

iSCSI Implementation and Best Practices on IBM Storwize Storage Systems

Jonathan Burton

Anuj Chandra

Jordan Fincher

Kushal Patel

Torsten Rothenwaldt

Subhojit Roy

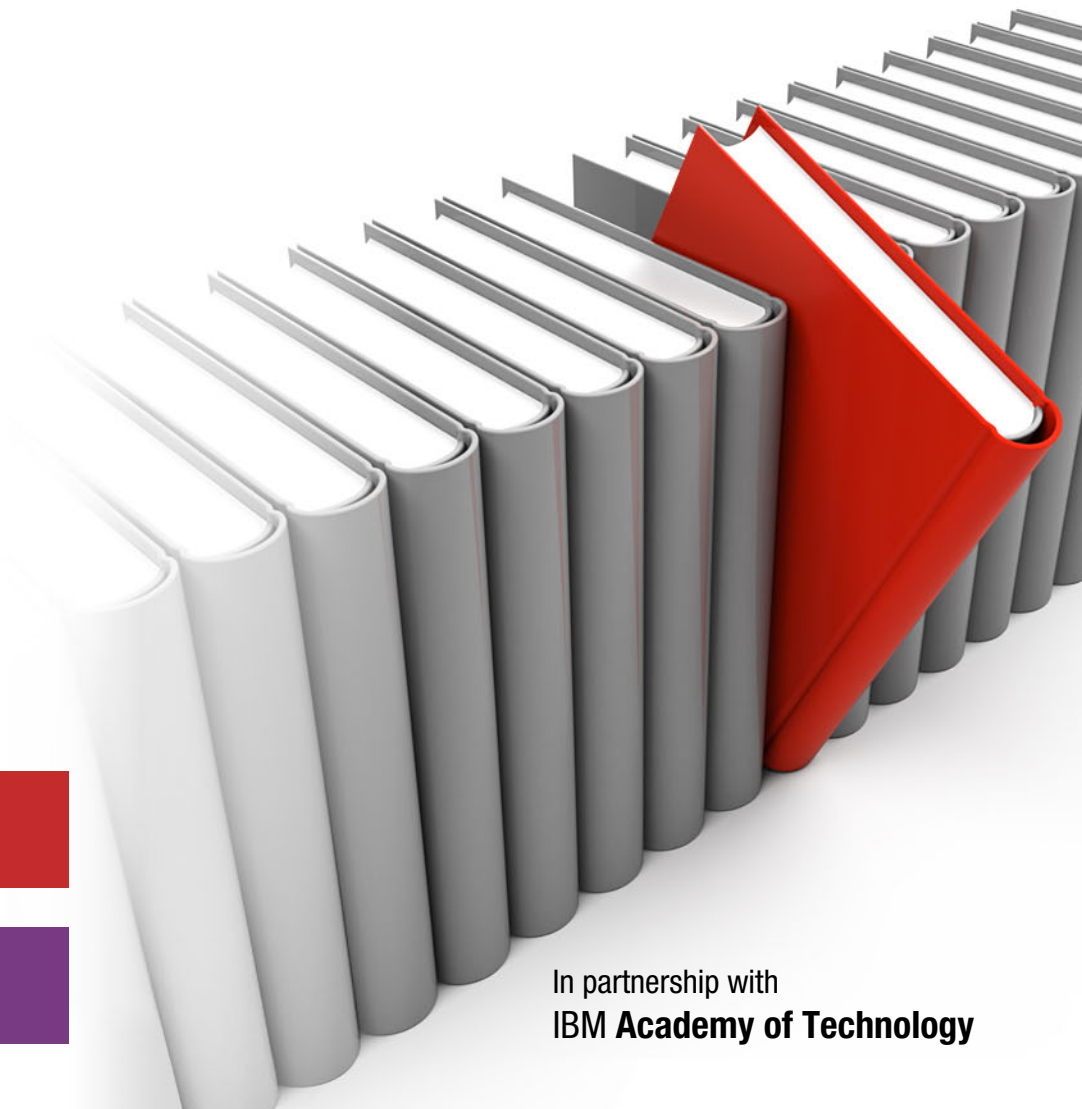
Bharti Soni

Shalaka Verma

Megan Gilge



Storage



In partnership with
IBM Academy of Technology



International Technical Support Organization

**iSCSI Implementation and Best Practices on IBM
Storwize Storage Systems**

October 2017

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

Second Edition (October 2017)

This edition applies to Version 7, Release 8, Modification 0 of IBM Spectrum Virtualize.

© Copyright International Business Machines Corporation 2016, 2017. 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	ix
Trademarks	x
Preface	xi
Authors	xi
Now you can become a published author, too!	xiii
Comments welcome	xiv
Stay connected to IBM Redbooks	xiv
Summary of changes	xv
October 2017, Second Edition	xv
Part 1. iSCSI overview	1
Chapter 1. Use cases for iSCSI virtualization	3
1.1 Consolidating iSCSI storage consolidation	4
1.2 Removing capacity silos from data centers	4
1.3 Improving the performance of iSCSI storage systems	5
Chapter 2. Introduction to iSCSI in IBM Storwize storage systems	7
2.1 What iSCSI is	8
2.2 iSCSI sessions	9
2.2.1 Components of an iSCSI session	9
2.2.2 The three phases of iSCSI login	10
2.3 iSCSI adapters	11
2.3.1 Ethernet card (network interface card)	11
2.3.2 TCP offload engine	12
2.3.3 iSCSI offload engine	13
2.4 iSCSI routing	13
2.5 Ethernet for iSCSI	14
2.5.1 Data Center Bridging	14
2.5.2 The future of Ethernet and its impact on iSCSI	17
2.6 Fibre Channel: FCoE terms and their iSCSI equivalents	18
2.6.1 Fibre Channel zoning	18
2.6.2 Virtual SAN	18
2.6.3 Buffer-to-Buffer credit	18
2.6.4 Worldwide name	18
2.6.5 Fabric name server	19
2.7 Comparison of iSCSI and FCoE	19
2.8 Why use iSCSI	20
2.8.1 iSCSI is cost-effective	20
2.8.2 No distance limitations	20
2.8.3 Good interoperability	20
2.8.4 Bandwidth usage and Converged Enhanced Ethernet benefits	21
2.8.5 Security	21
Chapter 3. External virtualization and host connectivity interface options for the IBM Storwize family	23
3.1 Connectivity options for the IBM Storwize V5000 Gen2 storage system	24

3.1.1	Connectivity options for the IBM Storwize V5010 storage system	24
3.1.2	Connectivity options for the IBM Storwize V5020 storage system	25
3.1.3	Connectivity options for IBM Storwize V5030 and IBM Storwize V5030F storage systems.	26
3.1.4	IBM Storwize V5010, IBM Storwize V5020, and IBM Storwize V5030 HIC options at a glance	28
3.2	Connectivity options for the IBM Storwize V7000 storage system	28
3.2.1	External connectivity options for the IBM Storwize V7000 Gen2+	29
3.2.2	External connectivity options for the IBM Storwize V7000 Gen2	30
3.3	The IBM Storwize V7000 Unified storage system.	32
3.4	SAN Volume Controller SV1 storage systems	33
3.5	Hardware terminology for the IBM Storwize disk systems	35
3.5.1	Control enclosures, nodes, and I/O groups.	35
3.5.2	Expansion enclosures.	36
3.5.3	IBM Storwize cluster system.	37
3.5.4	IBM Storwize virtualization	37
Chapter 4.	Planning considerations	39
4.1	General considerations	40
4.2	Network topology	42
4.2.1	Network topology with one Ethernet switch	42
4.2.2	Network topology with two VLANs	43
4.2.3	Network topology with four VLANs	43
4.2.4	Link aggregation between switches	44
4.2.5	iSCSI that uses a Layer 3 network topology	44
4.2.6	IP replication network topology	45
4.3	Planning for host access	46
4.3.1	Planning for IBM AIX.	46
4.3.2	Planning for Linux	50
4.3.3	Planning for VMware.	51
4.3.4	Planning for Windows	51
4.4	Planning considerations for external virtualization	52
4.4.1	Network security	52
4.4.2	iSCSI Protocol-specific considerations	52
4.4.3	Controller migration considerations.	52
4.5	IBM Storwize family and iSCSI limits	53
4.5.1	Version 7.8 configuration limits and restrictions for the IBM Storwize V3500 storage system	53
4.5.2	Version 7.8 configuration limits and restrictions for the IBM Storwize V3700 storage system	53
4.5.3	Version 7.8 configuration limits and restrictions for the IBM Storwize V5000 storage system	53
4.5.4	Version 7.8 configuration limits and restrictions for the IBM Storwize V7000 storage system	53
4.5.5	Version 7.8 configuration limits and restrictions for the SAN Volume Controller storage system	53
Chapter 5.	iSCSI storage connection security	55
5.1	iSCSI security model.	56
5.1.1	iSCSI network security	56
5.2	Configuring CHAP for an IBM Storwize storage system.	57
5.2.1	Configuring CHAP for the IBM Storwize storage system by using the GUI	57
5.2.2	Configuring CHAP for the IBM Storwize storage system by using the CLI.	60

5.3	Configuring CHAP authentication for the host	61
5.3.1	Setting up authentication for Linux hosts	61
5.3.2	Setting up authentication for Microsoft Windows hosts	63
5.3.3	Setting up authentication for AIX hosts	68
5.3.4	Setting up authentication for VMware hosts	69
5.4	iSCSI security	71
5.5	Mandatory security in real-world situations	71
Chapter 6.	IBM Storwize performance	73
6.1	Jumbo frames	74
6.2	VLAN separation	74
6.2.1	VLAN	74
6.2.2	Advantages of VLANs	75
6.2.3	VLAN and iSCSI performance	75
6.3	Subnetting	76
6.3.1	Network subnetting	76
6.3.2	Subnetting and iSCSI performance	77
6.4	Quality of service and traffic prioritization	78
6.5	iSCSI protocol digests and performance	79
Part 2.	iSCSI host attachment	81
Chapter 7.	Configuring the IBM Storwize storage system and hosts for iSCSI	83
7.1	Configuring the IBM Storwize storage system for iSCSI	84
7.1.1	Setting the IBM Storwize iSCSI IP address	84
7.1.2	Setting optional iSCSI settings on IBM Storwize storage systems	85
7.2	Configuring initiators for iSCSI	87
7.2.1	iSCSI discovery mechanisms	87
7.2.2	iSCSI operational parameters	88
7.2.3	Considerations for enabling TSO for host network adapters	91
7.2.4	Host configuration maximums for iSCSI with IBM Storwize storage systems	91
7.3	Configuring iSCSI on AIX 7.1	92
7.3.1	Ethernet network configuration	92
7.3.2	Selecting the discovery policy	92
7.3.3	Working with the IBM Storwize storage volume	96
7.4	Configuring iSCSI for SUSE Linux Enterprise Server	97
7.4.1	Prerequisites for mapping the iSCSI volume	97
7.4.2	Ethernet network configuration	98
7.4.3	Discovering and logging in to the iSCSI targets	98
7.4.4	Understanding iSCSI sessions for software-based initiators	103
7.4.5	Working with the IBM Storwize storage volume	105
7.5	Configuring iSCSI for Windows 2012	108
7.5.1	Prerequisites	108
7.5.2	Ethernet network configuration on Windows hosts	110
7.5.3	iSCSI target discovery for Windows hosts	111
7.6	Configuring iSCSI for VMware ESXi hosts	120
7.6.1	Configuring the Ethernet network on the VMware host	123
7.7	iSNS server configuration	143
7.7.1	Enabling iSNS server in Windows 2012	143
7.7.2	Configuring the iSNS server address on an IBM Storwize storage system	144
7.7.3	Configuring the iSCSI initiator with iSNS server details	146
7.8	Configuring Priority Flow Control for the IBM Storwize storage system	148
7.8.1	Requirements for PFC	148
7.8.2	Configuring Priority Flow Control on Brocade VDX	148

7.8.3 Verifying Priority Flow Control from the IBM Storwize storage system	149
7.9 Configuring the iSCSI host for the HyperSwap cluster	152
7.9.1 What HyperSwap is	152
7.9.2 Host site assignment.	153
7.9.3 Working with HyperSwap volumes	154
Chapter 8. IBM Spectrum Virtualize and IBM Storwize performance monitoring . . .	157
8.1 Manually gathering performance statistics	158
8.1.1 Statistics file naming	158
8.2 Real-time performance monitoring	159
8.2.1 Real-time performance monitoring with the CLI	160
8.2.2 Real-time performance monitoring with the GUI	163
8.3 Performance data collection with IBM tools	166
8.3.1 IBM Spectrum Control.	166
8.3.2 IBM Spectrum Control Storage Insights	167
Chapter 9. IBM Spectrum Virtualize and IBM Storwize storage systems on the OpenStack platform	169
9.1 Introduction to OpenStack components	170
9.2 Integrating the Cinder driver with IBM Spectrum Virtualize and IBM Storwize storage systems	171
9.2.1 Volume creation and host attachment with OpenStack	173
9.2.2 Volume attachment from Nova	174
Chapter 10. Troubleshooting	175
10.1 Storage tools on an IBM Storwize storage system	176
10.1.1 Management GUI	176
10.1.2 Service Assistant GUI	181
10.1.3 Command-line interface	182
10.1.4 Service CLI	189
10.1.5 USB.	190
10.1.6 Visual indicators (Ethernet port LED status)	190
10.2 Storage logs that are used for analysis.	191
10.2.1 Support Package on the IBM Storwize cluster	191
10.2.2 Event log on the IBM Storwize cluster	192
10.2.3 Audit log on the IBM Storwize cluster	193
10.2.4 Ethernet logs and statistics on IBM Storwize nodes	193
10.2.5 iSCSI logs on IBM Storwize nodes	195
10.3 Different IP addresses on the IBM Storwize storage system	196
10.3.1 Path failover mechanisms in iSCSI on an IBM Storwize storage system	197
10.3.2 iSCSI IP failover	197
10.4 Problem determination	198
10.4.1 Problem determination: Obtaining a basic configuration overview	199
10.4.2 Problem determination: Checking the network configuration	201
10.4.3 Problem determination: Checking the IBM Storwize configuration	204
10.4.4 Problem determination: Checking authentication	204
10.4.5 Problem determination: Checking active sessions from the IBM Storwize storage system	205
10.4.6 Problem determination: Checking a host configuration.	206
10.4.7 Problem determination: Checking for performance problems.	207

Part 3. iSCSI virtualization	211
---	------------

Chapter 11. iSCSI virtualization overview.	213
---	------------

11.1 Planning considerations for iSCSI virtualization	214
11.1.1 Fibre Channel versus iSCSI virtualization.	214
11.1.2 Storage port configuration model	215
11.1.3 Controller considerations	217
11.1.4 Stretched cluster and HyperSwap topology	218
11.1.5 Security	219
11.1.6 Limits and considerations	219
11.2 iSCSI external virtualization steps.	220
11.2.1 Port selection	220
11.2.2 Source port configuration	220
11.2.3 Target port configuration.	220
11.2.4 Host mapping and authentication settings on target controllers	221
11.2.5 Understanding the storage port model for a back-end controller	221
11.2.6 Discovering storage ports from the initiator.	221
11.2.7 Viewing the discovery results	222
11.2.8 Adding sessions to discovered storage ports	223
11.2.9 Viewing established sessions to storage ports	224
Chapter 12. External virtualization of IBM Storwize storage systems	227
12.1 Planning considerations	228
12.1.1 Limits and considerations	229
12.1.2 Performance considerations	229
12.2 Target configuration	229
12.2.1 System layer	229
12.2.2 Host mappings	230
12.2.3 Authentication	230
12.2.4 Port configuration	231
12.3 Initiator configuration.	232
12.3.1 Establishing connections and sessions.	232
12.4 Configuration validation.	235
Chapter 13. Virtualization of IBM Spectrum Accelerate storage systems	239
13.1 Planning considerations	240
13.1.1 Limits and considerations for IBM XIV, IBM FlashSystem A9000, and IBM Spectrum Accelerate.	241
13.1.2 Performance considerations	242
13.1.3 Migration considerations	242
13.2 Target configuration	243
13.2.1 Port configuration	245
13.2.2 Host mappings and authentication	247
13.2.3 Mapping XIV LUNs to the SAN Volume Controller or IBM Storwize system.	253
13.3 Initiator configuration.	254
13.3.1 Overview	255
13.3.2 Workflow that uses the CLI	255
13.3.3 Workflow with GUI.	257
13.3.4 Configuration validation	261
Chapter 14. External virtualization of Dell Equallogic PS Series	265
14.1 Planning considerations	266
14.1.1 Dell Equallogic PS Series connection considerations.	267
14.1.2 Migration considerations.	267
14.2 Target configuration	268
14.2.1 Port configuration	268
14.2.2 Setting up access policies.	271

14.2.3	Creating volumes and applying access policies	276
14.2.4	Modifying the settings of existing volumes	281
14.3	Initiator configuration	284
14.3.1	GUI workflow	285
14.3.2	CLI workflows	290
Chapter 15.	Configuration and administration of iSCSI	297
15.1	Changing the iSCSI port configuration	298
15.1.1	Changing the iSCSI initiator ports' IP addresses	298
15.1.2	Changing the iSCSI target ports' IP addresses	298
15.1.3	Enabling or disabling iSCSI on IP ports	300
15.2	Adding or removing nodes or I/O groups	300
15.2.1	Adding nodes	300
15.2.2	Removing nodes from a SAN Volume Controller initiator cluster	302
15.2.3	Adding ports to the SAN Volume Controller initiator	302
15.2.4	Removing ports	303
15.3	Changing the system name or node name	304
15.3.1	Changing the system name or node name of the initiator (SAN Volume Controller system)	304
15.4	Changing the CHAP configuration	312
15.4.1	General considerations	312
15.4.2	Instructions for a SAN Volume Controller or IBM Storwize initiator system with an IBM Storwize target system	313
15.5	Changing the number of LUNs, ports, and IQNs in an IBM Storwize system	315
15.5.1	Adding and removing LUNs exposed from IBM Storwize or XIV controllers	315
15.5.2	Adding LUNs from a Dell EqualLogic controller	316
15.5.3	Removing LUNs from a Dell EqualLogic controller	317
Chapter 16.	Troubleshooting iSCSi virtualization	319
16.1	Troubleshooting iSCSI target discovery	320
16.1.1	Problems with initial discovery	320
16.1.2	Problems adding a storage port	320
16.2	Troubleshooting a degraded or offline status	321
16.2.1	Restoring an offline MDisk or storage controller	321
16.2.2	Restoring degraded MDisk or storage controllers	321
16.3	Performance issues	322
Related publications		323
IBM Redbooks		323
Online resources		323
Help from IBM		324

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.

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®	IBM Spectrum™	Redbooks®
Easy Tier®	IBM Spectrum Accelerate™	Redbooks (logo)  ®
FlashCopy®	IBM Spectrum Control™	Storwize®
HyperSwap®	IBM Spectrum Virtualize™	System Storage®
IBM®	Insight®	Tivoli®
IBM FlashSystem®	Real-time Compression™	XIV®

The following terms are trademarks of other companies:

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

Intel, 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.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

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

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

Preface

This IBM® Redbooks® publication helps administrators and technical professionals understand Internet Small Computer System Interface (iSCSI) and how to implement it for use with IBM Storwize® storage systems. iSCSI can be used alone or with other technologies.

This publication provides an overview of the iSCSI protocol and helps you understand how it is similar to and different from Fibre Channel (FC) technology. It helps you plan and design your network topology. It explains how to configure your IBM Storwize storage systems and hosts (including IBM AIX®, Linux, VMware, and Microsoft Windows hosts) to interact with it. It also provides an overview of using IBM Storwize storage systems with OpenStack.

This book describes configuring iSCSI for IBM Storwize and SAN Volume Controller storage systems at Version 7.6 or later.

In addition to configuration, this publication provides information about performance and troubleshooting.

Authors

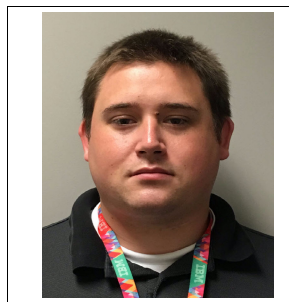
This book was produced by a team of specialists from around the world working in Pune, India.



Jonathan Burton is a software engineer for IBM UK. He works on the SAN Volume Controller and Storwize development team, which he joined as in 2015. Most recently, he worked on the first release of the IBM Spectrum™ Virtualize Software only project. Before joining IBM, he attained degrees both in physics and philosophy.



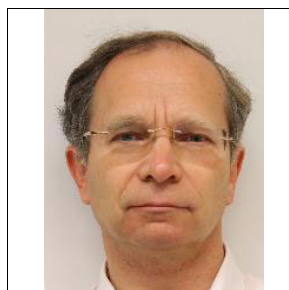
Anuj Chandra works as a development lead for the IBM Spectrum Virtualize™ product family working out of IBM Storage Labs in Pune, India. After joining IBM in 2010, he managed development for various features, and currently leads iSCSI protocol development. He has a bachelor's degree in Computer Engineering and has worked on various storage products in his career over 17+ years.



Jordan Fincher is a Product Field Engineer working in Storage Support at IBM. He received his Bachelor of Science degree in Information Security from Western Governor's University. Jordan first started his IBM career in 2012 as a Systems Engineer for the IBM Business Partner e-TechServices doing pre-sales consulting and implementation work for many IBM accounts in Florida. In 2015, Jordan started working in his current role as a Product Field Engineer for IBM Spectrum Virtualize storage products.

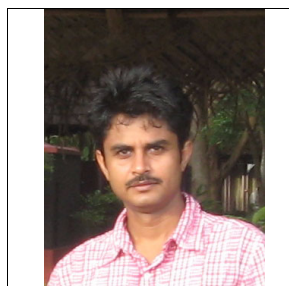


Kushal Patel is a Test Specialist at the IBM India Software Defined Systems and Storage Lab. He has a master's degree in Computer Engineering from the University of Pune, and holds an invention plateau for creative contributions to IBM. He works for the IBM SAN Volume Controller and IBM Storwize team, and possesses expertise in the validation of the iSCSI protocol on SAN Volume Controller and IBM Storwize storage systems.

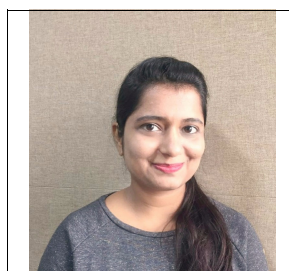


Torsten Rothenwaldt is an IT Specialist at the IBM European Storage Competence Center with a focus on virtualization and high availability (HA) solutions with IBM Spectrum Virtualize products. He is an author or co-author of several IBM Redbooks publications and journal articles.

Torsten's IBM career includes positions in the Software Group and in the IBM System x Server Division. He holds a master's degree in Mathematics. Before joining IBM in 1996, he worked in industrial research, as a software developer, and as a system administrator.



Subhojit Roy is a Senior Technical Staff Member working at IBM India Labs, Pune. He works as a development architect for IBM Spectrum Virtualize products. He has worked on data storage, storage virtualization, and storage networking for 23 years for organizations such as IBM, Veritas, Brocade, and Symantec. Seven of those years have been at IBM, where he works on Ethernet and IP storage architectures, and strategy for IBM Spectrum Virtualize products. He is the lead architect for high-speed Ethernet interconnect to all flash storage that involves iSER and NVMeF. Before he worked at IBM, he developed several key features for enterprise storage products. He is a Master Inventor and Member of the Academy of Technology at IBM.



Bharti Soni is a Staff Systems Software Engineer at IBM India Storage Development Lab. She works on the development of features for IBM SAN Volume Controller and IBM Storwize. She has five years of experience in the storage domain. Her areas of expertise include iSCSI storage protocol, iSer, and networking protocols, such as VLAN and PFC. She has a masters degree in Computer Application from the Indian Institute of Technology, Roorkee.



Shalaka Verma currently heads the IBM Storage Technical Sales Team for the Asia Pacific region. She has experience working with major FSS/Retail customers in India and South Asia on end-to-end infrastructure stack solutions, and drives architectural transformation of enterprise data.



Megan Gilge was a Project Leader at the IBM International Technical Support Organization. Before joining the ITSO, she was an Information Developer in the IBM Semiconductor Solutions and User Technologies areas.

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

Jon Tate

International Technical Support Organization

Anuj Chandra, Carlos Fuente, Virendra Kucheriya, Puja Leekha, Subhojit Roy, Bill Scales, Jeetendra Sonar

IBM Systems

Thanks to the authors of the previous editions of this book.

- Authors of the first edition, *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, published in July 2016, were:

Megan Gilge, Christopher Bogdanowicz, Shweta Kulkarni, Shanmuganathan Kumaravel, Saiprasad Parkar, Mario Rodriguez, Jeetendra Sonar

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 ITSO 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, International Technical Support Organization
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- Follow us on Twitter:

<http://twitter.com/ibmredbooks>

- Look for us on LinkedIn:

<http://www.linkedin.com/groups?home=&gid=2130806>

- 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>

Summary of changes

This section describes the technical changes that were made in this edition of the book and in previous editions. This edition might also include minor corrections and editorial changes that are not identified.

Summary of Changes
for SG24-8327-01

for iSCSI Implementation and Best Practices on IBM Storwize Storage Systems
as created or updated on October 25, 2017.

October 2017, Second Edition

This revision includes the following new and changed information.

New information

- This book was updated to include new information about iSCSI virtualization and updated features in IBM Spectrum Virtualize.



Part 1

iSCSI overview

This part introduces iSCSI and provides information about how it can be implemented with IBM Storwize and IBM SAN Volume Controller.

This part describes the following topics:

- ▶ Chapter 1, “Use cases for iSCSI virtualization” on page 3
- ▶ Chapter 2, “Introduction to iSCSI in IBM Storwize storage systems” on page 7
- ▶ Chapter 3, “External virtualization and host connectivity interface options for the IBM Storwize family” on page 23
- ▶ Chapter 4, “Planning considerations” on page 39
- ▶ Chapter 5, “iSCSI storage connection security” on page 55
- ▶ Chapter 6, “IBM Storwize performance” on page 73



Use cases for iSCSI virtualization

This chapter discusses the scenarios in which Internet Small Computer System Interface (iSCSI) host connectivity and iSCSI virtualization capability can be effectively applied.

This chapter describes three uses cases:

- ▶ 1.1, “Consolidating iSCSI storage consolidation” on page 4
- ▶ 1.2, “Removing capacity silos from data centers” on page 4
- ▶ 1.3, “Improving the performance of iSCSI storage systems” on page 5

1.1 Consolidating iSCSI storage consolidation

This strategy can be used by clients that have much multivendor iSCSI storage and want to simplify management by consolidating it. For larger clients that offer cloud services, this strategy can be a step towards a highly resilient, service-level agreement (SLA)-driven SAN free storage environment.

Consider a data center where there are multivendor iSCSI storage systems. Every storage system has a different management console, different methods of IBM FlashCopy®, and different replication methods. This inconsistency leads to increased operations cost, and also capacity islands across storage.

Virtualizing this storage by using SAN Volume Controller can dramatically minimize the operations impact and improve the return on investment (ROI) through optimal usage.

For all users, SAN Volume Controller becomes a single storage system, which can manage back-end capacity from all heterogeneous storages, which provides a single view. All storage features are enabled at the SAN Volume Controller layer, so all FlashCopy, replication, and other functions can be managed from SAN Volume Controller only.

Because SAN Volume Controller uses native multipath, it also minimizes the impact of managing server environments and dramatically reduces the firmware management in the data center. Using SAN Volume Controller, there is no multipathing conflict across storage systems from the host side, and users can use capacity from any storage in the datacenter. Additionally, SAN Volume Controller enables FlashCopy across storages, so it becomes easy to create multiple copies of data for testing, development, and so on.

This method virtually takes away the need for investing in large Intel nodes, and gives a way to create an SLA-driven, reliable, scalable, and feature-rich iSCSI storage pool with minimal operations impact.

1.2 Removing capacity silos from data centers

This is a useful use case for clients that have mix of both iSCSI and SAN storage systems. Currently, iSCSI storage systems can be used by iSCSI hosts, and SAN storage systems can be used by Fibre Channel (FC) hosts only. This situation can result in suboptimal usage. iSCSI virtualization that uses SAN Volume Controller can enable cross-usage of capacity because FC hosts can use capacity from iSCSI storage systems and vice versa.

Most data centers are run with a mix of iSCSI and SAN storage. Traditionally, certain hosts are connected to iSCSI, and certain hosts are connected to SAN.

When additional capacity is required, for example, a SAN host, and SAN storage does not have that capacity, it must be procured, even if free capacity is available in an iSCSI storage system and vice versa.

Virtualizing an entire storage system under SAN Volume Controller enables hosts to use capacity regardless of the storage type. Basically, iSCSI hosts can use SAN storage, and SAN hosts can use iSCSI storage. This situation enables ROI and gives unprecedented flexibility for storage administrators to manage, monitor, and allocate.

Flash storage can also be under the same SAN Volume Controller and used as a fast tier for any of the other storage systems to improve the performance of workloads that are running on any storage.

iSCSI has more latency than FC storage systems, so it is important to make sure that a response to time-sensitive and performance-hungry critical workloads that operate on FC SAN is possible.

1.3 Improving the performance of iSCSI storage systems

Traditionally, iSCSI storage systems are considered to be high capacity, low-performance storage systems. There is a tendency to put many new/pilot projects on these storage systems to minimize initial investments. However, as these new workloads are migrated to production and become more demanding, the deployed storage system cannot meet the demand. iSCSI virtualization can help with the tiering of these storage systems by using FC flash storage capacity to improve performance and protect investments.

There is unprecedented data growth, and multiple analytics applications come online every day and process data in an attempt to discover useful insights, which can be used for a business advantage.

The data itself uses much capacity, grows every day, and might have limited value. The applications generally are pilots and tests to check the viability and outcome of a certain theory or thought process. Out of these multiple applications, few become relevant to a business, and when they do, the characteristics can change.

When the new workloads are in the pilot or testing phase, there is a tendency to minimize the investment because the risk is high. So, many of these workloads are placed on iSCSI storage systems, which is inexpensive compared to FC arrays. There is hardly any SLA or expectations when applications are in the pilot phase.

However, the workloads that mature quickly become important to a business, and have many SLA and performance expectations regarding them.

iSCSI virtualization can help augment the performance and capabilities of iSCSI storage systems by enabling the tiering of FC flash storage with existing iSCSI storage. iSCSI virtualization also helps enhance existing, low-cost iSCSI storage with enterprise storage features such as disaster recovery (DR), FlashCopy or snapshots, IBM HyperSwap®, IBM Real-time Compression™, and volume mirroring.



