figure 13-25 Recommended Actions

At any time from any GUI window, you can directly go to this menu by clicking the **Status Alerts** icon at the top of the GUI (see Figure 13-26).

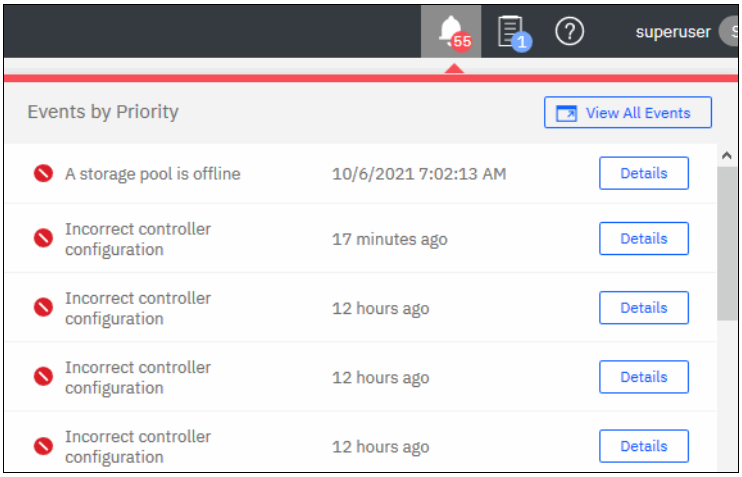


Figure 13-26 Status alerts

13.6.2 Running a fix procedure

If an error code exists for a logged event of type Alert, run a fix procedure to help resolve the problem. Fix procedures analyze the system and provide more information about the problem.

They also suggest actions to take and guide you through the actions that automatically manage the system where necessary while ensuring availability. Finally, they verify that the problem is resolved and reset internal error counters to set a baseline.

The Fix Procedure also checks that any other problem does not result in volume access being lost. For example, if a PSU 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 the Run Fix function, which resolves the most serious issues first. Often, lower-priority alerts are corrected automatically because they were the result of a more serious issue. In some documentation or in communication with IBM Support, the Fix Procedure also might be referred to as Directed Maintenance Procedure (DMP).

In this example, we show how to deal with the error 1220 that is related to remote FC port that was excluded, likely because of errors along the link between this specific back-end storage controller and the SAN switch to which it is attached.

Complete the following steps:

1. From the GUI menu on the left, select **Monitoring** → **Events**, and list only the Recommended Actions. In the **Actions** menu (see Figure 13-27) click **Run Fix Procedure**.

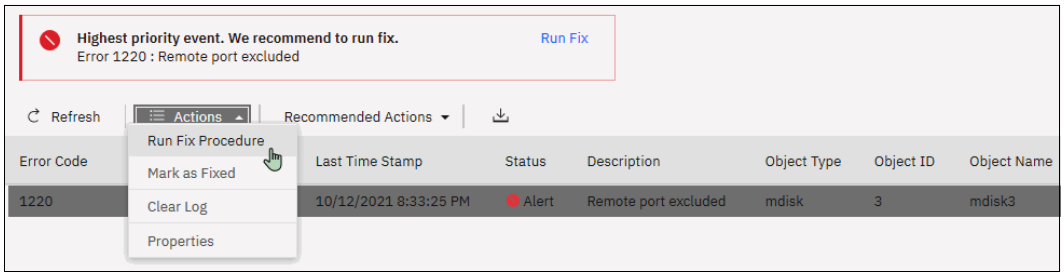


Figure 13-27 Initiate Run Fix procedure from the management GUI

2. The next window (see Figure 13-28) shows you details about the affected remote FC port. Click **Next** to proceed.

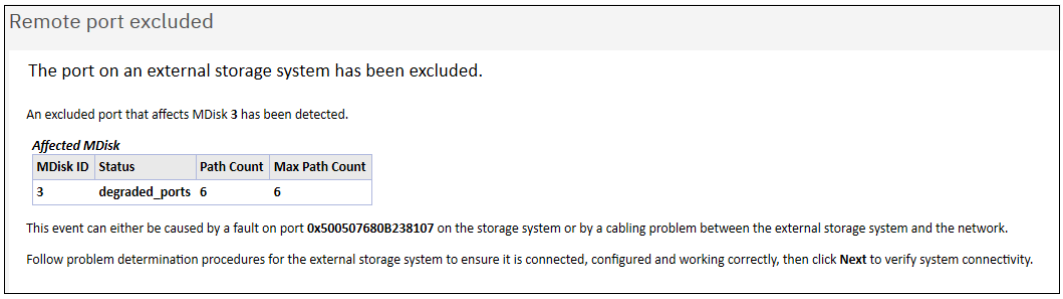


Figure 13-28 Error details in the Fix Procedure

3. In the next window (see Figure 13-29), the Fix Procedure proceeds with checking the configuration and the configuration action that is conducted automatically: Adding the MDisk to the configuration and detecting MDisks. The CLI commands **`includemdisk`** and **`detectmdisk`** are being issued in the background and the event is marked as fixed. This status also is visible in the system's audit log.

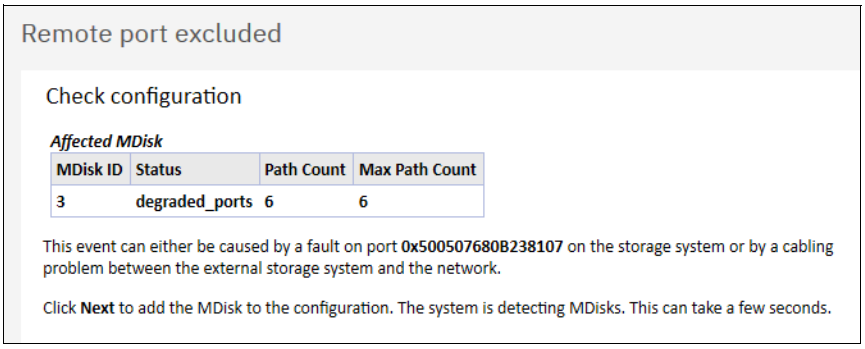


Figure 13-29 Fix Procedure continued

4. The next window (see Figure 13-30) shows a confirmation that the event was marked as fixed. **Close** to exit.



Figure 13-30 Correctly completed fix procedure

5. The event is marked as fixed, and you can safely finish the fix procedure. Click **Close** and the event is removed from the list of unfixed events in the Recommended Actions view (see Figure 13-31).

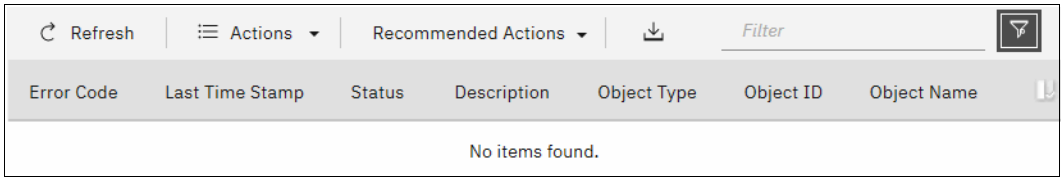


Figure 13-31 Pane showing that no excluded local logins exist

13.6.3 Resolving alerts in a timely manner

To minimize potential impact to your environment, always perform the recommended actions as quickly as possible after a problem is reported. Your system is resilient to most of the 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 volume data that is unavailable. If several unfixed alerts exist, fixing any one alert might become more difficult because of the effects of the others.

13.6.4 Event log details

Events that might require an interaction feature a four-digit Error Code in the range of 1001 - 3081, which is associated with a six-digit Event ID. The lower the error code, the higher the priority.

Multiple views of the events and recommended actions are available. The GUI works much like Windows menu, so the event log grid is manipulated by using the column headings row (see Figure 13-32).

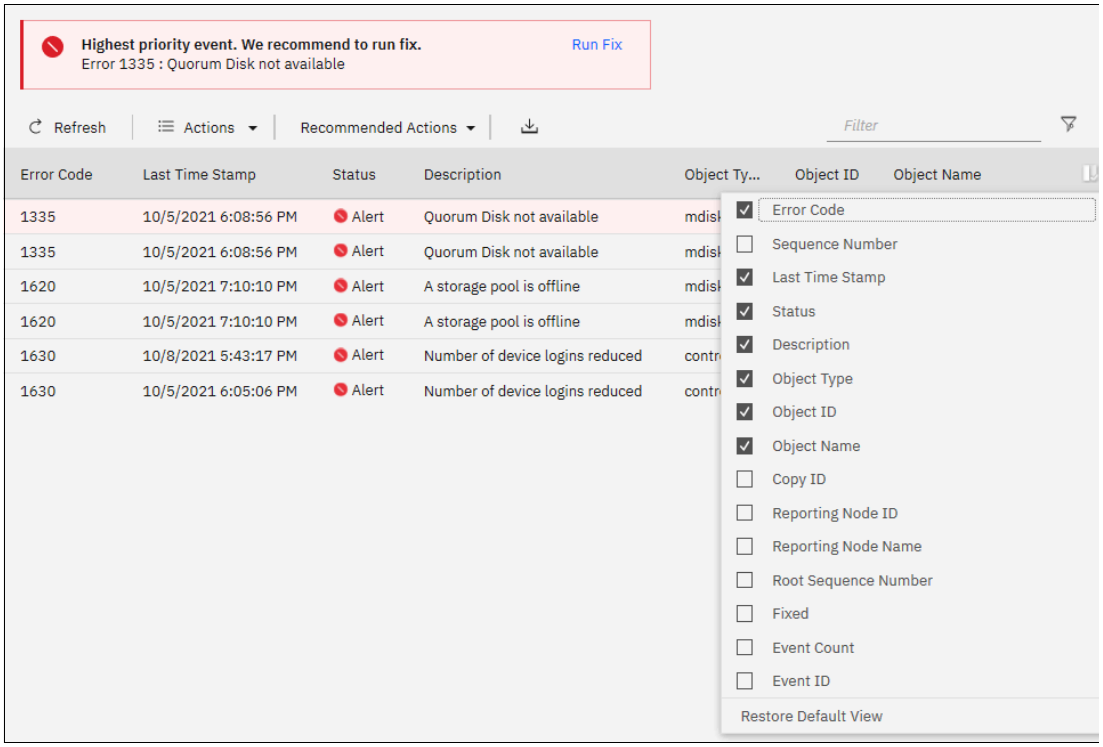


Figure 13-32 Grid options of the event log

As with most of the other tables in the Spectrum Virtualize GUI, you can configure the grid view according to your needs. Click the column icon at the right end of the table heading to open the menu for the selection of columns to be displayed, or right-click anywhere on the heading row.

Select or remove columns as needed. You can extend or shrink the width of the column to suit them to your needs. The sequence of the displayed columns can be changed easily by drag. Click and hold the column to be moved to another position and drag it to the wanted position in the heading row.

The example in Figure 13-33 shows the movement of the first column Error Code to its new position in between the columns Last time Stamp and Description.

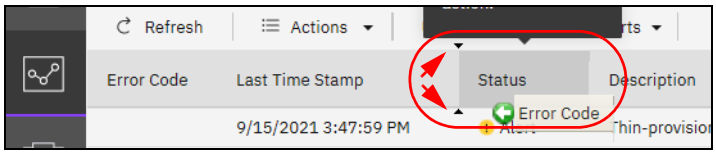


Figure 13-33 Moving first column

Every field of the textual event log is available as a column in the event log grid. Several fields are useful; for example, 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 feature the sequence number, event count, and the fixed state. The use of the Restore Default View sets the grid back to the defaults.

You might want to see more information about specific events to identify what caused an error to be logged. Some event details are not shown in the main grid. To access the event properties and more information, double-click the specific event anywhere in its row.

The properties window opens (see Figure 13-34 on page 888) with all the relevant sense data. This data includes the first and last time of an event occurrence, worldwide port name (WWPN), worldwide node name (WWNN), and whether an event was fixed automatically or by manual intervention as running a Fix Procedure.

In this example (the called out back-end storage controller with its WWNN), WWPN can assist to isolate the root of this problem; for example, by double-checking the SAN zoning configuration or its volume-to-host-mappings.

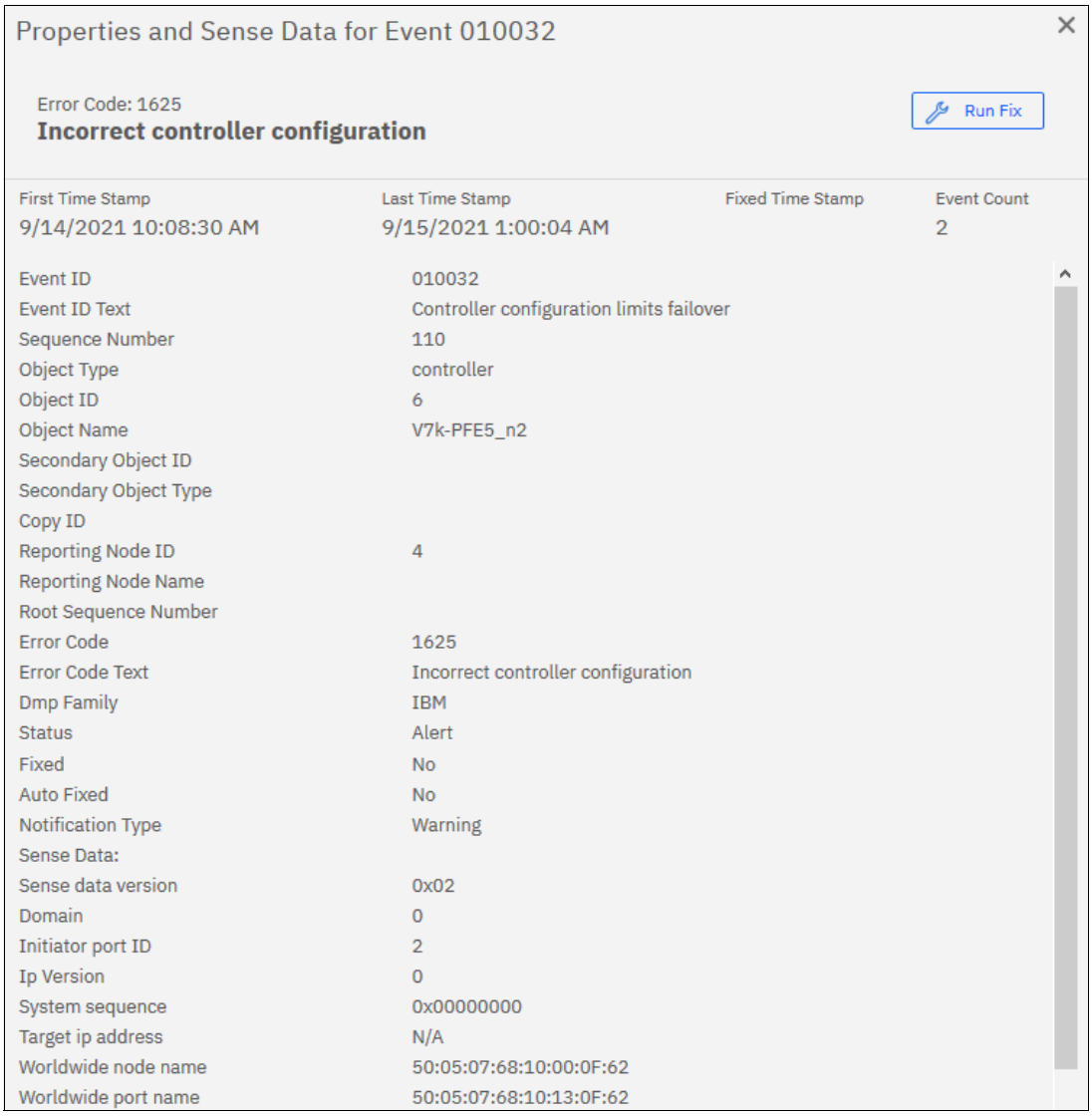


Figure 13-34 Event sense data and properties

For a general overview about event log management and more I information about troubleshooting options, see this IBM Documentation web page.

Note: Although the recommended practice is to use the Fix Procedure to address errors, you also can manually mark an event as fixed. For example, error code 2030 (Internal Error) is logged for a node that has warm-started *after* it returned to operation.

This event alone does not need any other action and can be safely marked as fixed (manually marking an event as fixed does not do any harm). If an error condition persists, the event is logged again instantly.

13.7 Monitoring and Event Notification

The Spectrum Virtualize software offers various means of sending event notifications to assist you monitoring of the system's health state. Therefore, you can respond to any reported issues in a timely manner. Event Notifications can be sent to IBM by way of Call Home, to individuals and to monitoring systems.

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, restrict notifications to only events that require action. The following event notification mechanisms are available:

Email Call Home An event notification can be sent to one or more individual email users and to the IBM Call Home address (IBM Call Home is enabled by default).

Cloud Call Home 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 under warranty or covered by a hardware maintenance agreement (alias HWMA), configure your IBM SAN Volume Controller 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 ticket and, if suitable, 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 current. Personnel changes can cause delays in IBM making contact.

The preferred method is Cloud Call Home. For more information about the initial setup of Cloud Call Home, see Chapter 3, "Initial configuration" on page 103.

Cloud Call Home requires an outbound https connection from the clustered system to the IBM Call Home environment. This connection is not possible or wanted in every customer environment. However, from Spectrum Virtualize V8.4.2 on, a system-wide proxy server can be configured so that the management network does not need to directly contact the internet.

Note: If the customer does not want to open their firewall, Cloud Call Home does not work. The customer can disable Cloud Call Home and Email Call Home is used instead. Also, if problems with the Cloud Call Home connectivity occur, Email Call Home is used as fallback method.

13.7.1 Call Home function and email notification

The Spectrum Virtualize Email Call Home function sends an email to the IBM Support center. Complete the following steps to configure Call Home and email notifications:

1. Prepare your contact information that you want to use for the Call Home function and verify its currency and accuracy. From the GUI menu, click **Settings** → **Support** (see Figure 13-35).

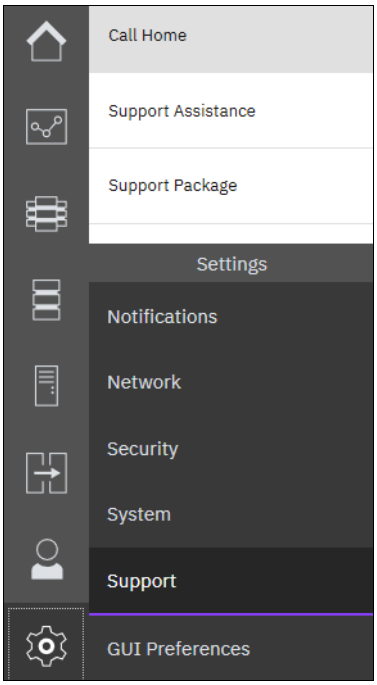


Figure 13-35 Support menu

2. Click **Call Home** and then, click **Enable Notifications** (see Figure 13-36).

   For the correct functions of email notifications, ask your network administrator if Simple Mail Transfer Protocol (SMTP) is enabled on the management network and granted to through the firewalls. Also, ensure that the destination domain `@de.ibm.com` is not blacklisted.

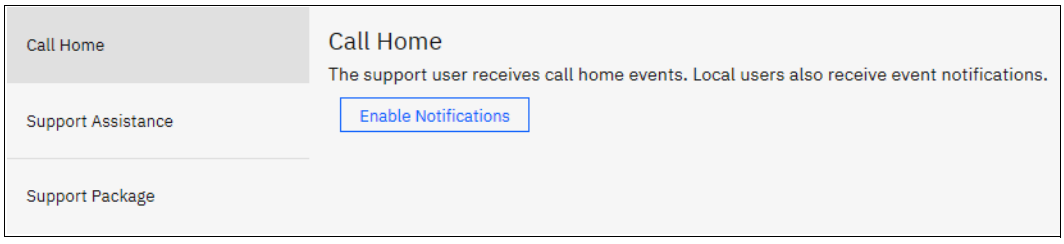


Figure 13-36 Configuration of Call Home function

Figure 13-37 shows the wizard that is used to configure the Call Home settings.

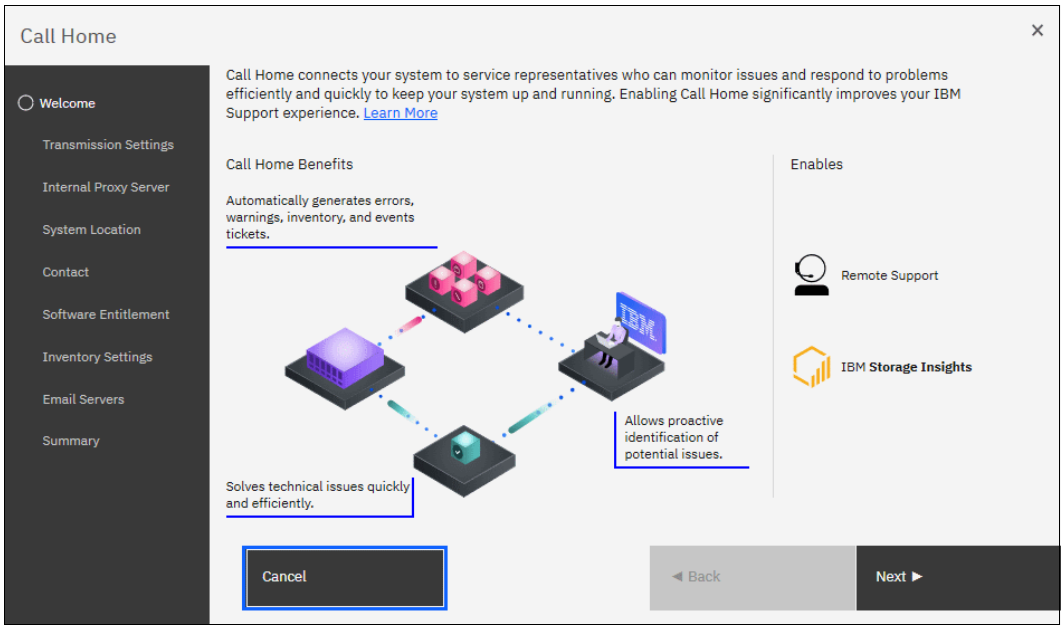


Figure 13-37 Call Home configuration wizard

3. After clicking **Next** in the welcome window, select the transmission type that you want to use: Cloud Services, Email Services or both (see Figure 13-38).

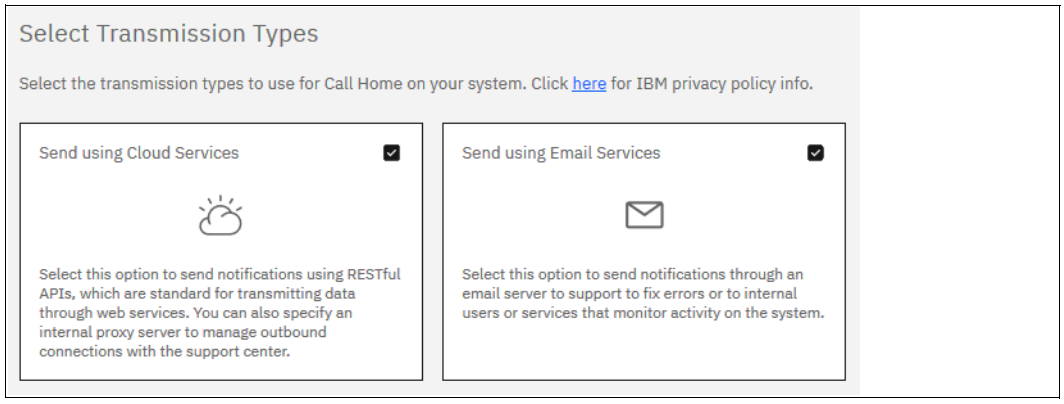


Figure 13-38 Call Home Transmission Types

4. If required, you can configure a Proxy Server next. In the Proxy Details window (see Figure 13-39) select the proxy authentication methods, if any. Click **Apply and Next**. A connection test is conducted automatically.

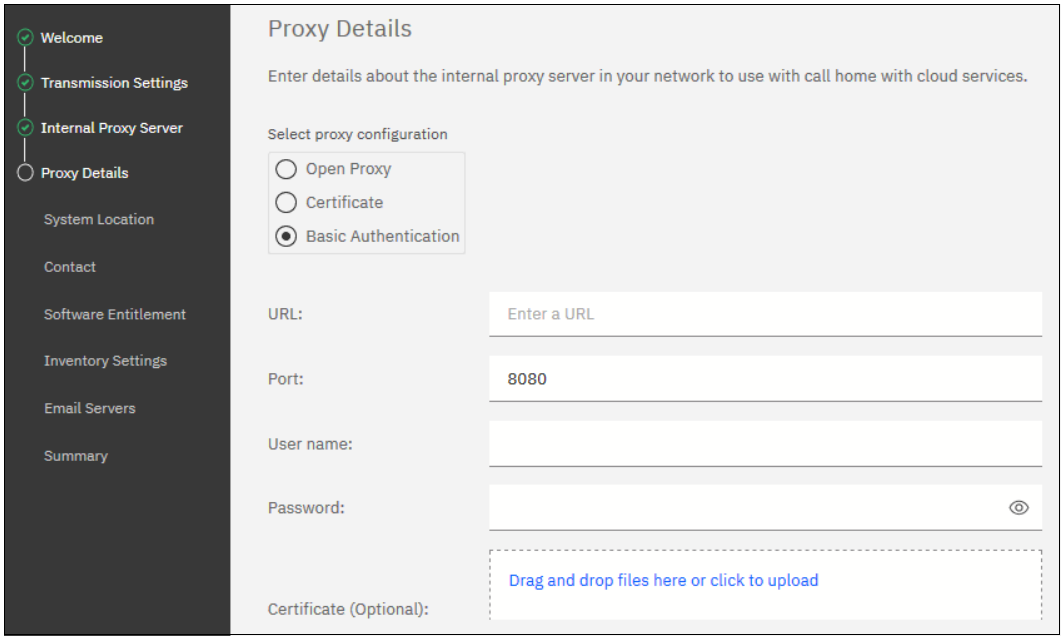


Figure 13-39 Call Home configuration - Proxy Details

5. The System Location (see Figure 13-40) and Contact (see Figure 13-41 on page 893) details that are configured in the two next steps must be correct and as precise as possible because they are used to contact the client when a Call Home case was opened with IBM Support.

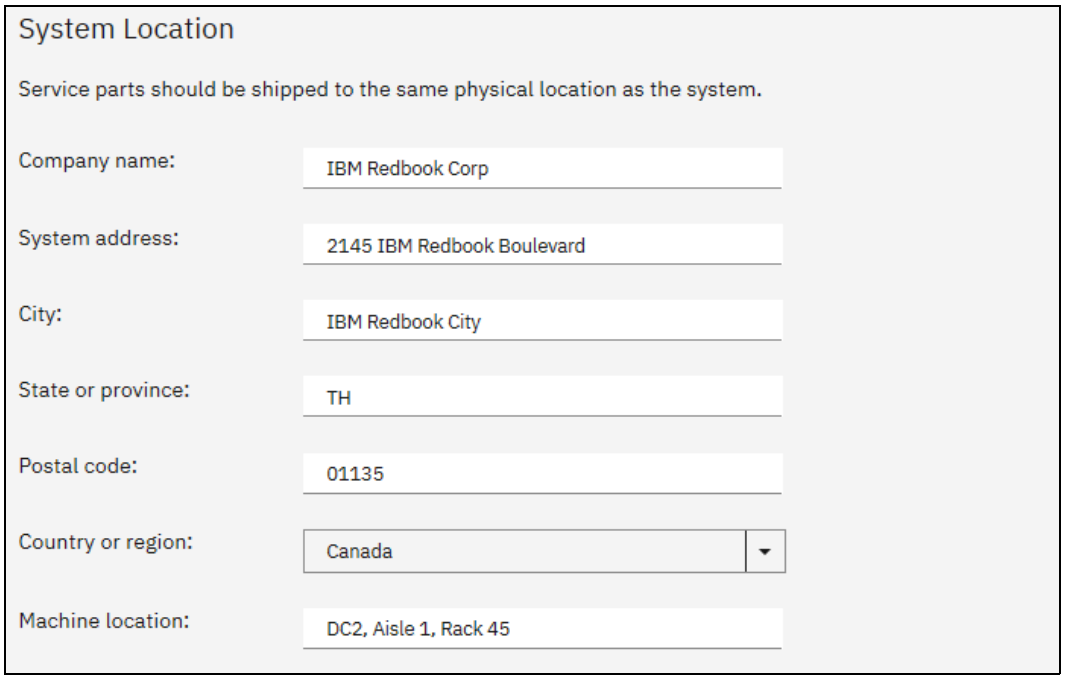


Figure 13-40 Call Home configuration - Location

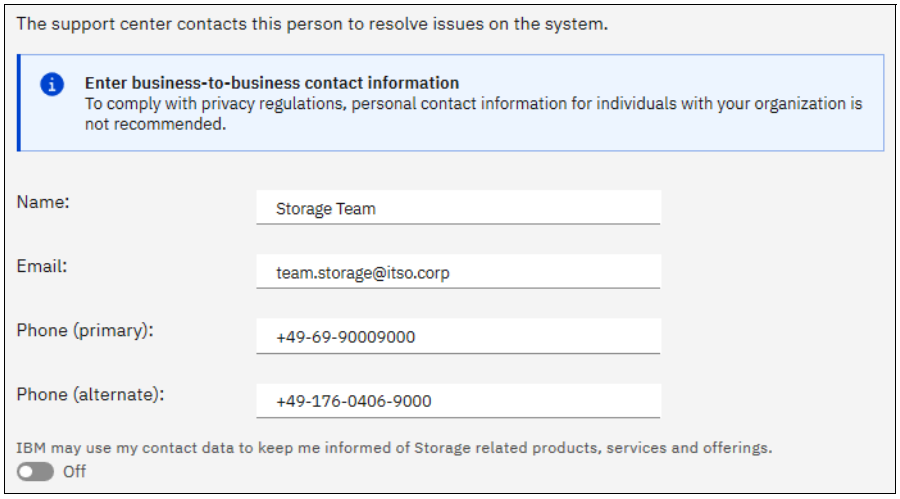


Figure 13-41 Call Home configuration - Contact information

6. If available, enter your IBM customer ID, which is a 7 - 10-digit numeric code in the Software Entitlement view, as shown in Figure 13-42. Providing the correct data helps to expedite processing of a support case and avoids the need for manual intervention.