Introduction to iSCSI in IBM Storwize storage systems

This chapter provides a beginner's perspective of Internet Small Computer System Interface (iSCSI). It includes considerations for implementing iSCSI storage. It also describes a few iSCSI keywords and their Fibre Channel (FC) and Fibre Channel over Ethernet (FCoE) equivalents. In conclusion, the benefits of deploying an iSCSI-based storage solution are listed, along with things to be considered.

This chapter describes the following topics:

- ▶ 2.1, "What iSCSI is" on page 8
- ▶ 2.2, "iSCSI sessions" on page 9
- ▶ 2.3, "iSCSI adapters" on page 11
- ▶ 2.4, "iSCSI routing" on page 13
- ▶ 2.5, "Ethernet for iSCSI" on page 14
- ▶ 2.6, "Fibre Channel: FCoE terms and their iSCSI equivalents" on page 18
- ▶ 2.7, "Comparison of iSCSI and FCoE" on page 19
- ▶ 2.8, "Why use iSCSI" on page 20

2.1 What iSCSI is

The Small Computer Systems Interface (SCSI) is a family of protocols for connecting and communicating with peripheral devices, such as printers, scanners, tape drives, and hard disk drives (HDDs). SCSI stands on the foundation of client/server architecture and both ends can send SCSI commands and receive responses. The individual I/O devices are called logical units (LUs) and they are identified by logical unit number (LUN). The SCSI target exposes the LUs to the SCSI initiator, which can then query or perform I/O operations on it. A SCSI initiator sends a command in a specific format, the Command Descriptor Block (CDB), and the SCSI target processes it.

iSCSI is a protocol that uses the Transmission Control Protocol and Internet Protocol (TCP/IP) to encapsulate and send SCSI commands to storage devices that are connected to a network. The detailed specification of the iSCSI standard is documented in [RFC3720](#).

iSCSI is used to deliver SCSI commands from a client interface, which is called an *iSCSI Initiator*, to the server interface, which is known as the *iSCSI Target*. The iSCSI payload contains the SCSI CDB and, optionally, data. The target carries out the SCSI commands and sends the response back to the initiator.

In summary, the way iSCSI works is that it encapsulates SCSI commands by adding a special iSCSI header. This header is forwarded to the TCP layer, which creates TCP segments. The TCP segments are further broken down into IP packets, which can be transferred over a local area network (LAN), wide area network (WAN), or the internet in general. Figure 2-1 shows the path that a SCSI command takes when it is transmitted by using iSCSI.

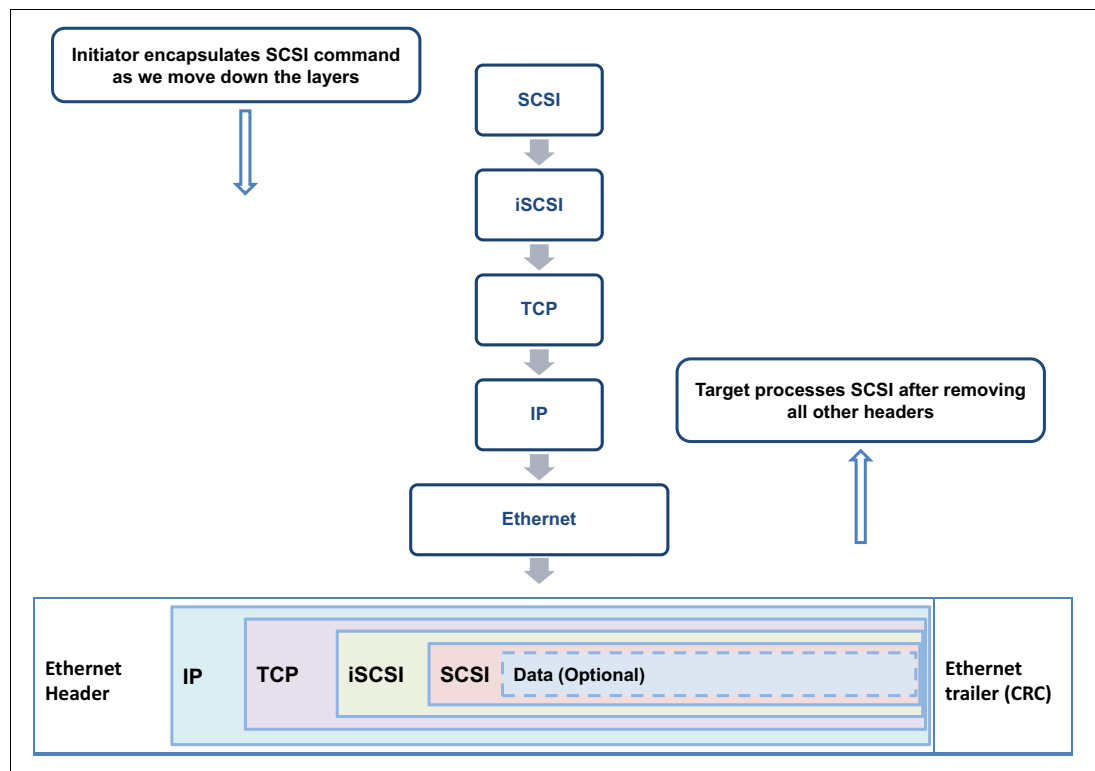


Figure 2-1 iSCSI through the layers and the packet format on the wire

2.2 iSCSI sessions

The iSCSI session is the basic block through which the entire iSCSI layer processing is done. An iSCSI session can be considered equivalent to a SCSI I_T nexus (that is, a path from the initiator to the target). An iSCSI session must be established between an iSCSI initiator and target before the initiator can send any SCSI commands to the target. To process SCSI I/O, the iSCSI initiator establishes an *iSCSI session* with the iSCSI target after agreeing on certain operational parameters. For more information, see 2.2.1, “Components of an iSCSI session” on page 9.

2.2.1 Components of an iSCSI session

An iSCSI session has three main components that help define it.

iSCSI names

iSCSI initiators and targets are identified by their iSCSI names, which can be specified in two formats: the iSCSI qualified name (IQN) and the iSCSI Enterprise Unique Identifier (EUI).

iSCSI qualified name

The IQN is the most commonly used naming mechanism for iSCSI. It has a maximum length of 256 bytes and has the following format beginning with the letters “iqn”:

`iqn.<yyyy-mm>.<reverse-domain-name>:<unique name>`

- ▶ `yyyy-mm` is the year and month when the naming authority was created.
- ▶ `reverse-domain-name` is the domain name of the naming authority in reverse.
- ▶ `unique-name` is a portion of the IQN that can be used by the naming authority to add some meaningful parameters.

`iqn.1986-03.com.ibm:2145.cluster.node1` is an example of an IQN.

iSCSI Enterprise Unique Identifier

The EUI starts with the letters “eui” and has the following format:

`eui.<16 hexadecimal digits>`

Sixteen hexadecimal digits must be used for assigning a globally unique identifier.

iSCSI discovery

iSCSI initiators must identify which targets are present in the system to serve I/O requests. For this purpose, an initiator can run a discovery. There are three supported mechanisms that can be used for discovery: static configuration, **SendTargets**, and iSCSI Name Server (iSNS).

Static configuration

In this mechanism, the initiator already knows the target IP address and port and no real discovery is done. The initiator can directly establish a session. This option can be selected for small, unchanging iSCSI configurations.

SendTargets

With this mechanism, the assumption is that the initiator knows the target’s IP address and the initiator sends a **SendTargets** command to the IP address and the response consists of a list of available targets. The SendTargets mechanism is for suitable for correlatively large configurations.

iSCSI Name Server

In an environment with many targets supporting iSCSI connectivity and many hosts that must connect to the target controllers, configuring target connectivity on each iSCSI initiator host can be cumbersome. The iSCSI protocol enables setting up a mechanism called iSNS. It is a name service mechanism to which all targets register. After a name server is configured on the initiator, the initiator can discover available targets from the name server and establish connectivity to the listed targets.

iSCSI login

iSCSI enables two kinds of logins.

Discovery session

This session is a restricted-access-only type of session in which initiators can discover only the targets. The initiator specifies that the session type is Discovery. The target accepts only requests with the SendTargets key to send back a list of targets to the initiator and log out requests to close the session.

Normal operational session

In this type of session, all iSCSI commands are accepted, and responded to, by the target.

2.2.2 The three phases of iSCSI login

After an iSCSI initiator discovers all the targets that it can communicate with, an iSCSI login must be done before data transfer can begin. An iSCSI login establishes a TCP session between the iSCSI initiator and the iSCSI target. The iSCSI target listens on a TCP port and the initiator begins the login by sending a connect request. Authentication and negotiation of supported session parameters are carried out. This process is done in three phases: security negotiation, operational parameter negotiation, and full feature phase.

Security negotiation

In this phase of the iSCSI login, the initiator sends its list of supported authentication methods and the one to be used is determined (the most popular one is CHAP). The initiator and target authenticate each other by using the selected method and the next phase can now begin.

Operational parameter negotiation

Operational parameter negotiation might be the first phase if security negotiation is skipped. This exchange goes on in the form of login requests and responses until both parties agree on the operational parameters, which include (but are not limited to) the following things:

- ▶ Header digest
- ▶ Data digest
- ▶ Immediate data
- ▶ MaxRecvDataSegmentLength
- ▶ MaxConnections

For the complete list of operational parameters, see the [IETF website](#).

Full feature phase

After authentication and parameter setting is done, the TCP connection is established and the iSCSI session can proceed. The initiator starts sending SCSI commands and data to LUNs in the form of iSCSI Protocol Data Units (PDUs).

SAN Volume Controller and IBM Storwize storage systems enable you to advance to the full-feature phase directly from operational negotiation.

2.3 iSCSI adapters

Three types of iSCSI adapter are most commonly deployed in iSCSI storage solutions. Latency and performance vary depending on which adapter is chosen. This section lists the impact of each type in terms of cost and performance.

2.3.1 Ethernet card (network interface card)

Figure 2-2 shows an Ethernet card.

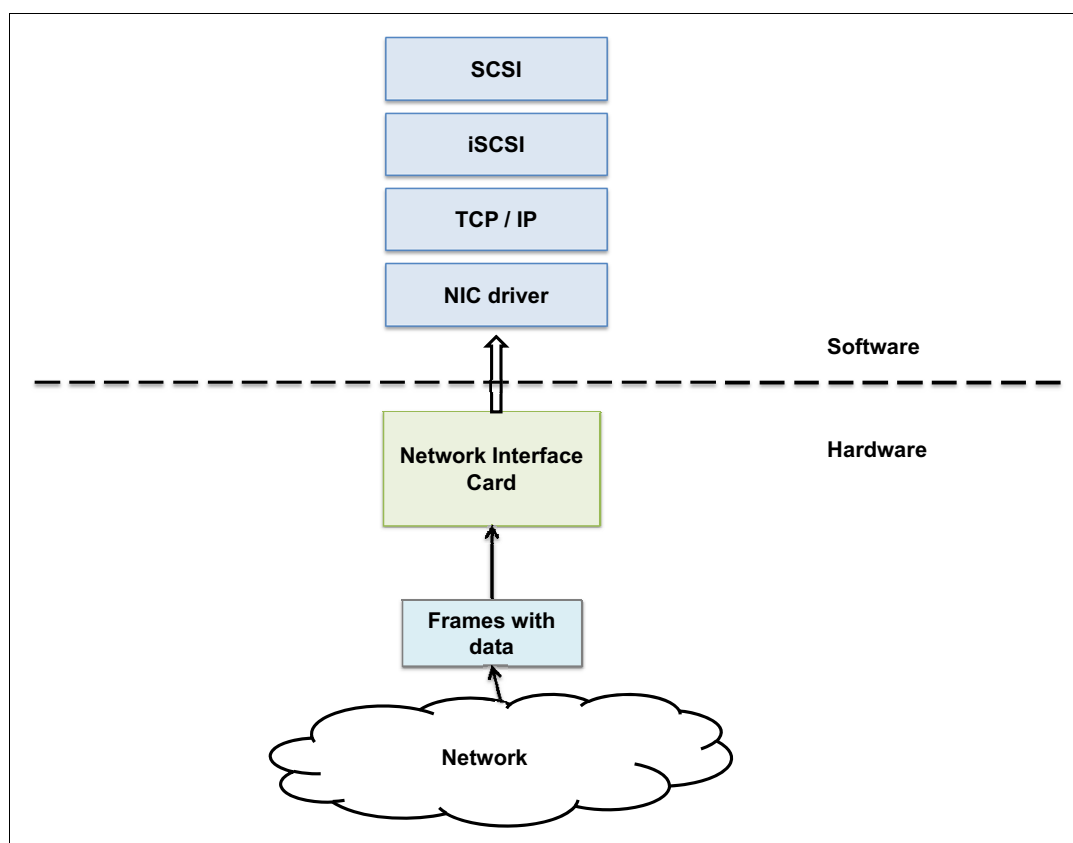


Figure 2-2 iSCSI implemented directly with an Ethernet card

Ethernet cards are designed to transfer IP packets. Therefore, to use Ethernet to send SCSI commands and data, it must first be packetized such that the adapter can process and forward it. The iSCSI protocol assists the server to achieve this task, but the entire iSCSI protocol processing is done in software before the operating system's TCP/IP handler code is called. This method is processor-intensive and reduces the overall performance of the server. It leads to increased latencies, and the performance also is affected in cases where the Ethernet card might be getting traffic from other applications that share the interface.

2.3.2 TCP offload engine

Figure 2-3 shows a TCP offload engine (TOE).

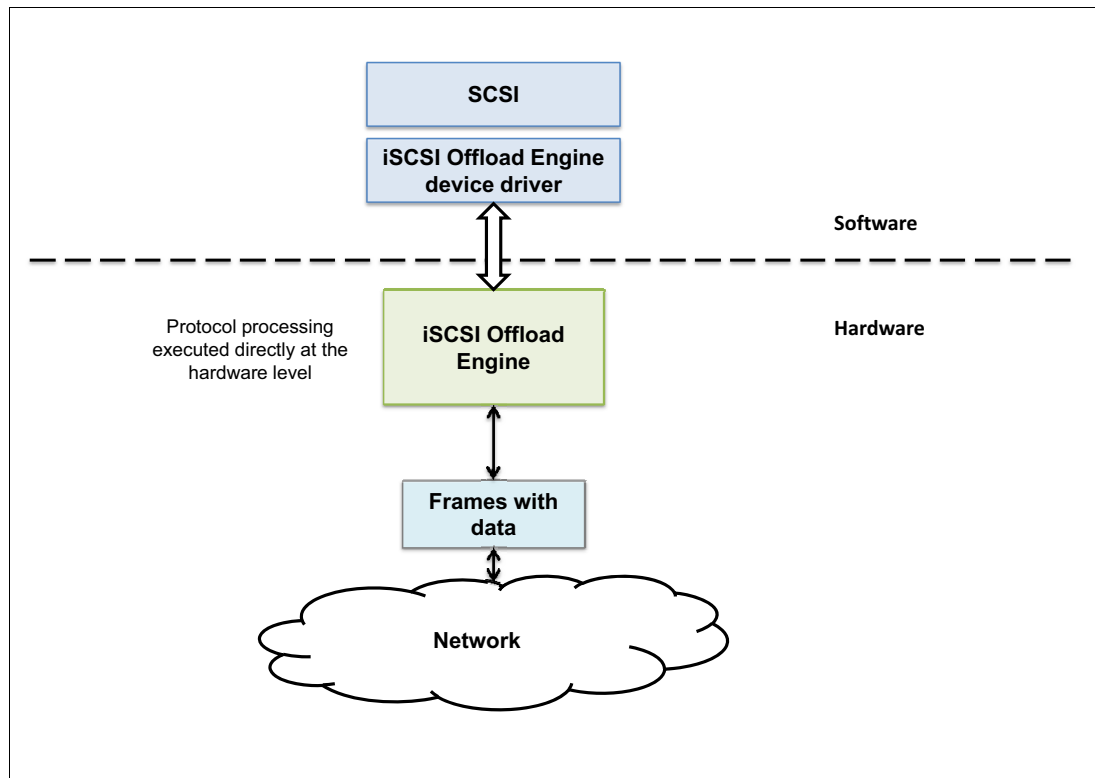


Figure 2-3 iSCSI implemented with a TCP offload engine

The TOE interface card is more sophisticated in terms of the processing capabilities, and most of the TCP packet processing is done by specialized hardware that is built into the adapter. This implementation means that the TOE is better than the NIC when compared on a performance or latency basis. The TOE is also more expensive than the NIC.

2.3.3 iSCSI offload engine

Figure 2-4 shows an iSCSI offload engine.

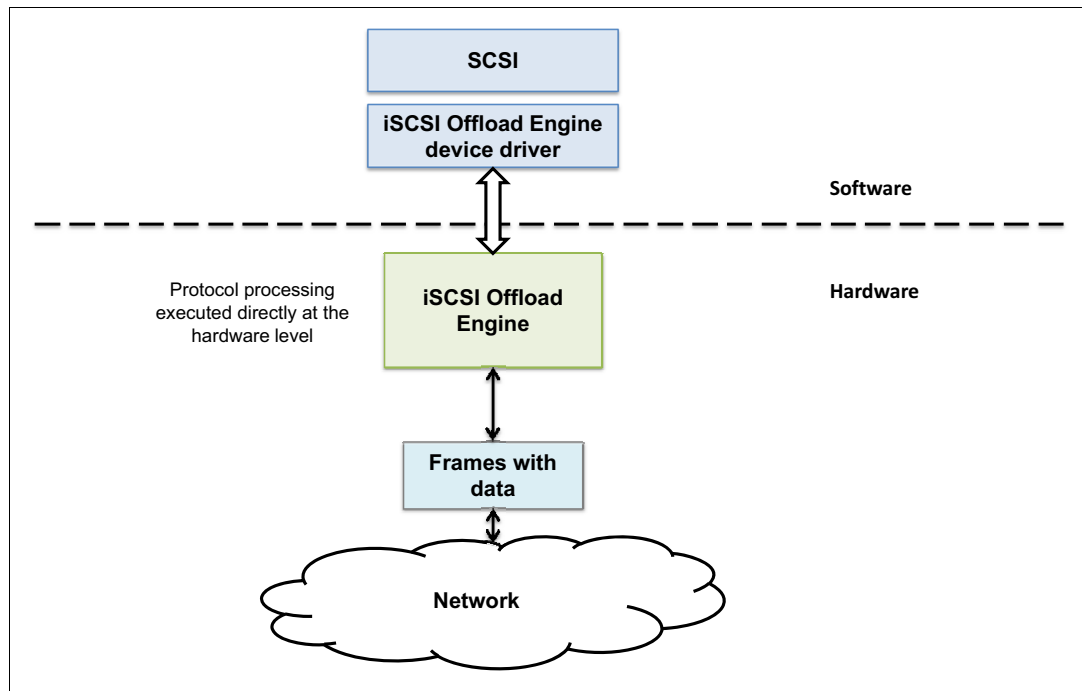


Figure 2-4 iSCSI offload engine

The iSCSI offload engine sends all of the iSCSI protocol processing to the iSCSI-specific hardware that is built into the iSCSI offload engine card. This implementation provides the least latency and best performance of the three. The iSCSI offload engine is expensive because most functions are implemented in hardware, but the latency is reduced because there is little or no processing in the operating system kernel.

2.4 iSCSI routing

Internet Protocol (IP) has defined and widely used routing standards. iSCSI relies on the IP protocol for its routing requirements.

FCoE and iSCSI can coexist in the same data center if an iSCSI gateway is used to route traffic. An iSCSI gateway is a device that facilitates conversion between FCoE and iSCSI. It has FCoE ports in addition to normal Ethernet ports that provide connectivity for the TCP/IP protocols. An iSCSI gateway exports FC LUNs as iSCSI targets to provide integration with use of fewer cables and host bus adapters (HBAs).

Figure 2-5 shows an example of how iSCSI routing can be done.

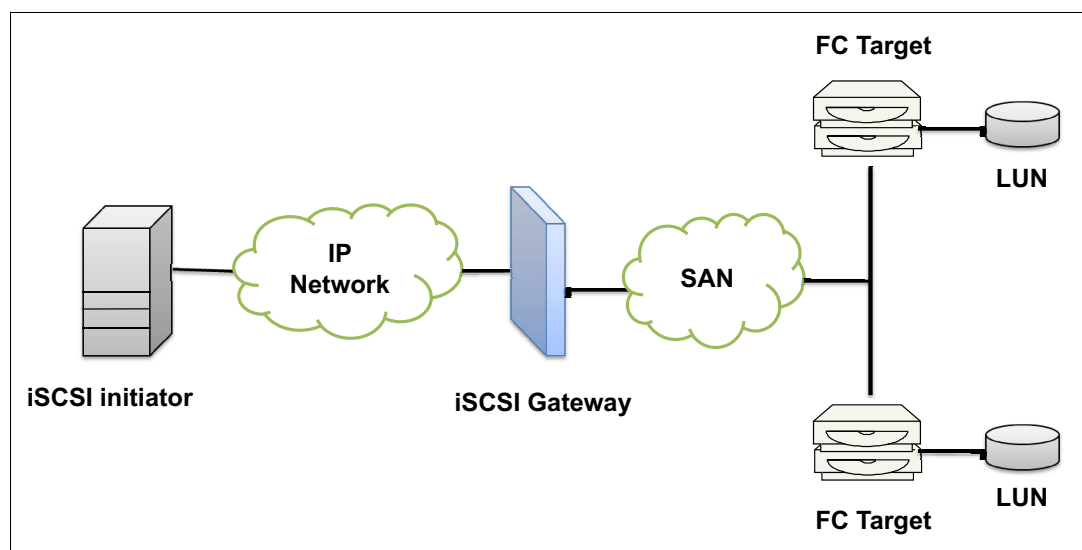


Figure 2-5 iSCSI gateway

2.5 Ethernet for iSCSI

In traditional deployments of Ethernet, frames that are lost (either because they are dropped or because a collision occurred) are retransmitted. This loss is acceptable in normal non-storage data networks because the application can usually afford to wait for the frame retransmission. For storage traffic, lost frames are more harmful because the application suffers from large I/O latencies. Thus, even though iSCSI can handle the best-effort delivery of Ethernet networks, reduced retransmission significantly improves iSCSI performance.

In its early days, iSCSI was deployed only on 1-Gigabit Ethernet (GbE). This implementation gave a theoretical maximum performance of 125 MBps. (This theoretical maximum performance cannot be achieved because there is processing impact for each of the three protocols, that is, iSCSI, TCP, and IP.) With 10 GbE, mechanisms can be used to prevent retransmission of frames. The Converged Enhanced Ethernet (CEE) (also known as Data Center Bridging (DCB)) standard was developed to consolidate FC and Ethernet networks to provide lossless connectivity in the data center with the fewest number of cables. DCB is not supported on 1 GbE.

The following section explains DCB and other trends in Ethernet networks at the time of writing. It also details a few challenges and opportunities as newer and better Ethernet technology emerges.

2.5.1 Data Center Bridging

DCB is a set of standards that are defined by the Institute of Electrical and Electronics Engineers (IEEE) Task Group to enhance existing 802.1 bridge standards. This enhancement is done by improving link robustness and enabling a 10 GbE link to support multiple traffic types simultaneously while preserving their respective traffic properties.

The goal of DCB is to improve the Ethernet protocol so it becomes lossless by eliminating packet loss due to queue overflow. This scenario is known as *lossless Ethernet*.

The DCB standards include Priority Flow Control (PFC), Enhanced Transmission Selection (ETS), Congestion Notification (CN), and Data Center Bridging Exchange (DCBx).

Priority Flow Control (IEEE standard 802.1 Qbb)

In traditional Ethernet networks, a transmitter can send frames faster than a receiver accepts them, which means that if the receiver runs out of available buffer space to store incoming frames for further processing, it is forced to drop all frames arriving, leading to retransmission. The solution to avoid retransmission is to pause traffic when it exceeds the receiver's capacity.

The traditional Ethernet flow control uses a PAUSE mechanism. If the port becomes busy, the switch manages congestion by pausing all the traffic on the port, regardless of traffic type.

PFC can individually pause traffic according to the tags that are assigned, and it facilitates lossless or no-drop behavior for a priority at the receiving port. Lossless behavior, when implemented end-to-end on a network, controls dropping frames during congestion by pausing traffic types that use PFC. Each frame that is transmitted by a sending port is tagged with a priority value (0 - 7) in the virtual local area network (VLAN) tag.

Figure 2-6 shows how PFC works. It divides the available 10 GbE bandwidth into eight different virtual lanes, with each lane assigned a priority level. If bursts of heavy congestion occur, lower priority traffic can be paused. In this example, lane 3 is paused while the rest of the lanes allow flow of traffic.

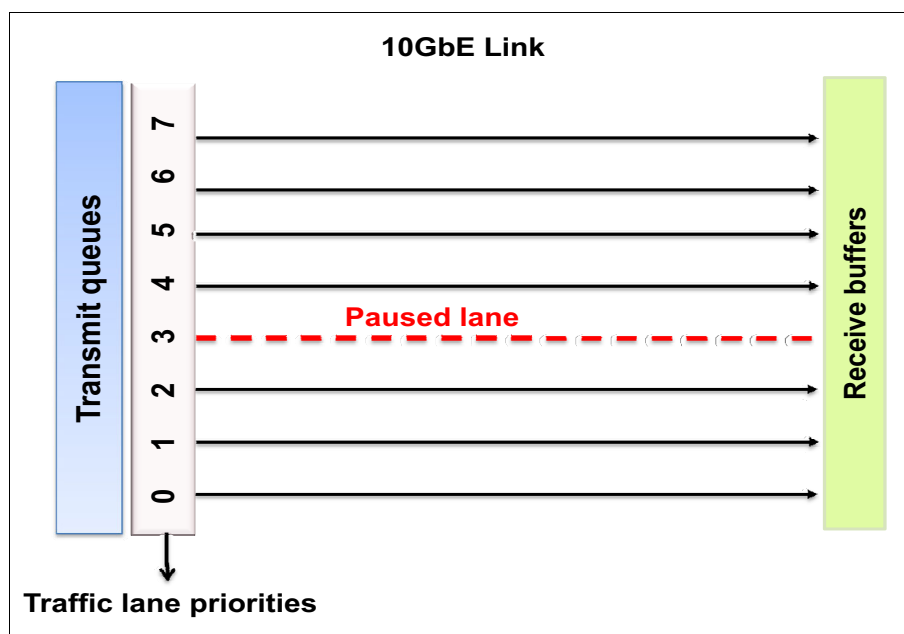


Figure 2-6 Priority Flow Control by pausing virtual lane 3

PFC provides fine-grained flow control. The switch pauses certain traffic types that are based on 802.1p Class of Service (CoS) values in the VLAN tag.

PFC works together with ETS, which is described in “Enhanced Transmission Selection (IEEE802.1 Qaz)” on page 16.

For more information about the PFC standard, see the [IEEE 802 website](#).

Enhanced Transmission Selection (IEEE802.1 Qaz)

ETS is used to allocate link bandwidth between different traffic classes. With ETS enabled, bandwidth allocation is carried out based on the 802.1p priority values in the VLAN tag. It is possible to combine multiple priority values into traffic groups or classes. The important traffic can be assigned high priorities and ensured bandwidths. To improve the overall network efficiency, ETS allows lower priority traffic to use unused bandwidth from the high-priority queues and to exceed their own bandwidth guarantees.

Figure 2-7 shows an example of adding priorities values for each type of traffic.

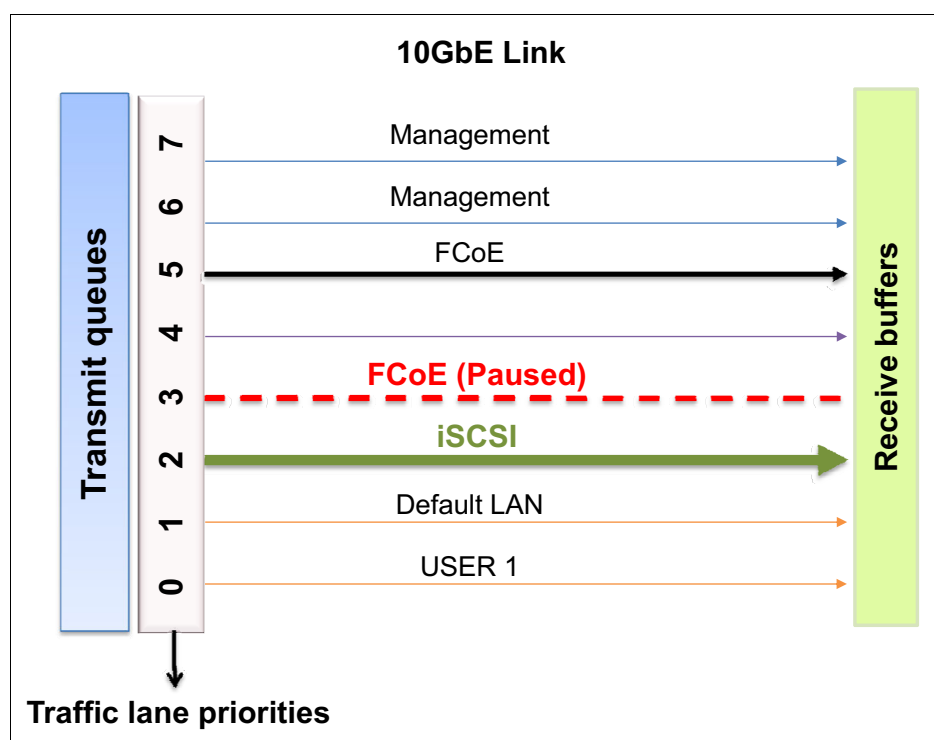


Figure 2-7 Enhanced Transmission Selection working with Priority Flow Control to pause specific types of traffic

For more information about ETS, see the [IEEE 802 website](#).

Congestion Notification (IEEE 802.1Qau)

CN is a flow control protocol for Layer 2 networks to eliminate heavy congestion due to long-lived traffic flows by throttling frame transmission. The congestion point, which can be a network switch or end-device port, can request that ingress ports limit their speed of transmissions when congestion is occurring. When the congestion ends, the ingress ports can increase their speed of transmission again. This process allows traffic flows to be throttled at the source to react to or prevent congestion by having a temporary reduction in transmission rate. This reduction is preferred to lost packets, which can cause long timeouts. This feature must be considered for large-scale environments with multi-hop networks.