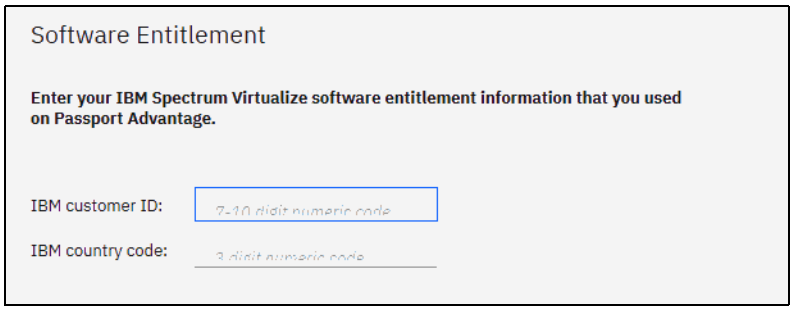


Figure 13-42 Call Home configuration - Software Entitlement

7. In the next window, enable Inventory Reporting and Configuration Reporting, as shown in Figure 13-43.

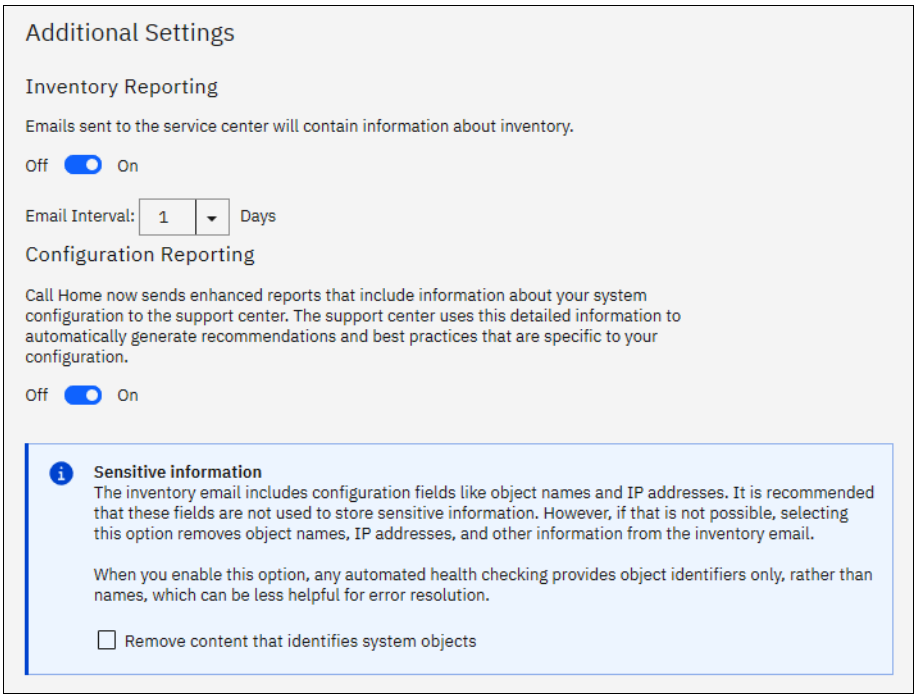


Figure 13-43 Call Home configuration - Inventory Reporting

Note: If data privacy is a concern, you can censor the system data that is sent to IBM by selecting the **Remove content that identifies system objects** option. Doing so removes some information, such as object names and IP addresses.

For more information, see this IBM Support web page. You can download a guide that explains how to manually check what data is transmitted to IBM.

8. Configure the IP address or the Fully qualified domain name (FQDN) of your company SMTP server if a DNS server is configured, as shown in Figure 13-44. When the correct SMTP server is configured, you can test the connectivity by pinging its IP address. You can configure more SMTP servers by clicking the Plus sign (**+**) at the end of the entry line. When you are done, click **Apply and Next**.

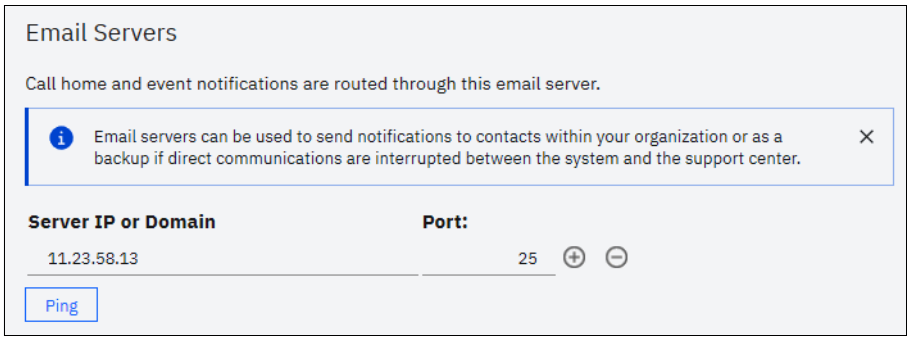


Figure 13-44 Call Home configuration: SMTP server

9. A summary window opens (see Figure 13-45). Verify the information, and then, click **Finish**.

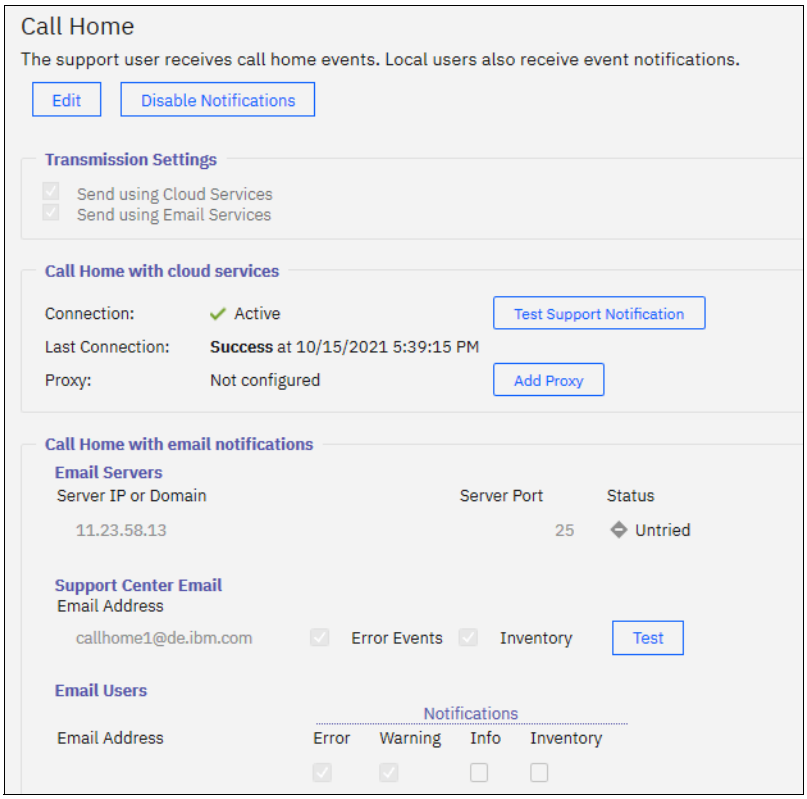


Figure 13-45 Call Home settings, email recipients, and alert types

You are then returned to the Call Home window where you can verify email addresses of IBM Support (`callhome1@de.ibm.com` for systems in the Americas and `callhome0@de.ibm.com` for all other regions). Optionally, you can add up to 12 addresses of individual email recipients and select the level of information that they receive.

Note: The default support email address `callhome0@de.ibm.com` and `callhome0@de.ibm.com` are predefined to receive Error Events and Inventory. Do not change these settings.

You can modify or add individual email 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 that describes your system hardware and critical configuration information. By default, the inventory email is sent weekly, which allows an IBM Cloud service to analyze and inform you whether the hardware or software that you use requires an update because of any known issue.

For more information, see 13.5, “Health checker feature” on page 880.

10. After completing the configuration wizard, test the email function. To do so, click **Edit**, as shown in Figure 13-46. In the same window, you can configure more email recipients or alter contact and location details as needed.

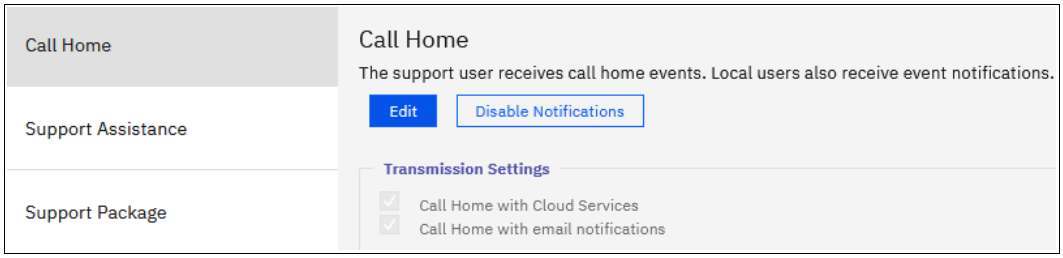


Figure 13-46 Entering Edit mode

We strongly suggest that you keep the sending inventory option enabled to IBM Support.

However, it might not be of interest to local users, although inventory content can serve as a basis for inventory and asset management.

11. In Edit mode, you can change any of the configured settings. After you are finished editing these parameters, adding recipients, or testing the connection, save the configuration to make the changes take effect (see Figure 13-47).

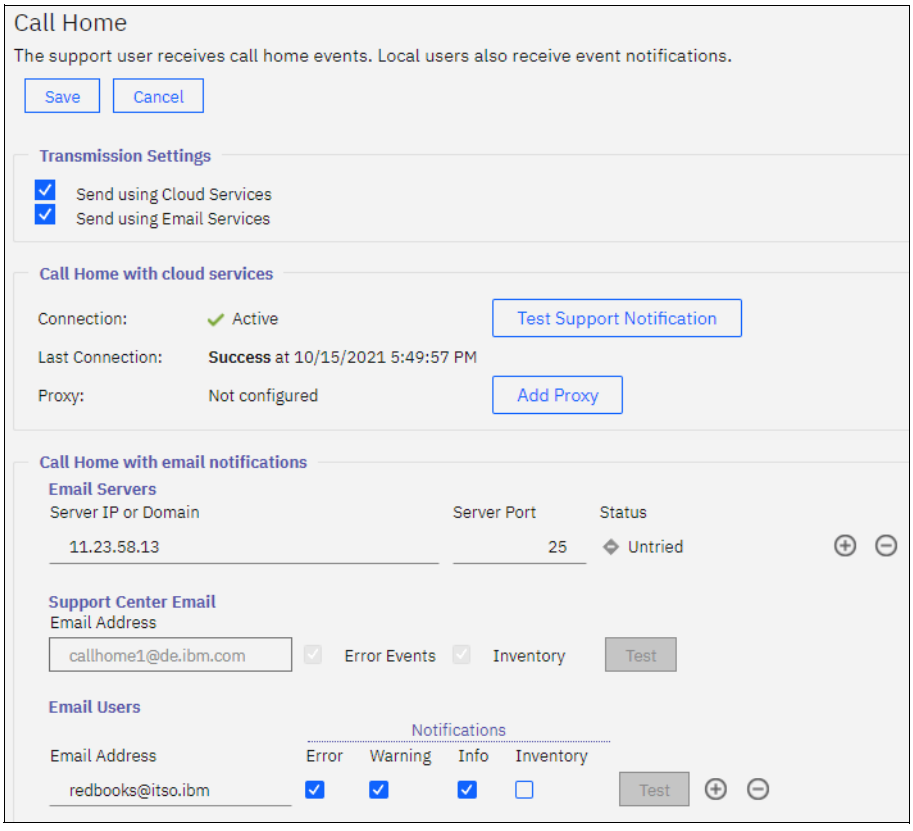


Figure 13-47 Saving modified configuration

Note: The **Test** button appears for new email users after first saving and then editing again.

13.7.2 Disabling and enabling notifications

At any time, you can temporarily or permanently disable email notifications, as shown in Figure 13-48. This best practice is useful when performing activities in your environment that might generate errors on your IBM Spectrum Virtualize cluster, 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 **`stopemail`** and **`startemail`** commands.

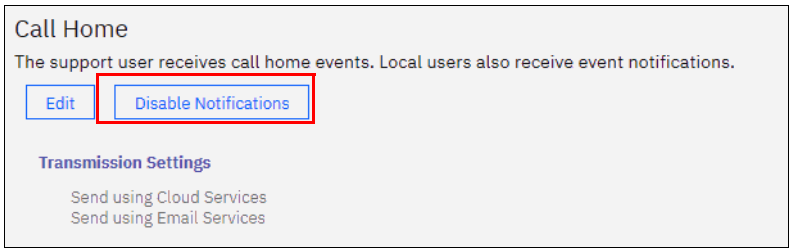


Figure 13-48 Disabling or enabling email notifications

As an alternative method, you can place any I/O group into the *maintenance mode*. By doing so, event notifications and thereby Call Home, are suspended for 30 minutes or until the maintenance mode is ended manually.

A Spectrum Virtualize system in Maintenance mode indicates that service actions are being run on the system. For example, it is used in the background whenever a Fix Procedure is being run. You can put a system into maintenance mode manually by using the CLI only (see Example 13-9).

Example 13-9

```
IBM_2145:SVC_ITSO:superuser>chiogrp -maintenance yes 0
```

Note: CLI command **`chiogrp`** (change I/O group) is used to modify an I/O group's properties (for example, its name) as bitmap memory allocation, NPIV mode, and others. It requires an I/O group name or ID as last argument.

Setting one I/O group into maintenance mode does affect all I/O groups in the system, regardless of the I/O group ID or supplied name. It is sufficient to use I/O group ID `0`, or whatever I/O group IDs are used in your system.

13.7.3 Remote Support Assistance

Introduced with Version 8.1, *Remote Support Assistance allo*ws IBM Support teams to connect remotely to the Spectrum Virtualize system through a secure tunnel for Remote Access and Remote Code Load. It can be used to perform analysis, log collection, and software updates.

The tunnel can be enabled ad hoc by the customer, or the customer can enable a permanent connection if wanted. In previous Spectrum Virtualize product versions, you might find the Support Assistance feature also referred to as *Secure Remote Access* (SRA).

Only authorized IBM Support agents can use RSA to access your system for the previously mentioned purposes by using a Spectrum Virtualize shell so that CLI commands can be issued.

Note: Customers who purchased Enterprise Class Support (ECS) are entitled to IBM Support to use RSA to connect to your system and diagnose problems quickly. However, IBM Support also might choose to use this feature on non-ECS systems at their discretion. Therefore, we advise that RSA is configured on all of your Spectrum Virtualize systems.

If you are enabling RSA, ensure that the following prerequisites are met:

- Cloud Call Home or a valid email server is configured. Cloud Call Home is used as the primary method to transfer the access token when you start a session, with email as backup.
- A valid Service IP address is configured on each node of the IBM Spectrum Virtualize cluster.
- Your IBM SAN Volume Controller can connect to the internet.

  If your system does not have direct internet connectivity, a system-wide web proxy can be configured from V8.4.2 on, which is the recommended practice. A configured Remote Support Proxy can be used, but it is considered deprecated.
- If your IBM SAN Volume Controller is behind a firewall, request your network administrator to configure a rule to allow outbound connections on TCP port 22 to the following IP addresses:
  - 129.33.206.139
  - 204.146.30.139
- Uploading support packages to IBM and downloading software from IBM Fix Central requires direct connections to the internet. A DNS server must be defined on your IBM SAN Volume Controller for these functions to work.
- Ensure that support packages are uploaded correctly by configuring the firewall to allow connections to the following host IP addresses on TCP port 443:
  - esupport.ibm.com
  - 129.42.56.189
  - 129.42.54.189
  - 129.42.60.189
- Ensure that software can be downloaded by configuring a firewall rule to allow outbound connections to the following hosts respectively IP addresses on TCP port 22:
  - delivery04.dhe.ibm.com
  - 170.225.15.104
  - 170.225.15.105
  - 170.225.15.107
  - 129.35.224.104
  - 129.35.224.105
  - 129.35.224.107

Figure 13-49 shows the Support Assistance view that is found in **Settings** → **Support** → **Support Assistance**.

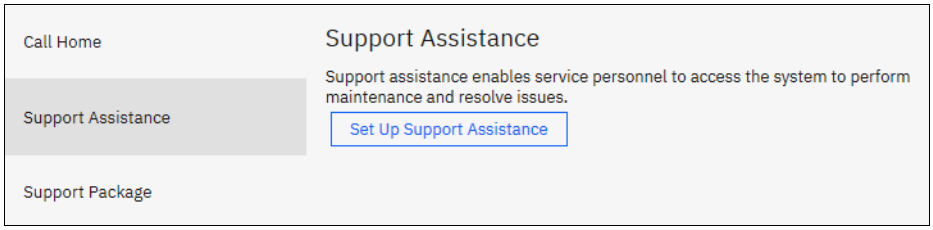


Figure 13-49 Support Assistance Setup

Choosing to set up support assistance opens a wizard to guide you through the configuration. Complete the following steps:

1. Figure 13-50 shows the initial wizard window. Select **I want support personnel to work on-site only** or enable remote assistance by selecting **I want support personnel to access my system both on-site and remotely**. Click **Next**.

Note: Selecting **I want support personnel to work on-site only** does not imply any entitlement for on-site support activities. The most hardware maintenance contracts are for customer-replaceable unit (CRU) support, in which IBM diagnoses your problem and sends a replacement component for you to replace, if required. If you prefer to have IBM perform replacement tasks for you, contact your local sales person about upgrading your maintenance contract.

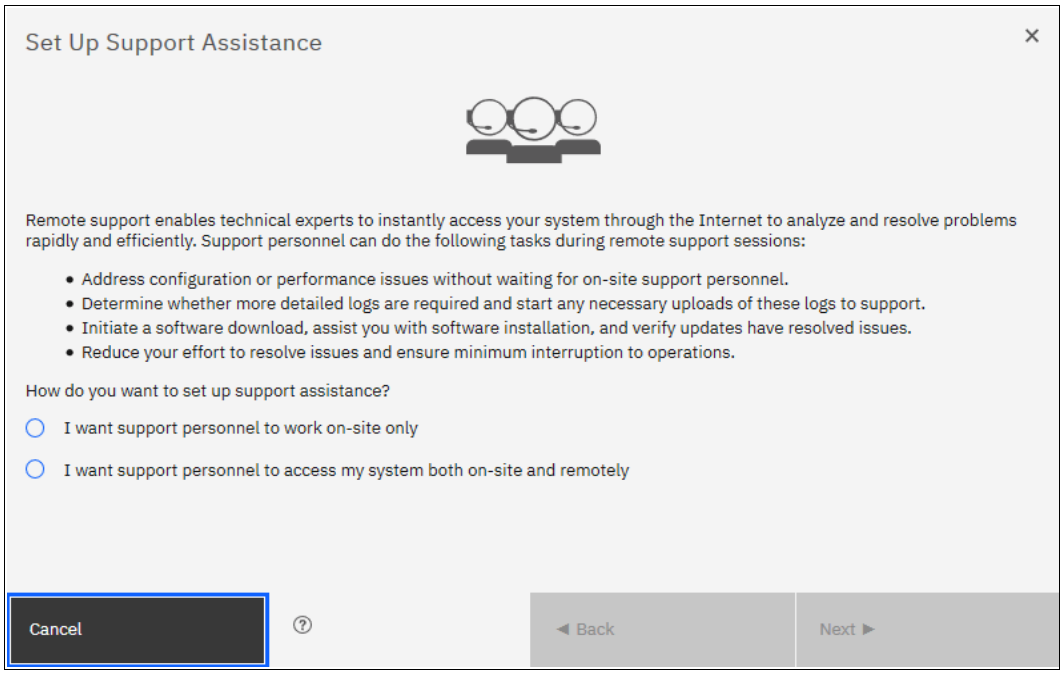


Figure 13-50 Remote Support wizard enable or disable

2. After selecting the second option, the next window, which is shown in Figure 13-51, lists the IBM Support center's IP addresses and Secure Shell (SSH) port that must be open in your firewall. You also can define a Remote Support Assistance Proxy if you have multiple FlashSystems or IBM SAN Volume Controller systems in the data center so that the firewall configuration is required for only the proxy server rather than every storage system.

   Because we do not use a proxy server, we leave the field blank and click **Next**.

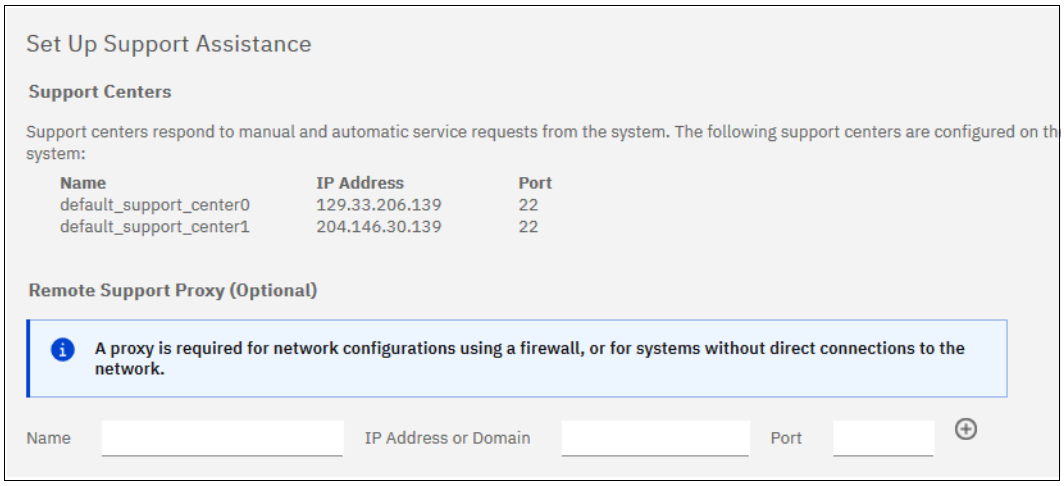


Figure 13-51 Support Centers and Proxy

3. The next window opens and prompts you to open a tunnel to IBM permanently so that IBM can connect to your IBM SAN Volume Controller. Your options are **At Any Time** or **On Permission Only**, as shown in Figure 13-52. On Permission Only requires a storage administrator to log on to the GUI and enable the tunnel when required. Click **Finish**.

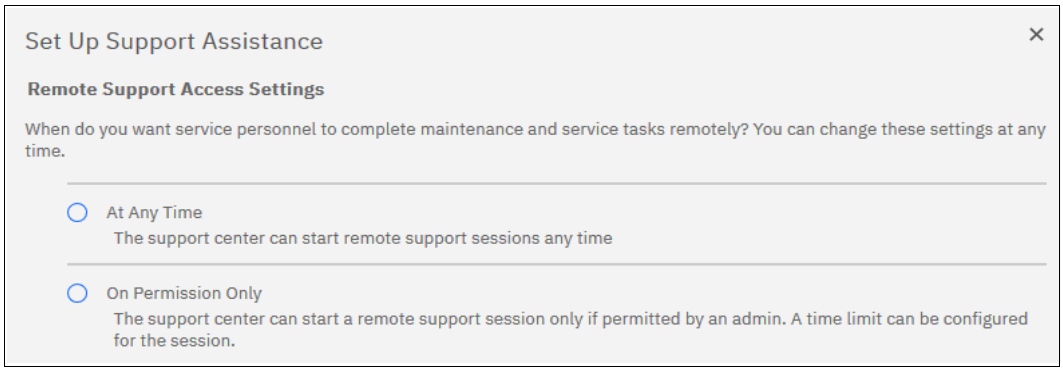


Figure 13-52 Support Assistance setup - Remote Support Access Settings

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.

Tip: More information about Support Assistance and setup instructions, see this IBM Documentation web page and this IBM Support web page.

13.7.4 SNMP configuration

SNMP is a standard protocol for managing networks and exchanging messages. An SNMP message alias SNMP Trap is being sent out to a configured SNMP management station upon every event logged in the system event log.

IBM Spectrum Virtualize supports SNMP protocol versions v2 and v3; SNMPv3 adds encryption and improved authentication.

Note: Spectrum Virtualize solely does support outbound SNMP by sending SNMP traps. Inbound SNMP requests (for example, **`snmpget`**, **`snmpwalk`** or **`snmpset`**) are *not* supported. Alternative methods are available for gathering system information or changing configuration settings; for example CLI, REST API, and CIM.

An SNMP management station in your environment can be defined as SNMP server object in your Spectrum Virtualize system; therefore, it can receive notifications upon informational, error, warning, or a combination of these events (see Figure 13-53).

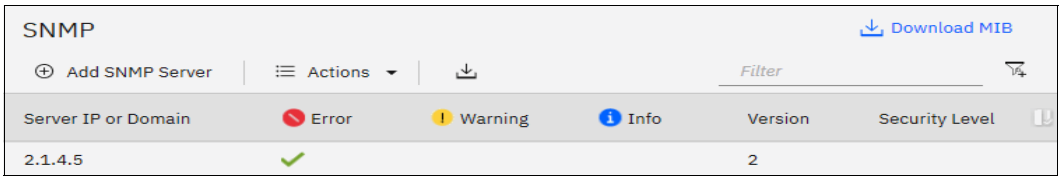


Figure 13-53 SNMP configuration

In the same window, you can also download the SNMP MIB file from the system to make it known to the SNMP server by clicking **Download MIB** in the top right corner. To configure an SNMP server, the following information must be supplied:

- IP address or host name

  The IP address or FQDN (fully qualified domain name) of the SNMP server.

- Server Port

  The port number for the SNMP server. The remote port number must be a value of 1 - 65535, where the default port for SNMP is 162.

- Community

  The SNMP community is the name of the group to which monitored devices and SNMP management stations belong. Typically, the default of `public` is used. Edit it to match the community name that is used in your environment.

- Events

  Consider the following points about the type of events being logged that trigger a notification to be sent as SNMP trap:

  – Error: An SNMP trap is sent about problems, such as hardware failures, that must be resolved immediately.

  – Warning: An SNMP trap is sent about problems and unexpected conditions. Investigate the cause immediately to determine any corrective action; for example, a space-efficient volume is running out of space.

  Important: Select **Recommended Actions** to run the fix procedures upon receiving an Error or Warning notification.

– Info: An SNMP trap is sent upon expected events. No action is required for these events.

To configure an SNMP server, click **Add SNMP Server** and complete the Add SNMP Server window, as shown in Figure 13-54. To remove an SNMP server, click the line with the server you want to remove, and click **Actions** → **Remove**.

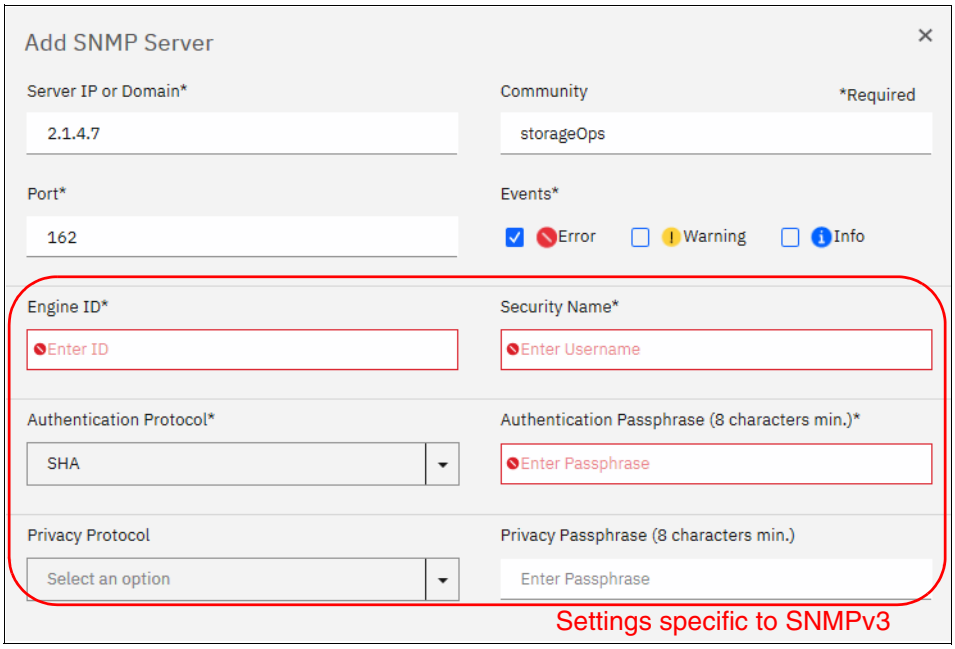


Figure 13-54 Add SNMP Server

Note: The following properties are specific to environments that use SNMP v3 protocol:

- Engine ID: Indicates the unique identifier (UID) in hexadecimal that identifies the SNMP server. It must be provided by the SNMP server administrator.
- 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.

The following CLI commands are relevant to the SNMP server configuration:

- `mksnmpserver`
- `lssnmpserver`
- `chsnmpserver`
- `rmsnmpserver`

Tip: You might want to verify that SNMP traps can be received by your SNMP management station. Although technically possible, we strongly discourage testing event creation and SNMP trap sending by deliberately causing hardware error by unplugging cables, drives, or similar potentially disruptive activities.

Instead, you can issue a nonharming configuration change; for example, creating a not existing email (SMTP) server and test sending an email. This test should generate an event log entry of type alert with an error code `2601/009153 - The system has failed to send a required email.`

13.7.5 Syslog notifications

The syslog protocol is a standard protocol for forwarding log messages from various devices as hosts, switches, and storage devices to a receiver (syslog server) over an IP network. The IP network can be Internet Protocol Version 4 (IPv4) or Internet Protocol Version 6 (IPv6). The Spectrum Virtualize system can send syslog messages that notify personnel about events that are logged in the Event log.

You can configure a syslog server object that receives log messages from various systems and store them in a central repository by entering the following information in the Syslog Servers view that is shown in Figure 13-55. You can reach it in the GUI by clicking **Settings** → **Notifications** → **Syslog**:

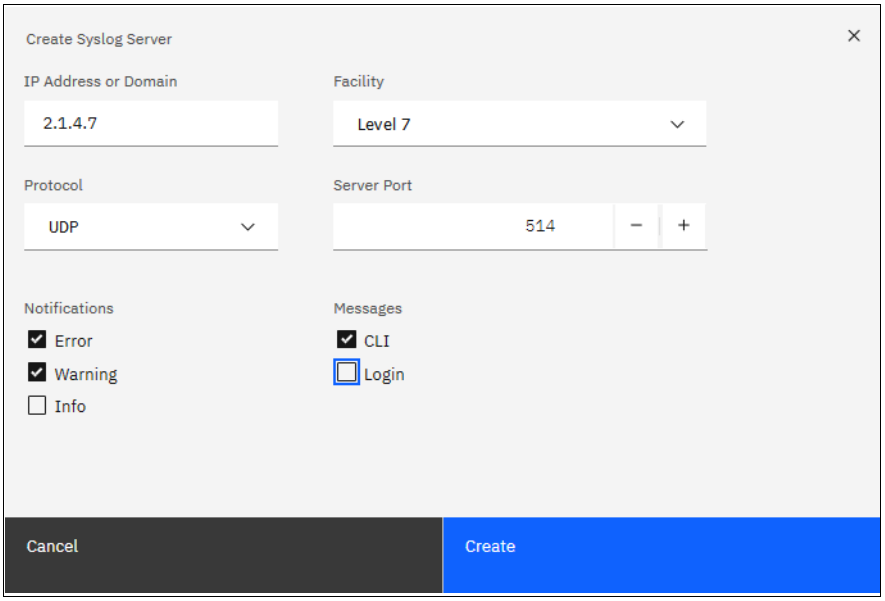


Figure 13-55 Syslog configuration

- IP Address or Domain

  The IP address or FQDN (fully qualified domain name) of the syslog server.

- Facility

  The facility determines the format and the level of details for the syslog messages. For example, facility levels 0 - 3 send syslog messages in concise format that contain only the basic data of a logged event. Facility levels 4 - 7 send the messages in full format that contain all details of an event, including the sense data. Adjust this information to your needs.

- Protocol

  The protocol to be used, UDP or TCP. It must match the setting of the syslog server.

- Server Port

  The port to communicate with the syslog server; that is, the port (UDP, TCP, or both) om which the syslog server is listening.

- Notifications

  The kind of logged events being forwarded as syslog message:

  – Error: Events about problems, such as hardware failures, that must be resolved immediately.
  – Warning: Events about problems and unexpected conditions.
  – Info: Events about expected events that do not require an intervention.

- Messages

  This feature was introduced with Spectrum Virtualize V8.3.1 as Enhanced Audit Log:

  – CLI

    This option enables forwarding all configuration commands that issued through the CLI or GUI, which also are locally logged to the audit log.

  – Login

    This option sends successful and failed authentication attempts to the specified syslog servers.

The syslog messages are sent in concise message format or expanded message format. depending on the Facility level that is selected.

Example 13-10 shows a compact format syslog message (line breaks are included for better readability).

Example 13-10 Compact syslog message example

```
Oct 11 10:39:36 9.42.162.160 ITSO_SVC local0 err IBM2145[18163] 2145 3.0 Record
Type = WARNING
# Error ID = 9183 : Unable to connect to LDAP server
# Error Code = 2251 : Unable to connect to LDAP server
# System Name = ITSO_SVC
# Error Sequence Number = 1008
# Timestamp = Mon Oct 11 10:39:22 2021
# Object Type = ldap_server
# Object ID = 1
# Object Name =
# Cluster Alias = 0x2042230ca96
# Copy ID =
```

Example 13-11 shows an expanded format syslog message, including all event details, such as sense data.

Example 13-11 Full format syslog message example

```
Oct 11 10:38:51 9.42.162.160 ITSO_SVC local7 err IBM2145[14664] 2145 3.7
Record Type = WARNING
# Machine Type = 2145SV2
# Serial Number = 78E35HW
# Error ID = 9183 : Unable to connect to LDAP server
```

```
# Error Code = 2251 : Unable to connect to LDAP server
# System Version = 8.4.2.0 (build 154.20.2109031944000)
# FRU = None
# System Name = ITSO_SVC
# Node ID = 2
# Error Sequence Number = 1007
# Timestamp = Mon Oct 11 10:38:42 2021
# Object Type = ldap_server
# Object ID = 1
# Object Name =
# Cluster Alias = 0x2042230ca96
# Copy ID =
# Additional Data (0 -> 63) =
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000
# Additional Data (64 -> 127) =
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000
# Machine Part Number =
```

13.8 Audit log

The Spectrum Virtualize *audit log* tracks all action commands that were issued through an SSH session, the management GUI, REST API or RSA.

The audit log is useful when past configuration events are analyzed, especially when you are attempting to determine, for example, how a volume was shared by two hosts or why the volume was overwritten. The audit log also is included in the `svc_snap` support package to aid in problem determination.

To keep a complete record of all action commands that are issued and to comply with regulatory rules, no option is available to clear the audit log.

The audit log is part of the hardened system data, in memory and saved to the internal boot drives. Starting with Spectrum Virtualize V7.7.0, the in-memory audit log was limited to 500 entries. The in-memory audit log is dumped to a file on the current configuration node in `/dumps/audit`.

The audit log provides the following entries:

- Sequence number
- Time stamp of when the actionable command was run on the configuration node
- Identity of the user who ran the action command
- The IP address of the workstation the command was issued from
- The run action command
- An object ID if new objects (for example hosts and VDisks) were created
- Parameters that were run with the actionable command
- The origin of the command; for example, GUI or CLI

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.

The following specific service commands are not included in the audit log:

- `cpdumps`
- `cleardumps`
- `finderr`

Figure 13-56 shows the access to the audit log in the GUI by clicking **Access** → **Audit Log** in the left menu to see which configuration CLI commands ran on the IBM SAN Volume Controller system.

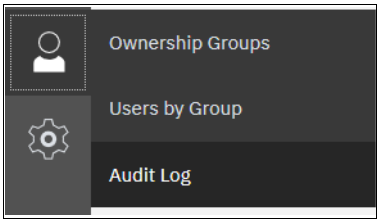


Figure 13-56 Audit Log from Access menu

Figure 13-57 shows an example of the audit log after several volumes were created.

9/21/2021 11:03:30 PM	superuser	svctask chmdisk -gui -tier tier_enterprise 0
9/21/2021 11:03:30 PM	superuser	svctask addmdisk -gui -mdisk 0 Pool1_nodrp
9/21/2021 11:03:14 PM	superuser	svctask mkmdiskgrp -unit mb -datareduction no -easytier auto -encrypt no -ext 1024 -gui...
9/21/2021 11:02:49 PM	superuser	svctask rmmdisk -force -gui -mdisk 0 Pool0
9/21/2021 10:57:06 PM	superuser	svctask chvdisk -gui -syncrate 150 13

Figure 13-57 Audit log entries

Changing the view of the Audit Log grid also is possible by right-clicking column headings (see Figure 13-58). The grid layout and sorting is under the user's control. Therefore, you can view everything in the audit log, sort different columns, and reset the default grid preferences.

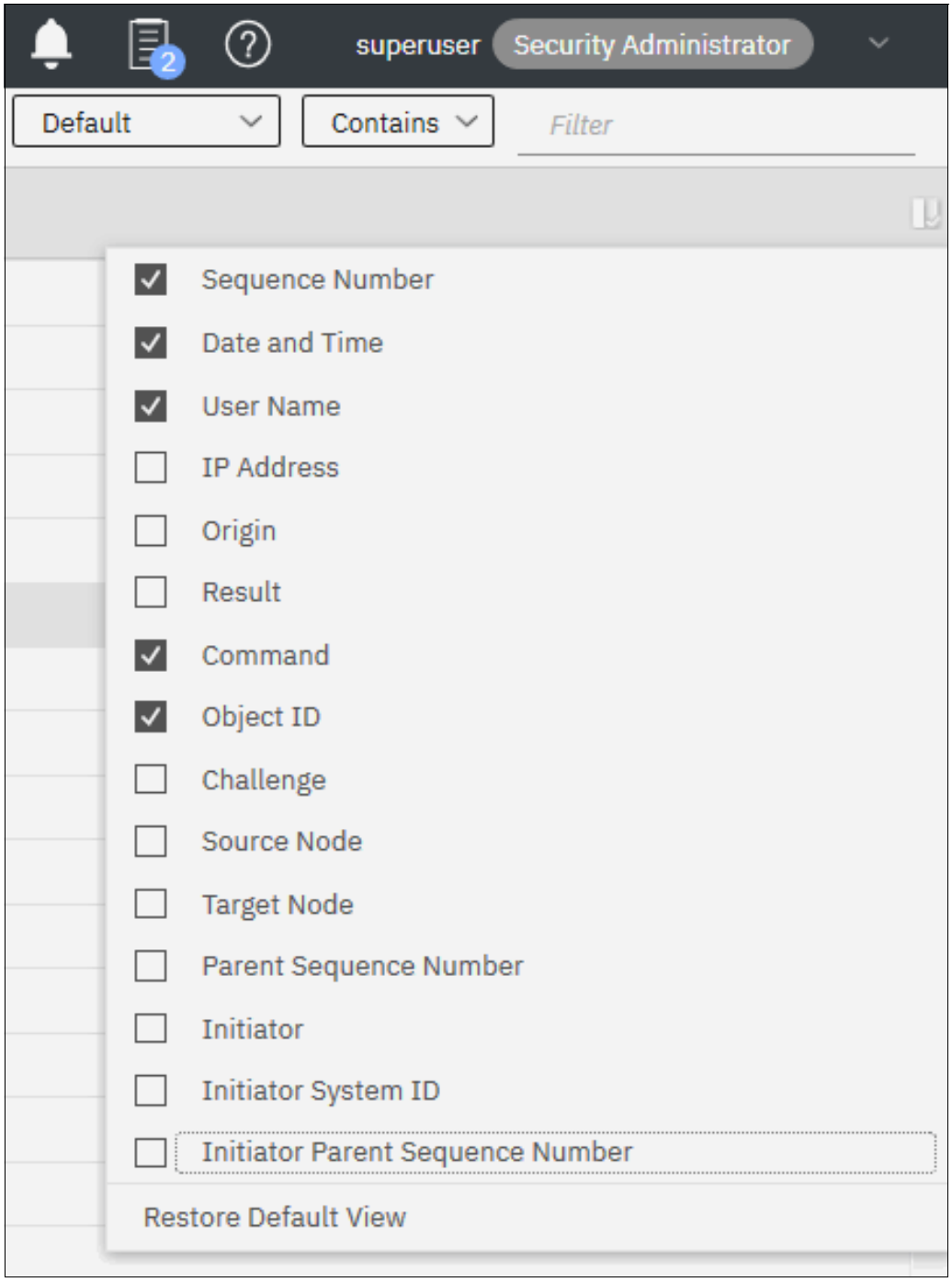


Figure 13-58 Right-click audit log column headings

13.9 Collecting support information by using the GUI, CLI, and USB

At times, if you are facing a problem and raise a ticket with the IBM Support Center, you most likely are asked to provide a support package, which is commonly referred to as a *snap*. You can collect a snap from within the Spectrum Virtualize management GUI by clicking **Settings** → **Support**.

The generated snap file is a **`.tgz`** archive.

For more information, see this IBM Support web page.

13.9.1 Collecting information by using the GUI

To collect information by using the GUI, browse to **Settings** → **Support** → **Support Package** (see Figure 13-59).

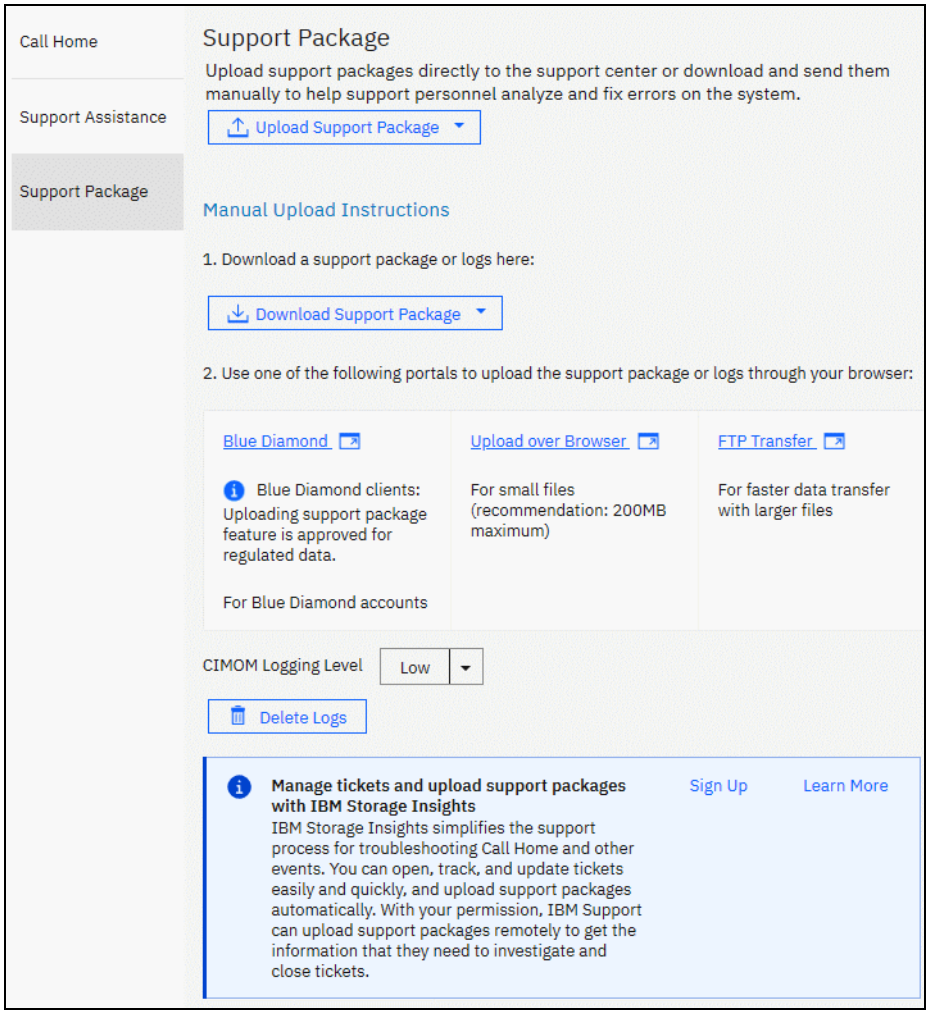


Figure 13-59 Support Package options

Various options are available for how to collect a snap and how to submit it to IBM Support Center.

Snap creation options

A new snap can be generated, and an existing snap or a specific file that is on any node's internal drive can be collected. When creating a snap, one of the following options must be selected (see Figure 13-60):

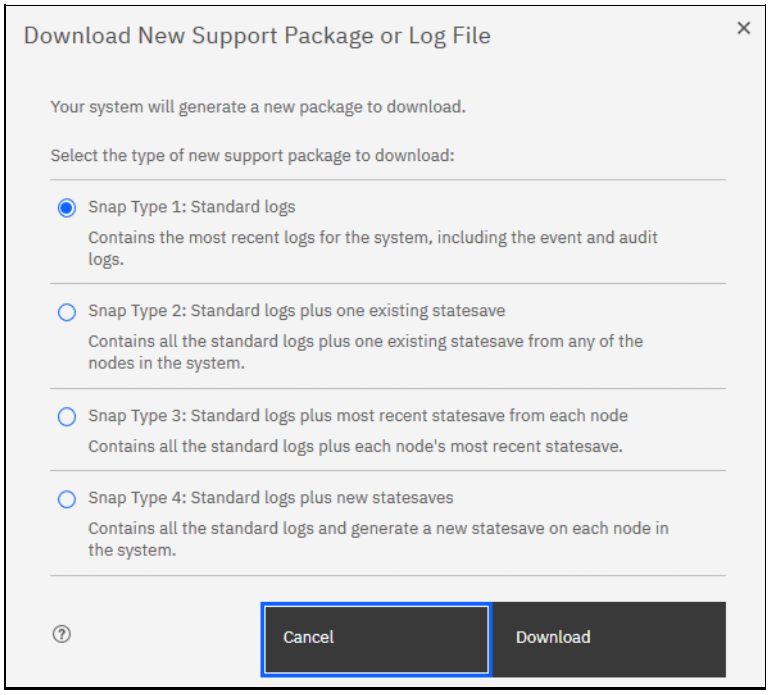


Figure 13-60 Snap type selection

- Snap Type 1: Standard logs
  - Smallest package, shortest creation time
  - Contains the most recent logs for the system, including the event and audit logs
  - Suitable in most cases
- Snap Type 2: Standard logs plus one existing statesave
  - Same as Snap Type 1, plus the most recently stored dump file from any of the nodes in the system
  - Moderately larger snap file for containing one node's dump file
- Snap Type 3: Standard logs plus most recent statesave from each node
  - Same as Snap Type 1, plus the most recently stored dump file from each node in the system
  - Snap file can be large
- Snap Type 4: Standard logs plus new statesaves
  - Same as Snap Type 1, a new statesave, also referred to as *livedump*, is created for each node in the system
  - Can take considerable time for creation, snap file can be large
  - Used in situations, where deeper investigation is needed

Snap transfer options

Assuming the Spectrum Virtualize system can connect to the internet, a new or existing snap can be uploaded directly to the IBM Support Center by using the **Upload Support Package** option (see Figure 13-59).

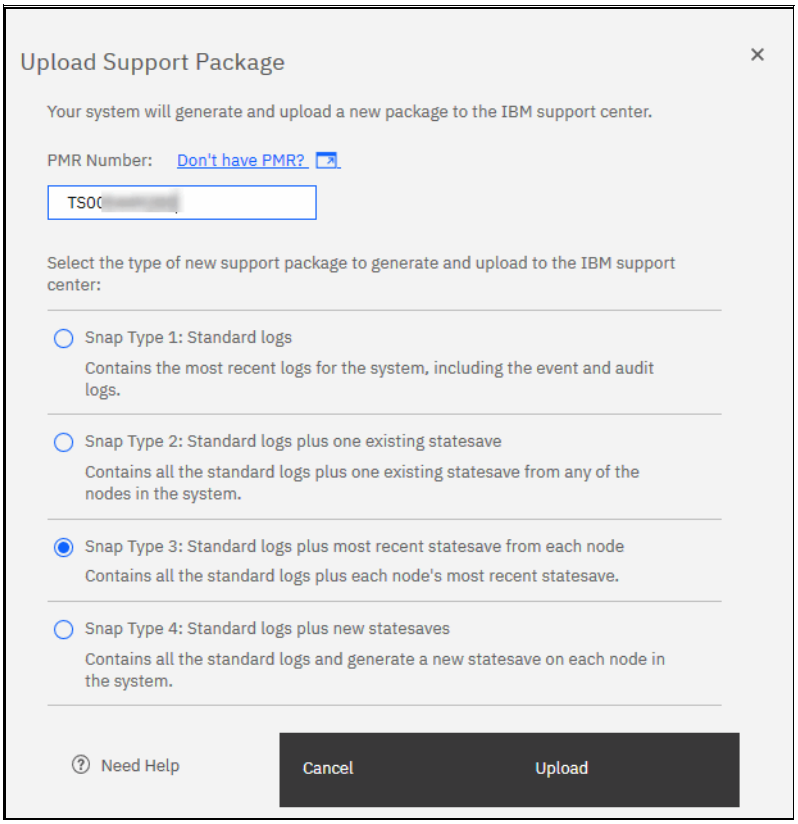


Figure 13-61 Upload Support Package window

In this case, you ideally raised a case with IBM Support already and received a case number to enter into the PMR Number field. Doing so helps IBM Support teams to associate uploaded support data with the case. Instead of creating a snap, you might be asked by IBM Support to upload a support package or dump file, see Figure 13-61.

Note: Although the Problem Management Record (PMR) Number field still follows the previous case number scheme of *ppppp,bbb,ccc*. you also can enter the new support case number format, such as `TS001234567`.

If you do not have a ticket number, you leave this field empty so that the support data is uploaded into the IBM Enhanced Customer Data Repository (ECuRep) and filed under the device serial number. ECuRep is IBM's strategic solution for managing client diagnostic data in an integrated and safe approach on global scale.

If the system does not have internet connectivity, choose the **Download Support Package** option to download the snap to your local workstation and then upload it manually to IBM ECuRep (log in required).

Example

Assuming the encountered problem was an unexpected node restart that logged a 2030 error, collect the default logs and the most recent statesave alias dump from each node to capture the most relevant data for support.

When a node restarts unexpectedly, it dumps its current statesave information to a dump file on the internal node boot drive before it restarts to recover from an error condition. This statesave is critical for support to analyze what caused the restart. Collecting a snap type 4 generates new statesaves at the time of the collection, which is not useful in this situation for understanding the restart event.

If you are contacted by IBM Support because your system sent a call home or you manually opened a case with IBM Support, you are provided a case number. Enter that case number into the **PMR** field and select the snap type, often referred to as an `option 1, 2, 3, 4 snap` or `snap/1(234)`, as requested by IBM Support (see Figure 13-62).

In our example, we enter our PMR number, select **Snap Type 3** (option 3) because this option automatically collects the statesave that is created at the time the node restarted, and then, click **Upload**.

Note: To open a support case online, see this IBM Support web page.

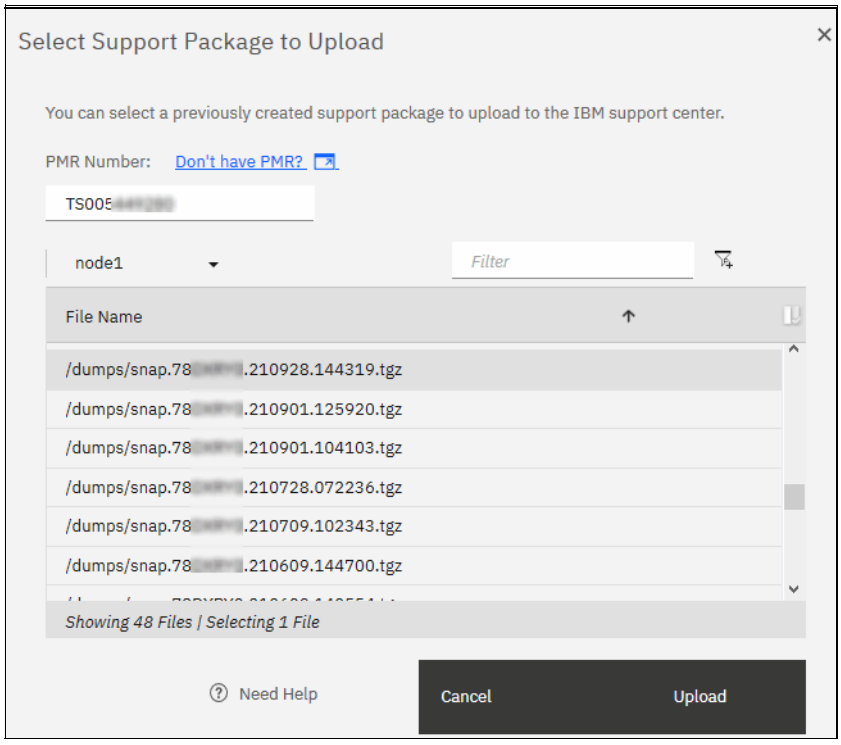


Figure 13-62 Upload existing Support Package window

The procedure to create the snap on an IBM Spectrum Virtualize system, including the latest statesave from each node, begins. This process might take a few minutes (see Figure 13-63).

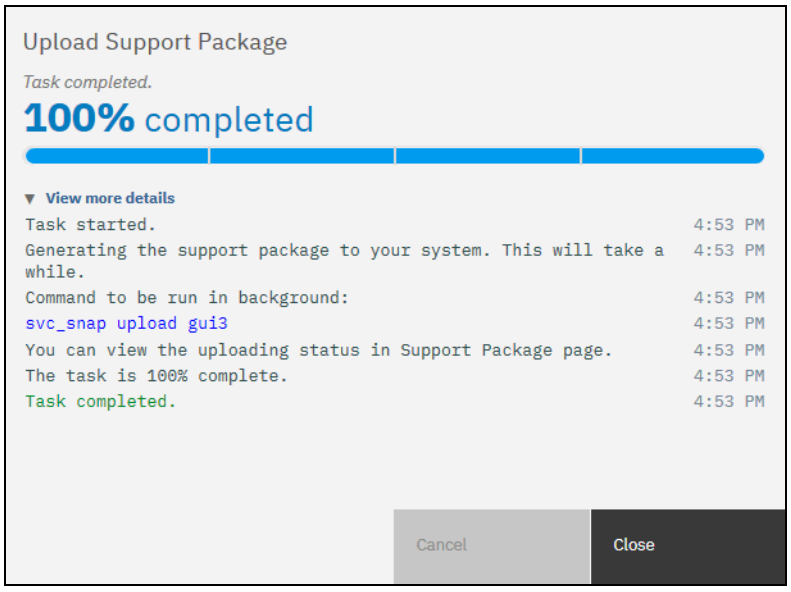


Figure 13-63 Snap Creation Progress

The time that it takes to generate the SNAP and the size of the file that is generated depends mainly on two factors: the SNAP option that you selected, and the size of your system.

A Snap/1 takes much less time than a Snap/4; for example, because no node live dumps are created for a Snap/1, whereas a Snap/4 requires the system to collect new statesaves from each node. In an 8-node cluster, this process can take quite some time; therefore, always collect the SNAP option that IBM Support recommends to you.

The approximate file sizes for each SNAP option are listed in Table 13-10.

Table 13-10 Snap file sizes

Option	Description	Approximate size (1 I/O group, 30 volumes)	Approximate size (4 I/O groups, 250 volumes)
1	Standard logs	10 MiB	340 MiB
2	Standard logs plus one existing statesave	50 MiB	520 MiB
3	Standard logs plus most recent statesave from each node	90 MiB	790 MiB
4	Standard logs plus new statesaves	90 MiB	790 MiB

Note: The snap sizes can vary depending on the number of objects that are in the system. For large systems (for example, systems that consist of four I/O groups and several thousand volumes), it can be several GiB in size.

13.9.2 Collecting logs by using the CLI

The CLI can be used as an alternative means to collect and upload a support package as requested by IBM Support by completing 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 (snap/1):

     ```
     svc_snap upload pmr=TS001234567 gui1
     ```

   - Standard logs plus one existing statesave (snap/2):

     ```
     svc_snap upload pmr=TS001234567 gui2
     ```

   - Standard logs plus most recent statesave from each node (snap/3):

     ```
     svc_snap upload pmr=TS001234567 gui3
     ```

   - Standard logs plus new statesaves (snap/4):

     ```
     svc_livedump -nodes all -yes
     svc_snap upload pmr=TS001234567 gui3
     ```

   We collect the type 3 (option 3) information, which is automatically uploaded to the PMR number that is provided by IBM Support, as shown in Example 13-12.

   Example 13-12 The svc_snap command

   ```
   IBM_2145:ITSOSVC:superuser>svc_snap upload pmr=TS001234567 gui3
   ```

   If you do not want to upload automatically the snap to IBM, do not specify the **upload pmr=TS001234567** part of the commands. When the snap creation completes, it creates a file that is named with the following format:

   ```
   /dumps/snap.<node_panel_id>.YYMMDD.hhmmss.tgz
   ```

   It takes a few minutes for the snap file to complete (longer if it includes statesaves).

2. The generated file can then be downloaded by using the GUI. Browse to **Settings** → **Support** → **Manual Upload Instructions** → **Download Support Package** → **Download Existing Package**, as shown in Figure 13-64.

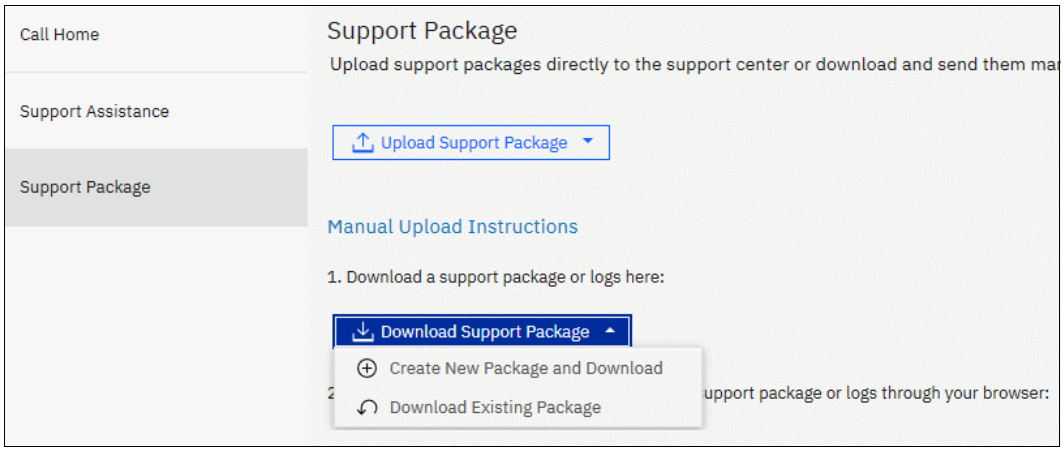


Figure 13-64 Downloaded Existing Package

3. Click **Filter** and enter `snap` to see a list of snap files, as shown in Figure 13-62 on page 911. Locate the name of the snap that was generated by the **svc_snap** command that ran earlier, select that file, and then, click **Download**.

4. Save the file to a folder of your choice on your workstation.
5. Upload the snap manually to IBM ECuRep.

Instead of using the GUI, snap files, dumps and other diagnostic files can be downloaded by using the SCP protocol that uses an SCP client as **scp** (command line tool available for all operating systems), **pscp.exe** (command line tool for Windows, part of the PuTTY suite).

Note: The permissions that are set for the node file system do not permit directory browsing. For example, using a GUI scp client, such as, `WinSCP` or `FileZilla` reports an error message when you attempt to navigate to the `/dumps` directory.

13.9.3 Collecting logs by using USB

In specific situations, for example, if a node cannot be reached over its IP address, it is possible to collect a instances snap from a node by attaching a USB flash drive to either of the node USB ports. Unless large dump files must be collected along with a single node snap, a USB flash drive capacity of 1 - 4 GiB often suffices.

Note: This procedure collects a *single SNAP* from the node, not a regular *clustered system SNAP*. It is useful for determining the state of the node canister.

When a USB flash drive is plugged into a node's USB, the Spectrum Virtualize code searches for a text file that is named `satask.txt` in the root directory. If the file exists, it runs the CLI command that is specified in the file. When the command completes, a file that is named `satask_result.html` is stored in the root directory of the USB flash drive.

If the 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, the status, and the configuration information from the node. The status and configuration information matches the detail that is shown on the service assistant home page panels.

Complete the following steps to collect a SNAP:

1. Ensure that your USB drive is formatted with a FAT32 file system on its first partition.
2. Create a text (.txt) file on the USB drive in the root directory called `satask.txt` (case-sensitive).
3. In the `satask.txt` file add a single-line **`satask snap -dump`** and save it.
4. Insert the USB drive 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 then be uploaded to the support center, as described in 13.9.4, "Uploading files to the Support Center" on page 915.

   Note: If a problem occurs with the procedure, the HTML file is still generated and includes the reasons why the procedure failed.

For more information about the commands that can be run by using USB, see this IBM Documentation web page.

13.9.4 Uploading files to the Support Center

If you choose to not have the system upload the support package automatically, it can still be uploaded for analysis from the ECuRep. Any uploads must be associated with a specific support case, which also is known as *PMR* or *ticket*.

To upload the information, complete the following steps:

1. Using a web browser, go to the IBM ECuRep web page (see Figure 13-65) and log in with your IBM ID. If you do not have a valid IBM ID, create one.

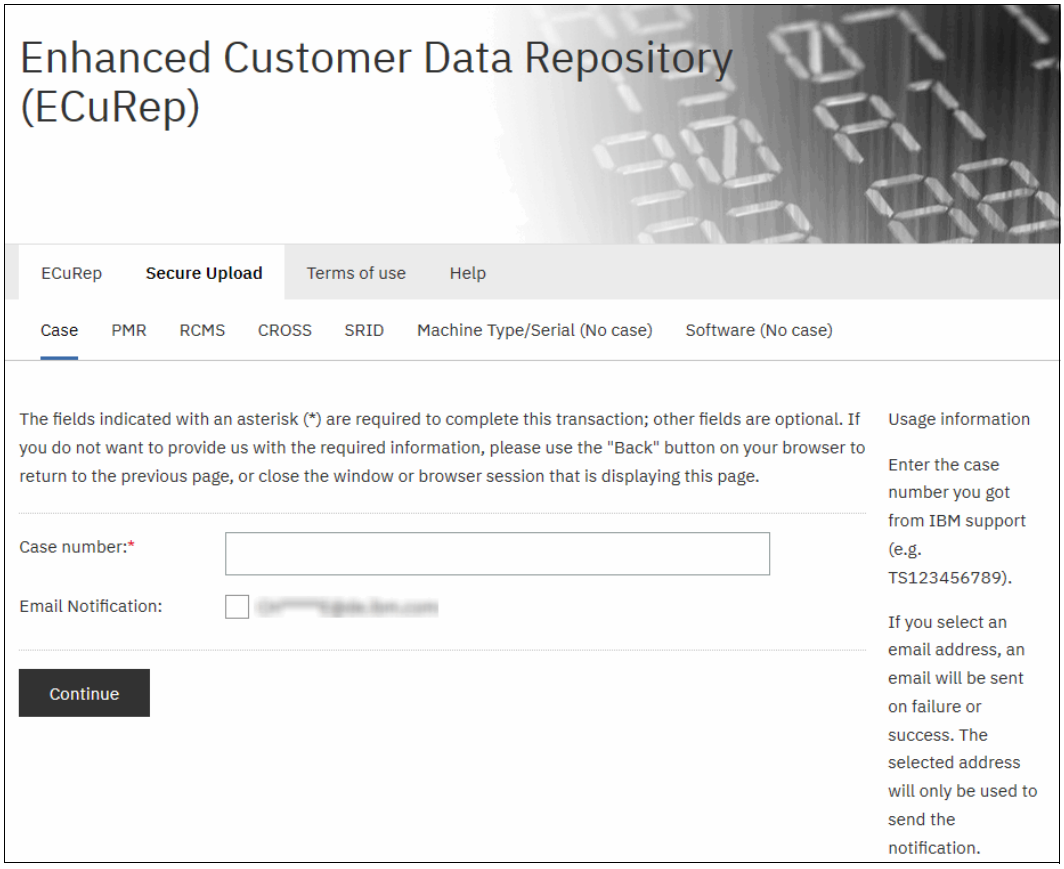


Figure 13-65 ECuRep details

2. Complete the **Case number** field by using the number that was provided by IBM Support for your specific case. This number should be in the format of `TS001234567`.