Figure 2-8 shows a basic example of what occurs when CN is used.

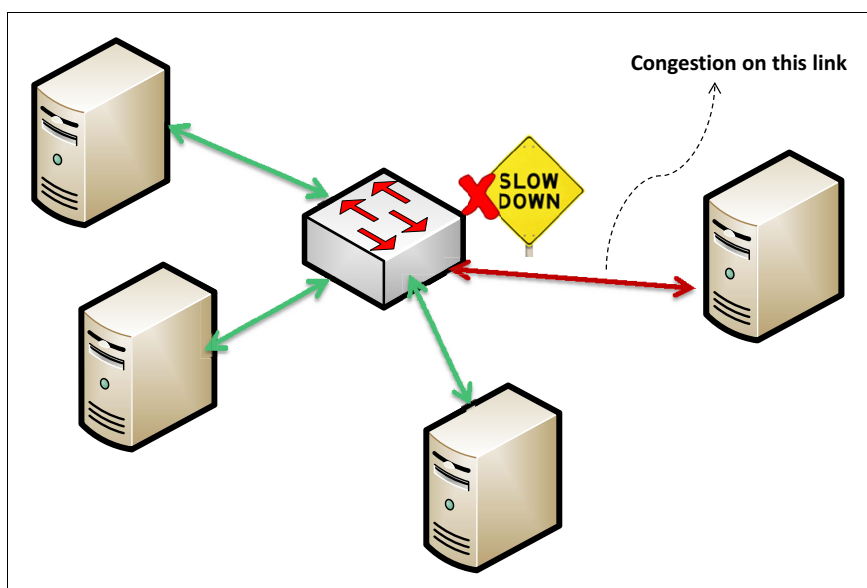


Figure 2-8 An example of Congestion Notification in a network

For more information, see the [IEEE 802 website](#).

Data Center Bridging Exchange

DCBx is a protocol that is used by DCB devices to exchange configuration information with directly connected peers to ensure a consistent configuration across the network.

DCBx uses Link Layer Discovery Protocol (LLDP) to exchange parameters between two link peers to learn about the capabilities of the other peer. For example, two link peer devices support PFC. PFC is described in “Priority Flow Control (IEEE standard 802.1 Qbb)” on page 15.

This protocol also can be used to detect misconfiguration of a feature between the peers on a link and to configure DCB features in its link peer.

For more information about this protocol, see the [IEEE 802 website](#).

2.5.2 The future of Ethernet and its impact on iSCSI

An IEEE study group has been formed to research 25 GbE and plans to achieve higher bandwidth in multiples of 25 Gbps. With Ethernet speeds increasing up to 40 Gbps and 100 Gbps and becoming commercially viable, new challenges for iSCSI storage must be addressed. The processor becomes a bottleneck for software-based iSCSI, necessitating the use of a TOE or mechanisms such as iSCSI over Remote Direct Memory Access (RDMA). The protocol-processing impact on I/O performance must be analyzed and resolved.

To address the need for higher bandwidths and lower latencies for iSCSI interconnects, new protocols that extend RDMA to Ethernet interconnects are being developed, such as internet Wide Area RDMA Protocol (iWARP) and RDMA over Converged Ethernet (RoCE). iSCSI Extensions for RDMA (iSER) is a new standard that enables iSCSI hosts and targets to take advantage of RDMA capabilities. iSER can run on top of any RDMA capable Network Interface Card (rNIC) regardless of the protocol, that is, iWARP or RoCE (V1 or V2). These new technologies are being adopted as higher Ethernet bandwidths, such as 25, 40, 50, and 100 Gbps, gain acceptance.

2.6 Fibre Channel: FCoE terms and their iSCSI equivalents

This section describes a few FC concepts and their iSCSI equivalents.

2.6.1 Fibre Channel zoning

FC zoning is a method to partition the switched fabric so that it restricts the visibility of certain FC endpoints and isolates devices into two zones. It is used to simplify security and management of the fabric.

iSCSI does not provide zoning.

2.6.2 Virtual SAN

Virtual fabric or Virtual SAN (vSAN) is a set of ports that is selected from a set of connected switches. Both FC zones and a vSAN can be used for isolation of traffic, but the key difference is that in a vSAN all the FC services are replicated within the switch so that it can act as a self-sufficient SAN.

This function is similar to VLAN in iSCSI and it helps for easier administration of a large SAN.

2.6.3 Buffer-to-Buffer credit

Buffer-to-Buffer credits (BB credits) are used in FC deployments for flow control, much like PFC is used in iSCSI. Each time an FC port transmits data, its BB credit is decremented by one and it is incremented when a recipient issues it some credits. A port with zero credits cannot transmit again until it obtains credits.

2.6.4 Worldwide name

A worldwide name (WWN) is a unique 64-bit identifier that is assigned to an FC device. It can either be a worldwide port name (WWPN) or worldwide node name (WWNN). A WWN consists of *Network Address Authority* (NAA) bits, usually followed by an *Organizationally Unique Identifier* (OUI). It can loosely be equated to an *iSCSI IQN* because an IQN uniquely identifies an iSCSI device.

2.6.5 Fabric name server

The fabric name server is a database of all the devices that are attached to a network fabric. An FC host can query the fabric name server to obtain information about a particular FC device. It is mandatory for operation. It is equivalent to iSNS. iSNS reduces the possibility of human error and provides for easier maintenance of the fabric.

2.7 Comparison of iSCSI and FCoE

Figure 2-9 shows the FCoE packet format.

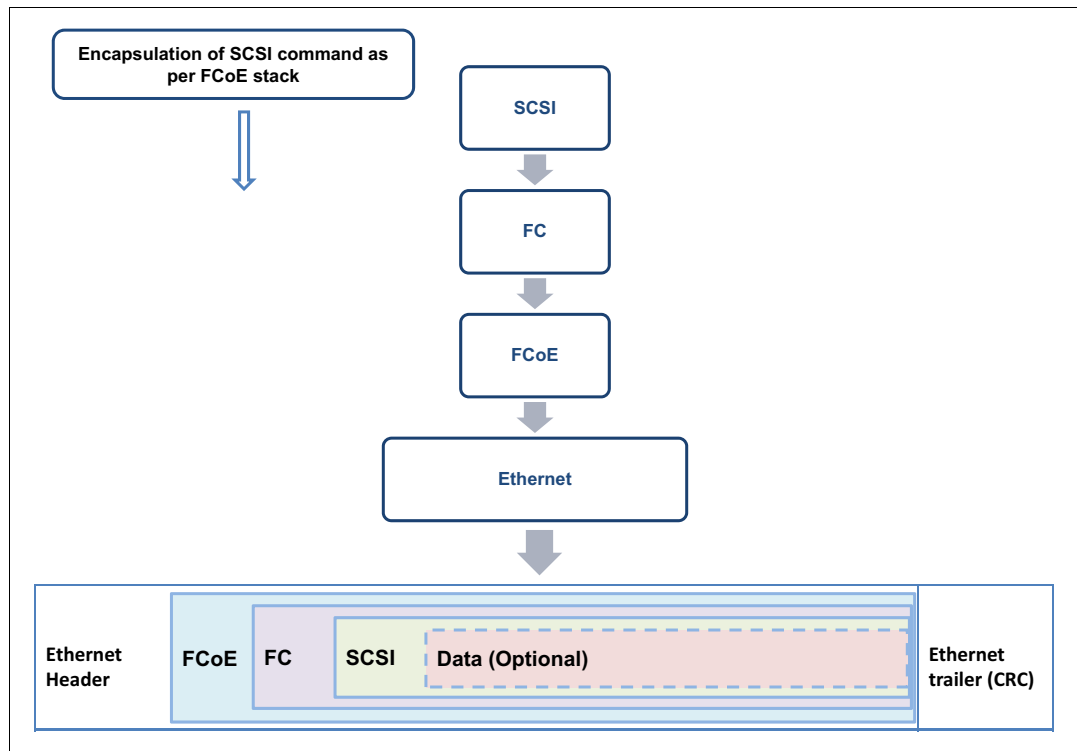


Figure 2-9 FCoE protocol stack and packet format

Although both FCoE and iSCSI operate on Ethernet, there are several differences in the manner in which they work. Table 2-1 lists the key differences between iSCSI and FCoE.

Table 2-1 Comparison of FCoE and iSCSI

FCoE	iSCSI
FCoE enables the encapsulation of FC frames over Ethernet networks. The underlying protocol is not TCP/IP.	iSCSI encapsulates SCSI into the TCP/IP format.
FCoE is not routable because it does not have IP headers. FCoE works within a subnet.	iSCSI can be routed based on IP headers, and iSCSI payloads can be carried beyond a subnet through a gateway.
Practical implementations of FCoE <i>require</i> DCBx and PFC.	Although they are good to have, iSCSI implementations can do without DCBx and PFC.

FCoE	iSCSI
In practice, for a usable FCoE solution, a fabric name server is <i>mandatory</i> .	Small and medium iSCSI solutions can work well without iSNS.
FCoE solutions that are commercially viable need firmware-based implementation in part.	iSCSI can be implemented completely in software.

2.8 Why use iSCSI

This section lists a few advantages of implementing an iSCSI-based storage solution.

2.8.1 iSCSI is cost-effective

iSCSI is a cost-effective storage solution because it uses existing hardware and network elements.

Cost of installation

iSCSI does not require expensive proprietary hardware on which to run. It does not need dedicated cabling and switches like FC. iSCSI can be implemented on standard Ethernet network hardware. Almost all organizations, including small and medium businesses, already have Ethernet network and cabling.

Maintenance and expansion costs

Replacement of parts of the network and burned-out hardware is inexpensive, which reduces the cost of maintaining the data center. Also, capacity expansion can be easily achieved by acquiring new disk arrays.

Administrative costs

Data center and network administrators are well-versed with TCP/IP configurations. Therefore, iSCSI has a natural advantage because it is implemented over IP. The cost of training staff in iSCSI can be lower than for other technologies.

2.8.2 No distance limitations

With the internet being so ubiquitous, it is possible to implement iSCSI storage such that the data center can be miles away from the application server. Also, iSCSI-based disaster recovery (DR) solutions over long distances are simplified, which are an affordable alternative to FC DR setups that require high-priced optical cables to be laid out from the primary site to the secondary site.

2.8.3 Good interoperability

iSCSI does not require specialized cabling or switches like FC. An iSCSI HBA provides Ethernet connectivity to storage devices and only the higher-level protocols are aware of iSCSI, and the transport layer (and layers lower than transport) treats the iSCSI packets as payload. iSCSI also provides good interoperability with equipment from multiple vendors because IP and Ethernet are common industry standards.

2.8.4 Bandwidth usage and Converged Enhanced Ethernet benefits

CEE provides a consolidated transport for both storage and networking traffic, which leads to better bandwidth usage in the data center. Implementation of lossless Ethernet over 10-Gigabit Ethernet, as described in 2.5, “Ethernet for iSCSI” on page 14, provides lower latency and improves performance. As servers deploy better processors for which the bus is no longer the bottleneck and commodity internet reaches higher speeds, converged Ethernet helps provide maximum performance with optimum resource usage.

2.8.5 Security

RFC 3720 lists six methods that are supported by iSCSI to provide security through authentication. The iSCSI initiator and target agree upon one of the six methods when the iSCSI session is established.

The most-widely used method that iSCSI uses to provide security is through Challenge Handshake Authentication Protocol (CHAP). CHAP limits an initiator’s access to volumes by using a challenge-response authentication mechanism. There are two ways CHAP can be set up: one-way CHAP and mutual CHAP.

For more information, see Chapter 5, “iSCSI storage connection security” on page 55.



External virtualization and host connectivity interface options for the IBM Storwize family

This chapter describes various external virtualization and host connectivity interface options of IBM Storwize products, and IBM SAN Volume Controller storage systems.

This chapter describes the following topics:

- ▶ 3.1, “Connectivity options for the IBM Storwize V5000 Gen2 storage system” on page 24
- ▶ 3.2, “Connectivity options for the IBM Storwize V7000 storage system” on page 28
- ▶ 3.3, “The IBM Storwize V7000 Unified storage system” on page 32
- ▶ 3.4, “SAN Volume Controller SV1 storage systems” on page 33
- ▶ 3.5, “Hardware terminology for the IBM Storwize disk systems” on page 35

3.1 Connectivity options for the IBM Storwize V5000 Gen2 storage system

The IBM Storwize V5000 Gen2 storage system is a flexible, scalable, and easy-to-use storage system, which is built with IBM Spectrum Virtualize software. There are multiple models, which support both all-flash and hybrid storage solutions, with varying performance and scalability.

The IBM Storwize V5000 Gen2 storage system has three hybrid models: IBM Storwize V5010, IBM Storwize V5020, and IBM Storwize V5030. IBM Storwize V5030F is an all flash model, which enables superior performance.

3.1.1 Connectivity options for the IBM Storwize V5010 storage system

The IBM Storwize V5010 storage system offers 16 GB of cache and up to 10 standard expansion enclosures or up to four high-density expansions. It can scale up to a maximum of 392 drives, and supports IBM Easy Tier®, FlashCopy, and Remote Mirroring.

Each IBM Storwize V5010 controller has by default two 1 Gbps iSCSI ports, so four ports per controller pair. One of the ports is marked as the T port, and should be used for the initial setup. After the initial setup, the T port can be used like any other iSCSI port. It is preferable to dedicate one port for management. Therefore, each controller pair has a total of three 1 Gbps iSCSI ports for host connectivity by default. Each controller also has one 12 Gbps SAS port that can be used for connecting expansion enclosures.

Each controller can be optionally configured with an additional interface card. The additional card can be quad-port 12 Gbps SAS, quad-port 16 Gbps FC, or quad-port 10 Gbps (optical) iSCSI/FCoE. Both controllers in each pair should be configured with identical interface cards.

External storage virtualization is not supported by the IBM Storwize V5010 storage system.

Figure 3-1 shows external connectivity interface options for the IBM Storwize V5010 storage system.

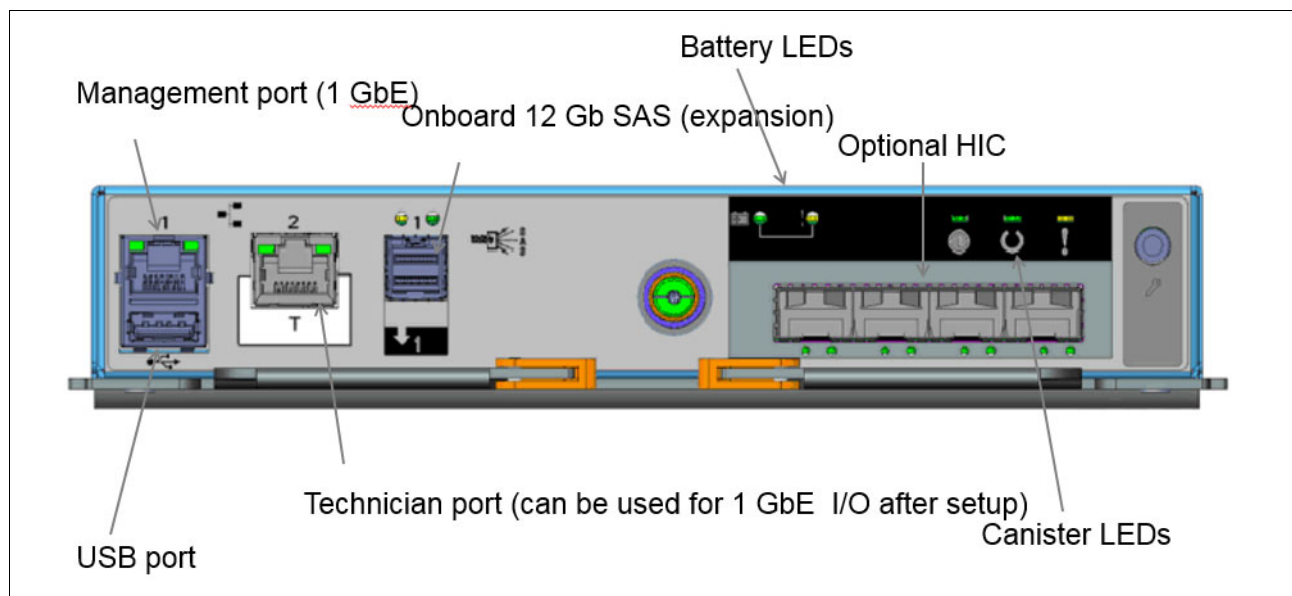


Figure 3-1 External connectivity interfaces for the IBM Storwize V5010 storage system

Table 3-1 provides details about the external connectivity interfaces for the IBM Storwize V5010 storage system.

Table 3-1 External connectivity interfaces per IBM Storwize V5010 Controller

Interface specification	Configuration type
Dual-port 1 Gbps iSCSI	Default
Quad-port 16 Gbps FC	Configurable ^a
Quad-port 10 Gbps iSCSI/FCoE	Configurable ^a
Quad-port SAS	Configurable ^a
Quad Port 1 Gbps iSCSI	Configurable ^a

a. Any one of the types of adapter can be configured. Both controllers must have an identical configuration.

3.1.2 Connectivity options for the IBM Storwize V5020 storage system

The IBM Storwize V5020 storage system offers up to 32 GB of cache, and up to 10 standard expansion enclosures *or* up to four high-density expansions. It can scale to a maximum of 392 drives, and supports encryption and higher performance in addition to IBM Storwize V5010 storage system capabilities.

Each IBM Storwize V5020 controller has by default two 1 Gbps iSCSI ports, so four ports per controller pair. One of the ports is marked as the T port, and should be used for the initial setup. After the initial setup, the T port can be used like any other iSCSI port.

It is preferable to dedicate one port for management. Therefore, each controller pair has a total of three 1 Gbps iSCSI ports for host connectivity by default. Each controller also has three 12 Gbps SAS ports. It is preferable to reserve one port for connecting expansion enclosures. The other two can be used for direct SAS host connectivity.

Each controller can be optionally configured with an additional interface card. The additional card can be quad-port 12 Gbps SAS, quad-port 16 Gbps FC, or quad-port 10 Gbps (optical) iSCSI/FCoE. Both controllers in each pair should be configured with identical interface cards.

External storage virtualization is not supported by the IBM Storwize V5020 storage system.

Figure 3-2 shows external connectivity interface options for the IBM Storwize V5020 storage system.

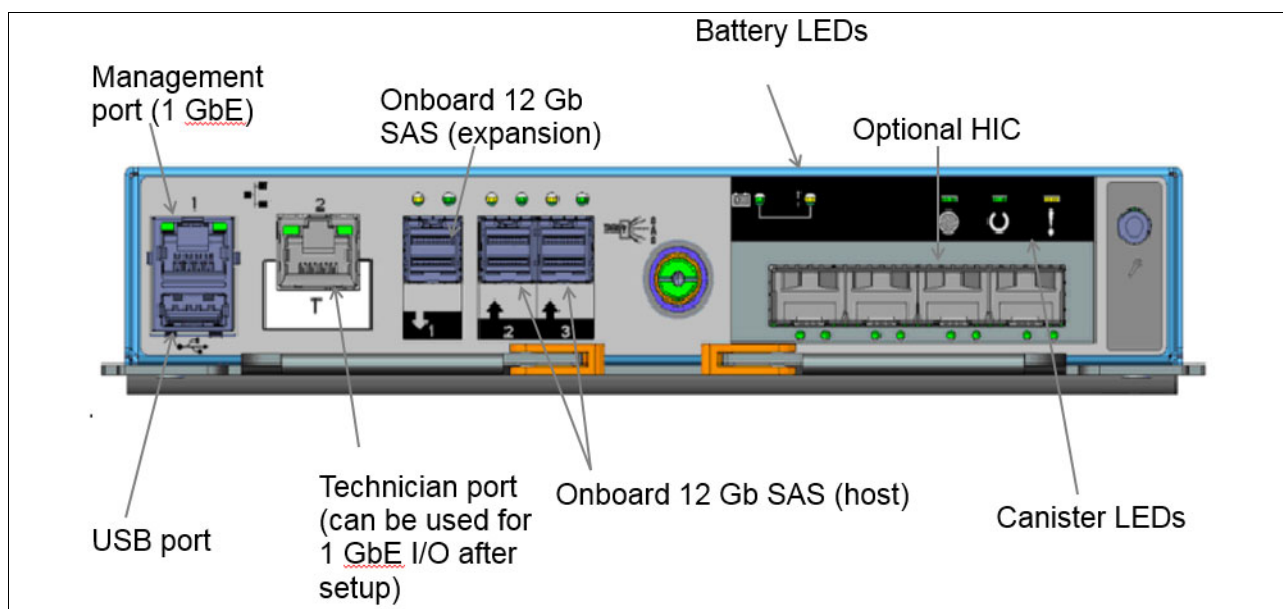


Figure 3-2 External connectivity interfaces for the IBM Storwize V5020 storage system

Table 3-2 provides details about the external connectivity interfaces for the IBM Storwize V5020 storage system.

Table 3-2 External connectivity interfaces per IBM Storwize V5020 controller

Interface Configuration	Configuration Type
Dual-port 1 Gbps iSCSI	Default
Three-port 12 Gbps SAS	Default
Quad-port 10 Gbps iSCSI/FCoE	Configurable ^a
Quad-port 12 Gbps SAS	Configurable ^a
Quad-port 16 Gbps FC	Configurable ^a
Quad-port 1 Gbps iSCSI	Configurable ^a

a. Any one of the types of adapter can be configured. Both controllers must have an identical configuration.

3.1.3 Connectivity options for IBM Storwize V5030 and IBM Storwize V5030F storage systems

The IBM Storwize V5030 and IBM Storwize V5030F storage systems offer up to 32 GB of cache, and up to 20 standard expansion enclosures *or* up to four high-density expansions per control enclosure. They support the clustering of two control enclosures. In a clustered configuration, they scale up to a maximum of 1,056 drives. The Storwize V5030 and IBM Storwize V5030F storage systems support external virtualization, compression, HyperSwap, and higher performance and scalability, in addition to the IBM Storwize V5020 storage system capabilities.

Each IBM Storwize V5030 or IBM Storwize V5030F controller has by default two 10 Gbps (copper) iSCSI ports, so four ports per controller pair. Each controller also has one 1 Gbps iSCSI port, also marked as the T port. It is preferable to dedicate the T port for management after the initial setup is complete.

Therefore, each controller pair has a total of four 10 Gbps iSCSI ports for external connectivity by default. Each controller also has two 12 Gbps SAS ports to be used for connecting expansion enclosures.

Each controller can be optionally configured with an additional interface card. The additional card can be quad-port 12 Gbps SAS, quad-port 16 Gbps FC, or quad port 10 Gbps (optical) iSCSI/FCoE. Both controllers in each pair should be configured with identical interface cards.

All the external connectivity ports can be used for host connectivity or external storage virtualization.

Figure 3-3 shows the external connectivity interface options for the IBM Storwize V5030 and IBM Storwize V5030F storage systems.

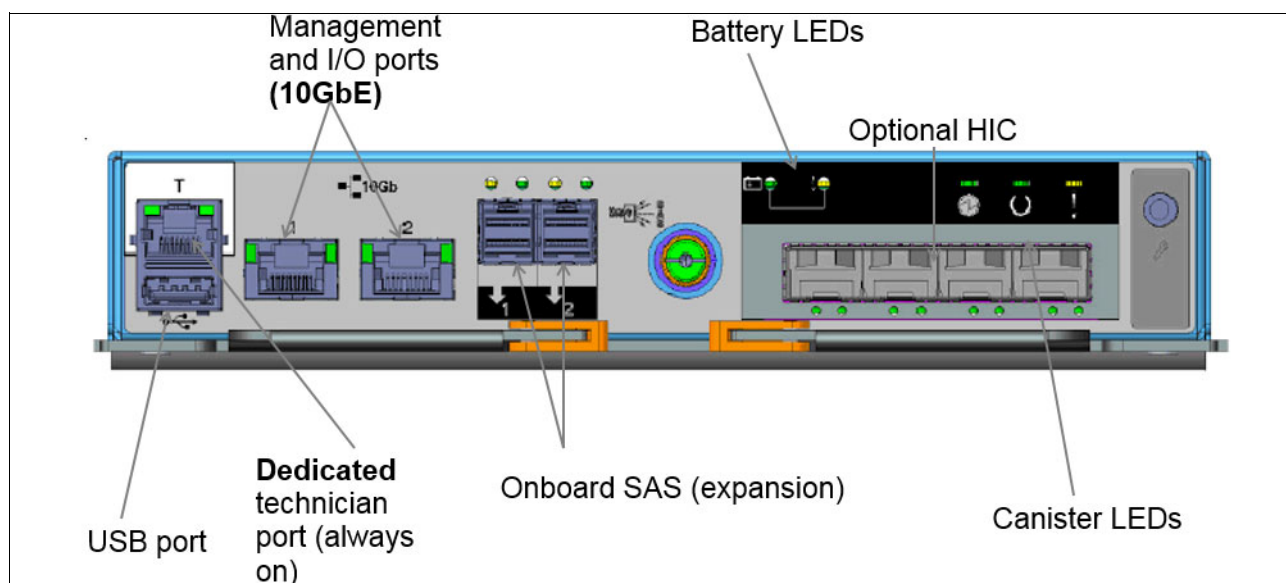


Figure 3-3 External connectivity interfaces for the IBM Storwize V5030 and IBM Storwize V5030F storage systems

Table 3-3 provides details about the external connectivity interfaces for the IBM Storwize V5030 and IBM Storwize V5030F storage systems.

Table 3-3 External connectivity interfaces for the IBM Storwize V5030 and IBM Storwize V5030F storage systems

Interface configuration	Configuration type
Dual-port 10 Gbps iSCSI	Default
Dual-port 12 Gbps SAS	Default
Single-port 1 Gbps iSCSI	Default
Quad-port 10 Gbps iSCSI/FCoE	Configurable ^a
Quad-port 12 Gbps SAS	Configurable ^a

Interface configuration	Configuration type
Quad-port 16 Gbps FC	Configurable ^a
Quad-port 1 Gbps iSCSI	Configurable ^a

a. Any one of these types of adapter can be configured. Both controllers must have an identical configuration.

3.1.4 IBM Storwize V5010, IBM Storwize V5020, and IBM Storwize V5030 HIC options at a glance

Table 3-4 shows the various external connectivity options for the IBM Storwize V5010, IBM Storwize V5020, and IBM Storwize V5030 storage systems.

Table 3-4 Overview of the external connectivity options

Storage system	1 Gbps iSCSI ports	10 Gbps copper iSCSI ports	10 Gbps optical SFP+ iSCSI/FCoE ports	16 Gbps FC ports	12 Gbps SAS ports
IBM Storwize V5030	8 (Configurable)	4 (Default)	8 (Configurable)	8 (Configurable)	8 (Configurable)
IBM Storwize V5020	4 (Default) and 8 (Configurable)	N/A	8 (Configurable)	8 (Configurable)	4 (Default) and 8 (Configurable)
IBM Storwize V5010	4 (Default) and 8 (Configurable)	N/A	8 (Configurable)	8 (Configurable)	8 (Configurable)

Note: For more information, see *Implementing the IBM Storwize V5000 Gen2 (including the Storwize V5010, V5020, and V5030)*, SG24-8162. For older generations of the IBM Storwize V5000 storage system, see [IBM Knowledge Center](#).

3.2 Connectivity options for the IBM Storwize V7000 storage system

The IBM Storwize V7000 storage system is a flexible, highly scalable, and easy-to-use storage system, which is built with IBM Spectrum Virtualize software. It has two variants: IBM Hybrid Storwize V7000 and all-flash IBM Storwize V7000F. The storage can be clustered with up to four control enclosures. Each control enclosure supports up to 20 expansion enclosures *or* up to eight high-density expansions. The total number of drives that are supported by a full four-way cluster is 3,040. It includes enterprise features, such as HyperSwap and compression, making it a perfect fit as a storage consolidation platform.

3.2.1 External connectivity options for the IBM Storwize V7000 Gen2+

Each IBM Storwize V7000 controller has four 1 Gbps iSCSI ports as the default. One port is marked as the T port, which is used during the initial setup. It is preferable to dedicate at least one T port for management after the initial setup is complete. Therefore, each controller pair has a total of seven 1 Gbps iSCSI ports for external connectivity, and one port for management, by default. Each controller also has two 12 Gbps SAS ports, which are used for connecting expansion enclosures.

Each controller can be optionally configured with an additional interface card. The additional card can be either quad-port 16 Gbps FC or quad-port 10 Gbps (optical) iSCSI/FCoE. Both controllers in each pair should be configured with identical interface cards. Therefore, per IBM Storwize V7000 control enclosure, you can have either eight 16 Gbps ports or eight 10 Gbps ports.

All of the external connectivity ports can be used for host connectivity or external storage virtualization.

Figure 3-4 shows the external connectivity interfaces, both default and configurable, on the IBM Storwize V7000 Gen2+ control enclosure.