   Click **Continue**. The Secure Upload tab is displayed.

3. Select **Click to Select Files** (see Figure 13-66), which opens the Operating System's file selection dialog, or drag and drop one or more files from your workstation onto the Drop Files Here area.

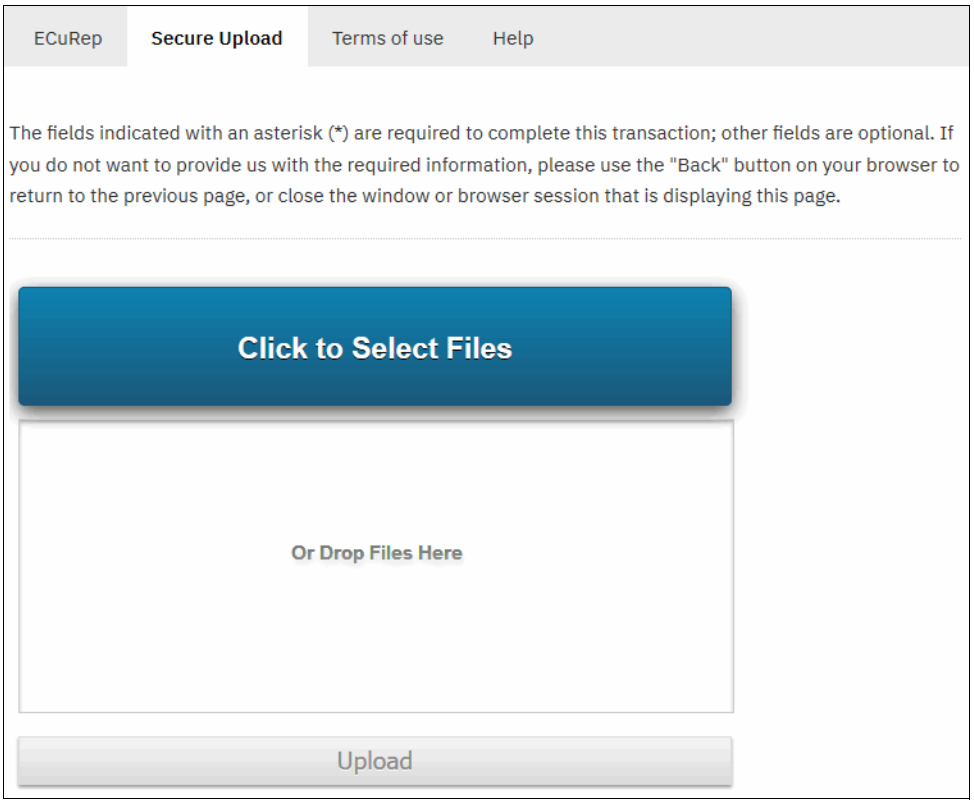


Figure 13-66 ECuRep File upload

4. After completed, click **Upload** to continue, and follow the directions.

13.10 Service Assistant Tool

The Service Assistant Tool (SAT) is a web-based GUI that is used to service individual node canisters, primarily when a node encounters 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, an IBM Spectrum Virtualize cluster is initially configured with the following IP addresses:

- One service IP address for each IBM SAN Volume Controller node.
- 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 is available and tasks can be accomplished by using the Service Assistance Tool:

- Status information about the connections and the IBM SAN Volume Controller nodes
- Basic configuration information, such as configuring IP addresses
- Service tasks, such as restarting the Web server service (Tomcat) or updating the node WWNN
- Details about node error codes

- Details about the hardware, such as IP address, FC port WWPNs and Ethernet port MAC addresses

The SAT GUI is available by using a service IP address that is configured on each node. It also can be accessed by using the cluster management IP addresses by appending `/service` to the system's management IP in your web browser's address bar.

If the clustered system is down, or during initial setup, the only method of communicating with the nodes is through the service IP address directly. A node can have a single service IP address that is configured on Ethernet port 1. A recommended practice is to configure a service IP address on all nodes of a clustered system, including any hot spare nodes.

To open the SAT GUI, enter one of the following URLs into any web browser, and then, click **Service Assistant Tool**:

- `https://<cluster_IP_address>/service`
- `https://<node_service_IP_address>/service`

To access the SAT, complete the following steps:

1. Enter either of the URLs above in your web browser. Log in to the SAT GUI or CLI is possible for the superuser only. Enter the Superuser Password, as shown in Figure 13-67.

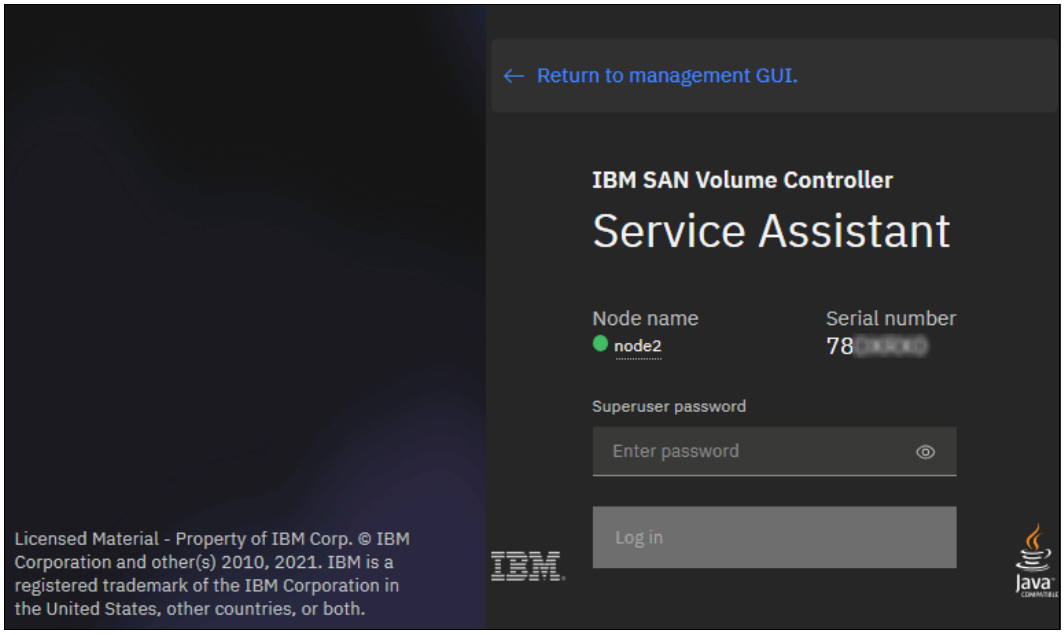


Figure 13-67 Service Assistant Tool GUI Login

After you are logged in, a list of Available Nodes is shown (see Figure 13-68). By using the SAT, you can view the status and run service actions on any available of the nodes that are listed here. From this SAT window, you can activate any node's blue indicator LED(*); for example, to help a service technician to identify the correct node in the rack. Also, you can power off, restart, reboot, or put a node into service state (**).

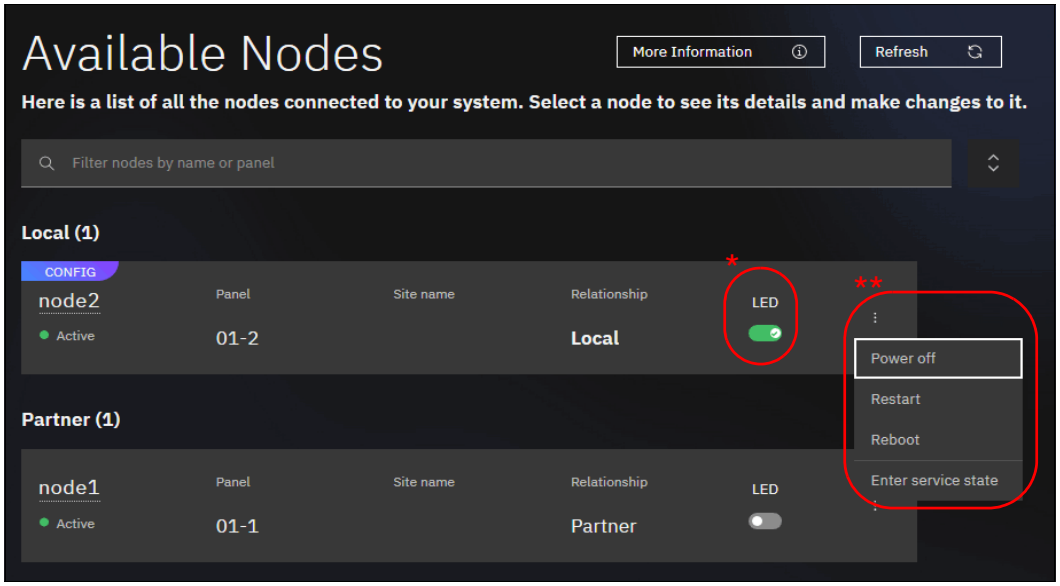


Figure 13-68 Service Assistant Tool GUI

2. To access all available service actions (see Figure 13-69 on page 919), select the node to act on by clicking its name.

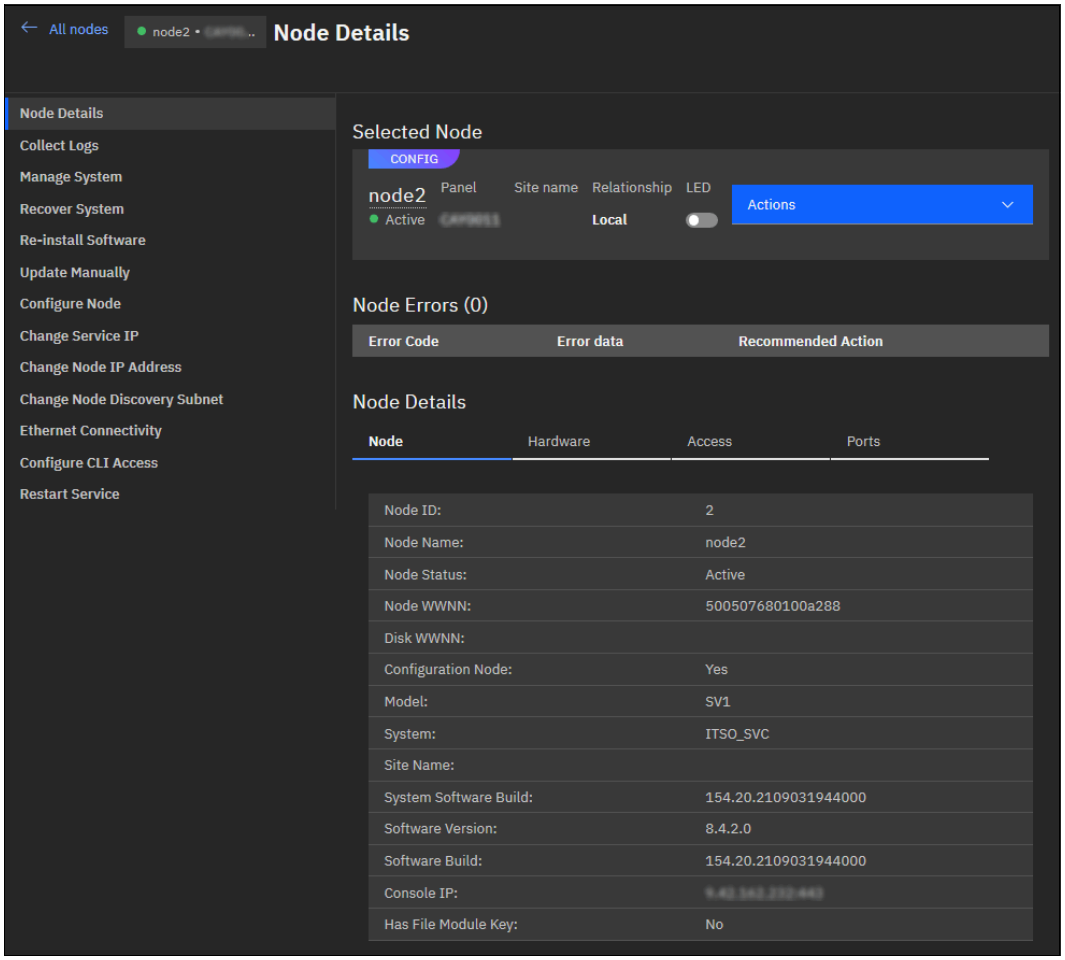


Figure 13-69 SAT GUI: Available options

Note: The SAT GUI provides access to service procedures, which can lead to access loss and even loss of data if applied incorrectly. It is advised that these procedures are carried out only if you are directed to do so by IBM Support.

For more information about how to use the SAT, see this IBM Documentation web page.

13.11 IBM Storage Insights Monitoring

IBM Storage Insights allows you to optimize your storage infrastructure by using this cloud-based storage management and support platform with predictive analytics. For more information about how to set up IBM Storage Insights and the features it offers, see Chapter 2, “Planning” on page 59.

The monitoring capabilities that IBM Storage Insights provides are useful for tasks, such as capacity planning, workload optimization, and managing support tickets for ongoing issues.

After you add your systems to IBM Storage Insights, you see the Dashboard in which you can select a system that you want to see the overview for, as shown in Figure 13-70.

Note: The system that is used to demonstrate IBM Storage Insights is a FlashSystem 9150. However, because you can add a large range of systems to IBM Storage Insights, the processes are standard across all, which gives you a single pane of glass to manage your devices.

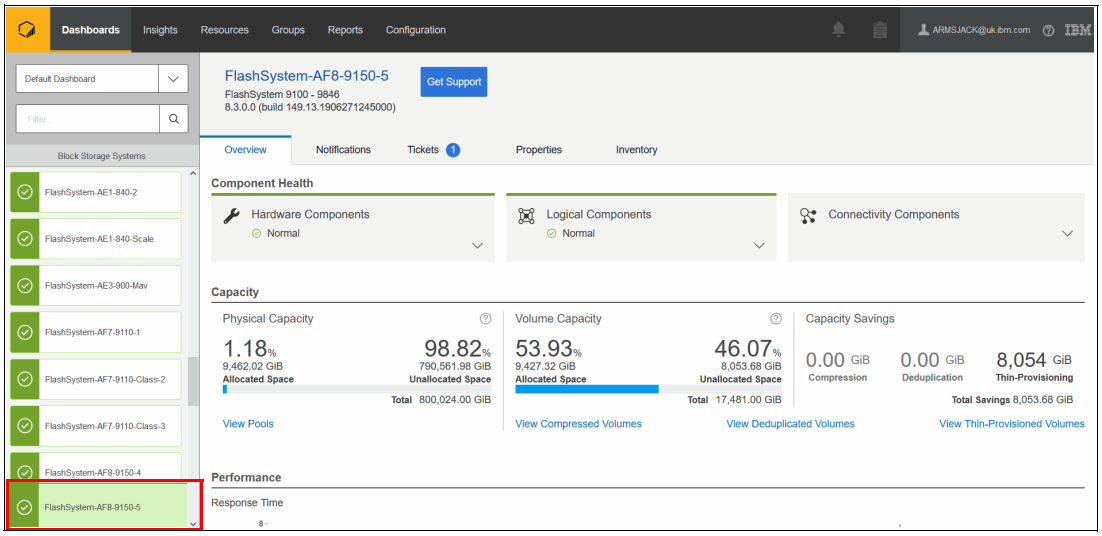


Figure 13-70 IBM Storage Insights System overview

Component health can be seen at the top in the most prominent part of the pane. Here, if a problem exists with one of the Hardware, Logical or Connectivity components, errors are displayed, as shown in Figure 13-71.

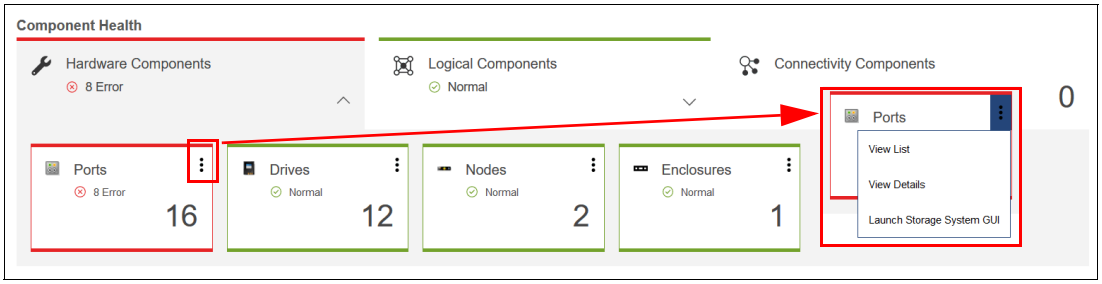


Figure 13-71 Component Overview

The errors can be expanded to obtain more details by selecting the three dots in 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, depending on which component has the error, as shown in Figure 13-72.

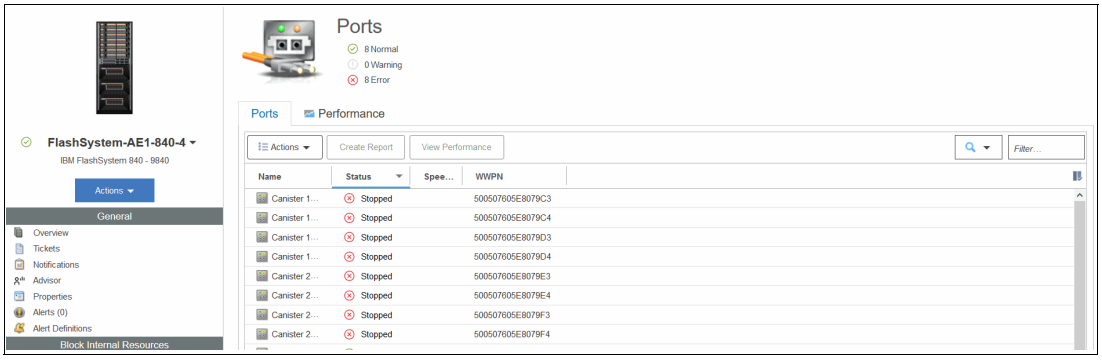


Figure 13-72 Ports in error

From here, it is now obvious which components have the problem and exactly what is wrong with them, which enables you to log a support ticket with IBM if necessary.

13.11.1 Capacity monitoring

You can see key statistics, such as Physical Capacity, Volume Capacity, and Capacity Savings, depending on what is configured after you select a system, as shown in Figure 13-73.

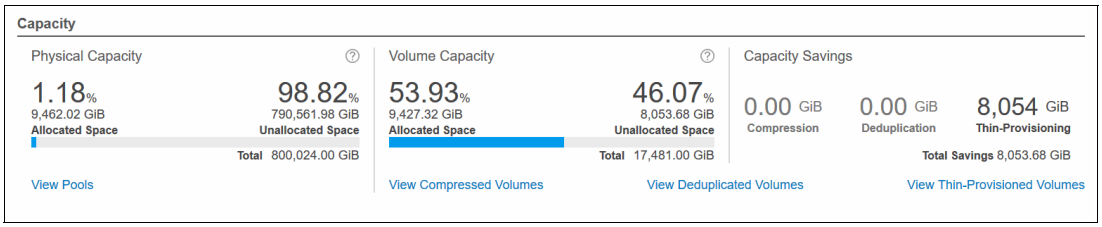


Figure 13-73 Capacity Area of the IBM Storage Insights System Overview

This menu allows the user to View Pools, View Compress Volumes, View Deduplicated Volumes, and View Thin-Provisioned Volumes. Clicking these options takes the user to the detailed System View for the selected option.

From there, Capacity can be selected to get a historical view of how system capacity changed over time, as shown in Figure 13-74 on page 922. At any time, the user can select the wanted timescale, resources, and metrics to be displayed on the graph by clicking any of the options that are around the graph.

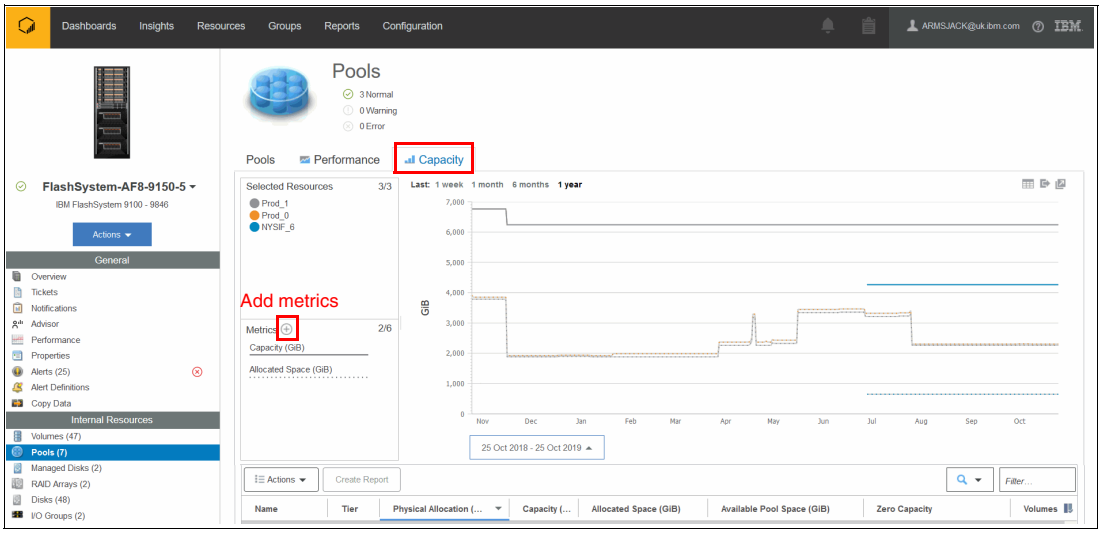
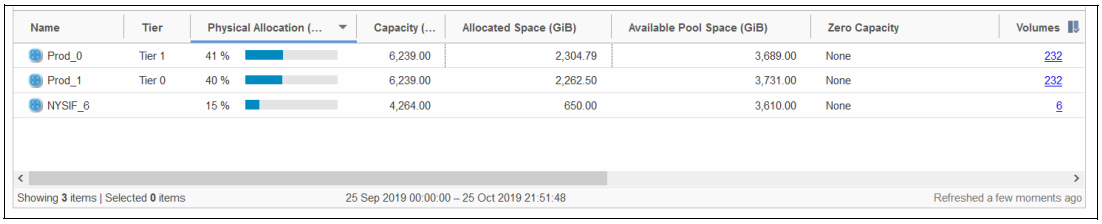


Figure 13-74 IBM Storage Insights Capacity View

If you scroll down, a list view of the selected option can be found below the graph. In this example, we selected **View Pools** so the configured pools are shown with the relevant key capacity metrics, as shown in Figure 13-75. Double-clicking a pool in the table display the pool's properties.



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 13-75 Pools List View

13.11.2 Performance monitoring

From the system overview, you can scroll down and receive the three key performance statistics for your system, as shown in Figure 13-76. For the performance overview, these stats are aggregated across the whole system, and you cannot drill down by Pool, Volume, and so on.

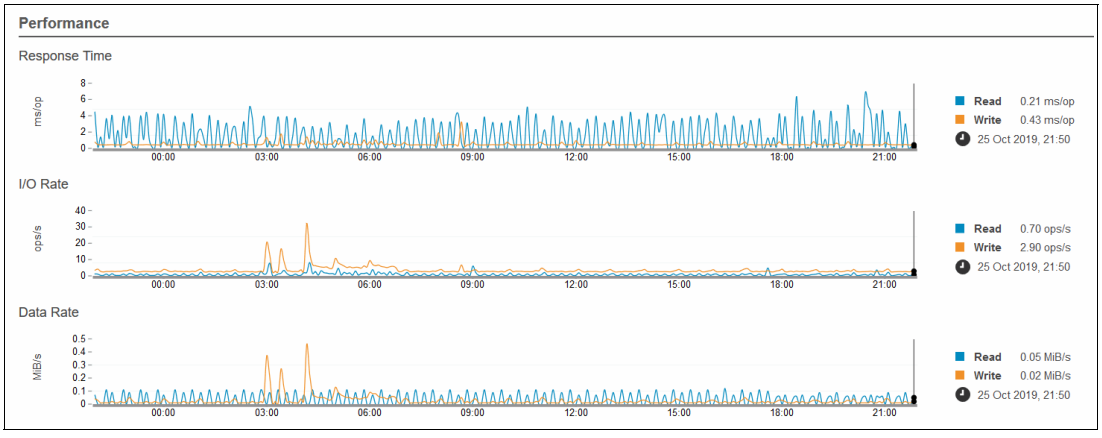


Figure 13-76 System overview Performance

For more information about performance statistics, we must enter the system view again as described in the example that is in 13.11.1, “Capacity monitoring” on page 921.

For this performance example, we again select **View Pools**, and then, select **Performance** from the System View window, as shown in Figure 13-77.

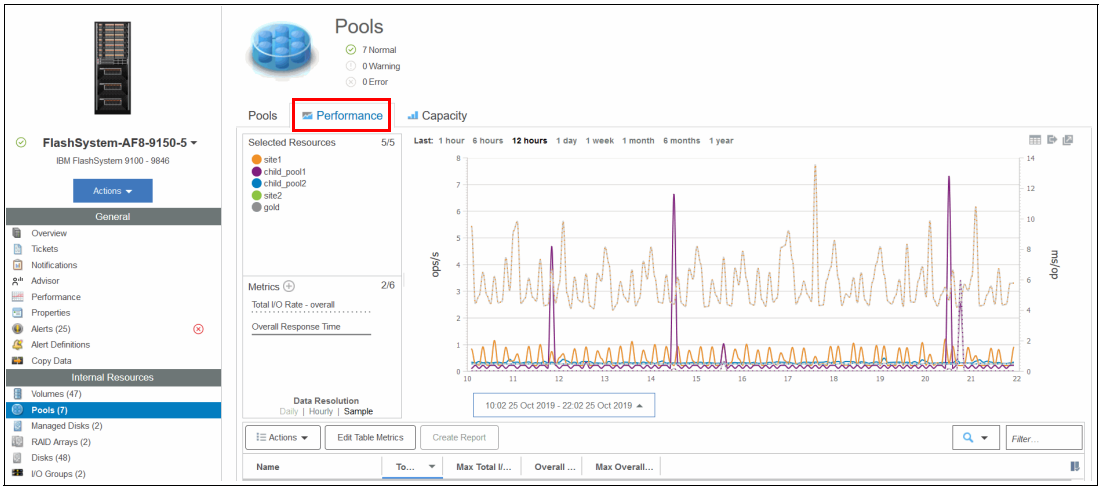


Figure 13-77 IBM Storage Insights Performance View®

It is then possible to customize what can be seen on the graph by selecting the wanted metrics and resources. In Figure 13-78, the Overall Response Time for one Pool over a 12-hour period is displayed.

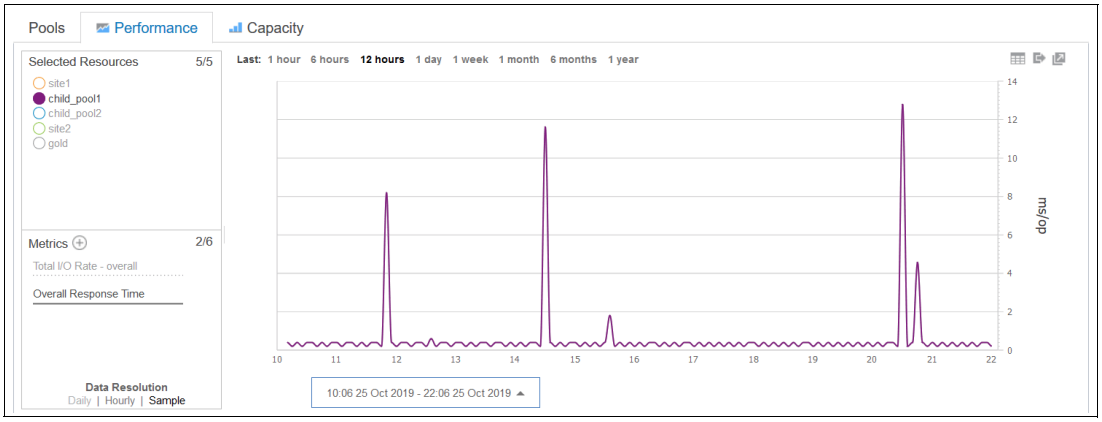


Figure 13-78 Filtered performance graph

Scrolling down below the graph, the Performance List View is visible, as shown in Figure 13-79 on page 924. Wanted metrics can be selected by clicking the filter button to the right of the column headers. If you select a row, the graph changes to be filtered for that selection only. Multiple rows can be selected by holding down the **Shift** or **Ctrl** keys**.**

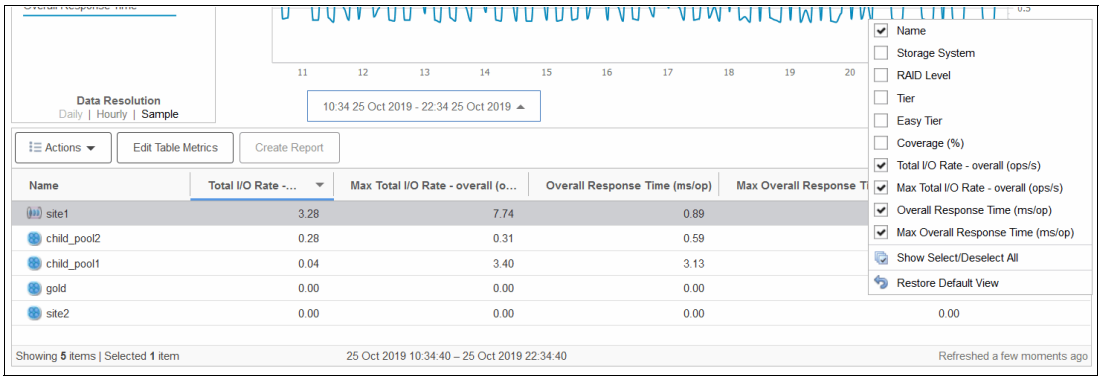


Figure 13-79 Performance List View

13.11.3 Logging support tickets by using IBM Storage Insights

IBM Storage Insights allows the user to log support tickets that greatly complements the enhanced monitoring opportunities that the software brings. When an issue is detected, and the user wants to engage IBM Support, complete the following steps:

1. Select the system to open the System Overview and click the **Get Support** button, as shown in Figure 13-80.

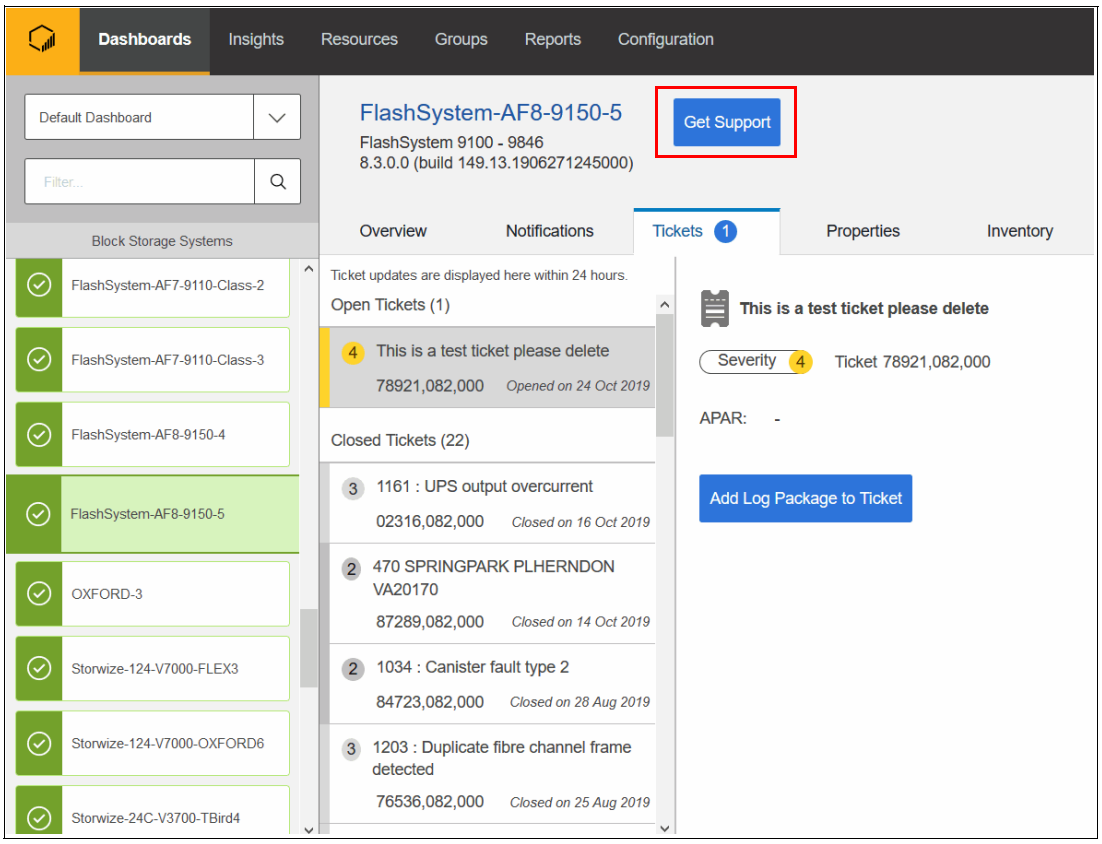


Figure 13-80 Get Support button

A window opens in which the user can opt to create a ticket or update a ticket, as shown in Figure 13-81.

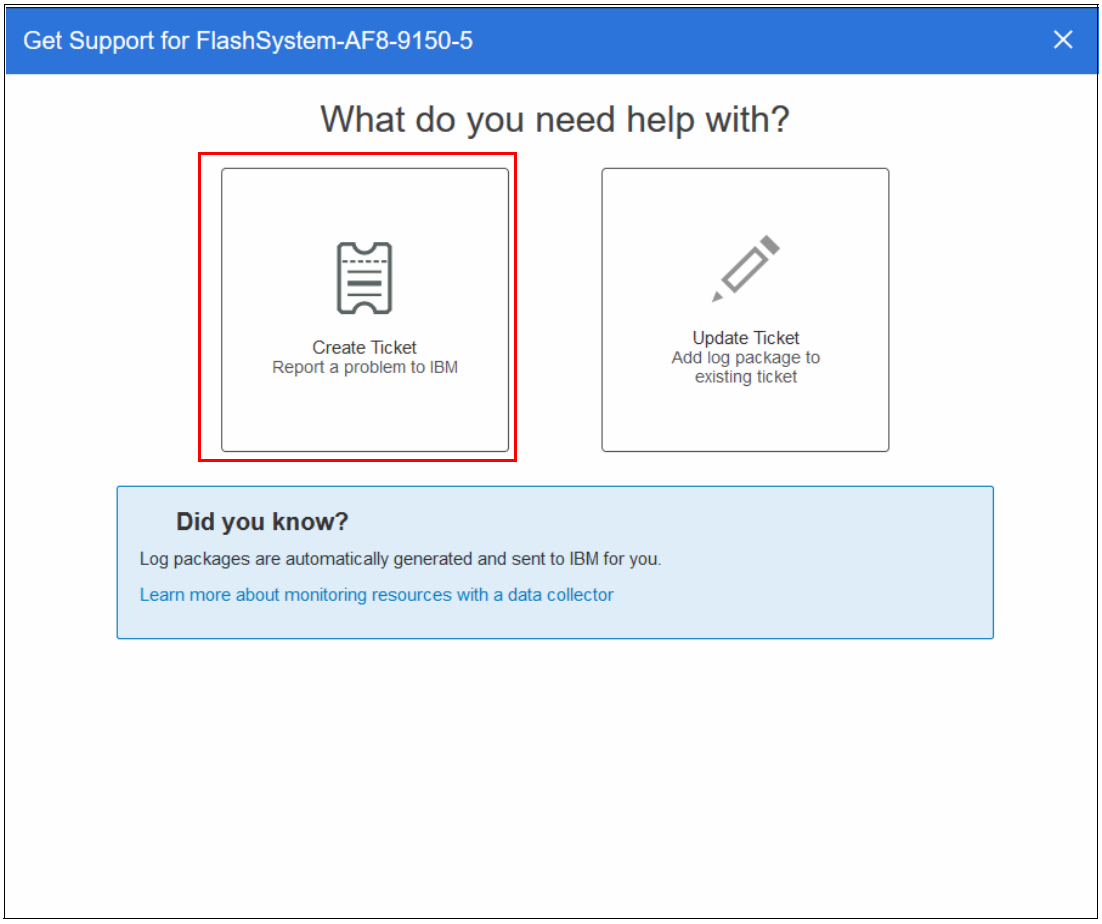


Figure 13-81 Get Support window

2. Select **Create Ticket** and the ticket creation wizard opens, as shown in Figure 13-82.

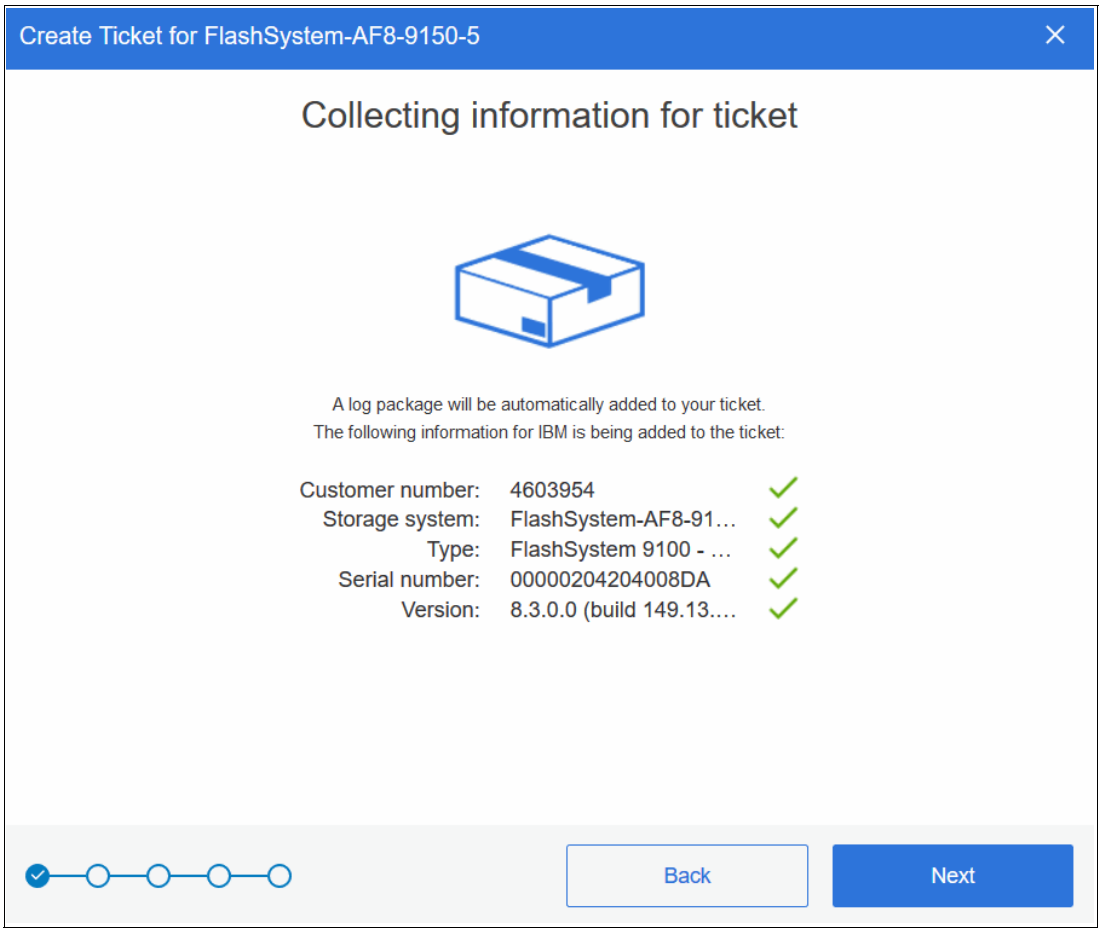


Figure 13-82 Create Ticket Wizard

3. Details of the system are automatically populated, including the customer number. Select **Next** (see Figure 13-83).

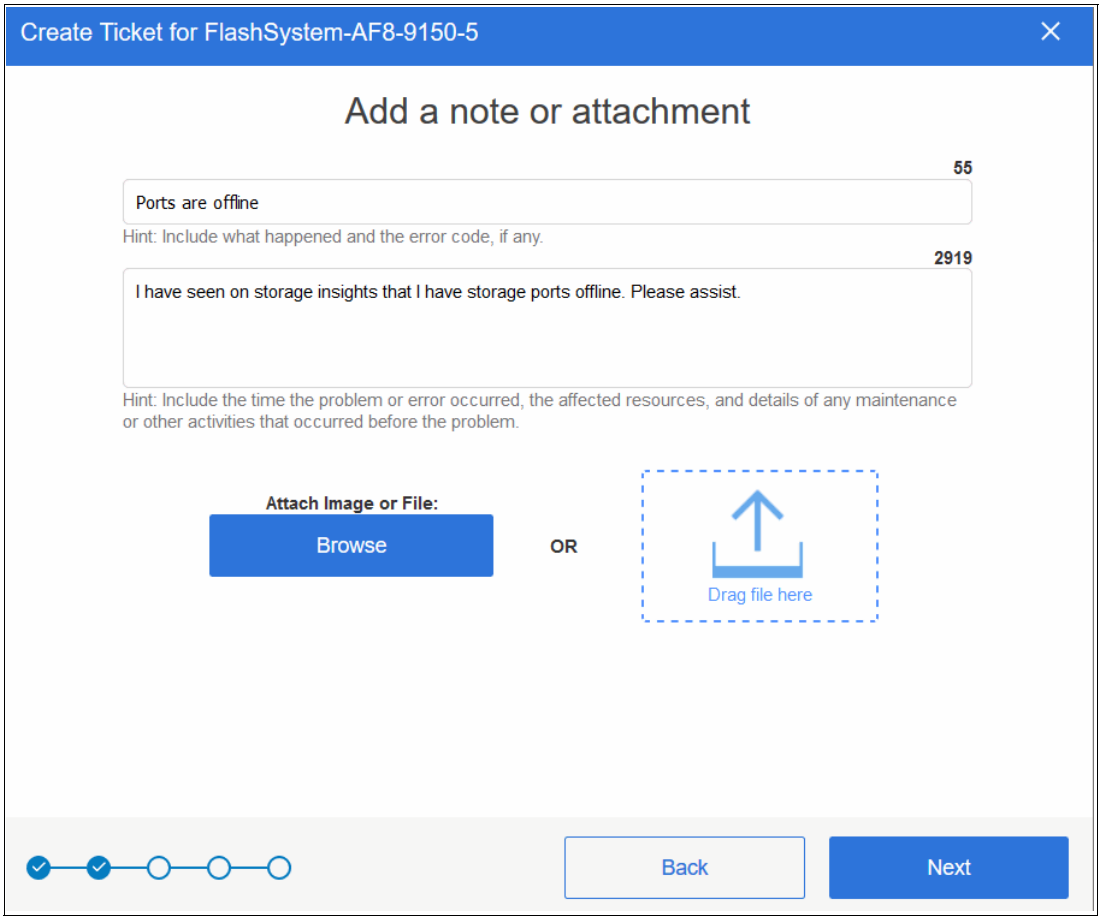


Figure 13-83 Add note or attachment

4. The user can enter relevant details about your problem into the ticket, as shown in Figure 13-83. It is also possible to attach images or files to the ticket, such as PuTTY logs and window captures. After you are done, select **Next**.

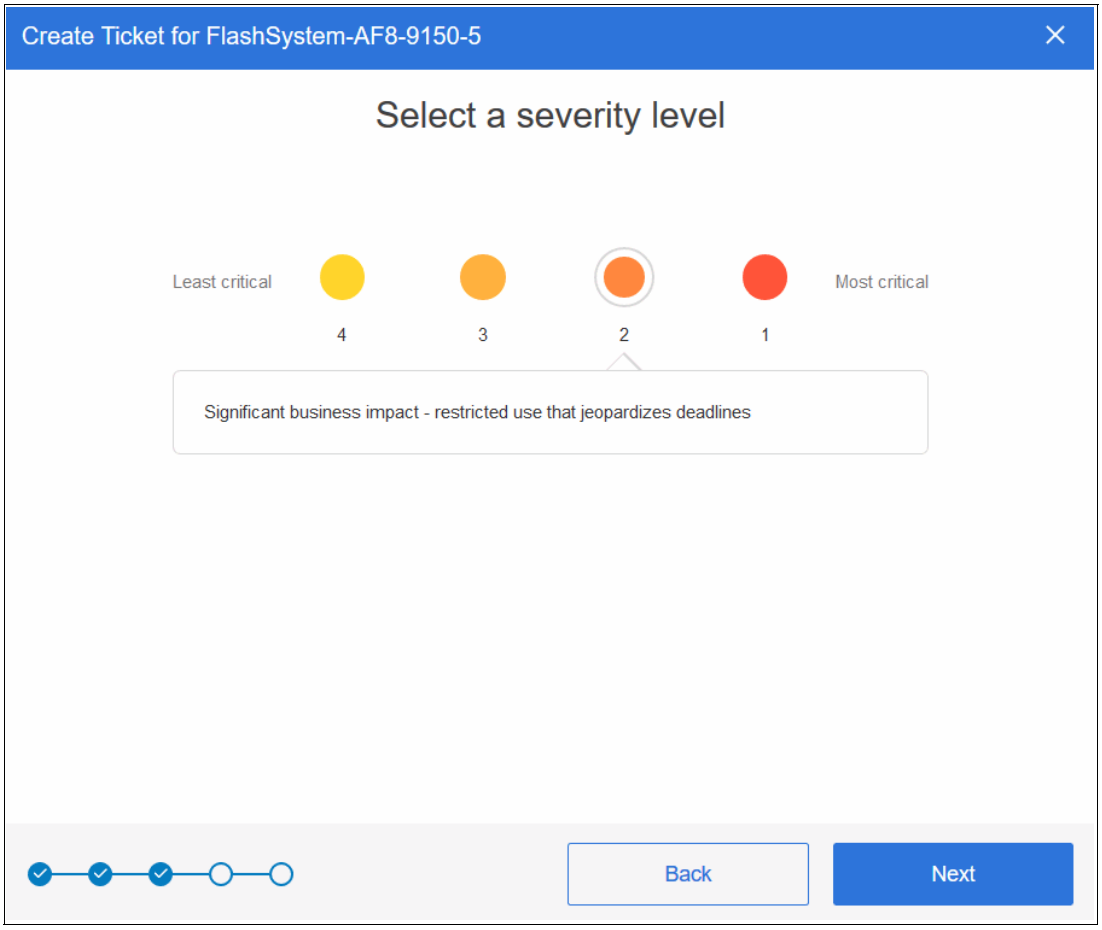


Figure 13-84 Select Severity Level

5. In the third window, you can select severity for the ticket; examples of what severity you should select are displayed, as shown in Figure 13-84. Because in our example storage ports are offline with no effect, we select **severity 2** because we lost redundancy.

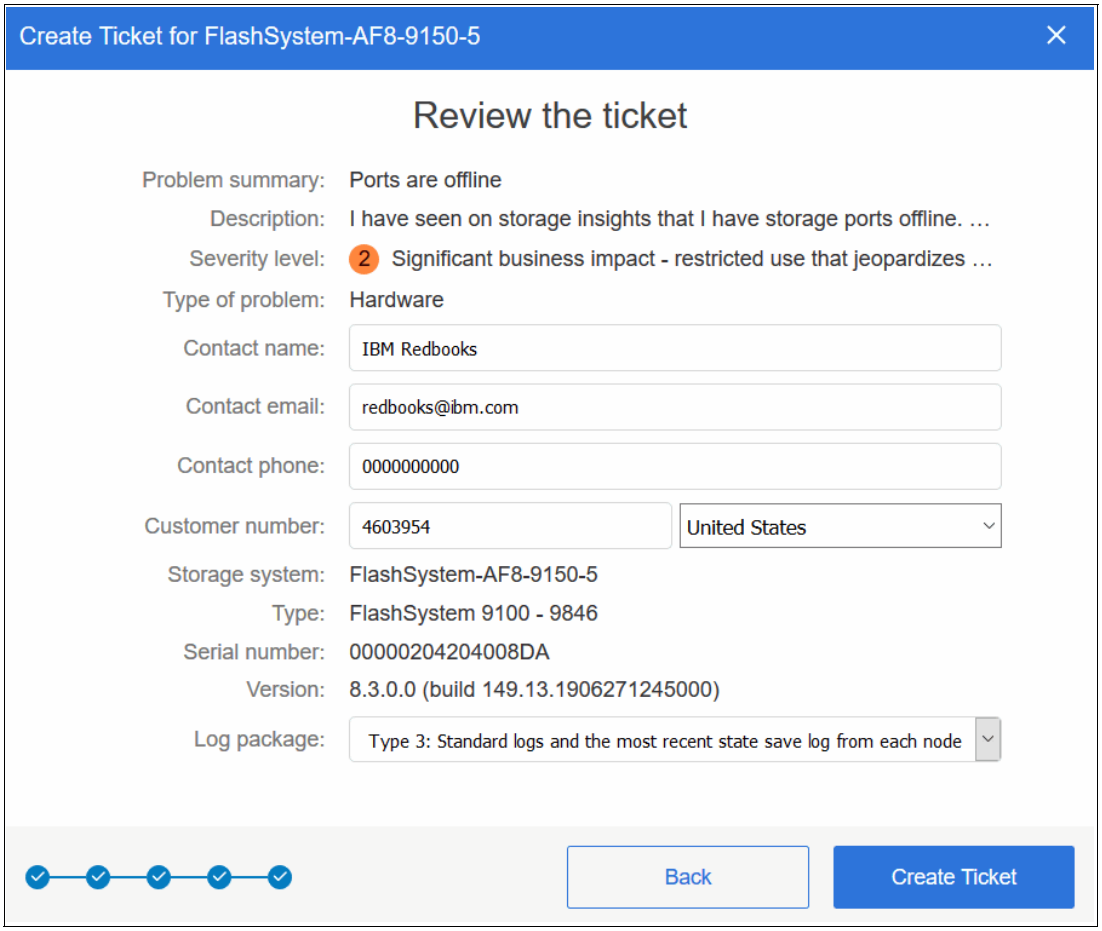


Figure 13-85 Review Ticket

6. The fourth window allows the user to choose whether this problem is with hardware or software. Select the relevant option (for this example, offline ports likely are caused by a physical layer hardware problem). After you are done, click **Next**.

   In the final pane, the user can review the details of the ticket that are logged with IBM, as shown in Figure 13-85. Contact details must be entered so that IBM Support can respond to the correct person. The user also must choose which type of logs should be attached to the ticket.

   For more information about the types of SNAP, see Table 13-10 on page 912.

7. After you are done, select **Create Ticket**. A confirmation window opens, as shown in Figure 13-86. IBM Storage Insights automatically uploads the SNAP to the ticket when it is collected, which requires no other intervention from the user.

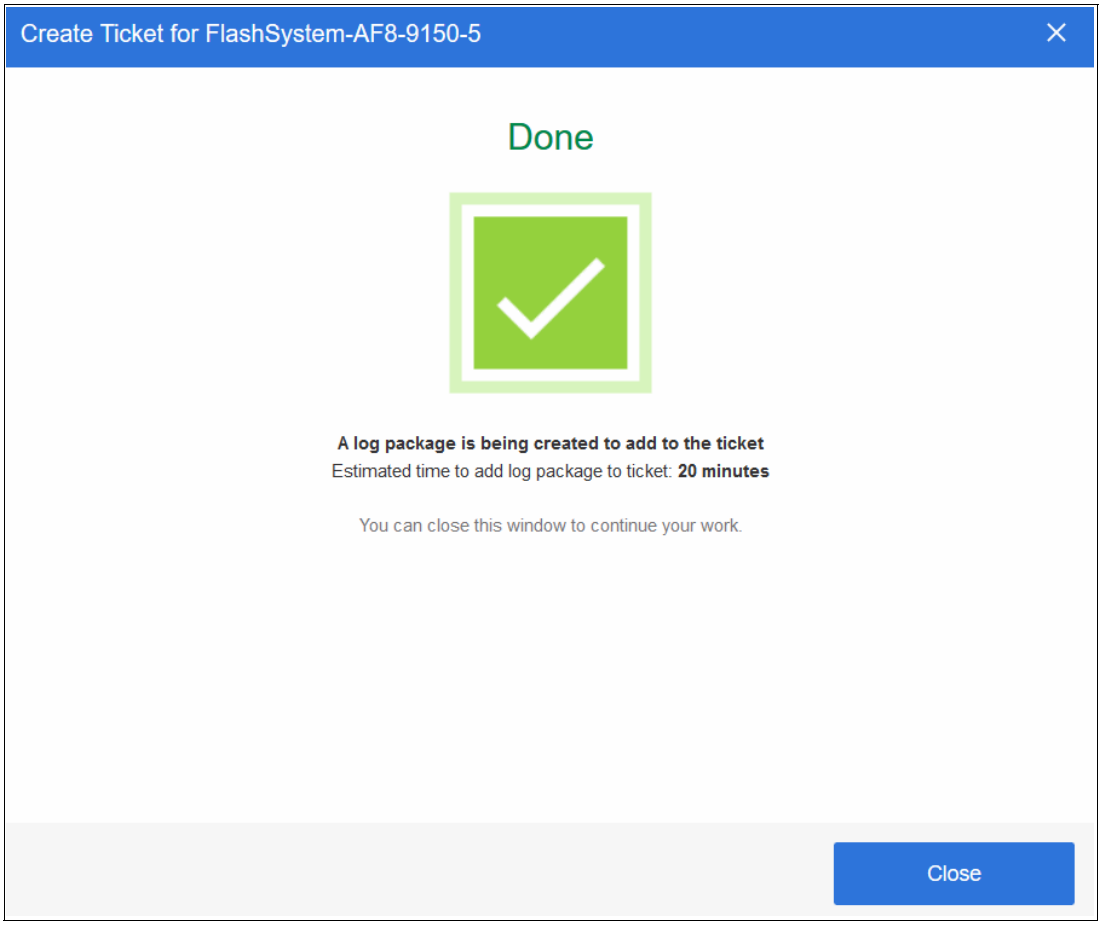


Figure 13-86 Ticket Creation confirmation

13.11.4 Managing support tickets by using IBM Storage Insights and uploading logs

IBM Storage Insights allows the user to track support tickets and upload logs to them. From the System Overview window, select **Tickets**, as shown in Figure 13-87.

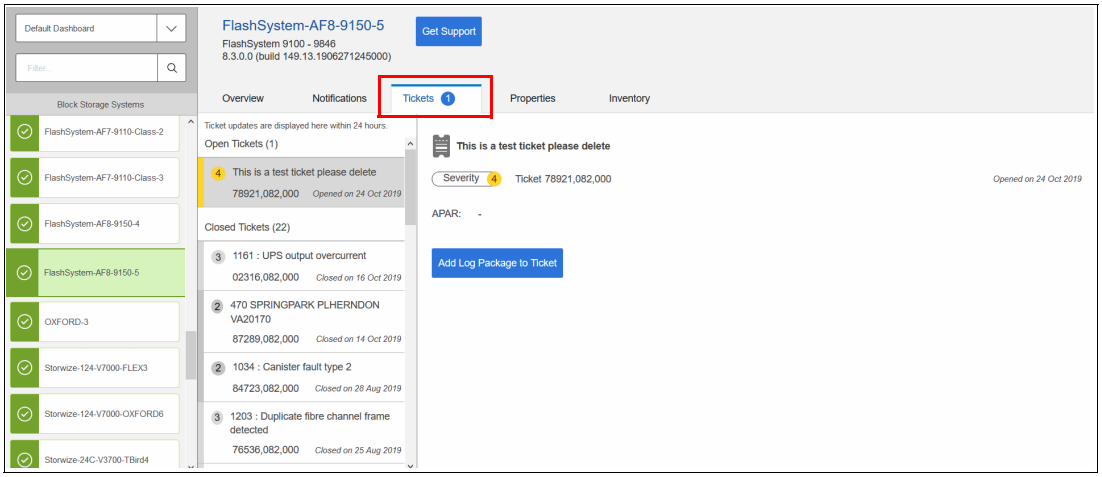


Figure 13-87 View Tickets

This window allows the user to see a large history of support tickets that were logged through IBM Storage Insights for the system. Tickets that are not open are listed under Closed Tickets; open tickets are listed under Open Tickets.

The user selects **Add Log Package to Ticket** to quickly add logs to a ticket without the need to browse to the system GUI or IBM ECuRep. A pop-up window opens that guides the user through the process, as shown in Figure 13-88. The user can select which type of log package they want and add a note to the ticket with the logs.

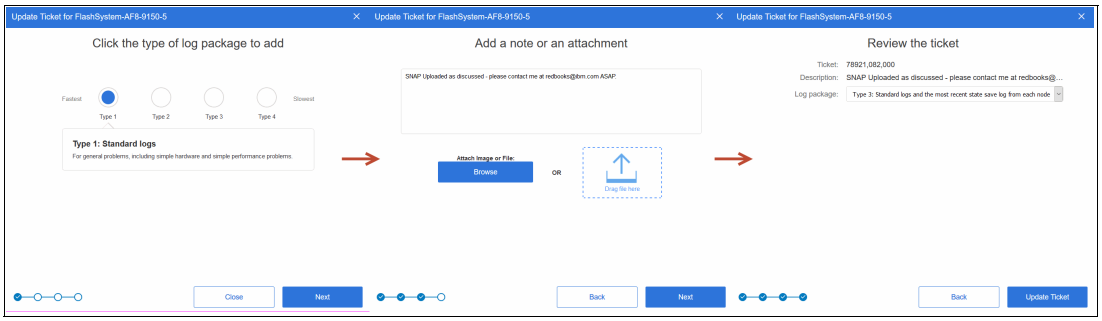


Figure 13-88 Adding a Log Package to the ticket

After clicking **Update Ticket**, a confirmation window opens, as shown in Figure 13-89, and the user can exit the wizard. IBM Storage Insights runs in the background to gather the logs and upload them to the ticket, which requires no other intervention from the user.

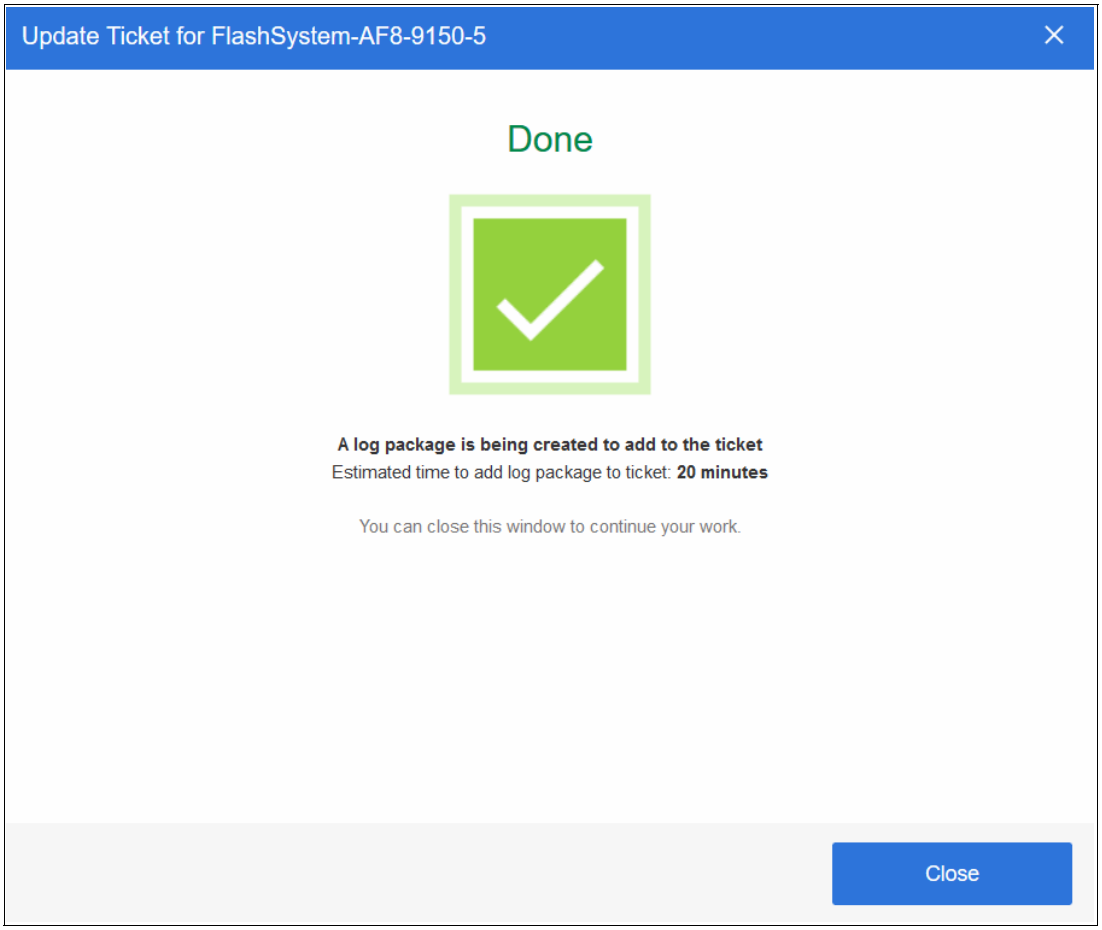


Figure 13-89 Confirmation of log upload

A

Performance data and statistics gathering

This appendix provides a brief overview of the performance analysis capabilities of the IBM SAN Volume Controller and IBM Spectrum Virtualize. It also describes a method that you can use to collect and process IBM Spectrum Virtualize performance statistics.