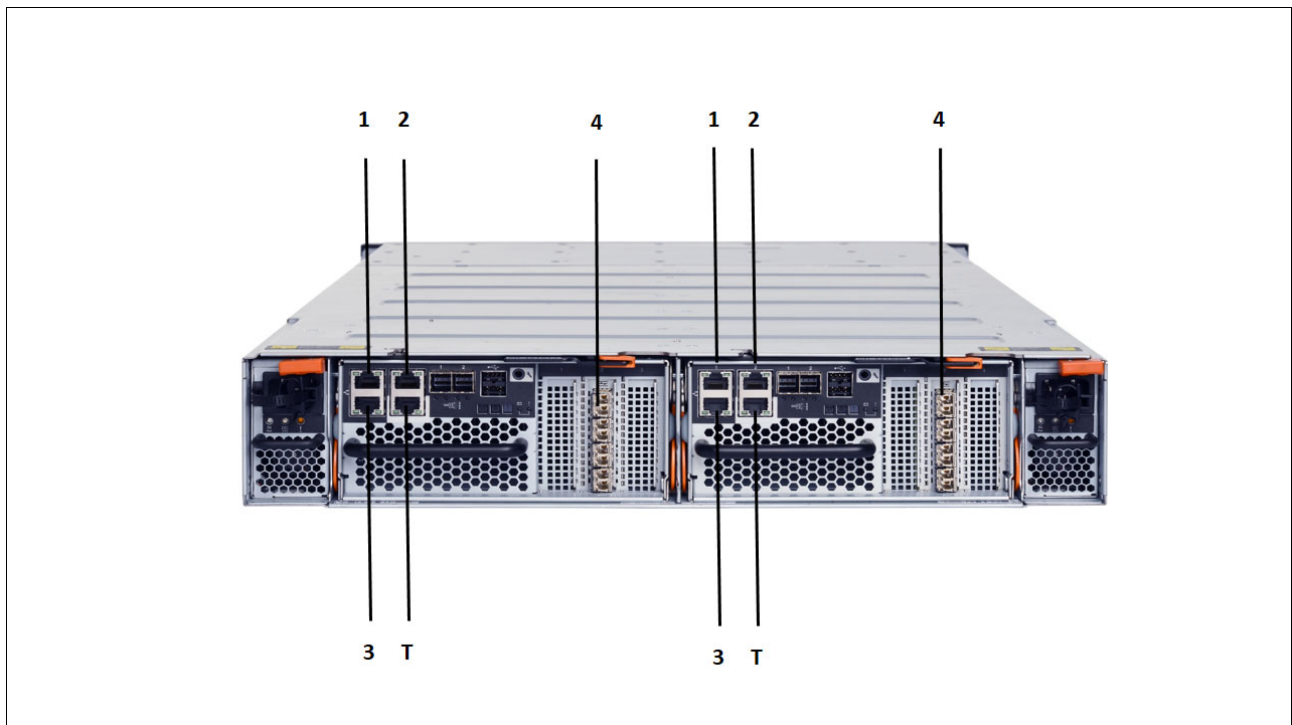


Figure 3-4 External connectivity interfaces on the IBM Storwize V7000 Gen2+ controller enclosure

Table 3-5 explains the slot numbers that are listed in Figure 3-4 on page 29 and their functions.

Table 3-5 Slot numbers and their functions

Label	Interface specification	Configuration type
1	1 Gbps iSCSI. This must be connected to system management, but optionally can be used for iSCSI.	Default
2	1 Gbps iSCSI connectivity.	Default
3	1 Gbps iSCSI connectivity.	Default
T	Technician port.	Default
4	4-port 10 Gbps iSCSI/FCoE OR 4-port 16 G FC adapter.	Configurable

3.2.2 External connectivity options for the IBM Storwize V7000 Gen2

The IBM Storwize V7000 storage systems have various host interface options, depending on the model of the controller enclosures.

Table 3-6 lists the HIC options for the IBM Storwize V7000 Gen2 storage system. SAS ports on the IBM Storwize V7000 storage system are used only for connecting with expansion enclosures.

Table 3-6 Host interface card options for the IBM Storwize V7000 Gen2 storage system

Slot	Host interface card options
1	Compression
2	Four 8 Gbps FC ports Two 16 Gbps FC ports Four 16 Gbps FC ports Four 10 Gbps Ethernet
3	Four 8 Gbps FC ports Two 16 Gbps FC ports Four 16 Gbps FC ports Four 10 Gbps Ethernet

Figure 3-5 shows a detailed view of the host interface ports at the rear of the IBM Storwize V7000 Gen2 storage system.

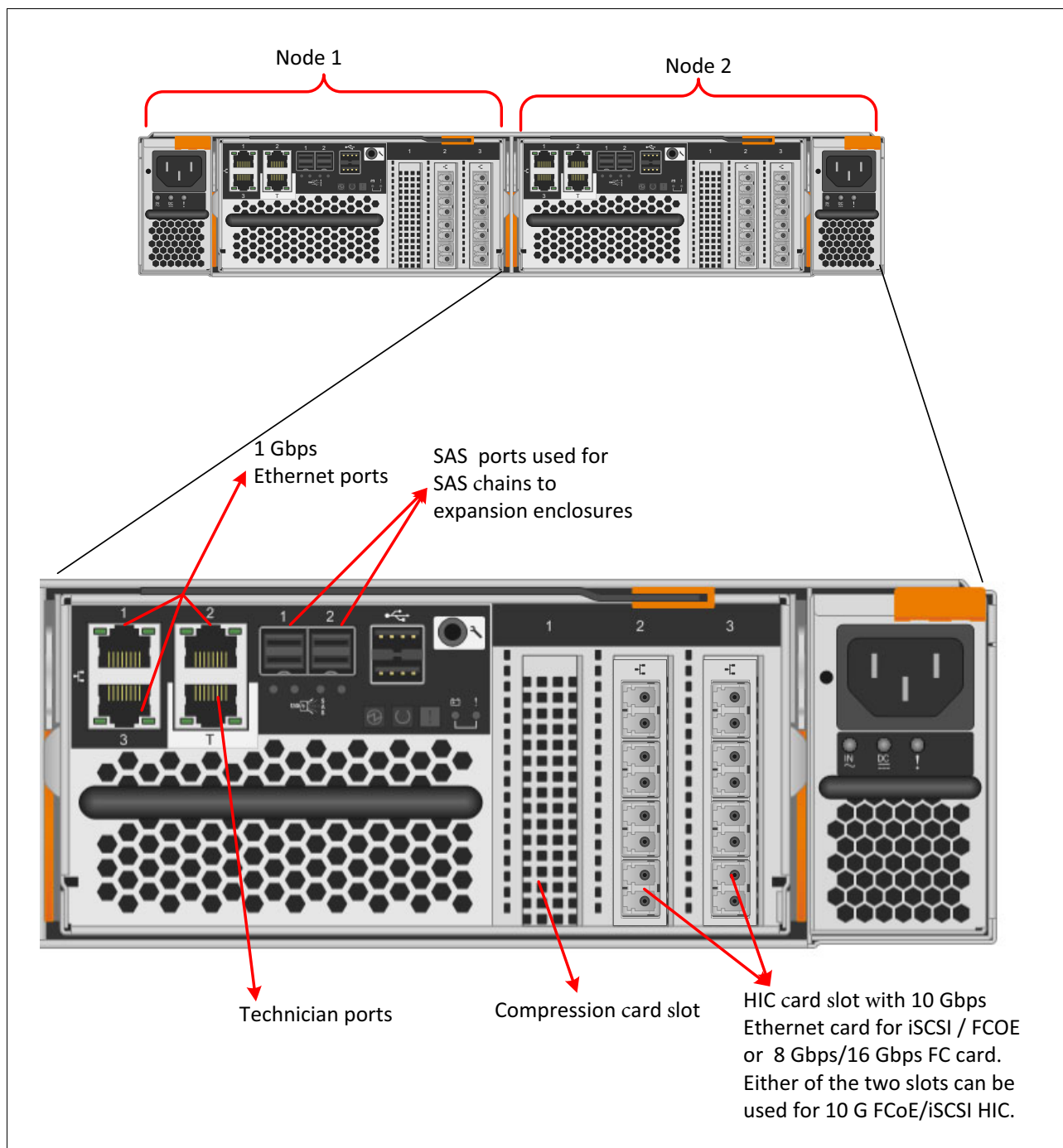


Figure 3-5 IBM Storwize V7000 Gen2 node ports

Note: For more information about the IBM Storwize V7000 Gen2 storage system, see *Implementing the IBM Storwize V7000 Gen2*, SG24-8244.

For more information about the IBM Storwize V7000 storage system, see [IBM Knowledge Center](#).

3.3 The IBM Storwize V7000 Unified storage system

The IBM Storwize Unified V7000 storage system is an integrated storage system that provides both file services and block service through FC and iSCSI. This storage uses the host interface of the IBM Storwize V7000 storage system for block-based storage, and uses the file module to provide the file services to the file client.

Figure 3-6 shows three server access methods that are provided by a single IBM Storwize V7000 Unified storage system. These methods are file-based access, block-level access with iSCSI, and block-level access with FC.

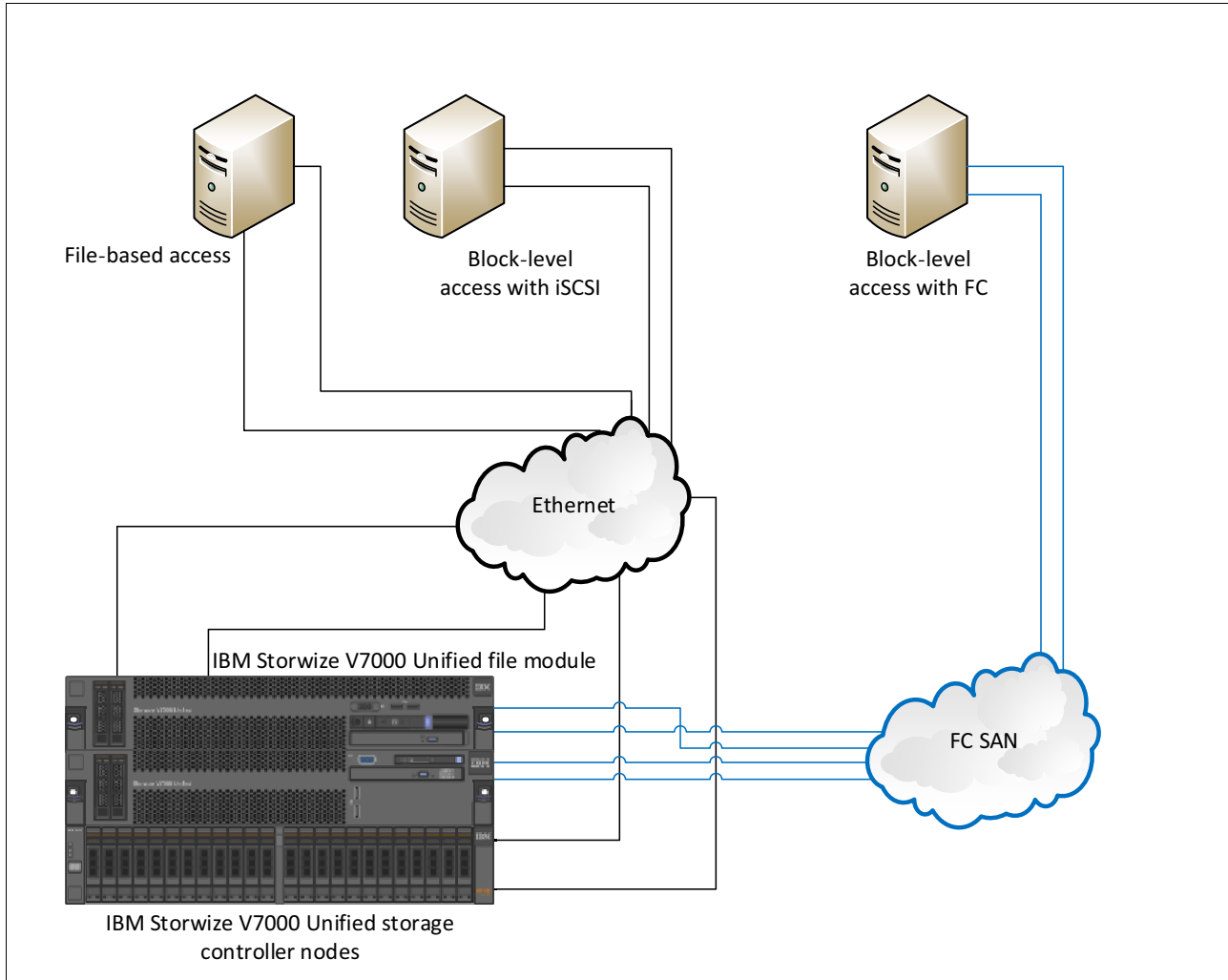


Figure 3-6 Server to storage access methods of an IBM Storwize V7000 Unified storage system

Table 3-7 provides information about roles for an IBM Storwize Unified V7000 storage system.

Table 3-7 Roles of the IBM Storwize Unified V7000 components

Access method	IBM Storwize Unified file module	IBM Storwize storage controller nodes
File-based access	Y	N
Block-level access with iSCSI	N	Y
Block-level access with FC	N	Y

Note: For more information about the IBM Storwize V7000 storage system, see *Implementing the IBM Storwize V7000 Unified Disk System*, SG24-8010.

3.4 SAN Volume Controller SV1 storage systems

The SAN Volume Controller storage system is a modular storage system that uses IBM Spectrum Virtualize. In Version 7.8 and later, the SAN Volume Controller SV1 storage system offers up to 256 GB of cache per node, and up to 20 standard expansion enclosures or up to eight high-density SAS expansions per node pair. In a clustered configuration, it can scale up to four node pairs.

Each SAN Volume Controller SV1 node has four 10 Gbps (copper) iSCSI ports by default. One port is marked as the T port, which is used during the initial setup. It is preferable to dedicate at least one T port for management after the initial setup is done. Therefore, each node has a total of three 10 Gbps iSCSI ports for external connectivity and one port for management by default. Each controller also has two 12 Gbps SAS ports, which are used for connecting expansion enclosures.

Each node must be configured with one additional interface card. This card can be either quad-port 10 Gbps iSCSI/FCoE or quad-port 16 Gbps FC. Optionally, three more quad-port 16 Gbps FC interface cards may be added to each node. Therefore, each node at maximum can have either of the following two configurations:

- ▶ 4 x 10 Gbps iSCSI (default) + 16 x 16 Gbps FC ports
- ▶ 4 x 10 Gbps iSCSI (default) + 12 x 16 Gbps FC ports + 4 x 10 Gbps iSCSI/FCoE ports

All of the external connectivity ports may be used for host connectivity or external storage virtualization.

Figure 3-7 shows the external connectivity ports, both default and configurable, on the SAN Volume Controller SV1 node engine.

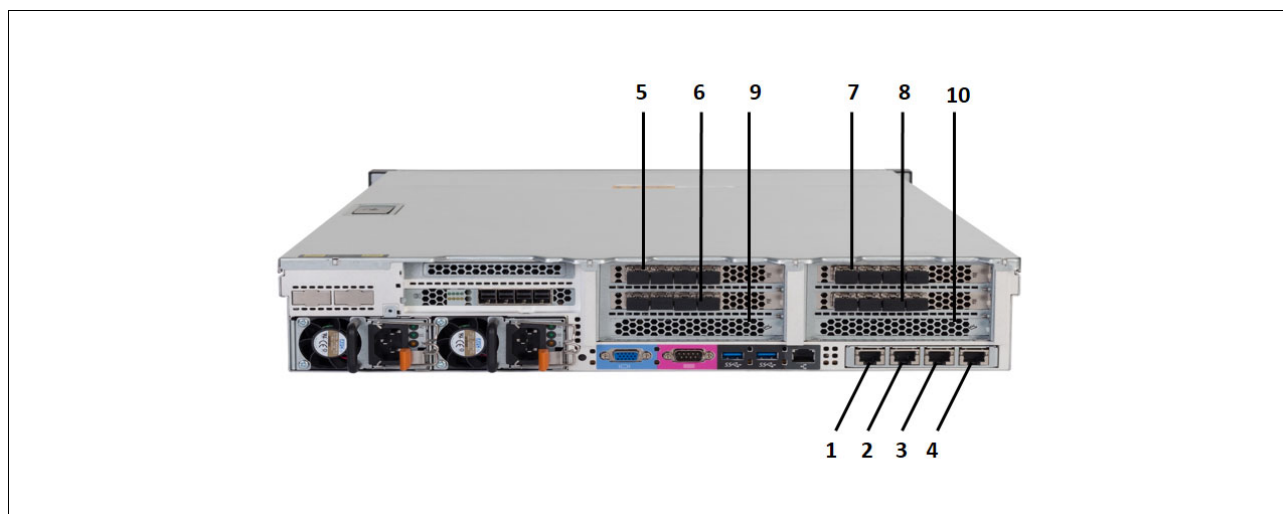


Figure 3-7 External connectivity options for the SAN Volume Controller SV1 node engine

Table 3-8 explains the slot numbers that are listed in Figure 3-7 and their functions.

Table 3-8 Slot numbers and their functions

Label	Interface specification	Configuration type
1	10 Gbps iSCSI/FCoE	Default
2	10 Gbps iSCSI/FCoE	Default
3	10 Gbps iSCSI/FCoE	Default
4	Technician Ethernet port	Default
5	4-port 16 Gbps FC OR 4-port 10 Gbps iSCSI/FCoE adapter	Configurable ^a
6	4-port 16 Gbps FC OR 4-port 10 Gbps iSCSI/FCoE adapter	Configurable ^a
7	4-port 16 Gbps FC OR 4-port 10 Gbps iSCSI/FCoE adapter	Configurable ^a
8	4-port 16 Gbps FC OR 4-port 10 Gbps iSCSI/FCoE adapter	Configurable ^a
9	IBM Real-time Compression (RtC) Accelerator	To be configured for RTC
10	SAS expansion adapter	To be configured for local storage expansion connectivity

a. Only one 4-port 10 Gbps iSCSI/FCoE can be configured per SAN Volume Controller SV1 node.

Note: For more information about SAN Volume Controller, see [IBM Knowledge Center](#).

3.5 Hardware terminology for the IBM Storwize disk systems

This section explains various hardware terminologies for the IBM Storwize disk systems.

3.5.1 Control enclosures, nodes, and I/O groups

Storage enclosures that have the controllers at the rear are called *control enclosures*. The control enclosures have disk drives at the front and storage controllers at the rear.

Each control enclosure has two node canisters. Each pair of nodes and the disks that are attached to the SAS expansion chain is called an *I/O group*.

Figure 3-8 shows the view of control enclosures, node canisters, and the I/O group.

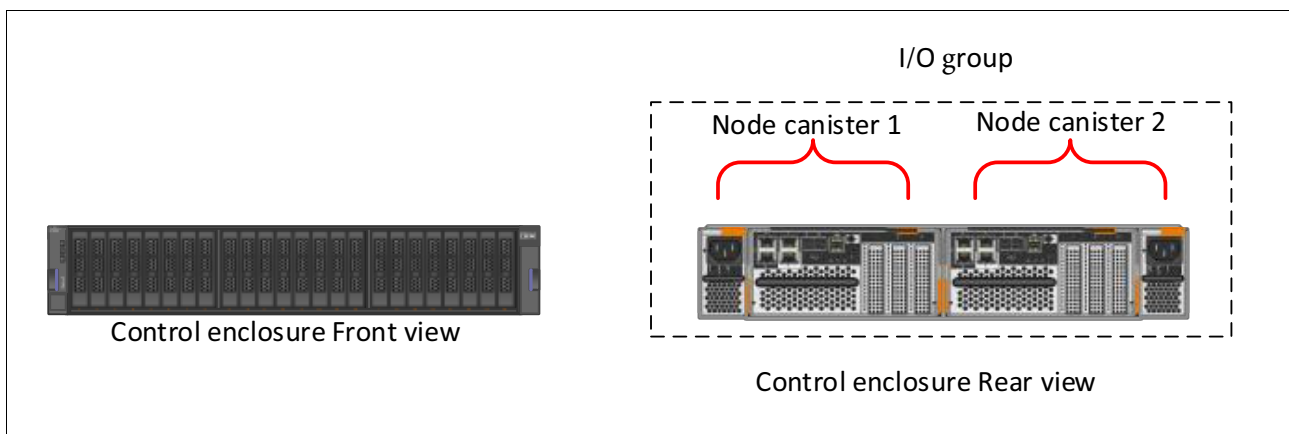


Figure 3-8 IBM Storwize control enclosures, node canister, and I/O group

3.5.2 Expansion enclosures

Expansion enclosures are additional disk enclosures that are connected to control enclosures to provide more storage capacity. These expansion enclosures are connected to control enclosures through SAS chains from control enclosures. The high-end IBM Storwize disk systems can connect 20 expansion enclosures through two SAS chains. Figure 3-9 shows an example configuration where a single control enclosure is connected to 20 expansion enclosures.

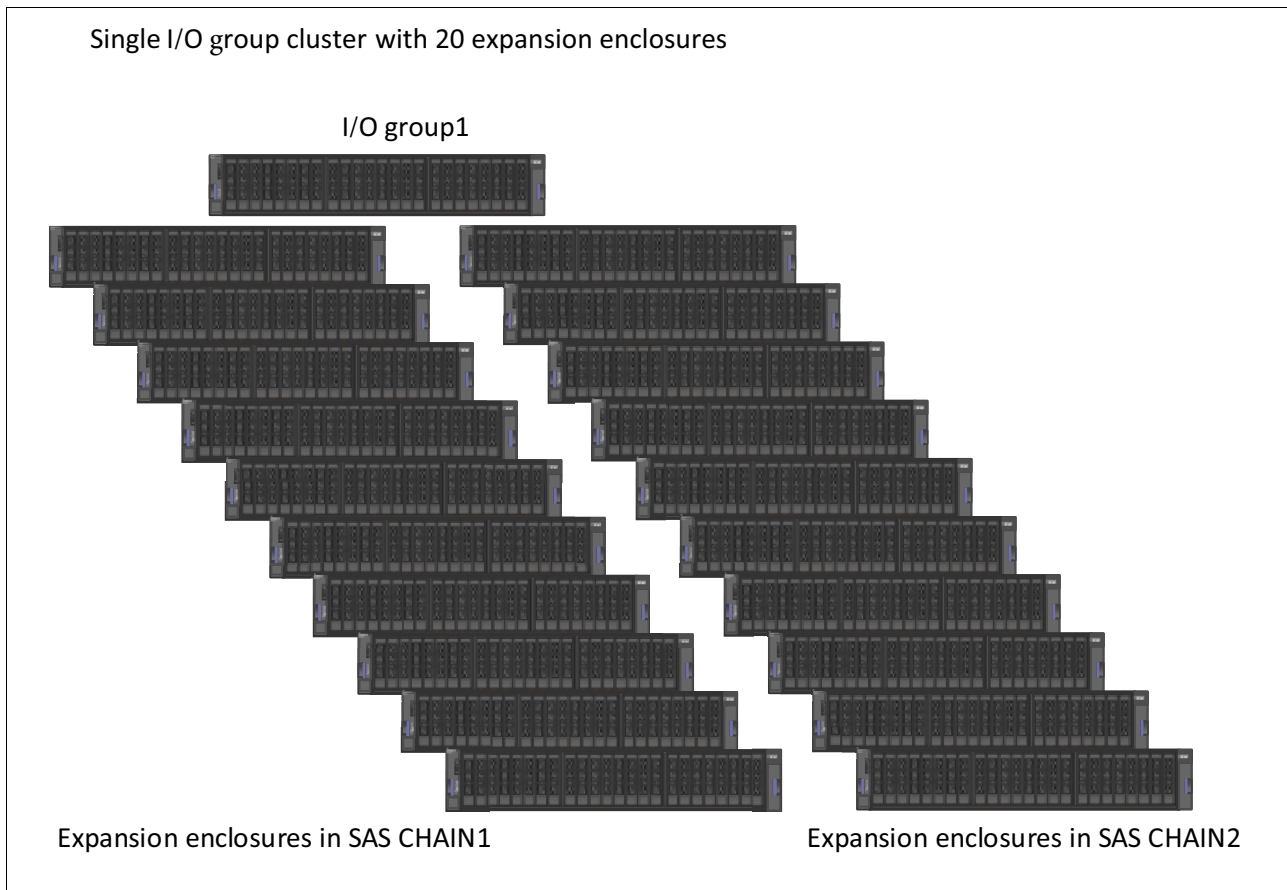


Figure 3-9 An IBM Storwize disk system with 20 expansion enclosures that are connected to a single I/O group

3.5.3 IBM Storwize cluster system

IBM Storwize storage systems can be clustered with 1 - 4 I/O groups. By default, a single I/O group is part of a cluster and can be expanded by adding more I/O groups. Figure 3-10 shows an example of an IBM Storwize cluster system with two I/O groups.

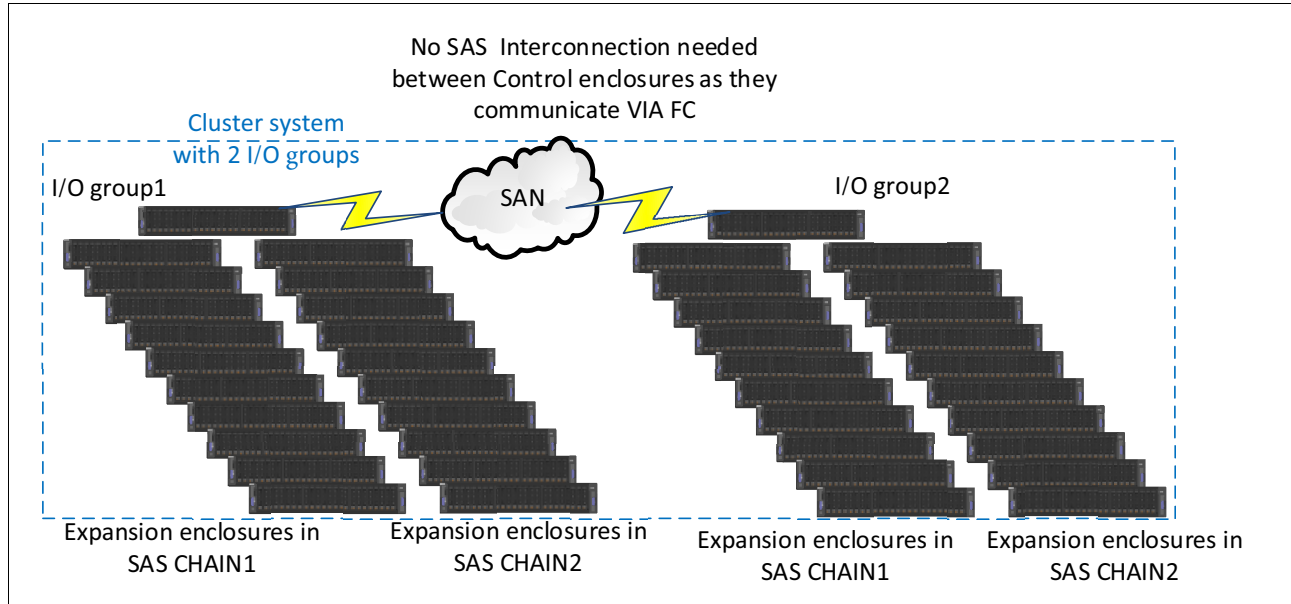


Figure 3-10 IBM Storwize cluster system with two I/O groups

3.5.4 IBM Storwize virtualization

The IBM Storwize storage system supports both internal virtualization with internal disks and external virtualization with LUNs that are mapped from the external storage systems. The virtualization concepts and terminology are described in the following topics.

MDisks

MDisks are logical volumes from an external storage array that is presented to the IBM Storwize storage system or a RAID array that is created from internal drives. The arrays that are created from internal drives are in array mode and are always associated with storage pools. Starting with Version 7.6, RAID arrays can have distributed RAID with distributed spares to allow faster rebuild and better I/O performance.

Storage pools

A storage pool is a collection of MDisks. The size of the storage pool is related to the number of MDisks, and can be expanded or shrunk by adding or removing MDisks dynamically. The MDisks in each storage pool are divided into several extents when storage pools are created. This extent size is defined by the administrator when the storage pools are created. It is a preferred practice to use the same extent size across all storage pools.

VDisks or volumes

VDisks or volumes are created by using available extents in a storage pool. VDisks are mapped to hosts according to the requirements of users.

Figure 3-11 depicts the virtualization terminology and concepts of IBM Storwize storage systems.

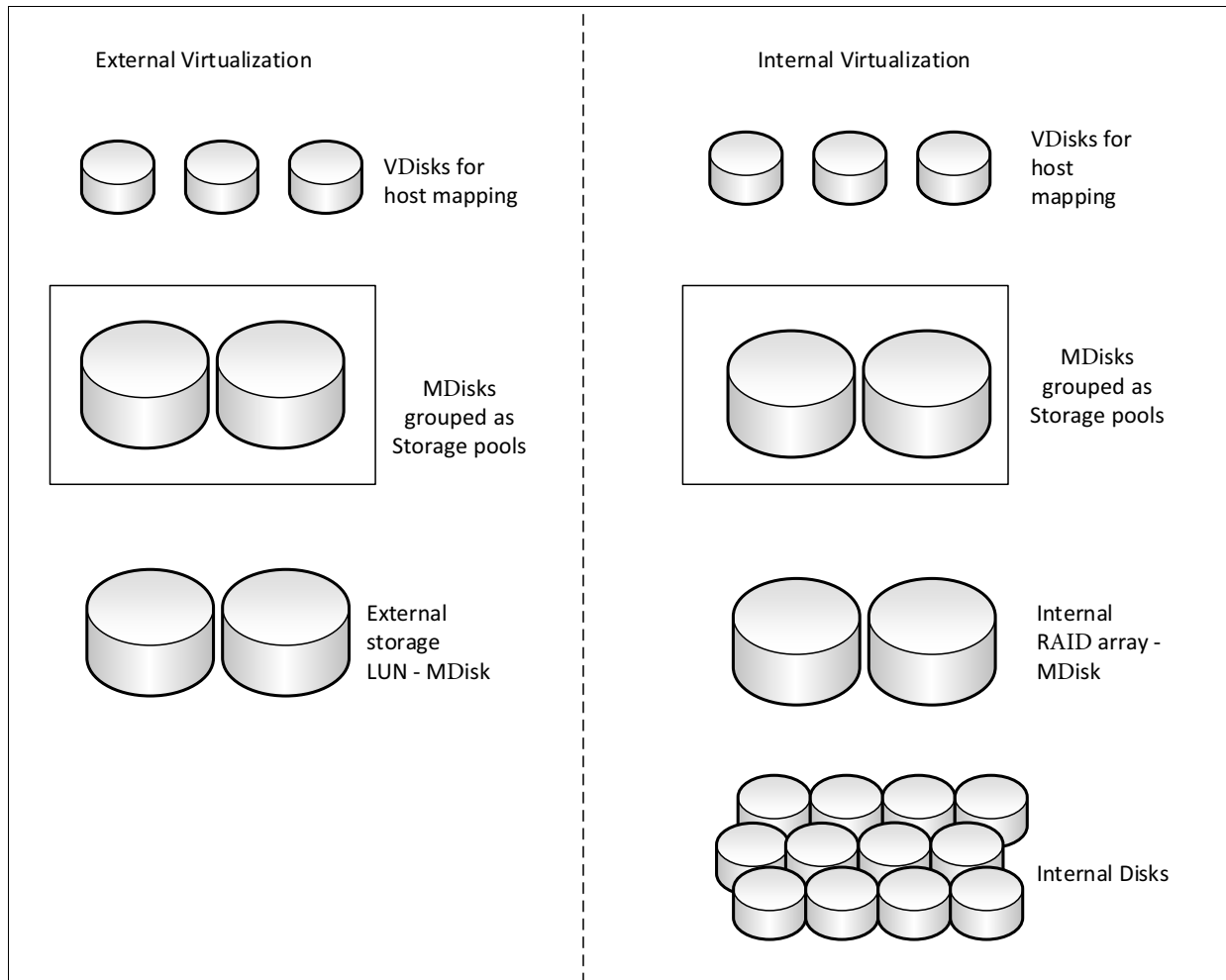


Figure 3-11 IBM Storwize storage system virtualization concepts and terminology



Planning considerations

This chapter describes the planning considerations and preferred practices for iSCSI implementation. It describes general considerations, network topology, and preferred practices based on specific operating systems. It lists limits for the iSCSI and IBM Systems IBM Storwize family.