It is beyond the intended scope of this book to provide an in-depth understanding of performance statistics or to explain how to interpret them. For more information about the performance of the SAN Volume Controller, see *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

This appendix includes the following topics:

- "IBM SAN Volume Controller performance overview" on page 934
- "Performance monitoring" on page 936

© Copyright IBM Corp. 2021. All rights reserved.

IBM SAN Volume Controller performance overview

Storage virtualization with IBM Spectrum Virtualize provides many administrative benefits. In addition, it can provide an increase in performance for some workloads. The caching capability of the IBM Spectrum Virtualize and its ability to stripe volumes across multiple disk arrays can provide a performance improvement over what can otherwise be achieved when midrange disk subsystems are used.

To ensure that the performance levels of your system are maintained, monitor performance periodically to provide visibility into potential problems that exist or are developing so that they can be addressed in a timely manner.

Performance considerations

When you are designing the IBM Spectrum Virtualize infrastructure or maintaining an infrastructure, you must consider many factors in terms of their potential effect on performance. These factors include dissimilar workloads that are competing for the same resources, overloaded resources, insufficient available resources, poor performing resources, and similar performance constraints.

Remember the following high-level rules when you are designing your storage area network (SAN) and IBM Spectrum Virtualize layout:

- Host-to-SAN Volume Controller inter-switch link (ISL) oversubscription.

  This area is the most significant input/output (I/O) load across ISLs. The recommendation is to maintain a maximum of 7-to-1 oversubscription. A higher ratio is possible, but it tends to lead to I/O bottlenecks. This suggestion also assumes a core-edge design, where the hosts are on the edges and the IBM SAN Volume Controller is the core.

- Storage-to-SAN Volume Controller ISL oversubscription.

  This area is the second most significant I/O load across ISLs. The maximum oversubscription is 7-to-1. A higher ratio is not supported. Again, this suggestion assumes a multiple-switch SAN fabric design.

- Node-to-node ISL oversubscription.

  This area is the least significant load of the three possible oversubscription bottlenecks. In standard setups, this load can be ignored. Although this load is not entirely negligible, it does not contribute significantly to the ISL load. However, node-to-node ISL oversubscription is referenced here in relation to the Stretched cluster capability.

  When the system is running in this manner, the number of ISL links becomes more important. As with the storage-to-SAN Volume Controller ISL oversubscription, this load also requires a maximum of 7-to-1 oversubscription. Exercise caution and careful planning when you determine the number of ISLs to implement. If you need assistance, contact your IBM representative and request technical assistance.

- ISL trunking and port channeling.

  For the best performance and availability, use ISL trunking or port channeling. Independent ISL links can easily become overloaded and turn into performance bottlenecks. Bonded or trunked ISLs automatically share load and provide better redundancy in a failure.

- Number of paths per host multipath device.

  The maximum supported number of paths per multipath device that is visible on the host is eight (with HyperSwap, you can have up to 16 active paths). Although, most vendor multipathing software can support more paths, the IBM SAN Volume Controller expects a maximum of eight paths. In general, you see only an effect on performance from more paths than eight. Although the IBM Spectrum Virtualize system can work with more than eight paths, this design is technically unsupported.

- Do not intermix dissimilar array types or sizes.

  Although the IBM Spectrum Virtualize supports an intermix of differing storage within storage pools, it is best to always use the same array model, which is Redundant Array of Independent Disks (RAID) mode, RAID size (RAID 5 6+P+S does not mix well with RAID 6 14+2), and drive speeds.

Rules and guidelines are no substitution for monitoring performance. Monitoring performance can provide a validation that design expectations are met, and identify opportunities for improvement.

IBM Spectrum Virtualize performance perspectives

The IBM Spectrum Virtualize software was developed by the IBM Research Group. It is designed to run on IMB SAN Volume Controller, IBM Storwize products, and commodity hardware (mass-produced Intel-based processors with mass-produced expansion cards). It is also designed to provide distributed cache and a scalable cluster architecture.

Currently, the IBM SAN Volume Controller cluster is scalable up to eight nodes and these nodes can be swapped for newer hardware while online. This capability provides a great investment value because the nodes are relatively inexpensive and a node swap can be done online.

For more information about replacing nodes nondisruptively, see IBM Documentation.

For more information about setting up Fibre Channel (FC) port masking when upgrading from nodes 2145-CF8, 2145-CG8, or 2145-DH8 to 2145-SV1, 2145-SV2 and 2145-SA2, see this web page.

The performance is near linear when nodes are added into the cluster until performance eventually becomes limited by the attached components. Although virtualization provides significant flexibility in terms of the components that are used, it does not diminish the necessity of designing the system around the components so that it can deliver the level of performance that you want.

The key item for planning is your SAN layout. Switch vendors have slightly different planning requirements, but the goal is that you always want to maximize the bandwidth that is available to the IBM SAN Volume Controller ports. The IBM SAN Volume Controller is one of the few devices that can drive ports to their limits on average, so it is imperative that you put significant thought into planning the SAN layout.

Essentially, performance improvements are gained by spreading the workload across a greater number of back-end resources and by more caching. These capabilities are provided by the IBM SAN Volume Controller cluster. However, the performance of individual resources eventually becomes the limiting factor.

Performance monitoring

This section highlights several performance monitoring techniques.

Collecting performance statistics

IBM Spectrum Virtualize is constantly collecting performance statistics. The default frequency by which files are created is 15-minute intervals. The collection interval can be changed by running the **startstats** command.

The statistics files for volumes, managed disks (MDisks), nodes, and drives are saved at the end of the sampling interval. A maximum of 16 files (each) are stored before they are overlaid in a rotating log fashion. This design then provides statistics for the most recent 240-minute period if the default 15-minute sampling interval is used. IBM Spectrum Virtualize supports user-defined sampling intervals of 1 - 60 minutes.IBM Storage Insights requires and recommends interval of 5 minutes.

For each type of object (volumes, MDisks, nodes, and drives), a separate file with statistic data is created at the end of each sampling period and stored in `/dumps/iostats`.

Run the **startstats** command to start the collection of statistics, as shown in Example A-1.

Example A-1 The startstats command

```
IBM_2145:IBM Redbook SVC:superuser>startstats -interval 5
```

This command starts statistics collection and gathers data at 5-minute intervals.

To verify the statistics status and collection interval, display the system properties, as shown in Example A-2.

Example A-2 Statistics collection status and frequency

```
IBM_2145:IBM Redbook SVC:superuser>lssystem
id 0000020320219A08
name IBM Redbook SVC
statistics_status on
statistics_frequency 5
-- The output has been shortened for easier reading. --
```

Starting with V8.1, it is not possible to stop statistics collection by running the **stopstats** command.

Collection intervals: Although more frequent collection intervals provide a more detailed view of what happens within IBM Spectrum Virtualize and IBM SAN Volume Controller, they shorten the amount of time that the historical data is available on the IBM Spectrum Virtualize system. For example, rather than a 240-minute period of data with the default 15-minute interval, if you adjust to 2-minute intervals, you have a 32-minute data period instead.

Statistics are collected per node. The sampling of the internal performance counters is coordinated across the cluster so that when a sample is taken, all nodes sample their internal counters at the same time. It is important to collect all files from all nodes for a complete analysis. Tools, such as IBM Spectrum Control and Spectrum Insight® Pro, perform this intensive data collection for you.

Statistics file naming

The statistics files that are generated are written to the `/dumps/iostats/` directory. The file name is in the following formats:

- `Nm_stats_<node_frontpanel_id>_<date>_<time>` for MDisks statistics
- `Nv_stats_<node_frontpanel_id>_<date>_<time>` for Volumes statistics
- `Nn_stats_<node_frontpanel_id>_<date>_<time>` for node statistics
- `Nd_stats_<node_frontpanel_id>_<date>_<time>` for drives statistics

The *node_frontpanel_id* is the pane name of the node on which the statistics were collected. The date is in the form *<yymmdd>* and the time is in the form *<hhmmss>*. The following example shows an MDisk statistics file name:

```
Nv_stats_78HYLDO_201026_112425
```

Example A-3 shows typical MDisk, volume, node, and disk drive statistics file names.

Example A-3 File names of per node statistics

```
IBM_2145:IBM Redbook SVC:superuser>lsdumps -prefix /dumps/iostats
id  filename
0   Nm_stats_78HYLDO_201026_112425
1   Nv_stats_78HYLDO_201026_112425
2   Nd_stats_78HYLDO_201026_112425
3   Nn_stats_78HYLDO_201026_112425
4   Nm_stats_78HYLDO_201026_113927
5   Nd_stats_78HYLDO_201026_113927
6   Nm_stats_78HXWYO_201026_113927
7   Nv_stats_78HYLDO_201026_113927
8   Nv_stats_78HXWYO_201026_113927
9   Nd_stats_78HXWYO_201026_113927
...
126 Nn_stats_78HYLDO_201026_152209
127 Nd_stats_78HXWYO_201026_152209
128 Nn_stats_78HXWYO_201026_152209
129 Nd_stats_78HYLDO_201026_152209
130 Nm_stats_78HXWYO_201026_152209
131 Nm_stats_78HYLDO_201026_152209
IBM_2145:IBM Redbook SVC:superuser>
```

Note: For more information about statistics files name convention, see IBM Documentation.

Tip: The performance statistics files can be copied from the IBM SAN Volume Controller nodes to a local drive on your workstation by using `pscp.exe` (included with PuTTY) from an MS-DOS command prompt, as shown in the following example:

```
C:\Program Files\PuTTY>pscp -unsafe -load IBM Redbook SVC
superuser@9.71.42.141:/dumps/iostats/* c:\statsfiles
```

Use the **-load** parameter to specify the session that is defined in PuTTY.

Specify the **-unsafe** parameter when you use wildcards.

PuTTY is available for download at this website.

Real-time performance monitoring

IBM SAN Volume Controller supports real-time performance monitoring. Real-time performance statistics provide short-term status information for the IMB SAN Volume Controller. The statistics are shown as graphs in the management GUI, or can be viewed from the CLI.

With system-level statistics, you can quickly view the CPU usage and the bandwidth of volumes, interfaces, and MDisks. Each graph displays the current bandwidth in megabytes per second (MBps) or input/output operations per second (IOPS), and a view of bandwidth over time.

Each node collects various performance statistics, mostly at 5-second intervals, and the statistics that are available from the config node in a clustered environment. This information can help you determine the performance effect of a specific node. As with system statistics, node statistics help you to evaluate whether the node is operating within normal performance metrics.

Real-time performance monitoring gathers the following system-level performance statistics:

- CPU usage
- Port usage and I/O rates
- Volume and MDisk I/O rates
- Bandwidth
- Latency

Real-time statistics are not a configurable option and cannot be disabled.

Real-time performance monitoring with the CLI

The **lsnodestats** and **lssystemstats** commands are available for monitoring the statistics through the CLI.

The **lsnodestats** command provides performance statistics for the nodes that are part of a clustered system, as shown in Example A-4. This output is truncated and shows only part of the available statistics. You can also specify a node name in the command to limit the output for a specific node.

Example A-4 The lsnodestats command output

```
IBM_2145:IBM Redbook SVC:superuser>lsnodestats
node_id node_name stat_name          stat_current stat_peak stat_peak_time
1       node 1    compression_cpu_pc 0            0         201026152448
1       node 1    cpu_pc             12           12        201026152448
1       node 1    fc_mb              0            0         201026152448
1       node 1    fc_io              1196         1220      201026152302
1       node 1    sas_mb             0            0         201026152448
1       node 1    sas_io             0            0         201026152448
1       node 1    iscsi_mb           0            0         201026152448
1       node 1    iscsi_io           0            0         201026152448
1       node 1    write_cache_pc     0            0         201026152448
1       node 1    total_cache_pc     1            1         201026152448
1       node 1    vdisk_mb           0            0         201026152448
1       node 1    vdisk_io           0            0         201026152448
1       node 1    vdisk_ms           0            0         201026152448
1       node 1    mdisk_mb           0            0         201026152448
1       node 1    mdisk_io           0            0         201026152448
1       node 1    mdisk_ms           0            0         201026152448
1       node 1    drive_mb           0            0         201026152448
1       node 1    drive_io           0            0         201026152448
1       node 1    drive_ms           0            0         201026152448
1       node 1    vdisk_r_mb         0            0         201026152448
1       node 1    vdisk_r_io         0            0         201026152448
1       node 1    vdisk_r_ms         0            0         201026152448
1       node 1    vdisk_w_mb         0            0         201026152448
1       node 1    vdisk_w_io         0            0         201026152448
1       node 1    vdisk_w_ms         0            0         201026152448
1       node 1    mdisk_r_mb         0            0         201026152448
1       node 1    mdisk_r_io         0            0         201026152448
1       node 1    mdisk_r_ms         0            0         201026152448
1       node 1    mdisk_w_mb         0            0         201026152448
1       node 1    mdisk_w_io         0            0         201026152448
1       node 1    mdisk_w_ms         0            0         201026152448
1       node 1    drive_r_mb         0            0         201026152448
1       node 1    drive_r_io         0            0         201026152448
1       node 1    drive_r_ms         0            0         201026152448
1       node 1    drive_w_mb         0            0         201026152448
1       node 1    drive_w_io         0            0         201026152448
1       node 1    drive_w_ms         0            0         201026152448
1       node 1    iplink_mb          0            0         201026152448
1       node 1    iplink_io          0            0         201026152448
1       node 1    iplink_comp_mb     0            0         201026152448
1       node 1    cloud_up_mb        0            0         201026152448
1       node 1    cloud_up_ms        0            0         201026152448
1       node 1    cloud_down_mb      0            0         201026152448
1       node 1    cloud_down_ms      0            0         201026152448
1       node 1    iser_mb            0            0         201026152448
1       node 1    iser_io            0            0         201026152448
2       node 2    compression_cpu_pc 0            0         201026152445
```

```
2       node 2    cpu_pc             16          16        201026152445
2       node 2    fc_mb              0           0         201026152445
2       node 2    fc_io              1147        1265      201026152300
2       node 2    sas_mb             0           0         201026152445
2       node 2    sas_io             0           0         201026152445
2       node 2    iscsi_mb           0           0         201026152445
2       node 2    iscsi_io           0           0         201026152445
2       node 2    write_cache_pc     0           0         201026152445
2       node 2    total_cache_pc     0           0         201026152445
2       node 2    vdisk_mb           0           0         201026152445
2       node 2    vdisk_io           0           0         201026152445
2       node 2    vdisk_ms           0           0         201026152445
2       node 2    mdisk_mb           0           0         201026152445
2       node 2    mdisk_io           0           0         201026152445
2       node 2    mdisk_ms           0           0         201026152445
2       node 2    drive_mb           0           0         201026152445
2       node 2    drive_io           0           0         201026152445
2       node 2    drive_ms           0           0         201026152445
2       node 2    vdisk_r_mb         0           0         201026152445
2       node 2    vdisk_r_io         0           0         201026152445
2       node 2    vdisk_r_ms         0           0         201026152445
2       node 2    vdisk_w_mb         0           0         201026152445
2       node 2    vdisk_w_io         0           0         201026152445
2       node 2    vdisk_w_ms         0           0         201026152445
2       node 2    mdisk_r_mb         0           0         201026152445
2       node 2    mdisk_r_io         0           0         201026152445
2       node 2    mdisk_r_ms         0           0         201026152445
2       node 2    mdisk_w_mb         0           0         201026152445
2       node 2    mdisk_w_io         0           0         201026152445
2       node 2    mdisk_w_ms         0           0         201026152445
2       node 2    drive_r_mb         0           0         201026152445
2       node 2    drive_r_io         0           0         201026152445
2       node 2    drive_r_ms         0           0         201026152445
2       node 2    drive_w_mb         0           0         201026152445
2       node 2    drive_w_io         0           0         201026152445
2       node 2    drive_w_ms         0           0         201026152445
2       node 2    iplink_mb          0           0         201026152445
2       node 2    iplink_io          0           0         201026152445
2       node 2    iplink_comp_mb     0           0         201026152445
2       node 2    cloud_up_mb        0           0         201026152445
2       node 2    cloud_up_ms        0           0         201026152445
2       node 2    cloud_down_mb      0           0         201026152445
2       node 2    cloud_down_ms      0           0         201026152445
2       node 2    iser_mb            0           0         201026152445
2       node 2    iser_io            0           0         201026152445
IBM_2145:IBM Redbook SVC:superuser>
```

Example A-4 on page 939 shows statistics for the two nodes members of `Cluster_10.155.19.123`. For each node, the following columns are displayed:

- `stat_name`: The name of the statistic field
- `stat_current`: The current value of the statistic field
- `stat_peak`: The peak value of the statistic field in the last 5 minutes
- `stat_peak_time`: The time that the peak occurred

The **lsnodestats** command can also be used with a node name or ID as an argument. For example, you run the **lsnodestats node1** command to display the statistics of node with name node1 only.

The **lssystemstats** command lists the same set of statistics that is listed by running the **lsnodestats** command, but representing all nodes in the cluster. The values for these statistics are calculated from the node statistics values in the following way:

- Bandwidth: Sum of bandwidth of all nodes.
- Latency: Average latency for the cluster, which is calculated by using data from the whole cluster, not an average of the single node values.
- IOPS: Total IOPS of all nodes.
- CPU percentage: Average CPU percentage of all nodes.

Example A-5 shows the resulting output of the **lssystemstats** command.

Example A-5 The lssystemstats command output

```
IBM_2145:IBM Redbook SVC:superuser>lssystemstats
stat_name          stat_current stat_peak stat_peak_time
compression_cpu_pc 0            0         201026152509
cpu_pc             14           14        201026152509
fc_mb              0            0         201026152509
fc_io              2423         2457      201026152258
sas_mb             0            0         201026152509
sas_io             0            0         201026152509
iscsi_mb           0            0         201026152509
iscsi_io           0            0         201026152509
write_cache_pc     0            0         201026152509
total_cache_pc     1            1         201026152509
vdisk_mb           0            0         201026152509
vdisk_io           0            0         201026152509
vdisk_ms           0            0         201026152509
mdisk_mb           0            0         201026152509
mdisk_io           0            0         201026152509
mdisk_ms           0            0         201026152509
drive_mb           0            0         201026152509
drive_io           0            0         201026152509
drive_ms           0            0         201026152509
vdisk_r_mb         0            0         201026152509
vdisk_r_io         0            0         201026152509
vdisk_r_ms         0            0         201026152509
vdisk_w_mb         0            0         201026152509
vdisk_w_io         0            0         201026152509
vdisk_w_ms         0            0         201026152509
mdisk_r_mb         0            0         201026152509
mdisk_r_io         0            0         201026152509
mdisk_r_ms         0            0         201026152509
mdisk_w_mb         0            0         201026152509
mdisk_w_io         0            0         201026152509
mdisk_w_ms         0            0         201026152509
drive_r_mb         0            0         201026152509
drive_r_io         0            0         201026152509
drive_r_ms         0            0         201026152509
drive_w_mb         0            0         201026152509
drive_w_io         0            0         201026152509
```

```
drive_w_ms          0             0          201026152509
iplink_mb           0             0          201026152509
iplink_io           0             0          201026152509
iplink_comp_mb      0             0          201026152509
cloud_up_mb         0             0          201026152509
cloud_up_ms         0             0          201026152509
cloud_down_mb       0             0          201026152509
cloud_down_ms       0             0          201026152509
iser_mb             0             0          201026152509
iser_io             0             0          201026152509
IBM_2145:IBM Redbook SVC:superuser>
```

Table A-1 lists the different counters that are presented by the **lssystemstats** and **lsnodestats** commands.

Table A-1 Counters in lssystemstats and lsnodestats

Value	Description
compression_cpu_pc	Displays the percentage of allocated CPU capacity that is used for compression.
cpu_pc	Displays the percentage of allocated CPU capacity that is used for the system.
fc_mb	Displays the total number of MBps for FC traffic on the system. This value includes host I/O and any bandwidth that is used for communication within the system.
fc_io	Displays the total IOPS for FC traffic on the system. This value includes host I/O and any bandwidth that is used for communication within the system.
sas_mb	Displays the total number of MBps for serial-attached Small Computer System Interface (SCSI) (SAS) traffic on the system. This value includes host I/O and bandwidth that is used for background RAID activity.
sas_io	Displays the total IOPS for SAS traffic on the system. This value includes host I/O and bandwidth that is used for background RAID activity.
iscsi_mb	Displays the total number of MBps for internet Small Computer Systems Interface (iSCSI) traffic on the system.
iscsi_io	Displays the total IOPS for iSCSI traffic on the system.
write_cache_pc	Displays the percentage of the write cache usage for the node.
total_cache_pc	Displays the total percentage for both the write and read cache usage for the node.
vdisk_mb	Displays the average number of MBps for read and write operations to volumes during the sample period.
vdisk_io	Displays the average number of IOPS for read and write operations to volumes during the sample period.
vdisk_ms	Displays the average amount of time in milliseconds that the system takes to respond to read and write requests to volumes over the sample period.
mdisk_mb	Displays the average number of MBps for read and write operations to MDisks during the sample period.

Value	Description
mdisk_io	Displays the average number of IOPS for read and write operations to MDisks during the sample period.
mdisk_ms	Displays the average amount of time in milliseconds that the system takes to respond to read and write requests to MDisks over the sample period.
drive_mb	Displays the average number of MBps for read and write operations to drives during the sample period.
drive_io	Displays the average number of IOPS for read and write operations to drives during the sample period.
drive_ms	Displays the average amount of time in milliseconds that the system takes to respond to read and write requests to drives over the sample period.
vdisk_w_mb	Displays the average number of MBps for read and write operations to volumes during the sample period.
vdisk_w_io	Displays the average number of IOPS for write operations to volumes during the sample period.
vdisk_w_ms	Displays the average amount of time in milliseconds that the system takes to respond to write requests to volumes over the sample period.
mdisk_w_mb	Displays the average number of MBps for write operations to MDisks during the sample period.
mdisk_w_io	Displays the average number of IOPS for write operations to MDisks during the sample period.
mdisk_w_ms	Displays the average amount of time in milliseconds that the system takes to respond to write requests to MDisks over the sample period.
drive_w_mb	Displays the average number of MBps for write operations to drives during the sample period.
drive_w_io	Displays the average number of IOPS for write operations to drives during the sample period.
drive_w_ms	Displays the average amount of time in milliseconds that the system takes to respond write requests to drives over the sample period.
vdisk_r_mb	Displays the average number of MBps for read operations to volumes during the sample period.
vdisk_r_io	Displays the average number of IOPS for read operations to volumes during the sample period.
vdisk_r_ms	Displays the average amount of time in milliseconds that the system takes to respond to read requests to volumes over the sample period.
mdisk_r_mb	Displays the average number of MBps for read operations to MDisks during the sample period.
mdisk_r_io	Displays the average number of IOPS for read operations to MDisks during the sample period.
mdisk_r_ms	Displays the average amount of time in milliseconds that the system takes to respond to read requests to MDisks over the sample period.

Value	Description
drive_r_mb	Displays the average number of MBps for read operations to drives during the sample period.
drive_r_io	Displays the average number of IOPS for read operations to drives during the sample period.
drive_r_ms	Displays the average amount of time in milliseconds that the system takes to respond to read requests to drives over the sample period.
iplink_mb	The total number of MBps for IP replication traffic on the system. This value does not include iSCSI host I/O operations.
iplink_comp_mb	Displays the average number of compressed MBps over the IP replication link during the sample period.
iplink_io	The total IOPS for IP partnership traffic on the system. This value does not include iSCSI host I/O operations.
cloud_up_mb	Displays the average number of Mbps for upload operations to a cloud account during the sample period.
cloud_up_ms	Displays the average amount of time (in milliseconds) it takes for the system to respond to upload requests to a cloud account during the sample period.
cloud_down_mb	Displays the average number of Mbps for download operations to a cloud account during the sample period.
cloud_down_ms	Displays the average amount of time (in milliseconds) that it takes for the system to respond to download requests to a cloud account during the sample period.
iser_mb	Displays the total number of MBps for iSCSI Extensions for RDMA (iSER) traffic on the system.
iser_io	Displays the total IOPS for iSER traffic on the system.

Real-time performance statistics monitoring by using the GUI

The IBM Spectrum Virtualize dashboard gives you performance at a glance by displaying important information about the system. You can see the entire cluster (the system) performance by selecting the information between Bandwidth, Response Time, IOPS, or CPU utilization. You can also display a Node Comparison by selecting the same information as for the cluster, and then switching the button, as shown in Figure A-1.

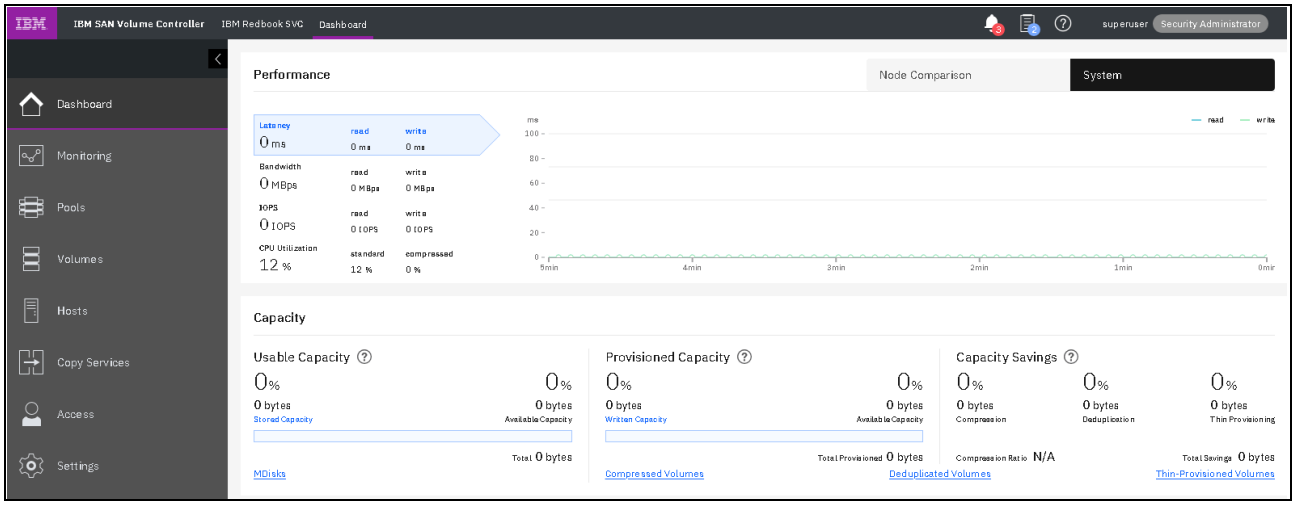


Figure A-1 IBM Spectrum Virtualize Dashboard displaying System performance overview.

Figure A-2 shows the display after switching the button.

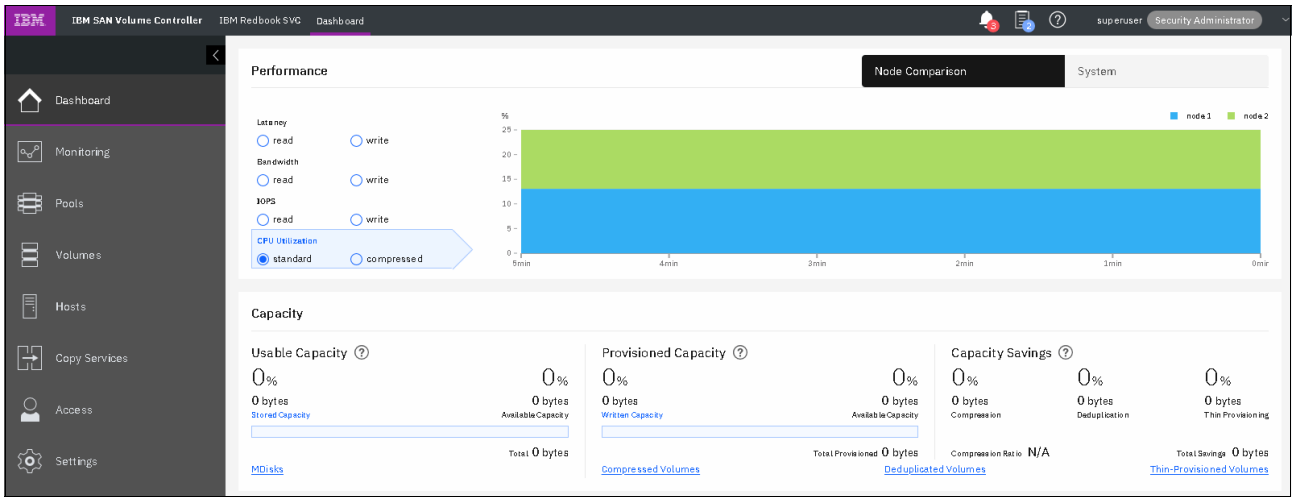


Figure A-2 IBM Spectrum Virtualize Dashboard displaying Nodes performance overview

You can also use real-time statistics to monitor CPU utilization, volume, interface, and MDisk bandwidth of your system and nodes. Each graph represents 5 minutes of collected statistics and provides a means of assessing the overall performance of your system.

The real-time statistics are available from the IBM Spectrum Virtualize GUI. Click **Monitoring** → **Performance** (as shown in Figure A-3) to open the Performance Monitoring window.

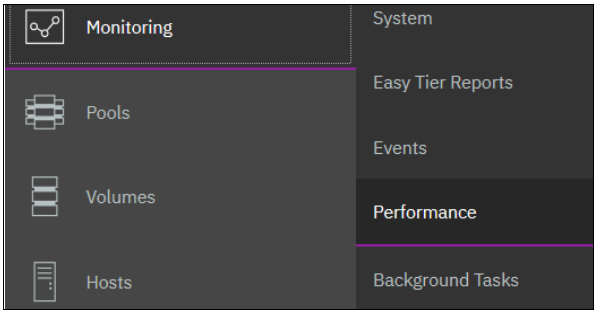


Figure A-3 Selecting Performance in the Monitoring menu

As shown in Figure A-4 on page 947, the Performance monitoring pane is divided into the following sections that provide usage views for the following resources:

- CPU Utilization: The CPU utilization graph shows the current percentage of CPU usage and peaks in utilization. It can also display compression CPU usage for systems with compressed volumes.
- Volumes: Shows four metrics for the overall volume utilization graphics:
  - Read
  - Write
  - Read latency
  - Write latency
- Interfaces: The Interfaces graph displays data points for FC, iSCSI, SAS, and IP Remote Copy (RC) interfaces. You can use this information to help determine connectivity issues that might affect performance:
  - FC
  - iSCSI
  - SAS
  - IP Remote Copy
- MDisks: Shows four metrics for the overall MDisks graphics:
  - Read
  - Write
  - Read latency
  - Write latency

  You can use these metrics to help determine the overall performance health of the volumes and MDisks on your system. Consistent unexpected results can indicate errors in configuration, system faults, or connectivity issues.

The system's performance is always visible in the bottom of the IBM Spectrum Virtualize window, as shown in Figure A-4 on page 947.

Note: The indicated values in the graphics are averaged on a 5-seconds based sample.