This chapter describes the following topics:

- ▶ 4.1, “General considerations” on page 40
- ▶ 4.2, “Network topology” on page 42
- ▶ 4.3, “Planning for host access” on page 46
- ▶ 4.4, “Planning considerations for external virtualization” on page 52
- ▶ 4.5, “IBM Storwize family and iSCSI limits” on page 53

4.1 General considerations

This section describes general topics to consider before you implement an iSCSI network solution. These considerations are described in the following list:

- ▶ Use a TCP offload engine (TOE) when possible.

Software-based iSCSI initiators use a significant amount of processor resources. If it is possible, use TOE cards. An iSCSI TOE card allows the network adapter to process iSCSI traffic locally to reduce further host server processor usage.
- ▶ Use a local area network (LAN) topology.

To improve performance, use iSCSI over LAN instead of using a wide area network (WAN). Latency might affect production performance and cause long delays to data transfers.
- ▶ Enable Priority Flow Control (PFC) when possible.

It is possible to prioritize iSCSI data over other data in the network. PFC is a link-level flow control mechanism that selectively pauses data traffic according to its previously configured class. Enable PFC on the hosts, switches, and the system storage.
- ▶ Use iSCSI security authentication.

Consider using any of the possible storage authentication methods. At the time of initiating the communication, the initiator sends a login request to the target system to begin an iSCSI session. Based on the authentication configuration, the storage system accepts or denies the login request.

At the time of writing, the supported authentication method for the IBM Storwize family is CHAP. The Radius and IPsec authentication methods are optional, but are not supported by IBM.
- ▶ IPv4 and IPv6.

It is possible to use IPv4 and IPv6 in the same node.
- ▶ Let the switch automatically negotiate the highest possible speed instead of manually setting the speed port on the Ethernet switch.
- ▶ Separate iSCSI traffic.

Isolate the iSCSI traffic onto separate physical switches or by using private virtual local area networks (VLANs) (IEEE802.1Q). VLAN tagging provides network traffic separation at the Layer 2 level for Ethernet transport. When using Fibre Channel over Ethernet (FCoE) also, you can prevent performance problems by not using the same physical ports as for iSCSI.
- ▶ Use the correct cabling for data transmission. For copper cabling, use CAT6 rated cables for gigabit networks and CAT 6a or CAT-7 cabling for 10-Gb implementations. For fiber, use OM3 or OM4 multimode fiber cabling.
- ▶ Enable Jumbo frames.

By default, iSCSI normally uses standard 1500-byte frames. It is possible to change the network to use other Ethernet frame sizes to adjust network performance. Jumbo frames are Ethernet frames with more than 1500 bytes of payload. They improve performance for the iSCSI network and must be configured on the initiator, target, and network switch. The maximum frame size is 9000.

- Multipathing.

When supported, install the correct driver to manage the paths to provide high availability and load balancing of storage I/O, automatic path-failover protection, and prevention of a single-point-failure that is caused by the host bus adapter (HBA), Fibre Channel (FC) cable, Ethernet cable, or host-interface adapter on supported storage.

- Separate management traffic from I/O traffic.

If you are using IBM Spectrum Control™ or an equivalent application to monitor the performance of your IBM SAN Volume Controller cluster, it is preferable to separate this management traffic from iSCSI host I/O traffic, such as using node port 1 for management IPs and node port 2 for iSCSI IPs.

- iSCSI host and system storage with the same link speed.

For performance reasons, use the same link speed on each Ethernet port for iSCSI hosts and the system storage.

- SCSI persistent reservations.

If the host application uses SCSI 3 persistent reservation, it must use only a single iSCSI session per host per node.

- Use a maximum of four sessions per node and one IQN per host.

When the iSCSI initiator discovers the iSCSI targets, it finds many different IP addresses. When the host establishes four sessions per SAN Volume Controller node and tries to establish the fifth one, this login is rejected. So, if more than four IPs are configured on the target, each host establishes logins with the first four IPs only unless IPs are explicitly mentioned on the host during session establishment.

Use a maximum of one session between an initiator port and a target port, which can be achieved through correct VLAN separation or IP subnetting.

- Use the correct queue depth value for the configuration.

The queue depth parameter represents the number of I/O operations that can be run in parallel on a device. Estimate the queue depth for each node to avoid application failures.

If a node reaches the maximum number of queued commands, it returns error codes and significantly degrades performance. The formula for queue depth calculation considers many factors. For a detailed explanation about queue depth in iSCSI networks, see [IBM Knowledge Center](#).

- Configure traffic storm control on the Ethernet switches.

A *traffic storm* is excessive network traffic in the network, which is caused by a packet flood. The result of this event is performance degradation. Configure traffic storm control in the Ethernet switches to prevent problems on the ports. Disable unicast storm control and enable broadcast and multicast storm control on switches where iSCSI traffic occurs.

- Use an iSCSI target name (IQN).

Take down all iSCSI sessions from the hosts to the target on which such changes are made, make the changes, and then reconfigure the iSCSI session.

- Refer to the IBM System Storage® Operation Center (SSIC).

To get the last compatibility matrix, see the [SSIC website](#).

4.2 Network topology

Network topology refers to the layout of a network. Its design is important to ensure that iSCSI works in the best conditions due to its high dependence on network health and usage. This section focuses on the different supported topologies, with example implementations of IBM Storwize family products.

4.2.1 Network topology with one Ethernet switch

Figure 4-1 shows a basic supported network topology in which there is only one switch with two configured VLANs. In this case, the first problem is that there is a single point of failure with the network switch. Additionally, it is not possible to use the round-robin algorithm for multipathing because there is only one connected port on each IBM Storwize node. Because the IBM Storwize storage system is active-passive, all the I/O iSCSI traffic for each volume is running only on one iSCSI target port.

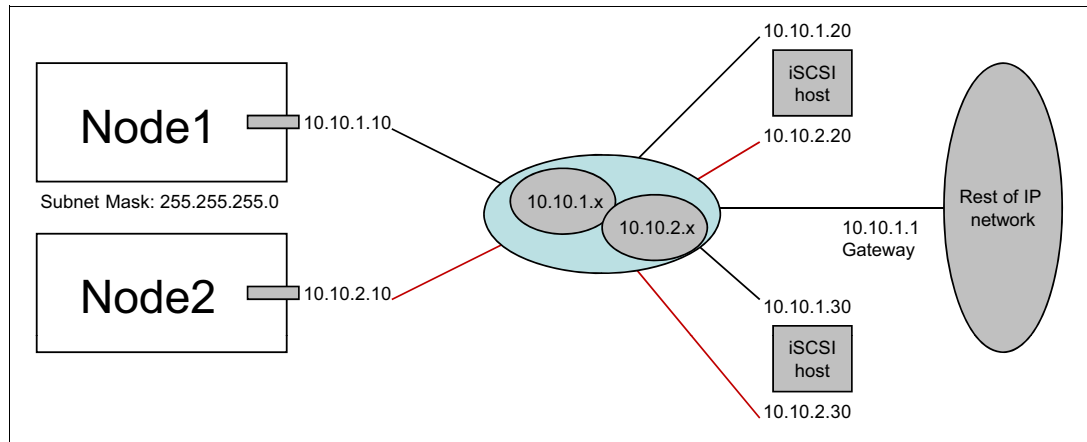


Figure 4-1 Network topology with one Ethernet switch

4.2.2 Network topology with two VLANs

Figure 4-2 shows a dual Ethernet switch network topology with one VLAN each. Use this network topology, for example, for an IBM Storwize V5010 storage system that uses 10-Gb adapters because it has a maximum number of four physical Ethernet ports.

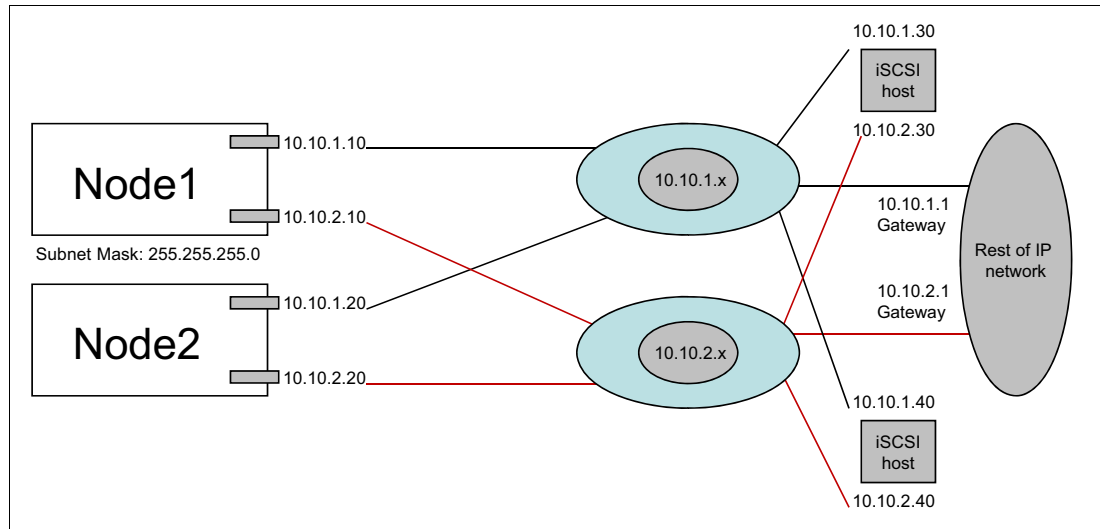


Figure 4-2 Network topology with two VLANs

4.2.3 Network topology with four VLANs

Figure 4-3 shows a network topology with two network switches that are configured with two VLANs each. This is the preferred network topology for iSCSI implementations. In this case, the iSCSI I/O traffic is more segmented and uses all the available IBM Storwize iSCSI ports.

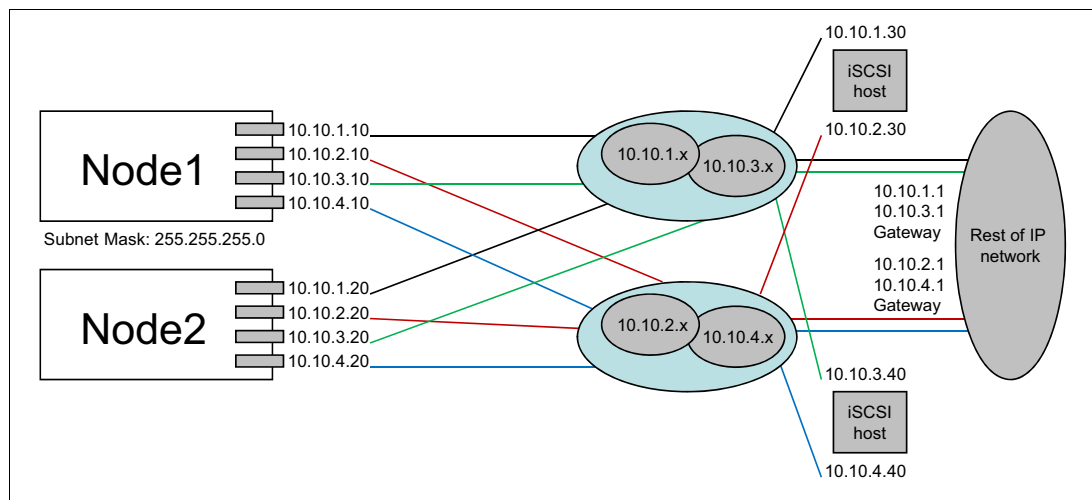


Figure 4-3 Network topology with four VLANs

4.2.4 Link aggregation between switches

Figure 4-4 shows a network topology in which iSCSI hosts and SAN Volume Controller storage systems are separated by two switches. In this case, consider using interswitch links (ISLs) that work as link aggregation. The number of Ethernet ports that are configured for each Ethernet switch in the link aggregation should be equal to the number of active Ethernet iSCSI target ports on the IBM Storwize storage system that is connected to the switch. In this example, it should be two ISLs per Ethernet switch.

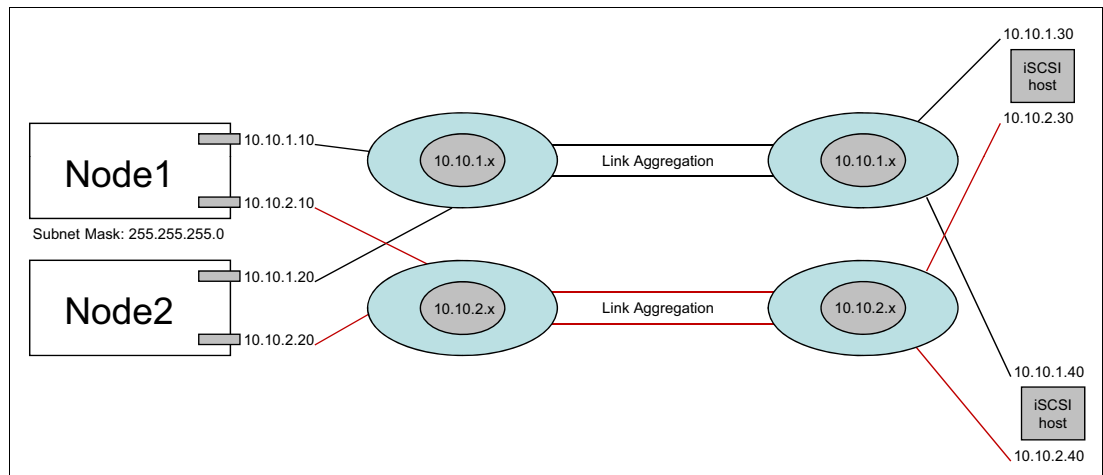


Figure 4-4 Link aggregation network topology

4.2.5 iSCSI that uses a Layer 3 network topology

Although it is supported, to prevent performance problems and eventual difficulties with troubleshooting, do not use a Layer 3 network topology. If it is the only option, be sure to have enough bandwidth between both sides of the network.

There is no specific preferred practice about any vendor routing protocols. Consult your vendor documentation.

Figure 4-5 shows a Layer 3 network topology in which four ports are connected to the IBM Storwize storage system, but there are only two links between both network segments. Therefore, there is an identified bottleneck at that point. In this case, consider a topology similar to Figure 4-4 on page 44.

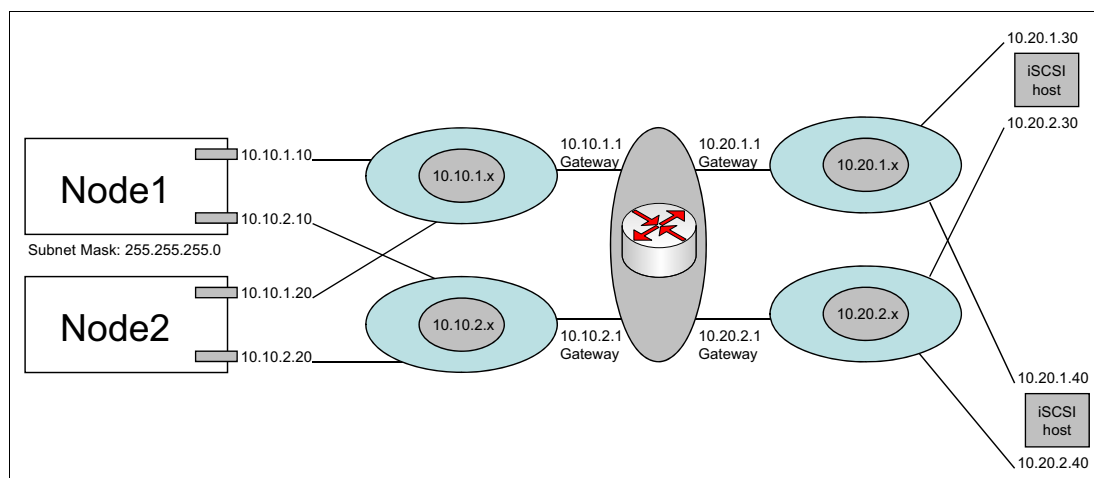


Figure 4-5 Layer 3 network topology

Important: IBM does not support network address translation (NAT) for iSCSI implementations. The only method to identify a network target on the IBM Storwize storage system is by its IP address. In that case, at the moment of the iSCSI discovery connection, the iSCSI initiator cannot map the IP address that is returned by the iSCSI target.

4.2.6 IP replication network topology

FCoE, iSCSI, and IP replication I/O traffic can travel on each of the IBM Storwize Ethernet ports. However, for performance reasons, it is important to separate each of them onto separate links.

Figure 4-6 shows a suggested topology that uses dual 4-port 10-Gb network Ethernet cards on the IBM Storwize storage system for iSCSI and FCoE I/O. The two onboard 1-Gb ports can be dedicated to IP replication. In this example, the I/O for IP replication is load balanced between the connected ports of each IBM Storwize node.

Avoid using the same physical port for management or maintenance.

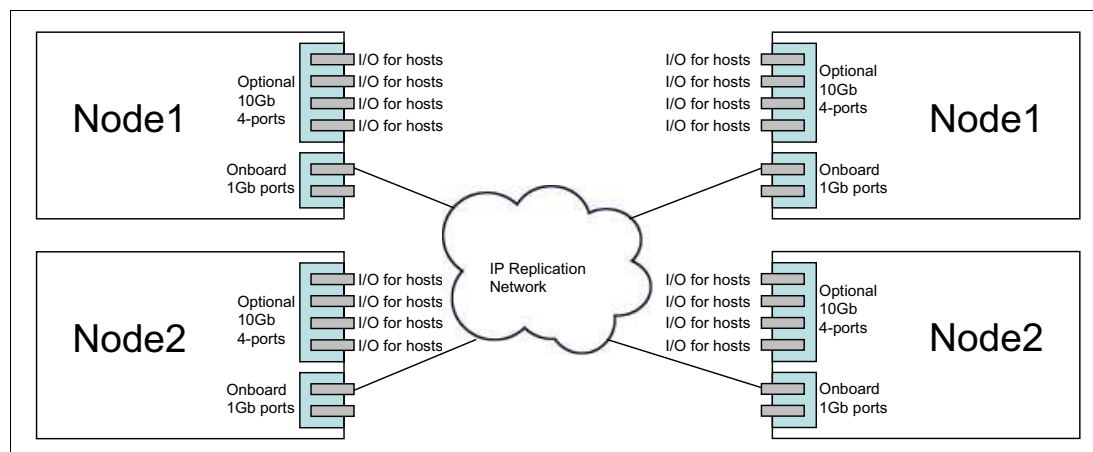


Figure 4-6 IP replication network topology

4.3 Planning for host access

This section describes topics to consider before you implement an iSCSI solution.

4.3.1 Planning for IBM AIX

This section describes considerations for AIX.

Single-path configuration

At the time of writing, the IBM Storwize storage family supports only a software initiator and a single path for AIX. This configuration means that AIX does not support multipathing. Depending on the configuration, the environment can use the automatic failover feature.

Figure 4-7 shows an example of a common environment with IBM AIX attached to a SAN Volume Controller storage system. The AIX host has a mapped volume (hdisk0) that is assigned from the SAN Volume Controller storage system.

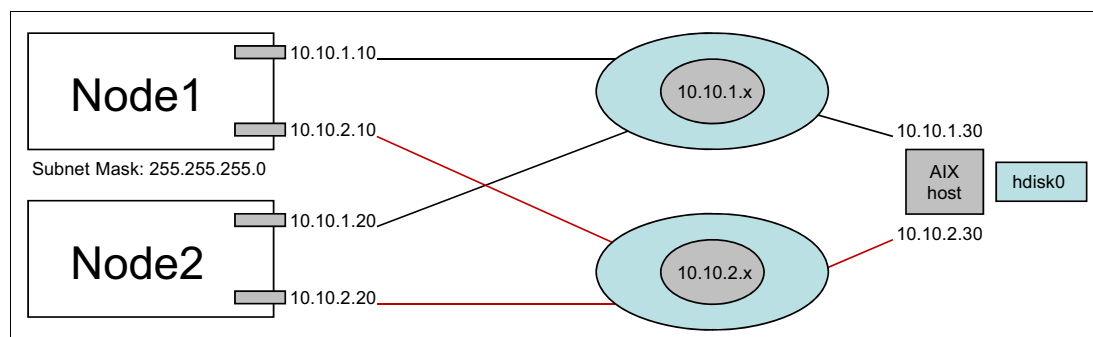


Figure 4-7 Basic single-path AIX environment

Figure 4-8 shows the state of the environment when one of the SAN Volume Controller nodes goes offline. In that case, AIX continues accessing hdisk0 because the SAN Volume Controller storage system uses the automatic failover feature in which node 2 takes the IP addresses from node 1. When the first node is reestablished, it automatically fails back after 5 minutes.

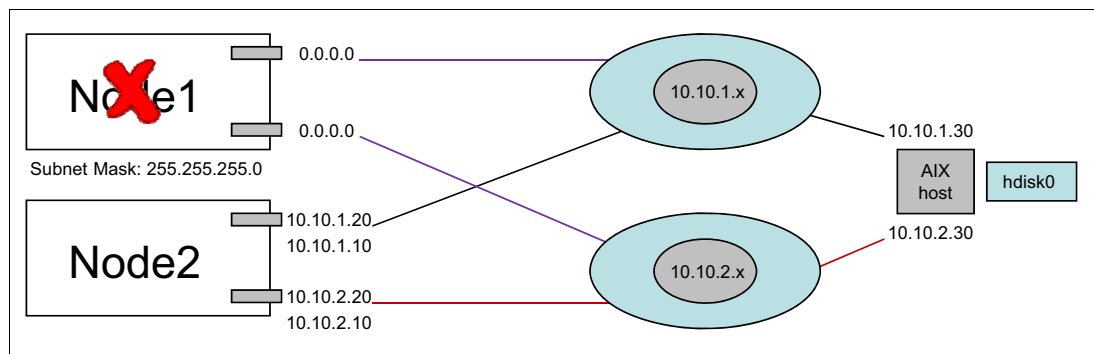


Figure 4-8 SAN Volume Controller node 1 failure

Figure 4-9 shows the worst-case scenario, in which the NIC with the primary I/O path goes offline. In that case, because IBM AIX does not support multipathing, hdisk0 is not accessible. To prevent this situation, consider other network port aggregation technologies, such as Etherchannel.

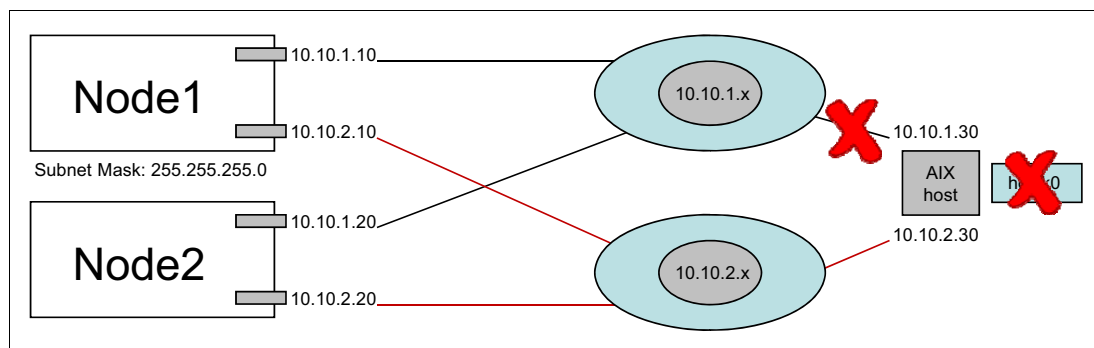


Figure 4-9 AIX host NIC failure

Considerations for starting

At the time of writing, starting from AIX with iSCSI is supported, but only with a software initiator and a single path. Multipathing, clustering, and boot from SAN are not supported over iSCSI in AIX currently.

iSCSI software target considerations

When you define an iSCSI software target and export logical unit numbers (LUNs), the iSCSI qualified name (IQN) of each virtual target is specified in SMIT when a software target is defined. The SMIT panel does not restrict the format of the name. However, some iSCSI initiators require that the IQN is specified in the format that is defined by the iSCSI protocol. Using an incorrect name format might prevent the initiator from logging to the target and accessing the disks that are exported by the target.

To display the current name of an iSCSI target device, complete the following steps:

1. Run a command similar to the following example. For this example, assume that the iSCSI target device is target0.

```
lsattr -E -l target0
```

2. Check the `iscsi_name` attribute.

The inquiry data that is returned for an exported LUN has the following values:

- ▶ Vendor ID: AIX
- ▶ Product ID: iSCSI_VDASD
- ▶ American National Standard version number: 3

iSCSI software initiator considerations

Consider the following things when you configure iSCSI software initiators:

- ▶ Target discovery

The iSCSI software initiator supports the following four forms of target discovery:

- File

A text file is used to configure each target.

- ODM

ODM objects are used to configure each target. When you use an iSCSI disk as a boot disk or as part of the rootvg boot, the ODM discovery method must be used.

- Internet Storage Name Service (iSNS)

Each target is registered in one or more iSNS servers.

- Service Location Protocol (SLP)

Each target is registered in one or more SLP service agents or directory agents.

- ▶ iSCSI Authentication

Only CHAP (MD5) can be used to configure initiator authentication. Target authentication is not implemented.

- ▶ Number of configured LUNs

The maximum number of configured LUNs that is tested by using the iSCSI software initiator is 128 per iSCSI target. The software initiator uses a single TCP connection for each iSCSI target (one connection per iSCSI session). This TCP connection is shared among all LUNs that are configured for a target. The software initiator's TCP socket send and receive space are both set to the system socket buffer maximum. The maximum is set by the `sb_max` network option. The default is 1 MB.

- ▶ Volume groups

To avoid configuration problems and error log entries when you create volume groups that use iSCSI devices, follow these guidelines:

- Configure volume groups that use iSCSI devices to be in an inactive state after a restart. After the iSCSI devices are configured, manually activate the iSCSI-backed volume groups. Then, mount any associated file systems. Volume groups are activated during a different boot phase than the iSCSI software driver. For this reason, it is not possible to activate iSCSI volume groups during the boot process.
- Do not span volume groups across non-iSCSI devices.

► I/O failures

If connectivity to iSCSI target devices is lost, I/O failures occur. To prevent I/O failures and file system corruption, stop all I/O activity and unmount iSCSI backed file systems before doing anything that causes long-term loss of connectivity to active iSCSI targets.

If a loss of connectivity to iSCSI targets occurs while applications are attempting I/O activities with iSCSI devices, I/O errors eventually occur. It might not be possible to unmount iSCSI backed file systems because the underlying iSCSI device stays busy.

File system maintenance must be done if I/O failures occur due to loss of connectivity to active iSCSI targets. To do file system maintenance, run the **fsck** command.

iSCSI security considerations

The `/etc/iscsi` directory, the `/etc/tmisci` directory, and the files in those directories are protected from non-privileged users through file permission and ownership.

CHAP secrets are saved in the `/etc/iscsi/targets` file and the `/etc/tmisci/autosecrets` file in clear text.

Important: Do not change the original file permission and ownership of these files.

iSCSI network considerations

Consider the following things when you configure the iSCSI network:

- Enable the TCP Large send offload option, TCP send and receive flow control, and Jumbo Frame features of the AIX Gigabit Ethernet card and the iSCSI Target interface. For a detailed explanation of these options, see [IBM Knowledge Center](#).
- Modify the network options and interface parameters for maximum iSCSI I/O throughput on the AIX system as follows:
 - Enable the RFC 1323 network option.
 - Set up the **tcp_sendspace**, **tcp_recvspace**, **sb_max**, and **mtu_size** network options and network interface options to the appropriate values. The iSCSI Software Initiator's maximum transfer size is 256 KB. Assuming that the system maximums for **tcp_sendspace** and **tcp_recvspace** are set to 262,144 bytes, an **ifconfig** command to configure a Gigabit Ethernet interface might look like the following example:

```
ifconfig en2 10.1.2.216 mtu 9000 tcp_sendspace 262144 tcp_recvspace 262144
```
 - Set the **sb_max** network option to at least 524288, and preferably 1048576.
 - Set the **mtu_size** to 9000.
 - For some iSCSI targets, the TCP Nagle's algorithm must be disabled for best performance. Use the **no AIX** command to set the **tcp_nagle_limit** parameter to 0, which disables the Nagle algorithm.

For more information about the global principles of communication tuning for AIX, see [IBM Knowledge Center](#).

4.3.2 Planning for Linux

This section describes considerations for Linux.

Considerations for booting

At the time of writing, for SUSE Linux Enterprise Server, SAN boot is supported on hardware initiators and network boot configurations, such as Preboot Execution Environment (PXE) for software initiators. For Red Hat Enterprise Linux, only hardware initiators are supported.

Multipath settings

The following settings are the preferred multipath settings for each specific Linux distribution and release.

Red Hat Linux Enterprise Linux Versions 5.x, 6.0, and 6.1

Example 4-1 shows a configuration example for `multipath.conf`.

Example 4-1 The multipath.conf file

```
vendor "IBM"
product "2145"
path_grouping_policy "group_by_prio"
path_selector "round-robin 0"
prio "alua"
path_checker "tur"
failback "immediate"
no_path_retry 5
rr_weight uniform
rr_min_io 1000
dev_loss_tmo 120
```

Red Hat Enterprise Linux Versions 6.2 and higher and 7.x

Example 4-2 shows a configuration example for `multipath.conf`.

Example 4-2 The multipath.conf file

```
vendor "IBM"
product "2145"
path_grouping_policy "group_by_prio"
path_selector "round-robin 0"
prio "alua"
path_checker "tur"
failback "immediate"
no_path_retry 5
rr_weight uniform
rr_min_io_rq "1"
dev_loss_tmo 120
```

SUSE Linux Enterprise Server versions 10.x and 11.0 and 11SP1

Example 4-3 shows a configuration example for `multipath.conf`.

Example 4-3 The multipath.conf file

```
vendor "IBM"
product "2145"
path_grouping_policy "group_by_prio"
path_selector "round-robin 0"
prio "alua"
path_checker "tur"
failback "immediate"
no_path_retry 5
rr_weight uniform
rr_min_io 1000
dev_loss_tmo 120
```

SUSE Linux Versions 11SP2 and higher

Example 4-4 shows a configuration example for `multipath.conf`.