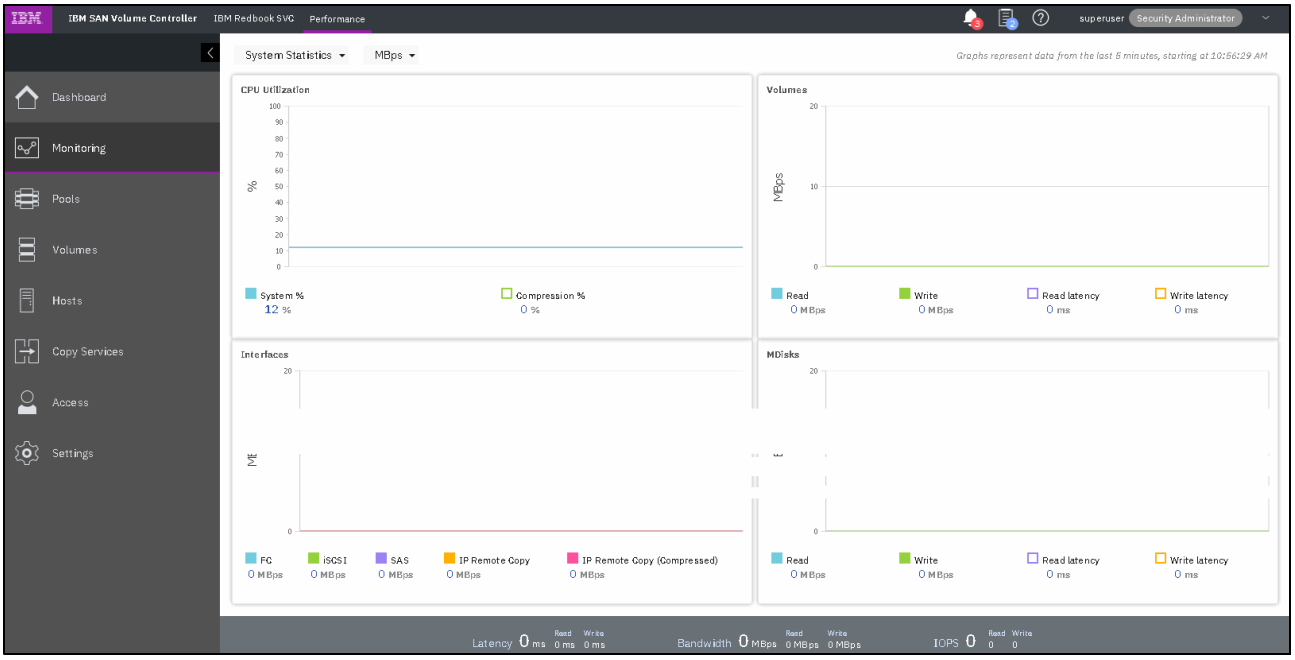


Figure A-4 IBM Spectrum Virtualize Performance window

You can also view the performance statistics for each of the available nodes of the system, as shown in Figure A-5.

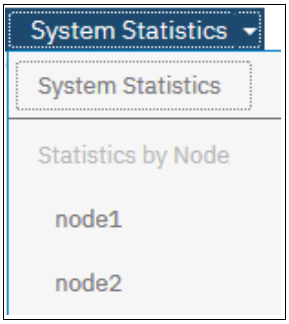


Figure A-5 View statistics per node or for the entire system

You can also change the metric between MBps or IOPS, as shown in Figure A-6.

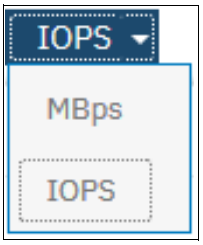


Figure A-6 View performance metrics by MBps or IOPS

On any of these views, you can select any point with your cursor to see the value and when it occurred. When you place your cursor over the timeline, it becomes a dotted line with the various values gathered, as shown in Figure A-7.

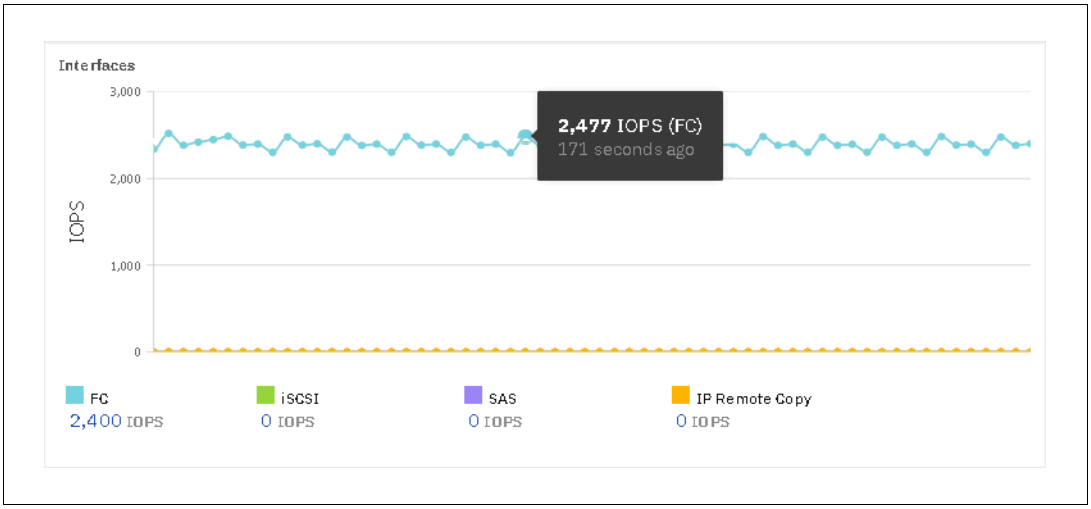


Figure A-7 Viewing performance with details

For each of the resources, various metrics are available and you can select which ones are shown. For example, as shown in Figure A-8, from the four available metrics for the MDisks view (Read, Write, Read latency, and Write latency), only Write and Write latency IOPS are selected.

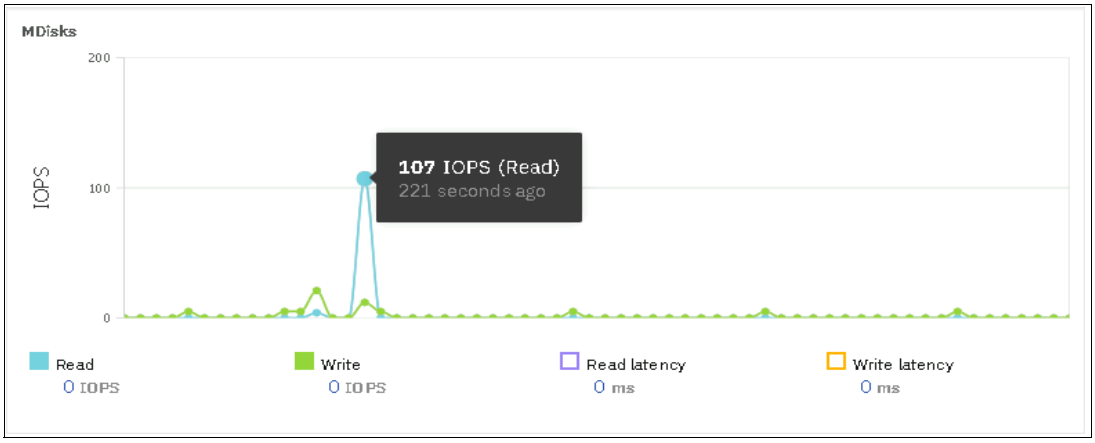


Figure A-8 Displaying performance counters

Performance data collection and IBM Spectrum Control

Although you can obtain performance statistics in standard `.xml` files, the use of `.xml` files is a less practical and more complicated method to analyze the IBM Spectrum Virtualize performance statistics. IBM Spectrum Control is the supported IBM tool to collect and analyze SAN Volume Controller performance statistics.

IBM Spectrum Control is installed separately on a dedicated system, and is not part of the IBM Spectrum Virtualize bundle.

For more information about the use of IBM Spectrum Control to monitor your storage subsystem, see this web page.

As an alternative to IBM Spectrum Control, a cloud-based tool is available that is called IBM Storage Insights. This tool provides a single dashboard that gives you a clear view of all your IBM block storage and shows performance and capacity information. You do not have to install this tool in your environment because it is a cloud-based solution. Only an agent is required to collect the data of the storage devices.

For more information about IBM Storage Insights, see this web page.

B

Command-line interface setup

This appendix describes access configuration to the command-line interface (CLI) by using the local Secure Shell (SSH) authentication method.

© Copyright IBM Corp. 2021. All rights reserved.

CLI setup

The IBM Spectrum 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 Spectrum 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 Spectrum Virtualize GUI, authentication is performed by supplying a user name and password. The CLI uses SSH to connect from a host to the IBM Spectrum Virtualize system. A private and a public key pair or user name and password is necessary.

Using SSH keys with a passphrase is more secure than a login with a user name 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 Spectrum 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 Spectrum 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 Spectrum Virtualize systems.

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 B-1.

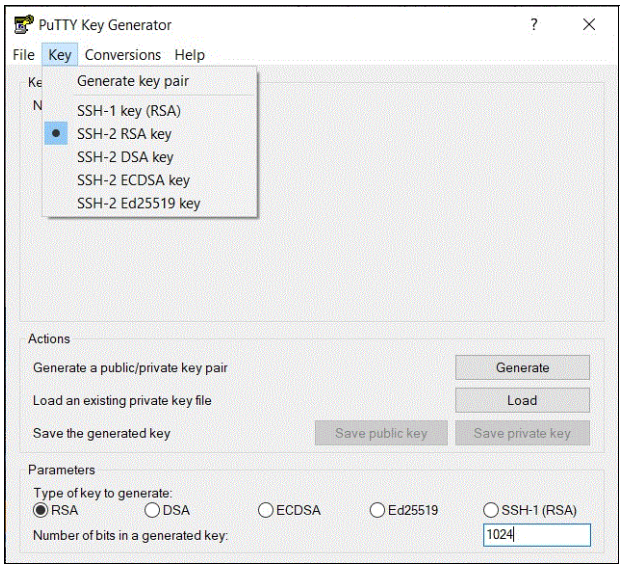


Figure B-1 PuTTY key generator

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 B-2).

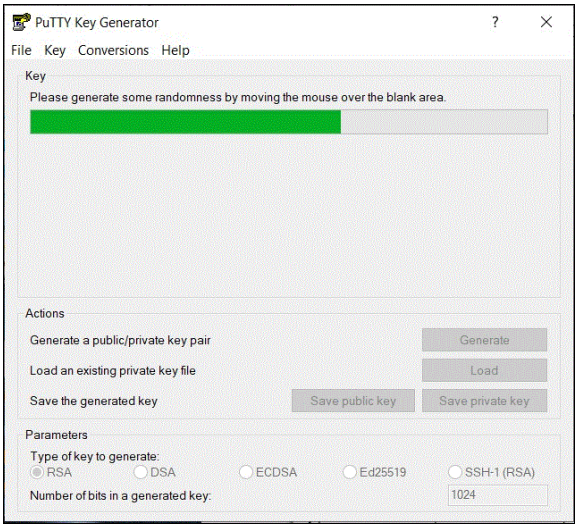


Figure B-2 Generating keys

Generating 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 that are 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 B-3). Click **Yes** to save the private key without a passphrase.

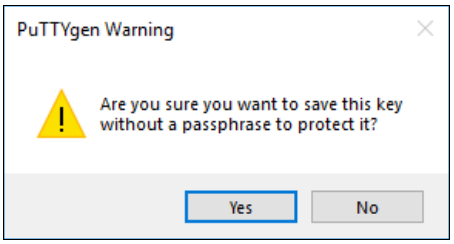


Figure B-3 Confirming the security warning

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 format is a proprietary format for PuTTY and the keys are not interchangeable with OpenSSH clients. A utility is available 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 B-4.

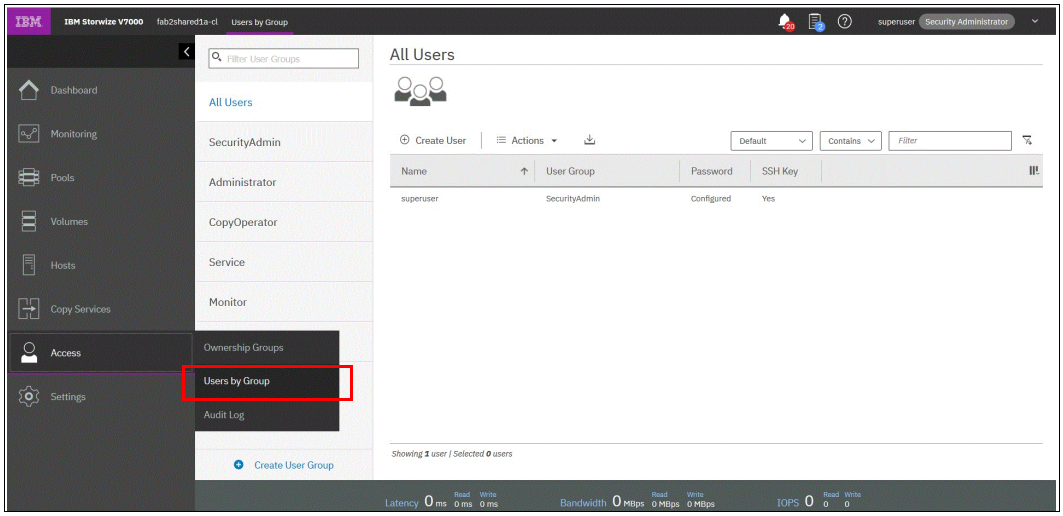


Figure B-4 Opening the user section

2. Right-click the user name for which you want to upload the key and click **Properties** (see Figure B-5).

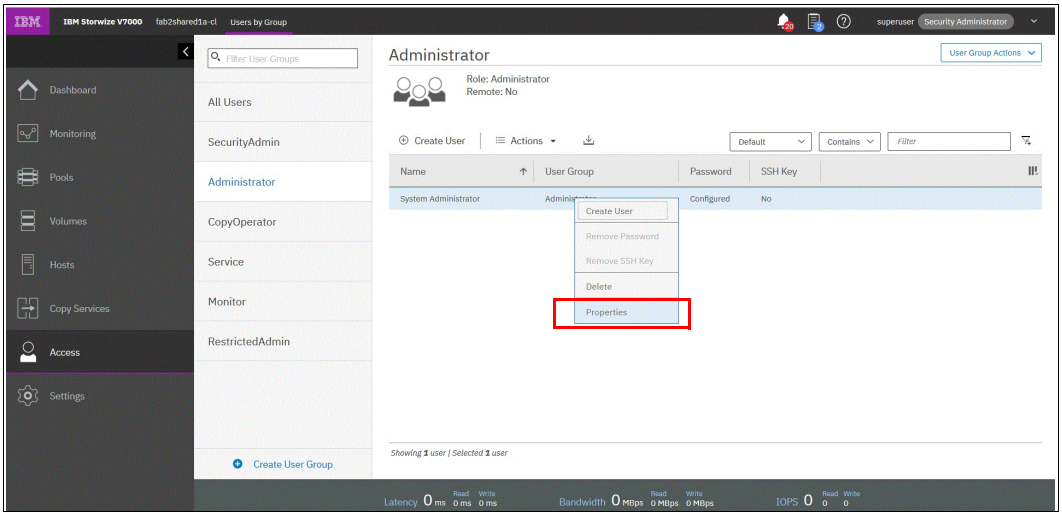


Figure B-5 User properties

3. To upload the public key, click **Browse**. Then, 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 B-6.

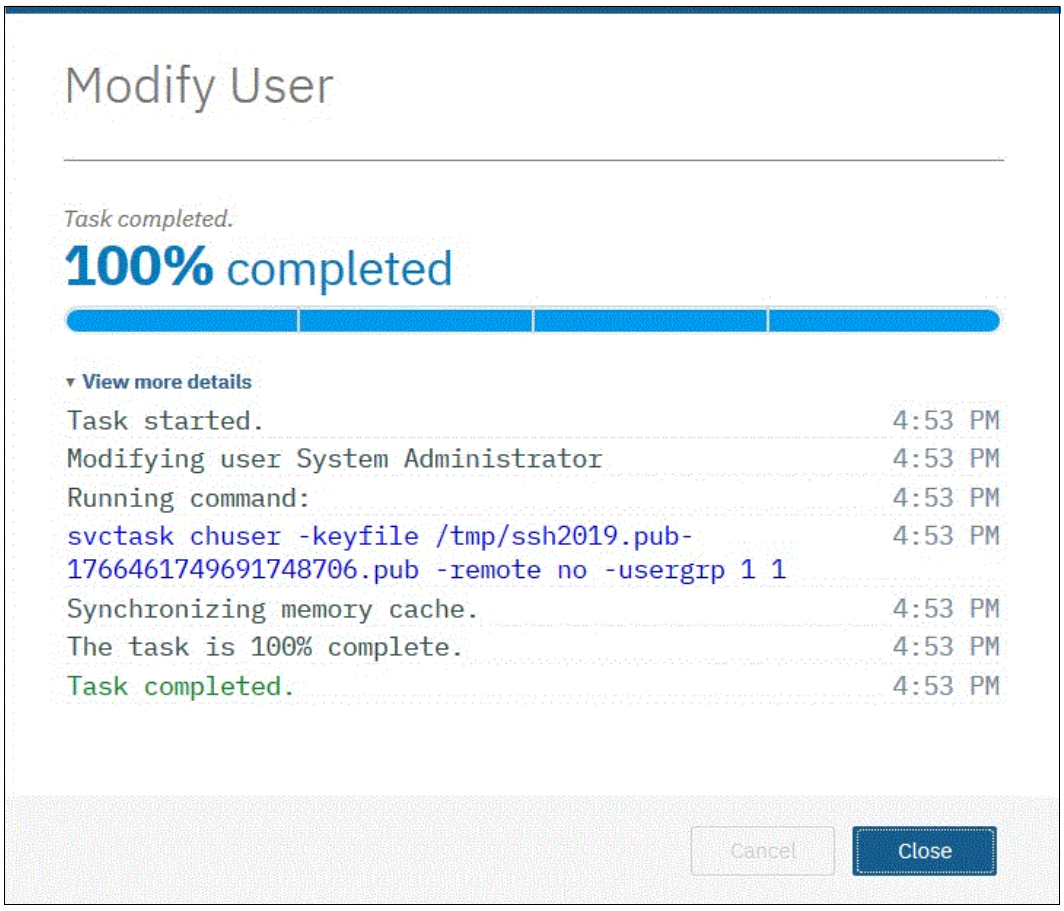


Figure B-6 Confirmation of SSH Key upload

5. Check in the GUI to ensure that the SSH key is imported successfully (see Figure B-7).

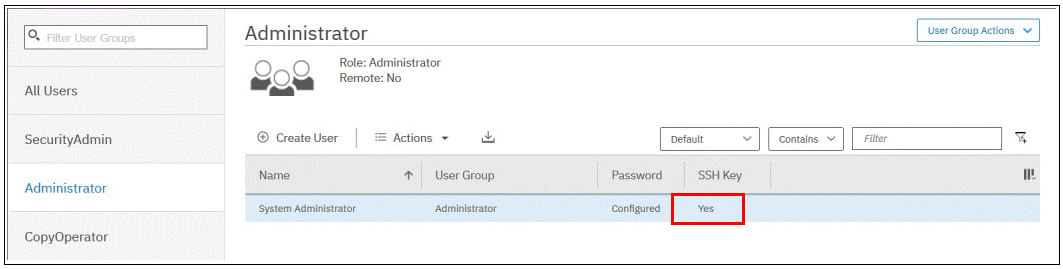


Figure B-7 Key successfully imported

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 B-8).

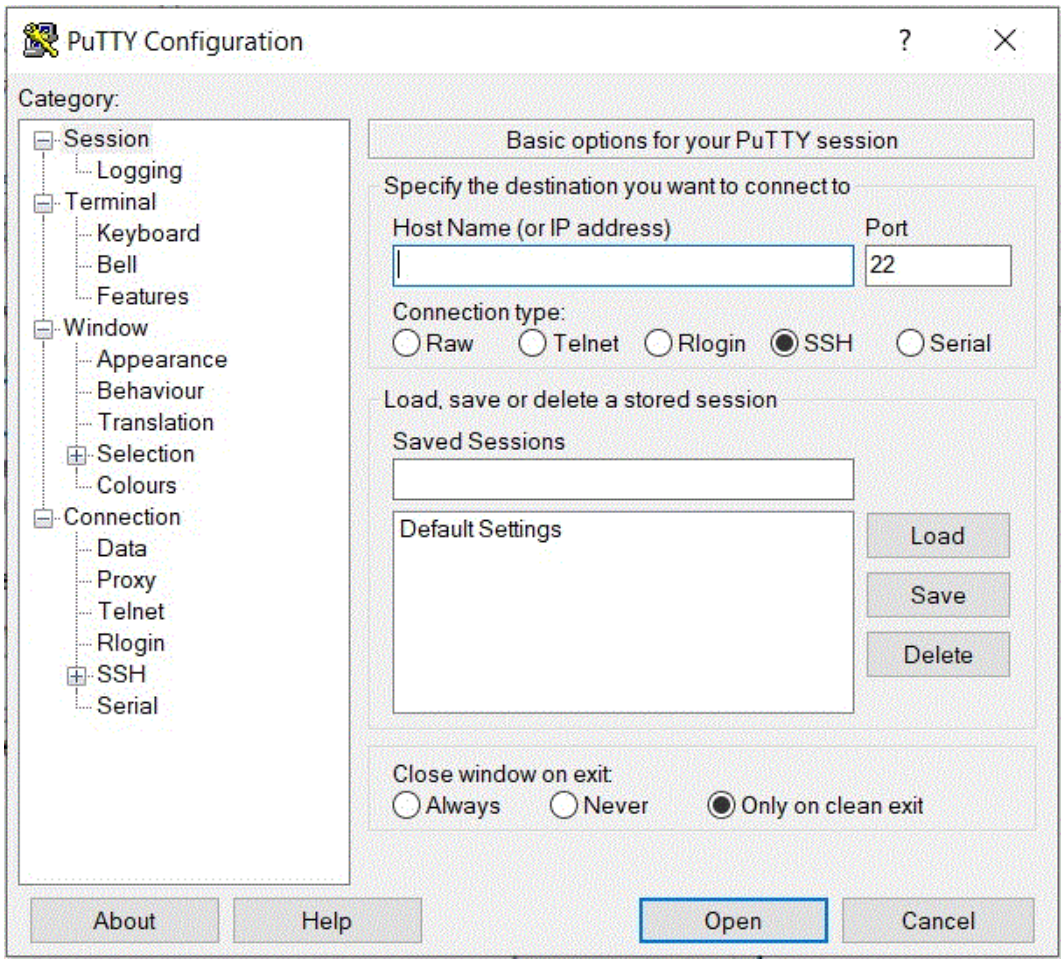


Figure B-8 PuTTY Configuration

2. In the upper right pane, select **SSH** as the connection type. In the “Close window on exit” section, select **Only on clean exit** (see Figure B-9 on page 958), which ensures that if any connection errors occur that they are displayed on the user’s window.

3. In the Category pane, on the left side of the PuTTY Configuration window, select **Connection** → **Data**, as shown on Figure B-9. In the “Auto-login username” field, enter the IBM Spectrum Virtualize user ID that was used when uploading the public key. The `admin` account was used in the example that is shown in Figure B-5 on page 956.

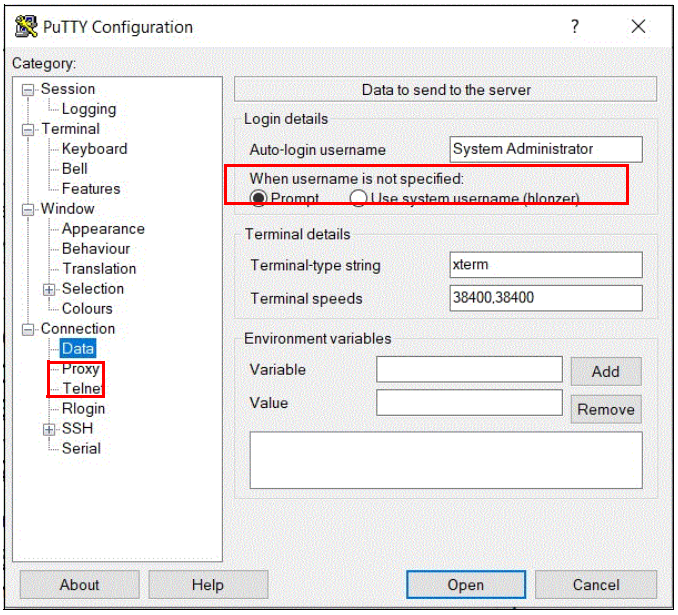


Figure B-9 PuTTY Auto-login user name

4. In the Category pane, on the left side of the PuTTY Configuration window (see Figure B-10), select **Connection** → **SSH** to open the PuTTY SSH Configuration window. In the SSH protocol version section, select **2**.

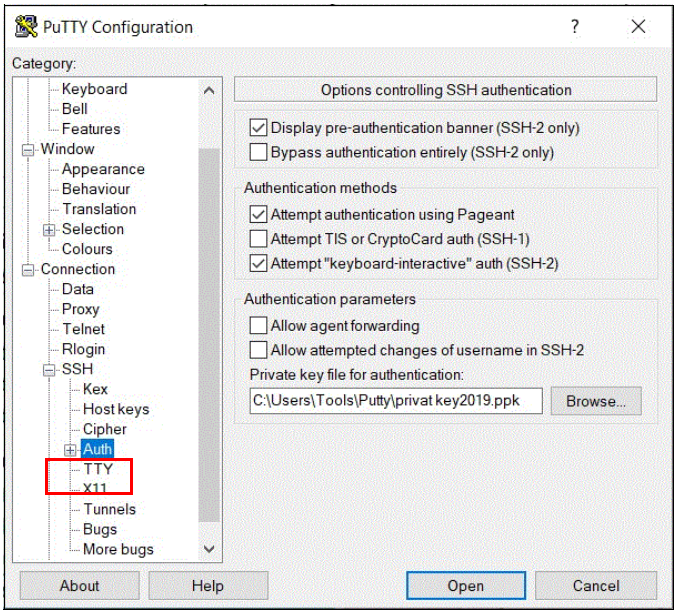


Figure B-10 SSH protocol Version 2

5. In the Category pane 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 B-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).

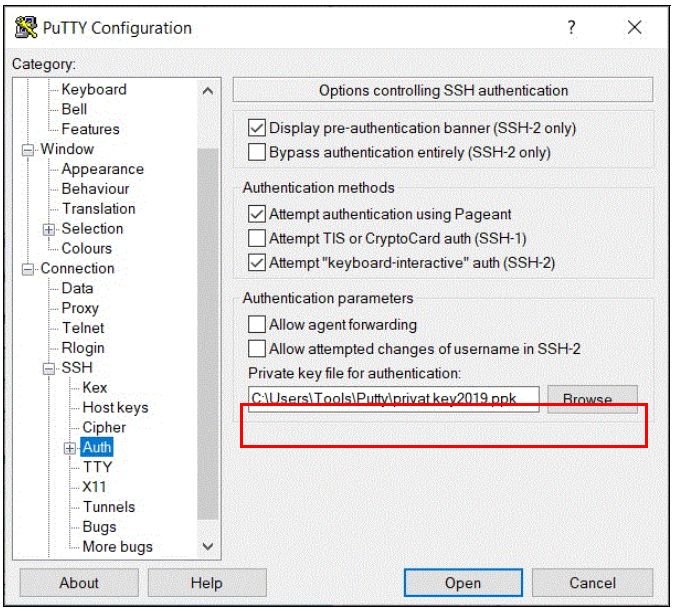


Figure B-11 SSH authentication

7. In the Category pane, 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 B-12):
   - Host Name (or Internet Protocol (IP) address): Specify the host name or system IP address of the IBM Spectrum Virtualize system.
   - Saved Sessions: Enter a session name.
9. Click **Save** to save the new session (Figure B-12).

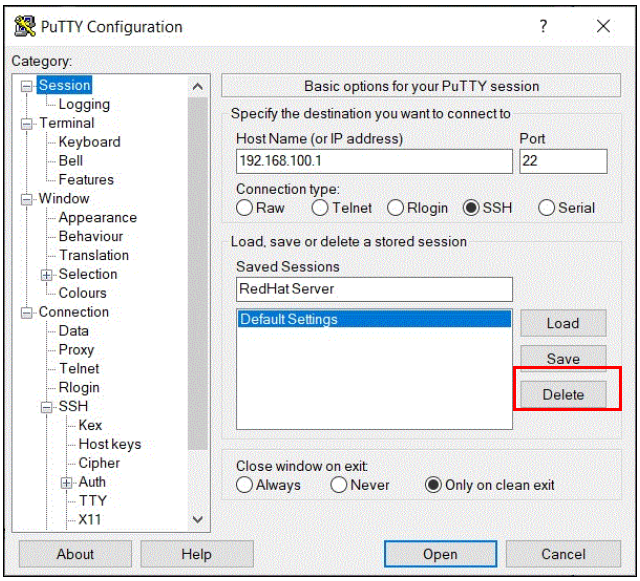


Figure B-12 Session information

10. Select the new session and click **Open** to connect to the IBM Spectrum Virtualize system, as shown in Figure B-13.

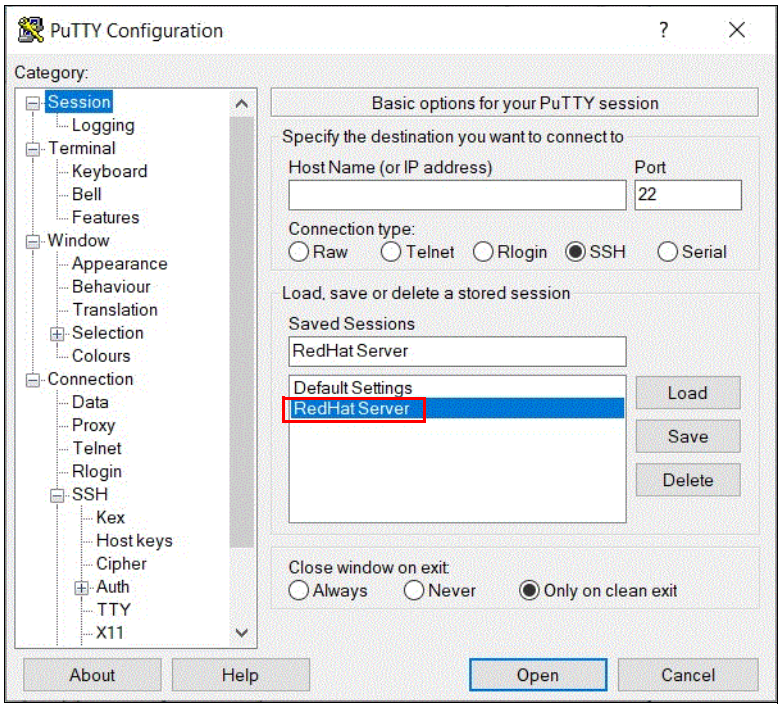


Figure B-13 Connecting to a system

11. If a PuTTY Security Alert opens as shown in Figure B-14, confirm it by clicking **Yes**.

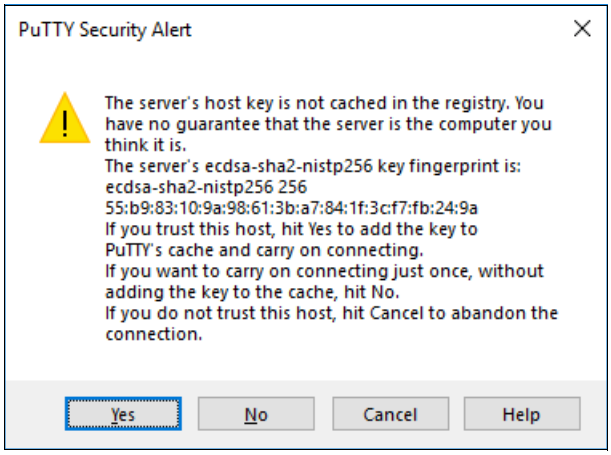


Figure B-14 Confirming the security alert

As shown in Figure B-15, PuTTY now connects to the system automatically by using the user ID that was specified earlier, without prompting for password.