Example 4-4 The multipath.conf file

```
vendor "IBM"
product "2145"
path_grouping_policy "group_by_prio"
path_selector "round-robin 0" # Used by SLES 11 SP2
prio "alua"
path_checker "tur"
failback "immediate"
no_path_retry 5
rr_weight uniform
rr_min_io_rq "1"
dev_loss_tmo 120
```

4.3.3 Planning for VMware

At the time of writing, mutual CHAP and SAN boot are not supported on VMware and the IBM Storwize family with iSCSI environments.

Ask your vendor whether you should use TCP segmentation offload (TSO). This parameter can cause poor network performance in certain cases.

For more information about VMware and iSCSI, see the [VMware vSphere website](#).

4.3.4 Planning for Windows

At the time of writing, SAN boot is supported only on hardware initiators, and Microsoft MPIO is the only supported multipath driver for the IBM Storwize storage system.

For more information about Microsoft Windows and iSCSI, see [Installing and Configuring Microsoft iSCSI Initiator](#).

4.4 Planning considerations for external virtualization

This section provides an overview of the various planning considerations that are specific to using iSCSI to virtualize external storage. Before you begin planning for virtualizing an external storage controller by using iSCSI, review the [SAN Volume Controller and Storwize Family iSCSI Storage Attachment Support Matrix](#) to ensure that it is supported.

4.4.1 Network security

iSCSI is a storage protocol that uses the TCP/IP stack for communication. Because of this, there is a possibility that the traffic being sent by the IBM Storwize product to the external storage is visible on the same physical devices as your normal Ethernet traffic. Also, there is a chance that traffic between the IBM Storwize storage system and the external storage is on the same logical topology as other communications in the network, such as inter-host traffic and host to IBM Storwize storage system traffic.

For security reasons, you might want to physically or logically separate this traffic from other parts of the network. This way, if a host is compromised, there is a reduced risk of the traffic between the IBM Storwize storage system and external storage being compromised as well. You can do this separation by provisioning a separate subnet or VLAN for traffic between the IBM Storwize storage system and external storage.

In addition to segmenting the traffic between the IBM Storwize storage system and the external storage away from the rest of the network, you might want to configure the authentication between the two systems. To do this, configure one-way CHAP. For more information this procedure, see the section that is specific to your external storage controller.

4.4.2 iSCSI Protocol-specific considerations

The iSCSI protocol has many different implementations across products and vendors. Because of this, there are some controller-specific considerations to be aware of when virtualizing storage over iSCSI. For more information about considerations for your specific controller, see the following sections:

- ▶ Chapter 12, “External virtualization of IBM Storwize storage systems” on page 227
- ▶ Chapter 13, “Virtualization of IBM Spectrum Accelerate storage systems” on page 239
- ▶ Chapter 14, “External virtualization of Dell Equallogic PS Series” on page 265

4.4.3 Controller migration considerations

In some systems, the maximum or configured MTU might not be the same as the maximum MTU of the SAN Volume Controller or Storwize system. In these cases, it is important to have the same MTU configured on all endpoints that are involved in the migration. As such, it is advisable to review the maximum MTU that is permitted by the hosts, SAN Volume Controller or IBM Storwize device, network devices, and existing storage controller. Then, make a comparison between these MTU sizes and configure the systems for the largest possible MTU size that is common to all components in the system.

4.5 IBM Storwize family and iSCSI limits

This section details the configuration limits and restrictions for each IBM Storwize family storage system.

4.5.1 Version 7.8 configuration limits and restrictions for the IBM Storwize V3500 storage system

For a detailed list of configuration limits and restrictions for the IBM Storwize V3500 storage system, see [V7.8.x Configuration Limits and Restrictions for IBM Storwize V3500](#).

4.5.2 Version 7.8 configuration limits and restrictions for the IBM Storwize V3700 storage system

For a detailed list of configuration limits and restrictions for the IBM Storwize V3700 storage system, see [V7.8.x Configuration Limits and Restrictions for IBM Storwize V3700](#).

4.5.3 Version 7.8 configuration limits and restrictions for the IBM Storwize V5000 storage system

For a detailed list of configuration limits and restrictions for the IBM Storwize V5000 storage system, see [V7.8.x Configuration Limits and Restrictions for IBM Storwize V5000 and V5030F](#).

4.5.4 Version 7.8 configuration limits and restrictions for the IBM Storwize V7000 storage system

For a detailed list of configuration limits and restrictions for the IBM Storwize V7000 storage system, see [V7.8.x Configuration Limits and Restrictions for IBM Storwize V7000 and V7000F](#).

4.5.5 Version 7.8 configuration limits and restrictions for the SAN Volume Controller storage system

For a detailed list of configuration limits and restrictions for the SAN Volume Controller storage system, see [V7.8.x Configuration Limits and Restrictions for IBM System Storage SAN Volume Controller](#).



iSCSI storage connection security

Internet Small Computer System Interface (iSCSI) is fundamentally a storage area network (SAN) protocol that is similar to Fibre Channel (FC). The key difference is that FC uses a specialized network and iSCSI uses TCP networks. iSCSI technology benefits from the low cost and familiarity of Ethernet and IP networking. The flexibility of Ethernet and IP networking enables iSCSI attached systems to share hardware, extend range, and increase bandwidth by adding hardware.

However, this familiarity and flexibility lead to a requirement for appropriate network security. Each of the different types of networks that are used by iSCSI-attached systems has its own security considerations.

This chapter describes the following topics:

- ▶ 5.1, “iSCSI security model” on page 56
- ▶ 5.2, “Configuring CHAP for an IBM Storwize storage system” on page 57
- ▶ 5.3, “Configuring CHAP authentication for the host” on page 61
- ▶ 5.4, “iSCSI security” on page 71
- ▶ 5.5, “Mandatory security in real-world situations” on page 71

5.1 iSCSI security model

Each of the different types of networks that are used by iSCSI-attached systems has its own security consideration.

5.1.1 iSCSI network security

Consider the following types of iSCSI network traffic:

- ▶ Storage security can involve one or more of the following mechanisms:
 - Network isolation and physical security
 - Firewalls
 - Challenge Handshake Authentication Protocol (CHAP)
- ▶ Virtual Ethernet security can involve one or more of the following mechanisms:
 - Network isolation and physical security
 - Firewalls
 - A Secure Sockets Layer (SSL) connection for sensitive data during user enrollment and remote command submission

Network isolation and physical security

Network isolation minimizes the risk of data being accessed by unauthorized devices and data being modified as it traverses the network. You can create an isolated network by using a dedicated Ethernet switch or a dedicated virtual local area network (VLAN) on a physical VLAN switch/network. When you configure VLANs on a network, you must configure the VLAN on switch ports and end points (hosts and targets).

Physical security involves physical barriers that limit access to the network equipment and the network endpoints at some level (locked rack enclosures, locked rooms, locked buildings, and so on).

Firewalls

A firewall can be used between a shared network and host node to protect a node from unwanted network traffic. iSCSI-attached system traffic has the following attributes that can be helpful when you configure a firewall.

IP Security

IP Security (IPSec) encrypts storage and virtual Ethernet traffic on the iSCSI network. A related protocol, Internet Key Exchange (IKE), ensures that the communicating IP end points are authentic. IPSec, a set of security extensions that are developed by the Internet Engineering Task Force (IETF), provides privacy and authentication services at the IP layer (by enabling a host to select the required security protocols that determine the algorithms to use for the services) and puts in place the hidden keys that are required for these services. To help protect the contents of IP datagrams, IPSec uses encryption algorithms to transform the data.

Challenge Handshake authentication Protocol

CHAP protects against the possibility of an unauthorized system by using an authorized system's iSCSI name to access storage. CHAP does not encrypt network traffic, but limits which system can access an IBM Storwize storage path.

Using CHAP authentication can facilitate the management of access controls because it restricts access through account names and passwords.

CHAP authentication

There are two types of CHAP authentication:

- **One-way CHAP**

The IBM Storwize storage system authenticates the initiator node.

- **Two-way CHAP**

In addition to the one-way CHAP authentication, the initiator node also authenticates the target IBM Storwize storage system.

Note: CHAP secrets that you select for one-way authentication and two-way authentication must be different.

5.2 Configuring CHAP for an IBM Storwize storage system

You can use the command-line interface (CLI) or the graphical user interface (GUI) to configure CHAP to authenticate the IBM Storwize clustered system with iSCSI-attached hosts. CHAP authentication must be configured both on the SAN Volume Controller storage system and iSCSI-attached hosts. The SAN Volume Controller storage system supports both one-way and two-way CHAP authentication. When CHAP authentication is configured for a host by using the **chhost** CLI, the same CHAP secret must be configured on the host. To isolate connectivity problems, CHAP authentication settings can be delayed until you establish that the iSCSI host can access and discover the iSCSI ports of a SAN Volume Controller storage system.

5.2.1 Configuring CHAP for the IBM Storwize storage system by using the GUI

To configure authentication between an IBM Storwize clustered system and the iSCSI attached hosts, complete one of the following tasks:

- “Configuring one-way CHAP on the IBM Storwize iSCSI host by using the GUI”
- “Configuring two-way CHAP for the IBM Storwize storage system by using a GUI” on page 60

Configuring one-way CHAP on the IBM Storwize iSCSI host by using the GUI

To configure one-way CHAP between an IBM Storwize iSCSI host and the iSCSI-attached hosts, complete the following steps:

1. Connect to the IBM Storwize system with a browser and create a host object. If the host object is already created, skip this step and go to step 2.

To create a host object, click **Hosts** → **Host** → **Add Host**.

A window opens, as shown in Figure 5-1.



Figure 5-1 Adding a host object in the IBM Storwize system

You must enter the IQN of the host into the iSCSI port field. You can choose to have a meaningful name for the host object by completing the Name field. Although this field is optional, it is a preferred practice to provide a meaningful name to the host object to identify the host in the future.

After providing the information, click **Add**, which creates a host object.

2. Select **Hosts** → **Hosts** to display the list of configured host objects. Right-click the host object for which the one-way CHAP secret must be set. Then, click **Properties**, as shown in Figure 5-2.

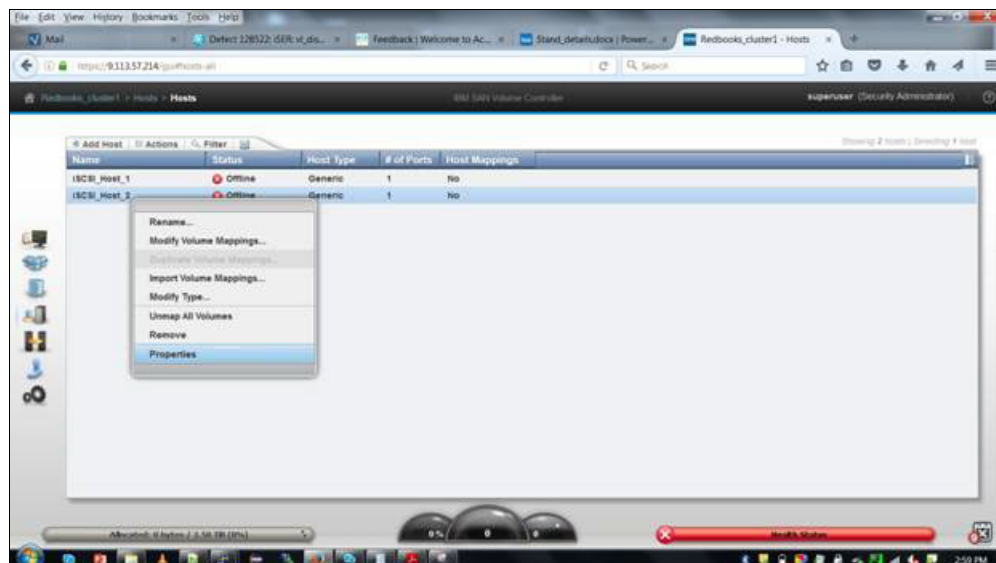


Figure 5-2 Modifying the host properties

3. Click the **Show Details** button in the lower left corner, which opens the iSCSI CHAP Secret field. Click **Edit**, complete the iSCSI CHAP Secret field, and click **Save**, as shown in Figure 5-3.

Host Details: host0

Overview Mapped Volumes Port Definitions

Host Name	host0
Host ID	1
Status	✓ Online
Host Type	Generic
# of FC Ports	
# of iSCSI Ports	
# of SAS Ports	
I/O Group	io_grp0, io_grp1, io_grp2, io_grp3
iSCSI CHAP Secret	

Edit

Show Details Close

Figure 5-3 Setting a one-way iSCSI CHAP secret for a host object

4. To disable the CHAP secret for a host object, complete steps 2 on page 58 and 3 and set the iSCSI CHAP Secret field to blank.

Configuring two-way CHAP for the IBM Storwize storage system by using a GUI

To configure authentication between an IBM Storwize clustered system and the iSCSI-attached hosts, complete the following steps:

1. Connect to the IBM Storwize clustered system with a browser by clicking **Settings** → **Network** → **iSCSI**, as shown in Figure 5-4.

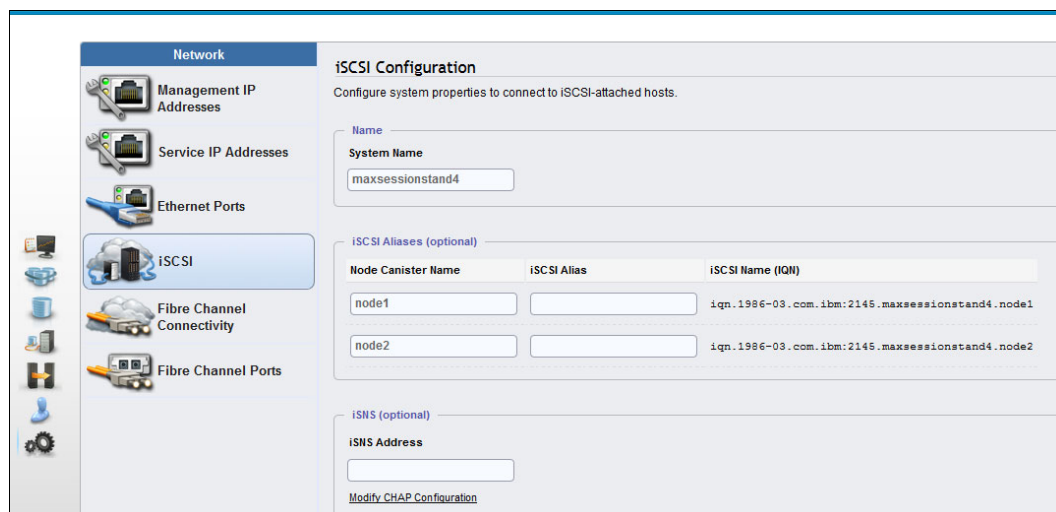


Figure 5-4 IBM Storwize iSCSI configuration

2. Click **Modify CHAP Configuration**. Figure 5-5 shows how to modify the system-wide CHAP secret.



Figure 5-5 Modifying the CHAP configuration

5.2.2 Configuring CHAP for the IBM Storwize storage system by using the CLI

Use the tasks in this section to configure CHAP for the IBM Storwize storage system.

Configuring one-way CHAP on IBM Storwize storage systems

To configure the CHAP configuration with the CLI, complete the following steps:

1. To enable the CHAP secret, run the following command:

```
chhost -chapsecret chap_secret
```

2. If wanted, to disable the CHAP secret, run the following command:

```
chhost -nochapsecret
```

For more information about the **chhost** command, see [IBM Knowledge Center](#).

Configuring two-way CHAP on IBM Storwize storage systems

To configure the CHAP configuration with the CLI, complete the following steps:

1. Configure a host object by using the **mkhost** CLI command. If the host object is already created, skip this step. Provide the host's IQN as `iscsiname` and optionally provide a name to the host object.

```
mkhost -iscsiname iqn.1994-05.com.redhat:eb5cdafabfe6 -name RHEL_host
```

2. To enable the CHAP secret, run the following command:

```
chsystem -chapsecret chap_secret
```

3. To disable the CHAP secret, run the following command:

```
chsystem -nochapsecret
```

For more information about the **chsystem** command, see [IBM Knowledge Center](#).

5.3 Configuring CHAP authentication for the host

A host system is an open systems computer that is connected to the switch through iSCSI. CHAP is a basic level of security. It is a protocol that is used to authenticate the peer of a connection and is based on the peer sharing a secret.

5.3.1 Setting up authentication for Linux hosts

This section provides instructions for setting up authentication for Linux hosts.

To set up CHAP authentication for a Linux host, complete the following steps:

1. Open `/etc/iscsi/iscsid.conf` or `/etc/iscsid.conf` with an appropriate editor.
2. Go to the CHAP settings paragraph. Example 5-1 shows the output.

Example 5-1 CHAP settings for a Linux host

```
# *****
# CHAP Settings
# *****

# To enable CHAP authentication set node.session.auth.authmethod
# to CHAP. The default is None.
node.session.auth.authmethod = CHAP

# To set a CHAP username and password for initiator
# authentication by the target(s), uncomment the following lines:
#node.session.auth.username = username
#node.session.auth.password = password
node.session.auth.username = iqn.host.test.com
node.session.auth.password = xxxxxxxxxxxx

# To set a CHAP username and password for target(s)
```

```
# authentication by the initiator, uncomment the following lines:
#node.session.auth.username_in = username_in
#node.session.auth.password_in = password_in
node.session.auth.password_in = yyyyyyyyyyyyyy

# To enable CHAP authentication for a discovery session to the target
# set discovery.sendtargets.auth.authmethod to CHAP. The default is None.
discovery.sendtargets.auth.authmethod = CHAP

# To set a discovery session CHAP username and password for the initiator
# authentication by the target(s), uncomment the following lines:
discovery.sendtargets.auth.username = username
discovery.sendtargets.auth.password = password

# To set a discovery session CHAP username and password for target(s)
# authentication by the initiator, uncomment the following lines:
#discovery.sendtargets.auth.username_in = username_in
discovery.sendtargets.auth.password_in = password_in
```

3. Make the following changes to the CHAP settings section:

a. Set up one-way authentication on the Linux host by completing the following steps:

i. Set a CHAP user name and password to the initiator name:

```
node.session.auth.authmethod = CHAP
node.session.auth.username = <initiator IQN name>
node.session.auth.password = <CHAP secret for host>
```

ii. Set a discovery session CHAP user name and password to your initiator name:

```
discovery.sendtargets.auth.authmethod = CHAP
discovery.sendtargets.auth.username = <initiator IQN name>
discovery.sendtargets.auth.password = <CHAP secret for host>
```

iii. Save these settings. You must log out of any current sessions, restart the `iscsi` service, and rediscover the system iSCSI target for the CHAP secret to be effective.

Note: In the previous example, `xxxxxxxxxxxxx` is the CHAP secret for the host, and `iqn.host.test.com` is the IQN name of the initiator. The IQN name must be the same name that is used to create a host object in the IBM Storwize storage system with the `mkhost` command.

b. Set up two-way authentication on the Linux host:

i. Edit the `password_in` to CHAP secret that you set up with the `chsystem` command on the IBM Storwize storage system.

ii. Set a CHAP user name and password for the target or targets:

```
node.session.auth.password_in = <CHAP secret for clustered system>
```

iii. Set a discovery session CHAP user name and password for the target or targets:

```
discovery.sendtargets.auth.password_in = <CHAP secret for clustered system>
```

Note: Do not set the user name (field `node.session.auth.username_in` and `discovery.sendtargets.auth.username_in`) for two-way CHAP.

- iv. Save these settings. You must log out of any current sessions and rediscover the system iSCSI target for the CHAP secret to be effective.

Note: It is not mandatory to set up two-way authentication. Before you configure two-way authentication, ensure that you have one-way authentication configured and working for your host.

5.3.2 Setting up authentication for Microsoft Windows hosts

This section describes the authentication methods that are available for Windows hosts.

IBM Storwize storage system supports two CHAP methods:

- ▶ One-way CHAP authentication (the IBM Storwize storage system authenticates the host iSCSI initiator).
- ▶ Two-way CHAP authentication (both the IBM Storwize storage system and the initiator authenticate each other).

Setting up authentication for discovery sessions for Windows hosts

This section provides instructions for setting up authentication for discovery sessions for Windows hosts.

After you install the initiator software, complete the following steps to configure one-way authentication for Windows hosts:

1. Go to the Control Panel and click **Administrative Tools** → **iSCSI Initiator**, as shown in Figure 5-6.

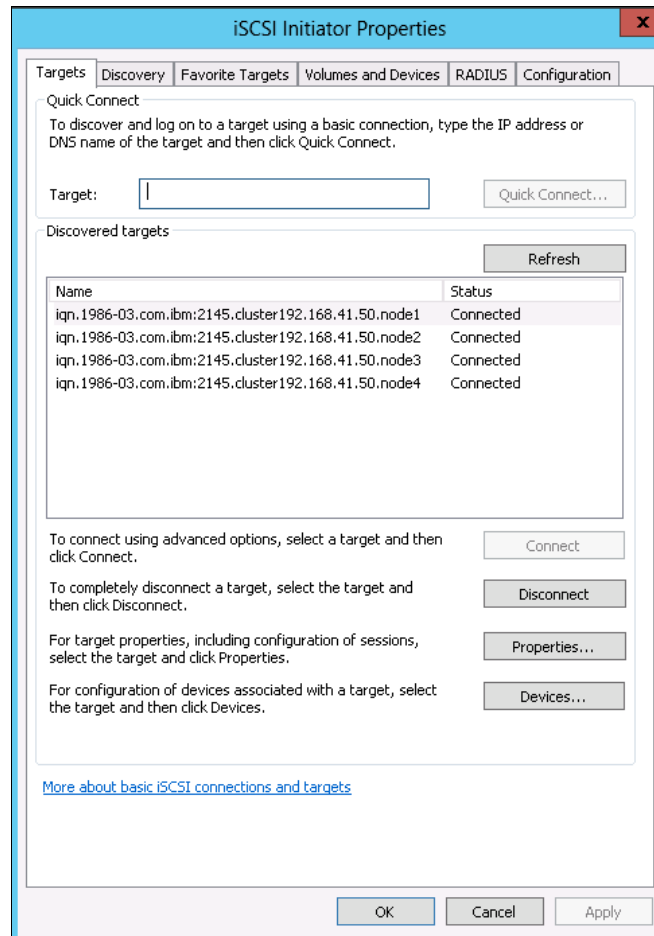


Figure 5-6 iSCSI Initiator Properties

2. Click the **Discovery** tab.

3. Click **Add** under the Target Portals section. You see the Add Target Portal dialog box, as shown in Figure 5-7.

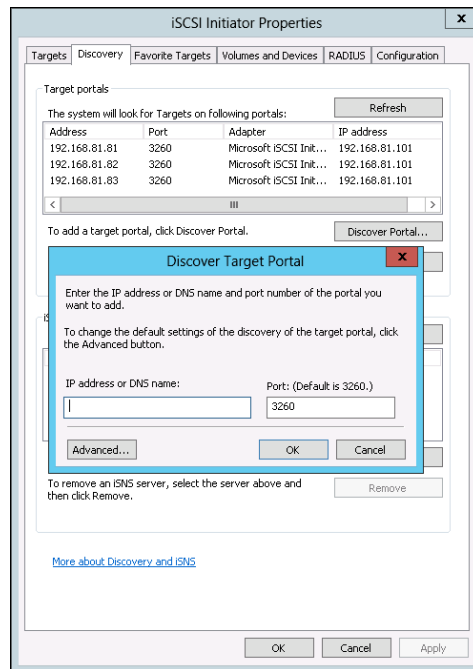


Figure 5-7 iSCSI Initiator Discover Target Portal

4. Click **Advanced**. You see the Advanced Settings window, as shown in Figure 5-8.

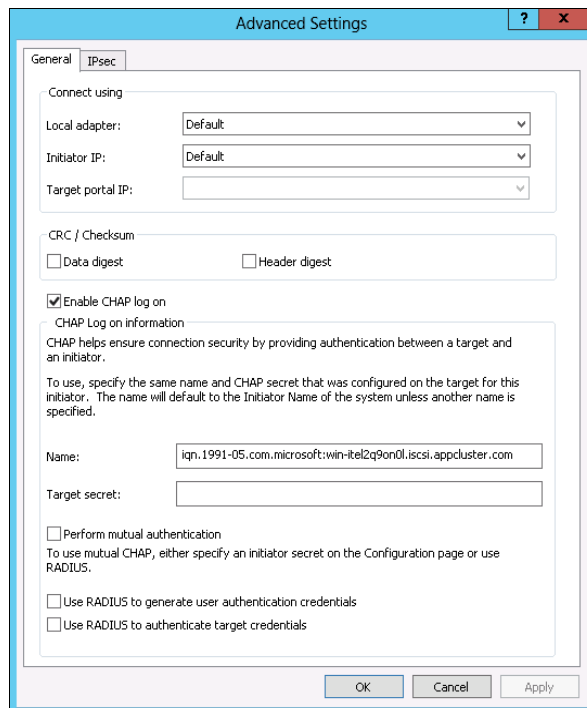


Figure 5-8 iSCSI Advanced Settings

5. Select **CHAP logon information**. For Windows Server 2012, select **Enable CHAP logon**.

6. Enter a value for the User name & Target secret, as shown in Figure 5-9. The user name must be the iSCSI qualified name (IQN) of the iSCSI host. The target secret must be a value of 12 characters. This value is the same value that you set with the **chhost** command on the IBM Storwize storage system for this host. After you enter all the required information, click **OK**.

Advanced Settings

General IPsec

Connect using

Local adapter: Default

Initiator IP: Default

Target portal IP:

CRC / Checksum

☐ Data digest ☐ Header digest

☒ Enable CHAP log on

CHAP Log on information

CHAP helps ensure connection security by providing authentication between a target and an initiator.

To use, specify the same name and CHAP secret that was configured on the target for this initiator. The name will default to the Initiator Name of the system unless another name is specified.

Name: iqn.1991-05.com.microsoft:win-itel2q9on0l.iscsi.appcluster.com

Target secret:

☐ Perform mutual authentication

To use mutual CHAP, either specify an initiator secret on the Configuration page or use RADIUS.

☐ Use RADIUS to generate user authentication credentials

☐ Use RADIUS to authenticate target credentials

OK Cancel Apply

Figure 5-9 CHAP configuration on Windows host

Note: Radius authentication is not supported for IBM Storwize storage systems.

Configuring two-way authentication for a Microsoft Windows host

Before you begin, verify that one-way authentication is working. Complete the following steps on the Microsoft Windows host:

1. Go to the Control Panel and click the **iSCSI Initiator** option.
2. From the iSCSI Initiator Properties window, click the **Configuration** tab, as shown in Figure 5-10.

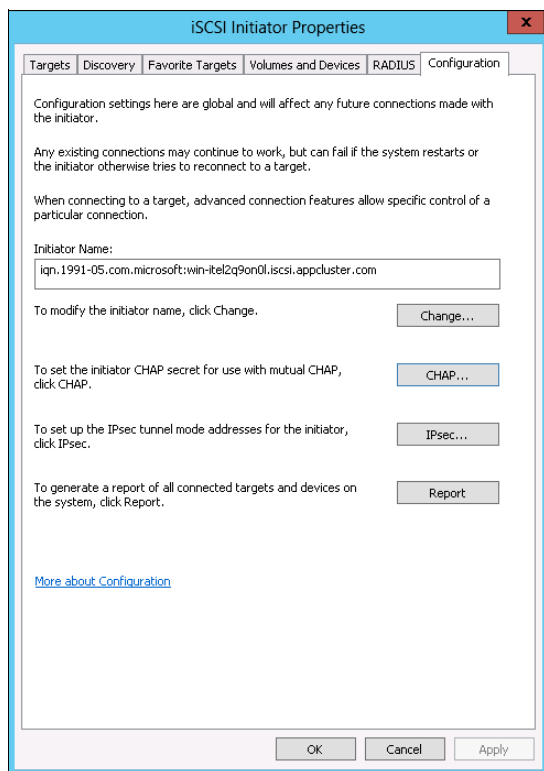


Figure 5-10 iSCSI initiator configuration tab

3. Click **CHAP**. On the iSCSI Initiator Mutual CHAP secret window, enter the IBM Storwize storage system CHAP secret and click **OK**, as shown in Figure 5-11.

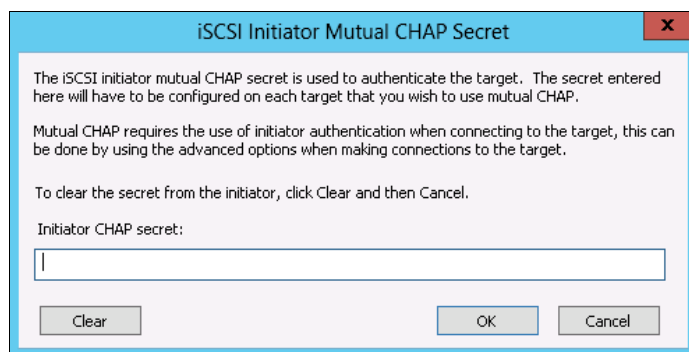


Figure 5-11 iSCSI Initiator Mutual CHAP Secret

Tips: Remember the following things.

- ▶ This setting applies to both the discovery session and normal session.
- ▶ The length restrictions in Example 5-2 on page 68 for CHAP secrets that apply to one-way authentication also apply to two-way authentication.
- ▶ The CHAP secrets for the IBM Storwize storage system and the host initiator cannot be the same.

To set up the two-way authentication, repeat the previous steps, but in this instance, select **Perform mutual authentication** from the Advanced Settings window. For more information, see “Setting up authentication for discovery sessions for Windows hosts” on page 63.

5.3.3 Setting up authentication for AIX hosts

This section describes how to set up CHAP authentication for AIX hosts.

Although the IBM Storwize storage system supports both one-way authentication and two-way authentication for iSCSI, the AIX software initiator supports only one-way authentication, where the IBM Storwize storage system target authenticates the initiator.

CHAP settings are defined in the `/etc/iscsi/targets` file on the host. The AIX initiator or host bus adapter (HBA) always uses its IQN as the CHAP user name.