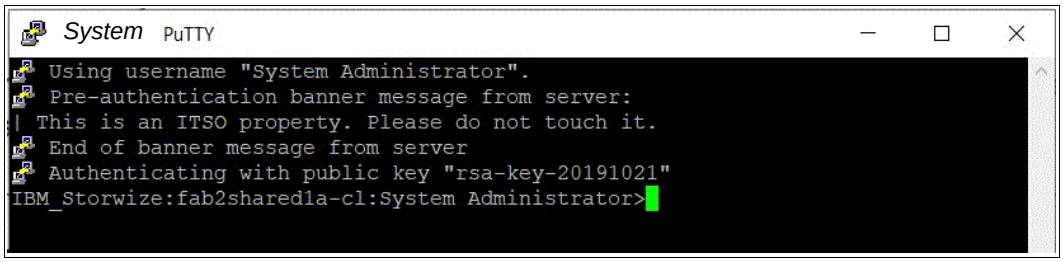


Figure B-15 PuTTY login

The CLI is now configured for IBM Spectrum Virtualize system administration.

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. It is installed by default on most of these types of operating systems. 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 IBM Spectrum 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 Spectrum Virtualize system without entering the user password, run the **`ssh-keygen`** tool, as shown in Example B-1.

Example: B-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 B-1 on page 961) and **`-b`** is the size of SSH key in bits (in Example B-1 on page 961, 1024 bits was used).

You also must 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 B-1 on page 961, 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

Uploading the new SSH public key to IBM Spectrum Virtualize by using the GUI is described in “Uploading the SSH public key to the IBM Storage System” on page 955.

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 Spectrum Virtualize active configuration node can be used to store the public key temporarily. Example B-2 shows the command to copy the newly generated public key to the IBM Spectrum Virtualize system.

   Example: B-2 SSH public key copy to 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 B-3) to associate the public SSH key with a user.

   Example: B-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 B-3, it is indicated that a user has a configured SSH key in the field `ssh_key`.

Connecting to an IBM Spectrum Virtualize system

Now that the SSH key is uploaded to the IBM Spectrum 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 B-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: B-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>
```

C

Terminology

This appendix summarizes the IBM Spectrum Virtualize and IBM SAN Volume Controller terms that are commonly used in this book.

To see the complete set of terms that are related to the IBM SAN Volume Controller, see the Glossary that is available at IBM Documentation.

© Copyright IBM Corp. 2021. All rights reserved.

Commonly used terms

This book uses the common IBM Spectrum Virtualize and IBM SAN Volume Controller terminology that are listed in this section.

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 980, “Managed mode” on page 983, and “Unconfigured mode” on page 994.

Activation key
See “License key” on page 982.

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 standard 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) that are used to define logical volumes or devices. An array is a group of drives designated to be managed with a Redundant Array of Independent Disks (RAID).

Asymmetric virtualization
Asymmetric virtualization is 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 993.

Asynchronous replication
Asynchronous replication is a type of replication in which control is given back to the application as soon as 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 993.

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 a particular authenticated user has to access specific resources or actions. See also “Authorization”.

Authorization
The mechanism by which a system determines what permissions 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
Automatic data placement mode is an IBM 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 983,and “Relationship” on page 989.

Available (usable) capacity
See “Capacity” on page 967.

Back end
See “Front end and back end” on page 977.

Caching input/output group
The caching I/O group is the I/O group in the system that performs the cache function for a volume.

Call Home
Call Home is 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 dump retrievals.

Canister
A canister is a single processing unit within a storage system.

Capacity
These are the definitions IBM applies 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 divided by the data reduction savings percentage.

- 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.

- Over provisioned 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. The over provisioned ratio is the ratio of provisioned capacity to usable capacity in the pool or system.

- Physical Capacity

  Physical capacity indicates the total capacity in all storage on the system. Physical capacity includes all the storage 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 - over provisioning limit

  In some storage systems, restrictions in the storage hardware or configured by the user that define a limit the maximum provisioned capacity allowed 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 completely uses storage at creation.

- Thin provisioning savings

  The total amount of usable capacity saved in a pool, system, or volume by using usable capacity when needed as a result of write operations. The capacity saved is the difference between the provisioned capacity minus the written capacity.

- Total capacity savings

  The total amount of usable capacity saved in a pool, system, or volume through thin-provisioning and data reduction techniques. This capacity 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 overhead capacity in a system, pool, array, or MDisk after data reduction techniques are applied.

- Written capacity

  The amount of usable capacity that would 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

Capacity licensing is a licensing model that licenses features with a price-per-terabyte model. Licensed features are FlashCopy, Metro Mirror (MM), Global Mirror (GM), and virtualization. See also “FlashCopy” on page 977, “Metro Mirror” on page 984, and “Virtualized storage” on page 995.

Capacity recycling

Capacity recycling means 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 percent of total usable physical capacity that used capacity must exceed before a notification is sent. See also “Total usable physical capacity” on page 994.

Certificate

A digital document that binds a public key to the identity of the certificate owner, thereby enabling the certificate owner to be authenticated. A certificate is issued by a certificate authority and is digitally signed by that authority.

Chain

A set of enclosures that are attached to provide redundant access to the drives inside the enclosures. Each control enclosure can have one or more chains.

Challenge Handshake Authentication Protocol

Challenge Handshake Authentication Protocol (CHAP) is an authentication protocol that protects against eavesdropping by encrypting the user name and password.

Change volume
A volume that is used in GM that holds earlier consistent revisions of data when changes are made.

Channel extender
A channel extender is 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
Administrators can use child pools 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 985.

Clone
A copy of a volume on a server at a particular point in time. 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
Cloud Container is a virtual object that include all of the elements, components, or data that are common to a specific application or data.

Cloud Provider
Cloud provider is the company or organization that provide off- and on-premises cloud services such as storage, server, network, and so on. IBM Spectrum Virtualize has built-in software capabilities to interact with Cloud Providers such as IBM Cloud, Amazon S3, and deployments of OpenStack Swift.

Cloud Tenant
Cloud Tenant is a group or an instance that provides common access with the specific privileges to an object, software, or data source.

Clustered system (SAN Volume Controller)
A clustered system, which was known as a cluster, is a group of up to eight SAN Volume Controller nodes that presents a single configuration, management, and service interface to the user.

Cold extent
A cold extent is an extent of a volume that does not get any performance benefit if it is moved from a hard disk drive (HDD) to a Flash disk. A cold extent also refers to an extent that needs to be migrated onto an HDD if it is on a Flash disk drive.

Compression

Compression is 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

A compression accelerator is 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 Consistency Group is 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 container is 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.

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 989.

Counterpart SAN

A counterpart SAN is a non-redundant portion of a redundant SAN. A counterpart SAN provides all of the connectivity of the redundant SAN, but without the 100% redundancy. SAN Volume Controller nodes are typically connected to a redundant SAN that is made up of two counterpart SANs. A counterpart SAN is often called a SAN fabric.

Cross-volume consistency

A consistency group property that ensures consistency between volumes when an application issues dependent write operations that span multiple volumes.

Customer-replaceable unit

A Customer-replaceable unit (CRU) is an assembly or part that can be replaced in its entirety by a user when any one of its components fails.

Data consistency

Data consistency is a characteristic of the data at the target site where the dependent write order is maintained to ensure the recoverability of applications.

Data deduplication

Data deduplication is 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

The data encryption key is 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 using the master access key.

Data migration

Data migration is the movement of data from one physical location to another physical location without the disruption of application I/O operations.

Data reduction

Data reduction is 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” and “Capacity” on page 967.

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 “Data reduction”

Data Reduction Pool

Data Reduction Pools are specific types of pools where more control over volumes capacity is given to specific hosts (for example VMware vStorage APIs for Array Integration (VAAI), vSphere APIs for Storage Awareness (VASA), and Microsoft Offloaded Data Transfer (ODX)). These hosts are able to return unused space for reuse. With standard pools, the system is not aware of any unused space on host-allocated volumes. See also “Data reduction”.

Data reduction savings

See “Capacity” on page 967.

Deduplication

See “Data deduplication” on page 972.

Dependent write operation

A write operation that must be applied in the correct order to maintain cross-volume consistency.

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 pane, 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 SAN Volume Controller 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 15,000 revolutions per minute (RPM) Fibre Channel (FC) or serial-attached Small Computer System Interface (SCSI) (SAS) disk, Nearline (NL) SAS, or Serial Advanced Technology Attachment (SATA), or even Flash Disks. 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 hence improves performance of the array. See also “Rebuild area” on page 989.

DNS

See “Domain Name System” on page 973. See also “Domain name server”.

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 (DNS) is the distributed database system that maps domain names to IP addresses.

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 Dual Inline Memory Module (DIMM) is small circuit board with memory-integrated circuits containing signal and power pins on both sides of the board.

- Memory terminology:
  - Channel

    The memory modules are installed into matching banks, which are usually color-coded on the motherboard. These separate channels allow the memory controller access each memory module. For the Intel Cascade Lake architecture, six DIMM Memory channels are available 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; however, for the purposes of clarity, slot DIMM0 or DIMM1 is used, which refers to the first or second slot within a channel on the motherboard. Two slots are available per memory channel on the SAN Volume Controller SV2 hardware. On the motherboard, 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 similar to 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

The Dynamic random access memory (DRAM) is a Storage in which the cells require repetitive application of control signals to retain stored data.

Easy Tier

Easy Tier is a volume performance function within the SAN Volume Controller that provides automatic data placement of a volume's extents in a multitiered storage pool. The pool normally contains a mix of Flash Disks and HDDs. Easy Tier measures host I/O activity on the volume's extents and migrates hot extents onto the Flash Disks to ensure the maximum performance.

Effective capacity

See "Capacity" on page 967.

Encryption key

The encryption key, also known as the master access key, 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 deadlock

The inability to access encryption keys to decrypt data. See also "Encryption recovery key".

Encryption key manger / server

An internal or external system that receives and then serves existing encryption keys or certificates to a storage system.

Encryption recovery key

An encryption key that allows a method to recover from an encryption deadlock situation where the normal encryption key servers are not available.

Encryption of data-at-rest

Encryption of data-at-rest is the inactive encryption data that is stored physically on the storage system.

Enhanced Stretched Systems

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 the following 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 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 additional policing rules to ensure that the correct configuration of a standard stretched system is used.

Evaluation mode

The evaluation mode is 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 event is 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

An event code is 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

An event ID is a value that is used to identify a unique error condition that was detected by the SAN Volume Controller. 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 SAN Volume Controller 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.

Extent

An extent is 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

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
Failback is the restoration of an appliance to its initial configuration after the detection and repair of a failed network or component.

Failover
Failover is 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 982.

Fibre Channel
FC is 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 996.

Fibre Channel port logins
FC port logins refer to the number of hosts that can see any one SAN Volume Controller node port. The SAN Volume Controller has a maximum limit per node port of FC logins that are allowed.

Fibre Channel Arbitrated Loop
Fibre Channel Arbitrated Loop (FC-AL) is an implementation of the FC standards that uses a ring topology for the communication fabric; refer to 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 Connection
A Fibre Channel connection is an FC communication protocol for IBM mainframe computers and peripheral devices.

Fibre Channel over IP
Fibre Channel over IP (FCIP) is network storage technology that combines the features of the Fibre Channel Protocol (FCP) and the IP to connect distributed SANs over large distances.

Fibre Channel Protocol
The serial SCSI command protocol used on FC networks.

Field-replaceable unit
Field-replaceable units (FRUs) are individual parts that are replaced entirely when any one of the unit’s components fails. They are held as spares by the IBM service organization.

File Transfer Protocol
In TCP/IP, 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

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

A FlashCopy mapping is a continuous space on a direct-access storage volume that is occupied by or reserved for a particular data set, data space, or file.

FlashCopy relationship

See “FlashCopy mapping” on page 977.

FlashCopy service

FlashCopy service is 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 986.

FlashCore Module

FlashCore Module (FCM) modules are a family of high-performance flash drives. The FCM design utilizes the Non-Volatile Memory Express (NVMe) protocol, a PCIe Gen3 interface, and high-speed NAND memory to provide high throughput and input/output operations per second (IOPS) and very low latency. FCM modules are available in different capacities. Hardware-based data compression and self-encryption are supported. The FCM modules are accessible from the front of the enclosure.

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 that contains flash memory, one or more flash controllers, and associated electronics. See also “Flash drive”.

Form factor

The industry-standard physical dimensions of a storage system drive enclosure. Possible values include “3.5 inch”, “2.5 inch”, and “1.8 inch.”

Front end and back end

The SAN Volume Controller takes MDisks to create pools of capacity from which volumes are created and presented to application servers (hosts). The volumes that are presented to the hosts are in the front end of the IBM SAN Volume Controller.

Full restore operation

A copy operation where a local volume is created by reading an entire a volume snapshot 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 the cloud.

General Parallel File System

General Parallel File System (GPFS) is a high-performance shared-disk file system that can provide data access from nodes in a clustered system environment.

GB

See“Gigabyte” on page 978.

Gigabyte

For processor storage, real and virtual storage, and channel volume, two to the power of 30 or 1,073,741,824 bytes. For disk storage capacity and communications volume, 1,000,000,000 bytes.

Global Mirror

Global Mirror (GM) is a method of asynchronous replication that maintains data consistency across multiple volumes within or across multiple systems. GM is generally used where distances between the source site and target site cause increased latency beyond what the application can accept.

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.

GPFS

See “General Parallel File System”.

GPFS cluster

A system of nodes that are defined as being available for use by GPFS file systems.

GPFS snapshot

A PiT copy of a file system or file set.

Grain

A grain is the unit of data that is represented by a single bit in a FlashCopy bitmap (64 kibibytes (KiB) or 256 KiB) in the SAN Volume Controller. 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 graphical user interface (GUI) is 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 computer system that hosts computer applications, with the host and the applications using storage.

Host bus adapter
A host bus adapter (HBA) is an interface card that connects a server to the SAN environment through its internal bus system, for example, PCI Express. Typically it is referred to the FC adapters.

Host Cluster
A configured set of physical or virtual hosts that share one or more storage volumes to increase scalability or availability of computer applications.

Host ID
A host ID is a numeric identifier that is assigned to a group of host FC ports or Internet Small Computer Systems Interface (iSCSI) host names 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
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 hot extent is a frequently accessed volume extent that gets a performance benefit if it is moved from an HDD onto a Flash Disk.

Hot-spare
Pertaining to redundant hardware (such as an adapter, a disk, a drive, or a server) that is installed and available in the event of a hardware failure.

Hot Spare Node
Hot Spare Node is an online SAN Volume Controller node defined in a cluster, but not in any I/O group. During a failure of any of online nodes in any I/O group of cluster, it is automatically swapped with this spare node. After the recovery of an original node has finished, the spare node returns to the standby spare status.

IBM HyperSwap
Pertaining to a function that provides continuous, transparent availability against storage errors and site failures, and is based on synchronous replication.

IBM Remote Support Server and Client

IBM Remote Support Client is 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

IBM SAN Volume Controller is 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 Spectrum Virtualize is a software engine of IBM SAN Volume Controller (and IBM Storage System family) that performs block-level virtualization of disk storage.

IBM Security Key Lifecycle Manager

IBM Security Key Lifecycle Manager centralizes, simplifies, and automates the encryption key management process to help minimize risk and reduce operational costs of encryption key management.

Image mode

Image mode is 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 983 and “Unconfigured mode” on page 994.

Image volume

An image volume is a volume in which a direct block-for-block translation exists from the MDisk to the volume.

I/O group

Each pair of 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 SAN Volume Controller 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 cloud 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

IOPS is a common performance measurement used to benchmark computer storage devices.

Input/output throttling rate

The maximum rate at which an I/O transaction is accepted for a volume.

Internal storage

Internal storage refers to an array of MDisks and drives that are held in enclosures and in nodes that are part of the SAN Volume Controller cluster.

Internet Protocol

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 is 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

The iSNS 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 hop

An inter-switch link (ISL) is 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” on page 981.

iSCSI alias

An alternative name for the iSCSI-attached host.

iSCSI initiator

An 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

iSCSI Qualified Name (IQN) refers to special names that identify both 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 a SAN Volume Controller node can be in the following format:

`iqn.1986-03.com.ibm:2145.<clustername>.<nodename>`

iSCSI session

The interaction (conversation) between an iSCSI Initiator and an iSCSI Target.

iSCSI target

An iSCSI target is a storage resource on an iSCSI server.

Just A Bunch Of Disks
A group of Hard disks that haven't been 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 VPN connection that is established between two endpoints.

See also “Encryption key manger / server” on page 974.

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
Least recently used (LRU) pertains to an algorithm used to identify and make available the cache space that contains the data that was LRU.

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
Lightweight Directory Access Protocol (LDAP) is 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 (DAP). 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 local fabric interconnect and the remote fabric interconnect are 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 (LW) gigabit interface converters (GBICs) or Small Form-factor Plugable (SFP) modules, or more sophisticated components, such as channel extenders or special SFP modules that are used to extend the distance between SAN components.

Local fabric
The local fabric is composed of SAN components (switches, cables, and so on) that connect the components (nodes, hosts, and switches) of the local cluster together.

Logical drive
See “Volume” on page 995.

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
An MDisk is a SCSI disk that is presented by a RAID controller and managed by the SAN Volume Controller. The MDisk is not visible to host systems on the SAN.

Managed disk group
See “Storage pool (managed disk group)” on page 992.

Managed mode
An access mode that enables virtualization functions to be performed. See also “Image mode” on page 980 and “Unconfigured mode” on page 994.

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 967, and “Relationship” on page 989.

Maximum replication delay
Maximum replication delay is the number of seconds that MM or GM replication can delay a write operation to a volume.

MDisk
See “Managed disk”.

MDisk group
See “Storage pool (managed disk group)” on page 992.

Media Access Control
In networking, the lower of two sub-layers 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 will be used.

Megabytes per second
Megabytes per second (MBps) is a unit of data transfer rate equal to 1024 * 1024 bytes.

Metro Global Mirror

Metro Mirror Global (MGM) is a cascaded 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.

Metro Mirror

Metro Mirror (MM) is a method of synchronous replication that maintains data consistency across multiple volumes within the system. MM is generally used when the write latency that is caused by the distance between the source site and target site is acceptable to application performance.

Mirrored volume

A mirrored volume is a single virtual volume that has two physical volume copies. The primary physical copy is known within the SAN Volume Controller as `copy 0` and the secondary copy is known within the SAN Volume Controller as `copy 1`.

N_Port ID Virtualization

N_Port ID Virtualization (NPIV) is a FC feature whereby multiple FC N_Port IDs can share a single physical N_Port.

Namespace Globally Unique Identifier

The Namespace Globally Unique Identifier (NGUID) is defined in the Identify Namespace data structure. The NGUID is composed of an IEEE OUI, an extension identifier, and a vendor specific extension identifier. 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 SAN Volume Controller node is a hardware entity that provides virtualization, cache, and copy services for the cluster. The SAN Volume Controller nodes are deployed in pairs that are called I/O groups. One node in a clustered system is designated as the configuration node.

Node canister

A node canister is a hardware unit that includes the node hardware, fabric and service interfaces, and SAS expansion ports. Node canisters are specifically recognized on IBM Storwize products. In SAN Volume Controller, all these components are spread within the whole system chassis, so we usually do not consider node canisters in SAN Volume Controller, but just the node as a whole.

Node rescue

The process by which a node that has no valid software installed on its HDD can copy software from another node connected to the same FC fabric.

Non-Volatile Memory Express

NVM Express (NVMe) or Non-Volatile Memory Host Controller Interface Specification (NVMHCIS) is an open logical-device interface specification for accessing non-volatile storage media attached via PCI Express (PCIe) bus.

NVMe Qualified Name

NVMe Qualified Names (NQNs) are used to uniquely describe a host or NVM subsystem for the purposes of identification and authentication. The NVMe Qualified Name for the NVM subsystem is specified in the Identify Controller data structure. An NQN is permanent for the lifetime of the host or NVM subsystem.

Object Based Access control

See “Ownership groups”.

Object storage

Object storage is a general term that refers to the entity in which Cloud Object Storage organizes, manages, and stores with units of storage, or just *objects*.

Overprovisioned

See “Capacity” on page 967.

Overprovisioned ratio

See “Capacity” on page 967.

Oversubscription

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

Ownership groups feature provides a method of implementing multi-tenant solution on the system. Ownership groups allow the allocation of storage resources to several independent tenants with the assurance that one tenant cannot access resources 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

Parent pools 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 970.

Partner node

The other node that is in the I/O group to which this node belongs.

Partnership
In MM or GM 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.

Performance group
A collection of volumes that is 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 “Performance group”.

Point-in-time copy
A PiT copy is an instantaneous copy that the FlashCopy service makes of the source volume. See also “FlashCopy service” on page 977.

Pool
See “Storage pool (managed disk group)” on page 992.

Pool pair
Two storage pools that are required to balance 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 “Relationship” on page 989.

Priority flow control
Priority flow control (PFC) is a link-level flow control mechanism, 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.

Private fabric
Configure one SAN per fabric so that it is dedicated for node-to-node communication. This SAN is referred to as a private SAN.

Provisioned capacity
See “Capacity” on page 967.

Provisioning Group
A provisioning group is 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.

Public fabric
Configure one SAN per fabric so that it is dedicated for host attachment, storage system attachment, and RC operations. This SAN is referred to as a public SAN. You can configure the public SAN to allow SAN Volume Controller node-to-node communication also. You can optionally use the **-localportfcmask** parameter of the **chsystem** command to constrain the node-to-node communication to use only the private SAN.

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 input/output (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 finally 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 assigned as upper limits for a specified user, group of users, or file set.

Random Access Compression Engine (RACE)
The Random Access Compression Engine compresses data on volumes in real time with minimal effect on performance. See “Compression” on page 971 or “Real-time Compression” on page 988.

RAID
See “Redundant Array of Independent Disks” on page 988.

RAID controller

See “Node canister” on page 984.

Raw capacity

See “Capacity” on page 967.

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 967.

Real-time Compression

IBM Real-time Compression (RtC) is an IBM integrated software function for storage space efficiency. RACE compresses data on volumes in real time with minimal effect on performance.

Redundant Array of Independent Disks

RAID 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 RAID (DRAID) see “Distributed Redundant Array of Independent Disks” on page 973.

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. It includes no data protection, however, so, strictly speaking, the appellation RAID is a misnomer. RAID 0 is also known as data striping.

RAID 1

RAID 1 is 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. In the event of a physical device failure, the data can be read or regenerated from the other drives in the RAID due to data redundancy.

RAID 5

RAID 5 is 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

RAID 6 is 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 of the array’s virtual disks (VDisks) in the presence of two concurrent disk failures.

Real capacity

The amount of storage that is allocated to a volume copy from a storage pool.

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, returning 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 973.

Reclaimable (or reclaimed) capacity

Reclaimable Data is 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 975.

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 971.

Redundant storage area network

A redundant SAN is a SAN configuration in which there is no single point of failure (SPOF). 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 971.

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 967, “Master volume” on page 983, “Primary volume” on page 986, and “Secondary volume”.

Reliability, availability, and serviceability

Reliability, availability, and serviceability (RAS) are 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” on page 978. See also “Metro Mirror” on page 984.

Remote fabric

The remote fabric is composed of SAN components (switches, cables, and so on) 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.

Safeguarded Copy

- 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 also can assign a volume to a Safeguarded volume group after creating a volume, or create a volume in a volume group and then, assign 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 initiates 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 written by the host application to the primary volume. See also “Relationship” on page 989.

Secure Copy Protocol

Secure Copy Protocol (SCP) is the secure transfer of computer files between a local and a remote host or between two remote hosts, using the Secure Shell (SSH) protocol.

Secure Sockets Layer certificate

Secure Sockets Layer (SSL) is the standard security technology for establishing an encrypted link between a web server and a browser. This link ensures that all data passed between the web server and browsers remain private. To create an SSL connection, a web server requires an SSL Certificate.

Sequential volume

A volume that uses extents from a single MDisk.

Serial-attached SCSI

SAS is a method that is used in accessing computer peripheral devices that employs a serial (one bit at a time) means of digital data transfer over thin cables. The method is specified in the American National Standard Institute standard called SAS. In the business enterprise, SAS is useful for access to mass storage devices, particularly external HDDs.

Service Location Protocol

Service Location Protocol (SLP) is 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 the 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

Small Computer System Interface (SCSI) is 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

A snapshot is an image backup type that consists of a PiT view of a volume.

Solid-state drive

A solid-state drive (SSD) or Flash Disk is 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 SAN Volume Controller as a disk tier generic_ssd.

Space efficient

See “Thin provisioning” on page 993.

Space-efficient virtual disk

See “Thin-provisioned volume” on page 993.

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

For more information about a space-efficient volume, see “Thin-provisioned volume” on page 993.

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.

Statesave

Binary data collection that is used for problem determination by service support.

Storage area network

A SAN is 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

Storage-class memory (SCM) is a type of NAND flash that includes a power source to ensure that data won't be 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 significantly quicker than access to data in local, PCI-connected solid state drives (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

Storage Capacity Unit (SCU) is a SAN Volume Controller license metric that measures the managed capacity in a way that the price is differentiated by the technology 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 storage pool is a collection of storage capacity, which is made up of 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 and a prerequisite of Easy Tier automatic data placement. Before SAN Volume Controller V6.1, this storage pool was known as a managed disk group (MDG).

Stretched system

A stretched system is an extended HA method that is supported by 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 storage system that provides a quorum disk.

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 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
Symmetric virtualization is a virtualization technique in which the physical storage, in the form of a RAID, is split into smaller chunks of storage known as extents. These extents are then concatenated, by using various policies, to make volumes. See also “Asymmetric virtualization” on page 966.

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 returned to the application. See also “Asynchronous replication” on page 966.

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.

Tie-break
In a case of a cluster split in 2 groups of nodes, tie-break is a role of a quorum device used to decide which group continues to operate as the system, handling all I/O requests.

Thin-provisioned volume
A thin-provisioned volume is a volume that allocates storage when data is written to it.

Thin provisioning
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. Before SAN Volume Controller V6.1, this thin-provisioned volume was known as *space efficient*.

Thin provisioning savings
See “Capacity” on page 967.

Throttles
Throttling 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 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 transmitted over a network in a given period of time. Throughput is generally 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
Transparent Cloud Tiering (TCT) is a separately installable feature of IBM Spectrum Scale that provides a native cloud storage tier.

Trial License
A temporary entitlement to use a licensed function.

Total capacity savings
See “Capacity” on page 967.

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 “Capacity threshold” on page 969, reserved capacity, stored capacity.

TSE for FlashCopy
A thin-provisioning method in which storage space is allocated from a TSE repository on an as needed basis. See also “TSE repository”.

TSE repository
The amount of capacity in a storage pool reserved for volumes that use a thin-provisioning method of TSE for FlashCopy. See also “TSE for FlashCopy”.

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 980 and “Managed mode” on page 983.

Unique identifier
A unique identifier (UID) is 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.

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.

Used capacity
The amount of usable capacity that is taken up by data or overhead capacity in a system, pool, array, or MDisk after data reduction techniques have been applied.

VDisk
See “Virtual disk”.

VDisk-to-host mapping
See “Host mapping” on page 979.

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.

Virtualized storage
Virtualized storage is physical storage that has virtualization techniques applied to it by a virtualization engine.

Virtual capacity
The amount of storage that is available. In a thin-provisioned volume, the virtual capacity can be different from the real capacity. In a standard volume, the virtual capacity and real capacity are the same.

Virtual disk
See “Volume” on page 995.

Virtual local area network
Virtual local area network (VLAN) tagging separates network traffic at the layer 2 level for Ethernet transport. The system supports VLAN configuration on both Internet Protocol Version 4 (IPv4) and Internet Protocol Version 6 (IPv6) connections.

Virtual storage area network
A virtual storage area network (VSAN) is a logical fabric entity defined within the SAN. It can be defined on a single physical SAN switch or across multiple physical switched or directors. In VMware terminology, the VSAN is defined as a logical layer of storage capacity built from physical disk drives attached directly into the ESXi hosts. This solution is not considered for the scope of our publication.

Vital product data
Vital product data (VPD or VDP) is information that uniquely defines system, hardware, software, and microcode elements of a processing system.

Volume
A volume is a SAN Volume Controller 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. Before SAN Volume Controller 6.1, this volume was known as a VDisk.

Volume copy
A volume copy is 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 remote-copy 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 on a cloud storage account that represents the data of a volume at a particular time.

Worldwide ID

A worldwide ID (WWID) is a name identifier that is unique worldwide and that is represented by a 64-bit value that includes the IEEE-assigned organizationally unique identifier (OUI).

Worldwide name

A worldwide name (WWN) is a 64-bit, unsigned name identifier that is unique.

Worldwide node name

A worldwide node name is a unique 64-bit identifier for a host containing a FC port. See also “Worldwide port name”.

Worldwide port name

A worldwide port name (WWPN) is a unique 64-bit identifier associated with a FC adapter port. The WWPN is assigned in an implementation-independent and protocol-independent manner. See also “Worldwide node name”.

Write-through mode

Write-through mode is a process in which data is written to a storage device at the same time that the data is cached.

Written capacity

See “Capacity” on page 967.

Zoning

The grouping of 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. See also “Fibre Channel” on page 976.

Related publications

The publications that are listed in this section are considered suitable for a more detailed description of the topics that are covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide more information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only:

- *IBM FlashSystem 5000 Family Products*, SG24-8449
- *IBM FlashSystem 9100 Architecture, Performance, and Implementation*, SG24-8425
- *IBM FlashSystem 9200 and 9100 Best Practices and Performance Guidelines*, SG24-8448
- *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521
- *Implementing the IBM Storwize V5000 Gen2 (including the Storwize V5010, V5020, and V5030) with IBM Spectrum Virtualize V8.2.1*, SG24-8162
- *Implementing the IBM Storwize V7000 with IBM Spectrum Virtualize V8.2.1*, SG24-7938
- *Implementing the IBM System Storage SAN Volume Controller with IBM Spectrum Virtualize V8.2.1*, SG24-7933
- *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430
- *Implementation Guide for SpecV/FlashSystem Safeguarded Copy*, REDP-5654
- *Using the IBM Block Storage CSI Driver in a Red Hat,* REDP-5613

You can search for, view, download, or order these documents and other Redbooks, Redpapers, web docs, drafts, and additional materials, at the following website:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

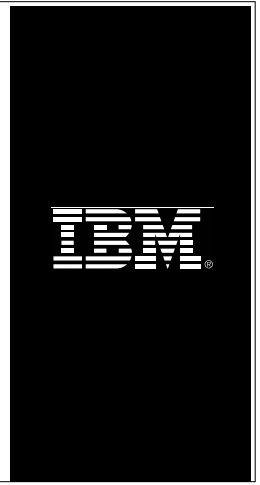
© Copyright IBM Corp. 2021. All rights reserved.



Implementing the IBM SAN Volume Controller with IBM Spectrum Virtualize Version 8.4.2

SG24-8507-00

ISBN 0738460206



(2.5” spine)
2.5”<->nnn.n”
1315<-> nnnn pages



SG24-8507-00

ISBN 0738460206

Printed in U.S.A.

Get connected