To configure authentication on AIX hosts, complete the following steps:

1. Open the `/etc/iscsi/targets` file with your preferred editor.
2. For each line that contains a target definition, append the CHAP secret of the initiator in quotation marks:

```
10.2.1.105 3260 iqn.com.ibm-K167-42.fc1a "my_chapsecret"
```

The CHAP secret value that you set here must match the value that was configured on the IBM Storwize V7000 clustered system for the host object that is associated with this host. Because the IBM Storwize V7000 storage system authenticates on a per-initiator basis, the CHAP secret is the same for all the IBM Storwize V7000 storage system targets on a particular clustered system.

An example of the `/etc/iscsi/targets` file is shown in Example 5-2.

Example 5-2 CHAP setting section for AIX hosts

```
# ChapSecret      = %x22 *( any character ) %x22
#                  ;      "                  "
#                  ; ChapSecret is a string enclosed in double quotation marks.
The
#                  ; quotation marks are required, but are not part of the
secret.
# EXAMPLE 1: iSCSI Target without CHAP(MD5) authentication
#      Assume the target is at address 192.168.3.2,
#      the valid port is 5003
#      the name of the target is iqn.com.ibm-4125-23WTT26
# The target line would look like:
# 192.168.3.2 5003 iqn.com.ibm-4125-23WTT26
# EXAMPLE 2: iSCSI Target with CHAP(MD5) authentication
#      Assume the target is at address 10.2.1.105
#      the valid port is 3260
```

```
#      the name of the target is iqn.com.ibm-K167-42.fc1a
#      the CHAP secret is "This is my password."
# The target line would look like:
# 10.2.1.105 3260 iqn.com.ibm-K167-42.fc1a "This is my password."
# EXAMPLE 3: iSCSI Target with CHAP(MD5) authentication and line continuation
#      Assume the target is at address 10.2.1.106
#      the valid port is 3260
#      the name of the target is iqn.2003-01.com.ibm:00.fcd0ab21.shark128
#      the CHAP secret is "123ismysecretpassword.fc1b"
# The target line would look like:
# 10.2.1.106 3260 iqn.2003-01.com.ibm:00.fcd0ab21.shark128 #
"123ismysecretpassword.fc1b"
192.168.1.41 3260 iqn.1986-03.com.ibm:2145.hostsessionstand1.node1
192.168.2.43 3260 iqn.1986-03.com.ibm:2145.maxsessionstand4.node1 "chapsecret"
```

The two targets in the previous example are members of different IBM Storwize clustered systems. One target is configured to authenticate the initiator, and the other target is not configured to authenticate the initiator.

Target `iqn.1986-03.com.ibm:2145.hostsessionstand1.node1` is not configured for authentication; therefore, the CHAP secret field is blank.

Target `iqn.1986-03.com.ibm:2145.maxsessionstand4.node1` is configured for authentication; therefore, the CHAP secret field is set to `chapsecret` for authentication.

5.3.4 Setting up authentication for VMware hosts

To set up authentication on the VMware host, complete the following steps:

1. From the vSphere Client window, click the **Inventory** tab, as shown in Figure 5-12.

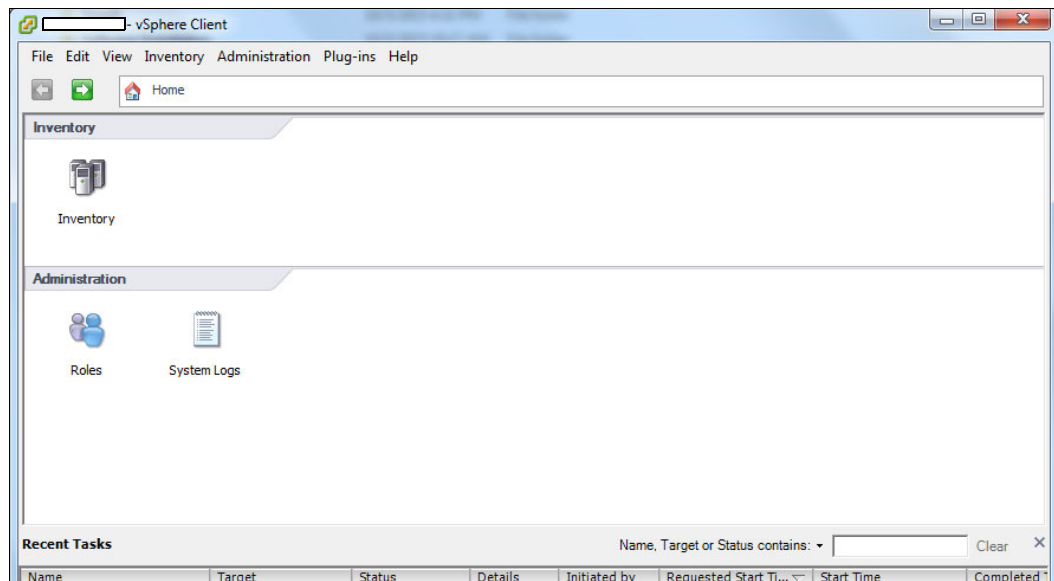


Figure 5-12 vSphere Client Console

2. Click **Physical Server** → **Configuration** → **Storage Adapters**, as shown in Figure 5-13.

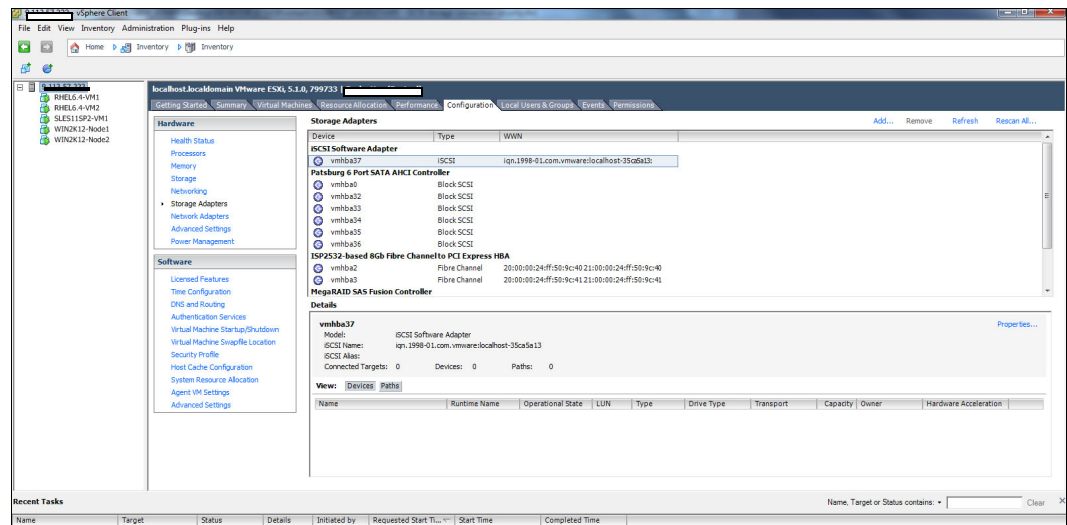


Figure 5-13 Showing the physical server storage adapter information

3. Select the iSCSI Storage Adapter, right-click, and select **Properties**, as shown in Figure 5-14.

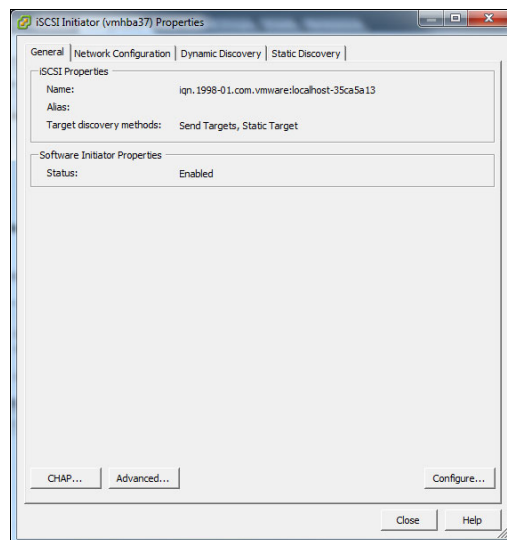


Figure 5-14 The vSphere Client iSCSI initiator properties

4. Click **CHAP**.

5. Select the preferred settings and enter the password, as shown in Figure 5-15.

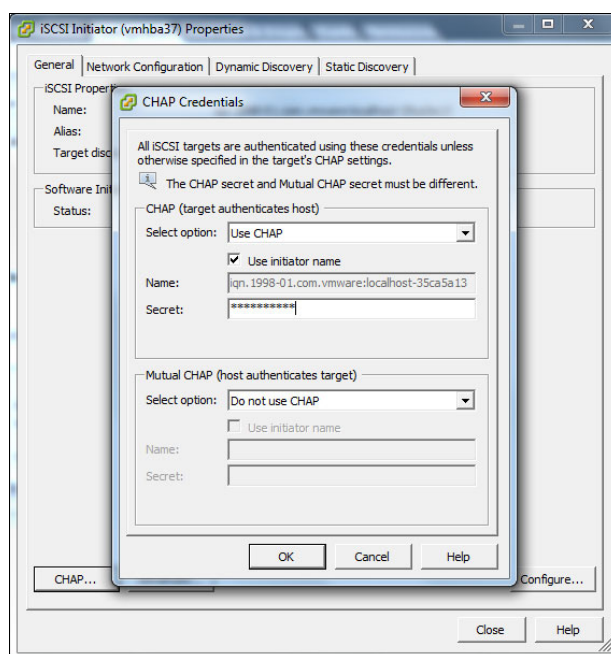


Figure 5-15 CHAP credentials

The IBM Storwize storage system supports only one-way CHAP with VMware hosts until release 7.7. Therefore, select the option **Do not use CHAP** in the Mutual CHAP field, as shown in Figure 5-15.

- From the CHAP (target authenticates host) menu, click **Use CHAP**.
- Click **Use initiator name**.
- Enter the CHAP secret.

5.4 iSCSI security

You can use the iSCSI protocol in networks where unauthorized data can be accessed, enabling different security methods. Encoding methods, such as IPSec, which use lower levels, do not require additional matching because they are transparent for higher levels and for the iSCSI. Various solutions can be used for authentication, for example, CHAP, Kerberos, or private keys exchange. An iSNS server can be used as a repository of keys.

5.5 Mandatory security in real-world situations

Security is not part of the design for the iSCSI protocol. The control packets and data packets are vulnerable to attack because messages are sent and received in plain text. The iSCSI protocol also enables configuration without any security measures. In this scenario, security is left to alternative protocols, such as CHAP and IPSec.

As a result, authentication for iSCSI that employs advance authentication methods to establish security, such as the CHAPv2, is important.

Security engineering principles, such as “security by obscurity”, are frequently adapted into iSCSI solutions. Improper design can cause security vulnerabilities.

The IETF considers security mandatory for iSCSI and mandates IPsec, as explained in [RFC 3723, Securing Block Storage Protocols over IP](#).

However, the appropriate method must be chosen based on the given, specific surrounding parameters of the environment. Where possible, it is a preferred practice to use a combination of Authentication Header (AH) and Encapsulated Security Payload (ESP) to ensure reliable authentication, guaranteed integrity, and confidentiality.

When iSCSI is used in the real world, security requires the utmost attention. iSCSI with added complements enables varying levels and complexity of security, depending on the security practices of an organization:

- ▶ Authentication
- ▶ CRC checksums
- ▶ Access control lists (ACLs)
- ▶ Firewall
- ▶ Encryption
- ▶ Isolation
- ▶ Segregation
- ▶ VLAN
- ▶ Virtual private network (VPN) for remote access

Consider the viability of system usage when you determine the correct security mechanisms for a solution. System usage and security must have a proportionate response. The more security a system has, the more complex a system becomes. The opposite also applies, where the less security a system has, the easier it is for a user to expose a system to security threats, increasing the potential for data theft, integrity issues, and confidentiality breaches. When security becomes inconvenient, it frequently leads to an infeasible system that can then lead to circumvention.

A good iSCSI security solution must have relative security benefits and few inconveniences. Therefore, a well-rounded solution involves physical security, operating system security, application security, network security, effective policies, and practical procedures.

For more information about iSCSI, see the following references:

- ▶ [RFC 3720, MPLS Support of Differentiated Services](#)
- ▶ [RFC 3723, Securing Block Storage Protocols over IP](#)
- ▶ *IBM BladeCenter iSCSI SAN Solution*, REDP-4153



IBM Storwize performance

This chapter provides a brief overview of the performance analysis capabilities of the IBM Storwize storage system. It also describes a method that you can use to collect and process performance statistics.

However, it is beyond the scope of this book to provide an in-depth understanding of performance statistics, or explain how to interpret them. For a more comprehensive look at the performance of the IBM Storwize storage system, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

For the IBM Storwize family, as with all other IBM storage subsystems, the official IBM tool for the collection of performance statistics, and to supply performance reporting is IBM Spectrum Control (formerly IBM Tivoli® Storage Productivity Center). For more information, see Chapter 8, “IBM Spectrum Virtualize and IBM Storwize performance monitoring” on page 157.

This chapter describes the following topics:

- ▶ 6.1, “Jumbo frames” on page 74
- ▶ 6.2, “VLAN separation” on page 74
- ▶ 6.3, “Subnetting” on page 76
- ▶ 6.4, “Quality of service and traffic prioritization” on page 78
- ▶ 6.5, “iSCSI protocol digests and performance” on page 79

6.1 Jumbo frames

Jumbo frame is a term that is applied to an Ethernet frame that carries more than the standard 1500-byte data payload. The most commonly quoted size for a jumbo frame is 9000 bytes, which is large enough for 8 KB of application data plus some upper layer protocol capacity.

Jumbo frames can improve performance in two ways:

- ▶ Packet assembly or disassembly in high-throughput environments can be an intensive operation. A jumbo frame decreases the number of packet processing operations by up to a factor of six.
- ▶ The protocol impact that is associated with the Ethernet packet when prepared for transmission is a smaller percentage of a jumbo frame than a regular sized frame.

Jumbo frames require the end points and all devices between them in the network to be configured to accept the larger packet size if they are not configured for them by default, including any network switching equipment.

For information about how to set jumbo frames, see 10.4.7, “Problem determination: Checking for performance problems” on page 207.

6.2 VLAN separation

Before you consider how virtual local area network (VLAN) separation contributes to enhancing the iSCSI performance, it is important to understand what a VLAN is and the advantages of implementing VLANs.

This section uses an example of a configuration that has iSCSI connectivity between ESX and an IBM Storwize storage system with the same subnet configuration.

For VLAN configuration guidelines for an iSCSI environment, see [IBM Knowledge Center](#).

6.2.1 VLAN

A VLAN can be described as a group of devices on one or more LANs that are configured to communicate as though they had the same physical connectivity, but actually might be on different LAN segments. It abstracts the idea of a LAN and might also comprise a subset of ports on a single switch or on multiple switches. VLANs help network administrators to partition the network to meet functional and security requirements. You can also refer VLANs as any broadcast domains, which are nothing but a partition of a network on the data link layer. Therefore, despite the geographical distribution of devices, you can have a logical group of workstations, servers, and network devices that are called a VLAN.

6.2.2 Advantages of VLANs

Here are some of the major advantages that can be obtained by implementing VLANs:

- ▶ **Increased Performance:** A logical grouping of servers, network devices, and workstations creates an administrative and authoritative boundary because the users do not have access to each group, which reduces collision. Less traffic must be routed and latency can decrease. Confinement of broadcast domains on a network reduces traffic. These factors contribute tremendously to achieve higher performance.
- ▶ **Security:** Network administrators have control over each port and user. A malicious or an anonymous user cannot log in to any switch port and sniff the network. Network administrators have the privilege to authorize each user to use specific resources and ports.
- ▶ **Reduced the need for routers:** You can reduce the need to deploy routers on a network to contain broadcast traffic. Flooding of a packet can be restricted to the switch port that is associated with the respective VLAN.
- ▶ **Improved manageability:** VLANs help in managing large networks more efficiently by allowing centralized configuration of devices in different geographical locations.

This section explains what a VLAN is and the major advantages it offers. 6.2.3, “VLAN and iSCSI performance” on page 75 explains how it enhances iSCSI performance.

6.2.3 VLAN and iSCSI performance

Figure 6-1 shows iSCSI connectivity that uses a 10 GbE link between the ESX and IBM Storwize storage systems. The main challenge here is a *traffic storm*. Consider a scenario where massive read operations are made from ESX. A single NIC on ESX can see all four storage paths. Therefore, all the storage paths send the data back. This process creates excessive load on the switch and it becomes a bottleneck. The switch might start dropping packets.

Also, in such scenarios, *latency* issues certainly are reported because the maximum amount of Ethernet broadcast is being observed.

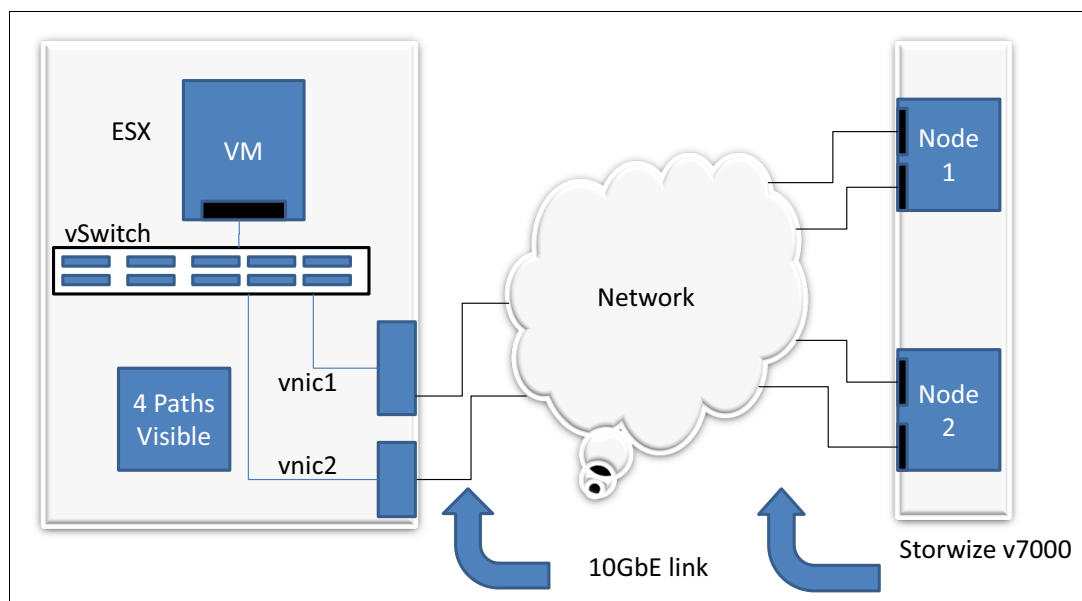


Figure 6-1 iSCSI configuration without a VLAN

These issues must be solved by reducing the number of paths, to be specific, by segregating the paths and restricting access so that the resources on the switch are not exhausted and it can handle the traffic efficiently. One solution to these problems is VLAN implementation and the other is a configuration with multiple subnets, which is described in 6.3, “Subnetting” on page 76.

Figure 6-2 shows a VLAN split out on the switch, which restricts access by reducing the number of paths sending traffic on the switch. This reduction helps achieve effective usage of the switch and reduces the traffic, which reduces the latency and effective usage of network resources to end the traffic storm. All these factors eventually contribute to enhancing the iSCSI performance.

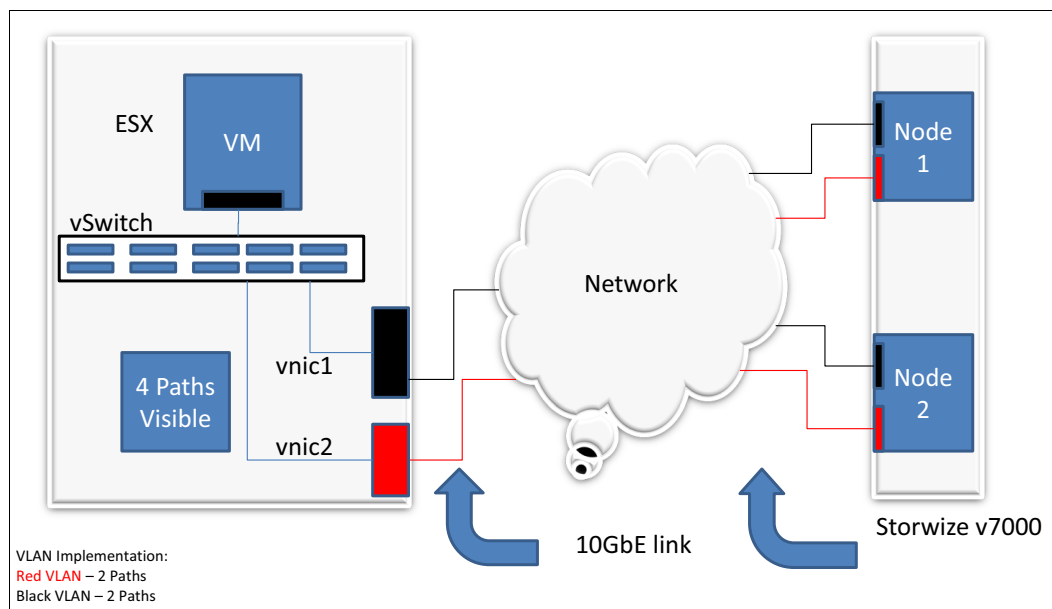


Figure 6-2 iSCSI configuration with a VLAN

6.3 Subnetting

Subnetting is one of the solutions to overcome network bottlenecks and latency issues. This section explains what subnetting is, its advantages, and how it helps enhance iSCSI performance.

6.3.1 Network subnetting

A division of an IP network is called a *sub network* or a *subnet*. The practice of dividing an IP network into further smaller networks is termed as *subnetting*.

Advantages

There are differences in the implementation of VLAN and subnetting. However, the advantages of implementing either strategy are almost the same. The reason is that both the strategies are focused on breaking the large network into small networks, restricting access by creating administrative boundaries, configuring hosts and storages to use only the assigned network, preventing bombardment of heavy traffic in the network, and preventing network bottlenecks, packet drop, and latency issues. Both strategies help network administrators to manage efficiently the network. For a detailed description of the advantages, see 6.2.3, “VLAN and iSCSI performance” on page 75.

6.3.2 Subnetting and iSCSI performance

This section uses the same example that is used in 6.2, “VLAN separation” on page 74, where ESX is connected to an IBM Storwize V7000 storage system by using iSCSI (10 GbE).

Figure 6-3 shows the same subnet configuration, where only one switch connects the ESX with the IBM Storwize V7000 storage system. This configuration has the following challenges and impact on iSCSI performance:

- ▶ Single point of failure: Because there is only one switch in this environment, if the switch observes a failure, it leads to a complete unavailability of storage.
- ▶ Network bottleneck (traffic storm): In Figure 6-3, a total of eight paths are visible. If massive reads are made from ESX, every path sends the data back, exhausting the network resources and potentially causing the switch to drop packets.
- ▶ Latency: The maximum amount of Ethernet broadcast causes latency issues, which contribute to deteriorating iSCSI performance.

Figure 6-3 shows a single subnet configuration.

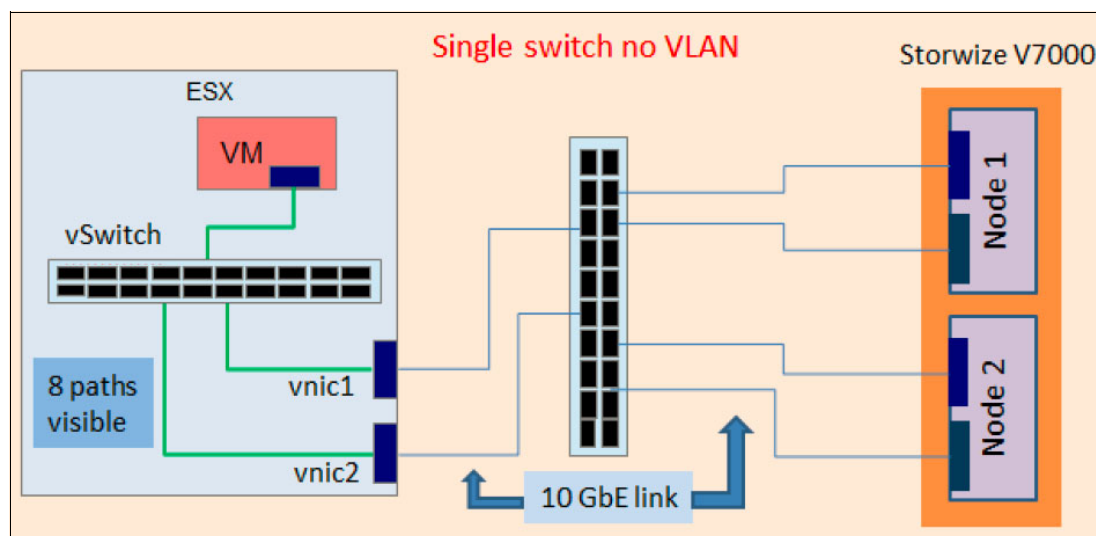


Figure 6-3 Single subnet configuration¹

¹ This figure was taken from *Networking best practices with IBM Storwize V7000 and iSCSI*, found at: <http://ibm.co/23Gfc8K>

Now, consider the modified configuration in Figure 6-4, where multiple subnets are configured.

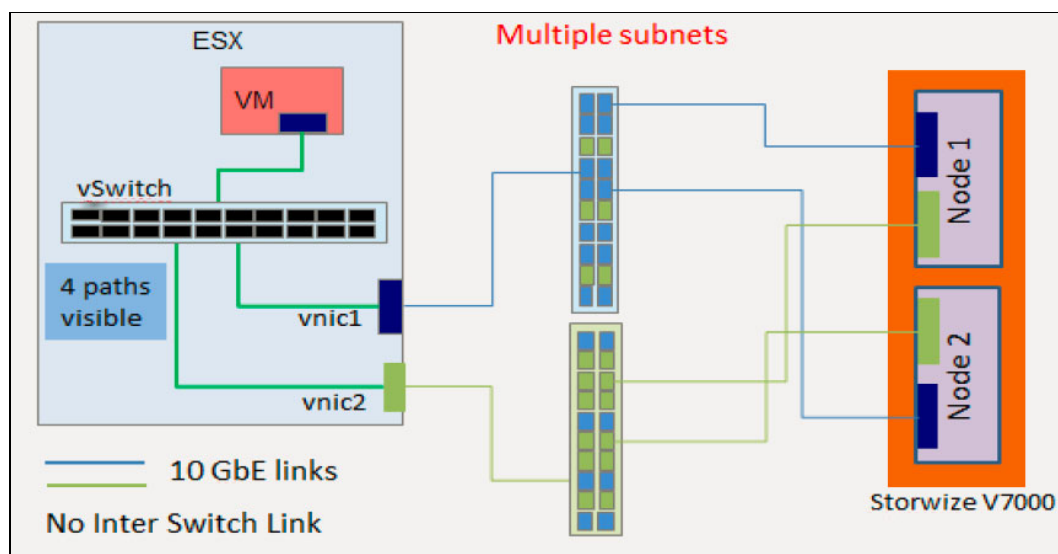


Figure 6-4 Multiple subnet configuration

The configuration in Figure 6-4 has a typical multipathing configuration, where there are two switches that are not connected to each other. This configuration helps reduce the number of paths. Now, four paths are visible because each vNIC is connected to one switch, which reduces traffic on each switch. The risk of a network bottleneck is mitigated and latency issues are addressed. Having two switches also prevents a single point of failure. To conclude, the network is properly managed, leading to a controlled environment that is more secure and optimized to improve iSCSI performance.

6.4 Quality of service and traffic prioritization

Certain Ethernet traffic can be prioritized relative to other traffic, specifically in 10 Gb enhanced Ethernet networks.

Priority Flow Control (PFC), as described in the IEEE 802.1Qbb standard, allows the network to control the flow of Ethernet traffic based on the class of traffic that is identified with its associated Priority Group. Network frames can be tagged with one of eight priority levels (described in the IEEE 802.1p priorities). Switches that are 802.1p compliant can give preferential treatment to priority values in terms of transmission scheduling. For example, priority can be assigned to iSCSI or Fibre Channel over Ethernet (FCoE) traffic to prioritize storage traffic if network congestion occurs.

Enhanced Transmission Selection (ETS): IEEE 802.1Qaz provides a mechanism to use 802.1p priority values to map traffic to defined bandwidth allocations on outbound switch links. Thus, iSCSI traffic can be given higher bandwidth allocation relative to other traffic. This higher allocation helps improve performance because in case the network gets saturated, the PFC pause mechanism pauses traffic with lower priority and prevents it from using bandwidth that is allocated to iSCSI storage traffic.

PFC and ETS must be configured on the Switch network to get guaranteed performance in a network congestion scenario. For more information about how to configure PFC and ETS, see 7.8, “Configuring Priority Flow Control for the IBM Storwize storage system” on page 148.

6.5 iSCSI protocol digests and performance

Digests can provide an extra measure of data integrity above TCP checksums. They might be useful if a WAN or known unreliable network is being used for iSCSI traffic. Most administrators prefer to run iSCSI over a network at least as reliable as a normal Network File System (NFS)/Common Internet File System protocol (CIFS)-level network, which they trust to deliver reliable file I/O requests and responses without any upper layer protocol checksums. If digests are enabled for the iSCSI protocol, the IBM Storwize iSCSI target must calculate and append checksum values to each outgoing packet, and calculate and verify values on each incoming packet.

There are two types of digests: header digests and data digests. Header digests performs checksums on only the 48-byte header of each iSCSI Protocol Data Units (PDU), and data digests perform checksums on the data segment that is attached to each PDU.

Enabling headers, data digest, or both is expected to have some impact on performance and CPU usage on the initiator and target systems because they require extra processing for calculation of checksums both for incoming and out going packets. For IBM Storwize storage systems, when performance measurements were done in the lab environment with header and data digests enabled, the IOPS workloads were virtually unaffected, and there was some drop seen in bandwidth usage. There was a marginal increase in CPU usage. Overall, enabling digests does not cause a significant decrease in performance.



Part 2

iSCSI host attachment

This part describes Internet Small Computer System Interface (iSCSI) host attachment and how to plan, configure, secure, and troubleshoot connections.

This part describes the following topics:

- ▶ Chapter 7, “Configuring the IBM Storwize storage system and hosts for iSCSI” on page 83
- ▶ Chapter 8, “IBM Spectrum Virtualize and IBM Storwize performance monitoring” on page 157
- ▶ Chapter 9, “IBM Spectrum Virtualize and IBM Storwize storage systems on the OpenStack platform” on page 169
- ▶ Chapter 10, “Troubleshooting” on page 175



Configuring the IBM Storwize storage system and hosts for iSCSI

This chapter describes the configuration of IBM Storwize and host operating systems for using the Internet Small Computer System Interface (iSCSI). It also describes some key considerations for configuring boot from SAN, Priority Flow Control (PFC), host-side clustering solutions, and Internet Storage Name Service (iSNS) for iSCSI.

This chapter describes the following topics:

- ▶ 7.1, “Configuring the IBM Storwize storage system for iSCSI” on page 84
- ▶ 7.2, “Configuring initiators for iSCSI” on page 87
- ▶ 7.3, “Configuring iSCSI on AIX 7.1” on page 92
- ▶ 7.4, “Configuring iSCSI for SUSE Linux Enterprise Server” on page 97
- ▶ 7.5, “Configuring iSCSI for Windows 2012” on page 108
- ▶ 7.6, “Configuring iSCSI for VMware ESXi hosts” on page 120
- ▶ 7.7, “iSNS server configuration” on page 143
- ▶ 7.8, “Configuring Priority Flow Control for the IBM Storwize storage system” on page 148
- ▶ 7.9, “Configuring the iSCSI host for the HyperSwap cluster” on page 152

7.1 Configuring the IBM Storwize storage system for iSCSI

This section describes how to configure iSCSI on IBM Storwize storage systems with the GUI and command-line interface (CLI). At the time of writing, an IBM Storwize V7000 storage system is used for showing the configuration steps. However, the steps are similar for other IBM Storwize models.

7.1.1 Setting the IBM Storwize iSCSI IP address

iSCSI is supported on 1-Gb and 10-Gb interfaces on IBM Storwize storage systems. The IP address for the wanted iSCSI interface can be set by using GUI or the CLI.

To configure the iSCSI IP with the GUI, complete the following steps:

1. Log in to the IBM Storwize storage system's GUI and click **Network** → **Settings**.
2. Click **Ethernet Ports**.
3. Right-click the port that you want to use for iSCSI and click **Modify**.
4. Enter the IP address, subnet mask, and gateway for the interface and make sure that the **iSCSI hosts** check box is selected.
5. For configuring an IPv6 address, select the **Show IPv6** option and enter the settings. Ensure that the **SCSI hosts** check box is selected.
6. Click **OK** to save the settings.

To configure the iSCSI IP with the CLI, complete the following steps:

1. Log in to the IBM Storwize storage system CLI.
2. To configure a new port IP address to a specified Ethernet port of a node with an IPv4 address, run the following command:

```
cfgportip -node node_name | node_id -ip ipv4addr -gw ipv4gw -mask subnet_mask -vlan vlan_id port_id
```

Where *node_name* | *node_id* is the name or ID of the node that is being configured, *ipv4addr* is the IPv4 address for the Ethernet port, *ipv4gw* is the IPv4 gateway IP address, *subnet_mask* is the IPv4 subnet mask, and *port_id* specifies the Ethernet port ID of the port on the node. To view a list of Ethernet ports on the system, run the **lspportip** command. The optional **-vlan** parameter sets the virtual local area network (VLAN) ID for an IPv4 address that is configured for iSCSI host attachment.

Here is an example command:

```
cfgportip -node node1 -ip 192.168.41.151 -mask 255.255.255.0 -gw 192.168.41.1 2
```

3. To configure a new port IP address to a specified Ethernet port of a node with an IPv6 address, run the following command:

```
cfgportip -node node_name | node_id -ip_6 ipv6addr -gw_6 ipv6gw -prefix_6 prefix -vlan_6 vlan_id port_id
```

Where *node_name* | *node_id* is the name or ID of the node that is being configured, *ipv4addr* is the IPv6 address for the Ethernet port, *ipv6gw* is the IPv6 gateway IP address, *prefix_6* is the IPv6 prefix, *port_id* specifies the Ethernet port ID of the port on the node. To view a list of Ethernet ports on the system, run the **lspportip** command. The optional **-vlan** parameter sets the VLAN ID for an IPv4 address that is configured for iSCSI host attachment.

7.1.2 Setting optional iSCSI settings on IBM Storwize storage systems

This section describes the configuration of optional iSCSI settings, such as iSCSI aliases, CHAP, VLAN, and iSNS for IBM Storwize storage systems.

Configuring optional iSCSI settings with the GUI

To configure optional iSCSI settings with the GUI, complete the following steps:

1. Log in to the IBM Storwize storage systems GUI and click **Network** → **Settings**.
2. Click **iSCSI**. The iSCSI configuration options are displayed in the right-side pane:
 - a. Enter the iSCSI alias for each node under iSCSI Aliases in the iSCSi Alias text box and click **Apply Changes**.

Tip: An alias cannot contain spaces. However, an underscore character can be used.

- b. Enter the IP address of the iSNS server under iSNS in the iSNS Address text box. Click **Apply Changes** to save the changes.
- c. To enter the CHAP settings, click **Modify CHAP Configuration**.
- d. Enter the CHAP secret in the System-wide CHAP secret text box and select the **Use for iSCSI-attached hosts** check box. Click **Modify** to save the changes.

The VLAN can be set only after the IP is configured for the port. To set the VLAN, complete the following steps:

- a. Click **Ethernet Ports**, right-click the port, and click **Modify VLAN**.
- b. In the new window, select the **Enable for VLAN** option to enable VLAN tagging.
- c. Enter the VLAN ID in the VLAN tag text box.
- d. Select the **Apply changes to failover port to** check box. To prevent access to the volumes in the event of a failover of nodes, make sure that the failover port has the same VLAN tag.
- e. Click **Modify** to save the settings.

Configuring the optional iSCSI settings with the CLI

This section explains how to configure or modify an iSCSI alias with the CLI.

Tip: An alias cannot contain spaces. However, an underscore character can be used.

Configuring a new port's IP address to a specified Ethernet port

To configure a new port's IP address to a specified Ethernet port of a node, run the following command:

```
chnode -iscsialias alias node_name | node_id
```

Where *iscsialias* is the iSCSI alias and *node_name | node_id* is the node name or ID for which the alias is configured.

For example, run the following command:

```
chnode -iscsialias iscsi-node-1 node1
```

Specifying an IPv4 or IPv6 address for the iSCSI storage naming service

To specify an IPv4 or IPv6 address for the iSCSI Storage Naming Service (iSNS), enter the following CLI commands:

```
chsystem -isnsip sns_server_address
chsystem -isnsip_6 ipv6_sns_server_address
```

Where *isnsip* is the IPv4 address and *isnsip_6* is the IPv6 address of the iSNS.

Here is an example command:

```
chsystem -isnsip 192.168.41.100
chsystem -isnsip_6 2001:0db8:85a3:0000:0000:8a2e:0370:7334
```

Configuring iSCSI authentication with the CLI

The SAN Volume Controller storage system supports one-way and mutual CHAP authentication. One-way CHAP authentication requires the administrator to configure a per-host CHAP secret that must be known to the SAN Volume Controller storage system when the host is configured. The same CHAP secret must be configured on the host. The host uses this CHAP to authenticate itself to the SAN Volume Controller storage system. The SAN Volume Controller storage system enables a different CHAP to be configured for each iSCSI host that is configured with the SAN Volume Controller storage system.

The SAN Volume Controller storage system also supports mutual CHAP authentication. This is two-way authentication that ensures that both the host and the SAN Volume Controller storage system authenticate themselves with each other. The mechanism to configure host CHAP is the same as mentioned above. In addition, a single CHAP secret can be configured for the entire SAN Volume Controller cluster by using the chsystem CLI. This CHAP must also be configured on the iSCSI hosts.

To configure iSCSI authentication with the CLI, use the following commands:

- To configure CHAP authentication for an iSCSI host, run the following command:

```
chhost -chapsecret chap_secret host_name
```

Where *chap_secret* is the CHAP secret to be used to authenticate the system through iSCSI and *host_name* is the name of the iSCSI host. The *chap_secret* value must be 12 characters.

- To set the authentication method for the iSCSI communications of the system, run the following CLI command:

```
chsystem -iscsiauthmethod chap -chapsecret chap_secret
```

Where *chap* specifies that CHAP is the authentication method and *chap_secret* is the CHAP secret to be used. The specified CHAP secret cannot begin or end with a space.

- You can run the **lsiscsiauth** command to display the CHAP secret that is configured.

7.2 Configuring initiators for iSCSI

The initiators must be configured to connect to iSCSI targets. The configuration varies for different operating systems. This section describes discovery types and iSCSI operational parameters that are supported by IBM Storwize storage systems.

Accessing SAN Volume Controller LUNs: The iSCSI initiator must be configured by running the `mkhost` command to access LUNs (vDisks) that are configured on a SAN Volume Controller storage system. Unconfigured hosts can log in, but LUN assignment cannot be done.

Each host that is configured on a SAN Volume Controller storage system can access only those vDisks that are mapped to it.

iSCSI hosts can set up iSCSI sessions with a SAN Volume Controller cluster even though they are not configured. They cannot access any of the LUNs (vDisks) that are configured on the SAN Volume Controller storage system, but can discover the iSCSI target that is configured on the SAN Volume Controller storage system that includes the per controller IQN and IP addresses.

Link local addressing is not supported: Link local addressing is a special range of IP addresses that are not supported. For IPV6, these addresses have the `fe:80.*` prefix, and for IPV4 they are in the block `169.254.0.0/16`.

7.2.1 iSCSI discovery mechanisms

The iSCSI protocol implements two session types: *discovery session* and *normal session*. A discovery session is used to discover the target's iSCSI name and network portal information. A normal session is used for all other purposes. The initiator can discover iSCSI targets only when it has information about the IP address, TCP port number, and iSCSI target name (also known as the *IQN* of the target). The iSCSI discovery mechanisms provide a way for iSCSI initiators to discover information about iSCSI targets. The discovery mechanisms that are described in this section vary in terms of the assumptions about the information that is already available to the initiators and the information that still must be discovered.

The iSCSI protocol supports the following discovery mechanisms:

- Static configuration

Static configuration assumes that the IP address, TCP port number, and target name or IQN is already available to the initiator. The initiator does not need to do any discovery. It uses the target's IP address and TCP port number to create a TCP connection with the target and uses a target name (IQN) to create an iSCSI session with the target. This method is suitable for small iSCSI implementations.

- SendTargets

This mechanism assumes that the target's IP address and TCP port number are already available to the initiator. The initiator uses this information to establish a discovery session with the target. After the discovery session is established, the initiator sends a **SendTarget** text command to the target to obtain a list of targets available. Thereafter, the initiator can establish sessions with the individual targets that are returned in the **SendTarget** query response. This method is suitable for large iSCSI implementations.

- **Zero configuration**

In this mechanism, the assumption is that the initiator does not have any information about the target. It can either send multicast discovery messages directly to the target or can send discovery messages to storage name servers. There are many discovery frameworks available. However, the most popular one is iSNS. In the iSNS-based discovery, initiators query the iSNS server to discover the iSCSI targets and their parameters.

IBM Storwize storage systems support static discovery, SendTarget discovery, and iSNS based discovery mechanisms.

7.2.2 iSCSI operational parameters

iSCSI implements two phases: the *login phase* and *full feature* phase. The login phase occurs first and is composed of two stages: the *security parameter negotiation* and *operational parameter negotiation*. These phases are optional but at least one phase must occur. If iSCSI authentication is implemented for iSCSI devices, the security negotiation phase must occur first. Although the operational parameter negotiation phase is optional, it is a practical requirement for real-world deployments. Each initiator and target must support the same parameters to communicate successfully. It is possible for the default parameter settings for each iSCSI device to match, but it is not probable. Thus, the operational parameter negotiation phase is implemented by almost all iSCSI devices to negotiate various parameters.

Here is a list of the operational parameters that are supported by IBM Storwize storage systems and a brief description of them:

- **HeaderDigest and DataDigest**

Digests provide another level of integrity check beyond the integrity check that is provided by the link layers. It covers the whole communication path, including all elements that might change network level Protocol Data Units (PDUs), such as routers, switches, and proxies. The header and data digest can have only two values: “None” and “CRC32C”. The default value is “None” for both HeaderDigest and DataDigest, which means that they are disabled. When digests are enabled on iSCSI devices, the parameter is set to “CRC32C”.

When the iSCSI initiator and target agree on a digest, this digest is used for every PDU. The system does a checksum over each iSCSI PDU’s header and data part and verifies them by using the CRC32C algorithm. On IBM Storwize storage systems, header and data digest support is provided only if the initiator is configured to negotiate and it is only supported for normal sessions. Header and data digest support for a discovery session is not supported.

- **FirstBurstLength**

This operational parameter limits the maximum number of bytes of unsolicited data that can be sent during the running of a single SCSI command by an iSCSI initiator to the target. It covers the immediate data and the sequence of unsolicited Data-out PDUs that follow the command.

This parameter is irrelevant in the following two cases:

- *SessionType=Discovery
- *InitialR2T=Yes and ImmediateData=No

Important: FirstBurstLength must not exceed MaxBurstLength.

IBM Storwize storage systems support **FirstBurstLength** values up to 32 KB, which is by default set on the target.

Immediate data is not supported on SAN Volume Controller or Storwize storage systems, so **FirstBurstLength** is irrelevant regarding immediate data.

► **Initial Request to Transfer (**InitialR2T**)**

Request to Transfer is a mechanism in iSCSI where the iSCSI target requests the iSCSI initiator to send data. It is of significance typically for write I/O operations where data is sent to the iSCSI targets. The **InitialR2T** parameter is used to allow an iSCSI initiator to start sending data to the iSCSI target as though it already received the initial ready to transfer request from the iSCSI target. By default, initial Request To Transfer is required by most iSCSI devices unless both the initiator and the target both negotiate not to use it by setting **InitialR2T=No**. For IBM Storwize devices, by default, **InitialR2T** is set to “1”, which means that the target does not accept data from the initiator until it send the R2T request.

► **ImmediateData**

The iSCSI initiator and target negotiate for support for immediate data during the operational negotiation phase. The initiator and target can turn on support for immediate data if both of them have the setting **ImmediateData=Yes**. When enabled, the initiator might send unsolicited data immediate data, one unsolicited burst of Data-Out PDUs to the target, or both. This parameter also depends on the support of **InitialR2T**. Here is the behavior with different permutations of the **InitialR2T** parameter:

- If **ImmediateData** is set to “Yes” and **InitialR2T** is also set to “Yes”, the initiator can send only immediate data on the first burst.
- If **ImmediateData** is set to “No” and **InitialR2T** is set to “Yes”, then the initiator does not send any unsolicited data and target rejects unsolicited data if it receives it with appropriate response code.
- If **ImmediateData** is set to “No” and **InitialR2T** is set to “No”, the initiator does not send unsolicited immediate data but it might send one unsolicited burst of Data-Out PDUs.
- If **ImmediateData** is set to “Yes” and **InitialR2T** is set to “No”, then the initiator can send unsolicited immediate data, one unsolicited burst of Data-Out PDUs, or both.

For IBM Storwize storage systems, **ImmediateData** is not supported (the **ImmediateData** parameter is by default set to “NO”).

► **Maxconnections**

The **Maxconnection** parameter is used by the iSCSI initiator and target to negotiate the maximum number of TCP connections that can be requested or are acceptable. The parameter value is numerical and can be 1 - 65535. The default value of this parameter for IBM Storwize storage systems is 1, which means that only one TCP connection per session is supported.

► **MaxRecvDataSegmentLength**

iSCSI initiators or targets declare the maximum data segment length in bytes that it can receive in an iSCSI PDU by using this parameter. The initiator or the target must not send PDUs with a data segment that does not exceed the **MaxRecvDataSegmentLength**. IBM Storwize storage systems support a **MaxRecvDataSegmentLength** length of 32 KB.

► **MaxBurstLength**

The **MaxBurstLength** parameter negotiates the maximum SCSI data payload in bytes that can be transferred in Data-In PDUs or a solicited Data-Out iSCSI sequence. A sequence consists of multiple Data-In or Data-Out PDUs. IBM Storwize storage systems support a **MaxBurstLength** of 32 KB.

► **DefaultTime2Wait**

The default time to wait negotiates the minimum time, in seconds, to wait before attempting an implicit or an explicit logout or an active task reassignment if there is an unexpected termination of connection or a connection reset. For IBM Storwize storage systems, the default time to wait is 2 seconds.

► **DefaultTime2Retain**

The default time to retain is the maximum time, in seconds, before which an active task reassignment is still possible after an initial wait (**DefaultTime2Wait**) if there is an unexpected termination of connection or a connection reset. If the value for this parameter is set to zero, the connection or task is immediately discarded by the target. On IBM Storwize storage systems, the default time to retain is 20 seconds.

► **MaxOutstandingR2T**

MaxOutstandingR2T is the maximum number of requests to transfer that can remain active per task. A request to transfer is considered to be active until the last PDU for that task is transferred. The next request for data transfer can be sent only after the previous one is completed. IBM Storwize storage systems support **MaxOutstandingR2T** of 1, which means that only one request to transfer can remain outstanding at any time.

► **DataPDUInOrder**

The **DataPDUInOrder** parameter is used to indicate that the PDUs within the sequence must be in order and overlays are forbidden. The IBM Storwize storage system requires data PDUs to be sent to it in order.

► **DataSequenceInOrder**

Data sequence is a sequence of Data-In or Data-Out PDUs that is terminated with a Data-In or Data-Out PDU with an 'F' bit set. If **DataSequenceInOrder** is set to 'Yes', Data Sequences must be transferred by using continuously non-decreasing sequence offsets. If it is set to 'No', then data PDUs can be transferred in any order. IBM Storwize storage systems require data sequence in order.

► **ErrorRecoveryLevel**

This parameter is negotiated by the iSCSI initiator and the target to identify the recovery level that is supported. The recovery level reflects the capability of the iSCSI device to recover from an error condition. There are three levels of error recovery:

– Session recovery (**ErrorRecoveryLevel**=0)

In session recovery, all TCP connections for a session are closed, all running and queued SCSI tasks are stopped, all outstanding SCSI commands are stopped with the appropriate SCSI service responses, and the session is restarted on new set of TCP connections.

– Digests Failure Recovery (**ErrorRecoveryLevel**=1)

This recovery class is composed of two subclasses, within a command (that is, without requiring a command restart) and within a connection (that is, without requiring a connection to be rebuilt, but perhaps requiring a command restart). This class provides the capability to recover from header and data digest errors plus the capability of recovery level 0.

– Connection recovery (**ErrorRecoveryLevel**=2)

This recovery class provides the capability to recover from TCP connection failures plus the capability that is provided by recovery level 1.

IBM Storwize storage systems support an error recovery level of 0.

7.2.3 Considerations for enabling TSO for host network adapters

The TCP segmentation offload (TSO) or large segment offload (LSO) configurable setting is available on many network adapters and is used to increase bandwidth for outbound traffic while reducing the usage of host processor cycles. The job of segmentation, as the name suggests, is offloaded to the network adapter hardware. Data and packet corruption has been observed in some cases when this option is enabled on some devices. Therefore, check with the network interface card and host operating system vendor for any known compatibility or other issues before configuring these options with IBM Storwize storage systems.

7.2.4 Host configuration maximums for iSCSI with IBM Storwize storage systems

When you select and configure your hosts for iSCSI, you must stay at or below the maximums that are supported by IBM Storwize storage systems. The limits that are presented in Table 7-1 and Table 7-2 represent tested limits that are supported on IBM Storwize storage systems. These limits are current at the time of writing. Check the latest IBM Storwize information in [IBM Knowledge Center](#) for about the latest configuration maximums for the IBM Storwize software release that is installed on your system.

Note: The maximum number of paths to a LUN per SAN Volume Controller node is four. Therefore, for two nodes, the maximum is eight, and for four nodes, the maximum is 16 if the same VDisk is accessible through multiple I/O groups.

Table 7-1 lists the iSCSI maximums for IBM Storwize software Version 7.6. and later.

Table 7-1 iSCSI maximums for IBM Storwize software Version 7.6 and later

IBM Storwize nodes	Maximum number of iSCSI host IQNs	Maximum number of paths to the same LUN	Maximum no. of paths per IQN per I/O group
2	512	2048	8
4	1024	4096	16
6	1536	6144	32
8	2048	8192	64

Table 7-2 lists the iSCSI maximums for IBM Storwize software Version 7.5 and earlier.

Table 7-2 iSCSI maximums for IBM Storwize software Version 7.5 and earlier

IBM Storwize nodes	Maximum number of iSCSI host IQNs	Maximum number of paths to the same LUN	Maximum no. of paths per IQN per I/O group
2	256	512	8
4	512	1024	16
6	768	1536	32
8	1024	2048	64

7.3 Configuring iSCSI on AIX 7.1

The iSCSI software initiator enables AIX to access storage devices by using iSCSI. The iSCSI software target enables AIX to export local storage that can be accessed by another iSCSI initiator. This publication focuses on how to configure the iSCSI software initiator to access an IBM Storwize storage volume. Before you configure iSCSI on AIX, read Chapter 4, “Planning considerations” on page 39.

In this section, it is assumed that an IBM Storwize volume is already mapped to the AIX server.

Note: At the time of writing, AIX and iSCSI are supported only with single-path configuration and a software initiator, which means that AIX does not support multipathing.

7.3.1 Ethernet network configuration

Example 7-1 shows the base network configuration for this example.

Example 7-1 Base network configuration

```
# ifconfig -a
en0:
flags=1e084863,c0<UP,BROADCAST,NOTRAILERS,RUNNING,SIMPLEX,MULTICAST,GROUPRT,64BIT,
CHECKSUM_OFFLOAD(ACTIVE),LARGESEND,CHAIN>
    inet 192.168.41.170 netmask 0xffffffff broadcast 192.168.41.255
    tcp_sendspace 131072 tcp_recvspace 65536 rfc1323 0
en1:
flags=1e084863,c0<UP,BROADCAST,NOTRAILERS,RUNNING,SIMPLEX,MULTICAST,GROUPRT,64BIT,
CHECKSUM_OFFLOAD(ACTIVE),LARGESEND,CHAIN>
    inet 9.113.57.41 netmask 0xffffffe0 broadcast 9.113.57.255
    inet6 2801::200/64
    tcp_sendspace 131072 tcp_recvspace 65536 rfc1323 0
lo0:
flags=e08084b,c0<UP,BROADCAST,LOOPBACK,RUNNING,SIMPLEX,MULTICAST,GROUPRT,64BIT,LAR
GESEND,CHAIN>
    inet 127.0.0.1 netmask 0xff000000 broadcast 127.255.255.255
    inet6 ::1%1/0
    tcp_sendspace 131072 tcp_recvspace 131072 rfc1323 1
```

The en1 Ethernet interface is used only for management traffic. In this case, en0 is used for iSCSI traffic.

It is necessary to change the default configurations, such as jumbo frames, **tcp_sendspace**, and **tcp_recvspace**. For information about these considerations for AIX, see 4.3.1, “Planning for IBM AIX” on page 46.

7.3.2 Selecting the discovery policy

To discover the iSCSI targets, it is necessary to choose the discovery policy. This section describes two methods: file and ODM.

File discovery policy

To configure the discovery policy with SMIT, complete the following steps:

1. Run `smit iscsi` at the AIX command line. The iSCSI menu opens. Select iSCSI Protocol Device from the iSCSI menu, as shown in Figure 7-1.

```
iSCSI

Move cursor to desired item and press Enter.

iSCSI Adapter
iSCSI Protocol Device
iSCSI Target Device Parameters in ODM
iSNS Discovery Configuration
```

Figure 7-1 iSCSI menu from SMIT

2. Select Change / Show Characteristics of an iSCSI Protocol Device, as shown in Figure 7-2.

```
iSCSI Protocol Device

Move cursor to desired item and press Enter.

List All iSCSI Protocol Devices
Change / Show Characteristics of an iSCSI Protocol Device
Generate Error Report
Trace iSCSI Protocol Device
Remove iSCSI Protocol Device
```

Figure 7-2 iSCSI Protocol Device menu in SMIT

3. Select `iscsi0`, as shown in Figure 7-3.

```
iSCSI Protocol Device

Move cursor to desired item and press Enter.

List All iSCSI Protocol Devices
Change / Show Characteristics of an iSCSI Protocol Device
Generate Error Report
Trace iSCSI Protocol Device
Remove iSCSI Protocol Device

+-----+
|                                     iSCSI Protocol Device                                     |
|                                     |
| Move cursor to desired item and press Enter. |
|                                     |
| iSCSI0 Available iSCSI Protocol Device |
|                                     |
| F1=Help           F2=Refresh           F3=Cancel |
| F8=Image          F10=Exit             Enter=Do  |
| F1 /-=Find        n=Find Next          |
| F9+-----+ |
|                                     |
```

Figure 7-3 Selecting `iscsi0` from the iSCSI Protocol Device menu

4. Select file in the Discovery Policy field, as shown in Figure 7-4.

```

Change / Show Characteristics of an iSCSI Protocol Device

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

iSCSI Protocol Device
Description
Status
iSCSI Initiator Name
Maximum number of commands to queue to driver
Discovery Policy
Maximum Targets Allowed
Apply change to DATABASE only

[Entry Fields]
iscsi0
iSCSI Protocol Device
Available
[iqn.localhost.hostid.0>
[200]
file
[16]
no

```

Figure 7-4 Change / Show Characteristics of an iSCSI Protocol Device menu

5. Press Enter and the panel in Figure 7-5 opens.

```

COMMAND STATUS

Command: OK          stdout: yes          stderr: no

Before command completion, additional instructions may appear below.

iscsi0 changed

```

Figure 7-5 iSCSI Protocol Device changed

6. Edit the /etc/iscsi/targets file to specify only one of the available iSCSI targets on the IBM Storwize storage system, as shown in the following example:

```
192.168.41.150 3260 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node1
```

ODM discovery policy

To configure ODM, complete the following steps:

1. Repeat steps 1 on page 93 - 3 on page 93 in “File discovery policy” on page 93. Instead of selecting file, select odm, as shown in Figure 7-6.

```

Change / Show Characteristics of an iSCSI Protocol Device

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

iSCSI Protocol Device
Description
Status
iSCSI Initiator Name
Maximum number of commands to queue to driver
Discovery Policy
Maximum Targets Allowed
Apply change to DATABASE only

[Entry Fields]
iscsi0
iSCSI Protocol Device
Available
[iqn.localhost.hostid.0>
[200]
odm
[16]
no

```

Figure 7-6 Change / Show characteristics of an iSCSI Protocol Device menu

2. Press Enter to accept the change, as shown in Figure 7-7.

```
COMMAND STATUS
Command: OK          stdout: yes          stderr: no
Before command completion, additional instructions may appear below.
iscsi0 changed
```

Figure 7-7 *iscsi0 changed*

3. Select iSCSI Target Device Parameters in ODM, as shown in Figure 7-8.

```
iSCSI
Move cursor to desired item and press Enter.
iSCSI Adapter
iSCSI Protocol Device
iSCSI Target Device Parameters in ODM
iSNS Discovery Configuration
```

Figure 7-8 *iSCSI main menu*

4. Select Add an iSCSI target Device in ODM, as shown in Figure 7-9.

```
iSCSI Target Device Parameters in ODM
Move cursor to desired item and press Enter.
List All iSCSI Target Devices in ODM
Add an iSCSI Target Device in ODM
Delete an iSCSI Target Device from ODM
Change existing iSCSI Target Device in ODM
```

Figure 7-9 *iSCSI target device parameters in the ODM menu*

5. Select Add a Statically Discovered iSCSI Target Device in ODM, as shown in Figure 7-10.

```
Add an iSCSI Target Device in ODM
Move cursor to desired item and press Enter.
Add a Statically Discovered iSCSI Target Device in ODM
Add iSCSI Target Device(s) into ODM from a File
Add Authentication Data for an Automatically Discovered iSCSI Target Device
```

Figure 7-10 *Adding an iSCSI Target Device in the ODM menu*

- Enter only one of the available iSCSI target names, IP addresses, and port numbers of the IBM Storwize storage system, as shown in Figure 7-11. All this information is in the IBM Storwize network ports' configuration.

Change / Show Characteristics of an iSCSI Adapter

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

	[Entry Fields]
iSCSI Adapter	iscsi0
iSCSI Target Name	[iqn.1986-03.com.ibm:2145>
IP Address of iSCSI Target	[192.168.41.150]
Port Number of iSCSI Target	[3260]
Password	[]

Figure 7-11 Change / Show Characteristics of an iSCSI adapter

7.3.3 Working with the IBM Storwize storage volume

When the iSCSI target device is added, you are ready to use the volume. Example 7-2 shows a basic AIX configuration to demonstrate that the use of the iSCSI volume is not different from other volumes.

Example 7-2 Basic configuration of an iSCSI volume in AIX

```
# cfgmgr -l iscsi0
# lspv
hdisk0          00f6996d1b845dc2          rootvg          active
hdisk1          00f6996da88ae81e          None
# lsattr -El hdisk1
clr_q           no                      Device CLEARS its Queue on error True
host_addr       192.168.41.150           Hostname or IP Address False
location        Location Label True
lun_id          0x10000000000000          Logical Unit Number ID False
max_transfer     0x40000                 Maximum TRANSFER Size True
port_num        0xcbc                   PORT Number False
pvid            00f6996da88ae81e00000000000000000 Physical volume identifier False
q_err           yes                     Use QERR bit True
q_type          simple                   Queuing TYPE True
queue_depth     8                       Queue DEPTH True
reassign_to     120                     REASSIGN time out value True
reserve_policy  no_reserve               Reserve Policy True
rw_timeout      60                      READ/WRITE time out value True
start_timeout   60                      START unit time out value True
target_name     iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node1 Target NAME False
unique_id       352136005076400AF0001A80000000000006504214503IBMisci Unique device identifier False
# mkvg -y testvg hdisk1
testvg
# lspv
hdisk0          00f6996d1b845dc2          rootvg          active
hdisk1          00f6996da88ae81e          testvg          active
# mklv -t jfs2 -y testlv testvg 3
testlv
# crfs -v jfs2 -d testlv -A yes -u fs -m /test
File system created successfully.
98096 kilobytes total disk space.
New File System size is 196608
# mount /test
# df -g /test
Filesystem      GB blocks      Free %Used      Iused %Iused Mounted on
/dev/testlv      0.09           0.09      1%           4         1% /test
```

7.4 Configuring iSCSI for SUSE Linux Enterprise Server

This section describes how to configure the IBM Storwize storage system with Linux iSCSI for SUSE Linux Enterprise Server 11 SP2.

7.4.1 Prerequisites for mapping the iSCSI volume

To prepare to configure the Linux host, complete the following steps:

1. Be sure that your host server is using the latest firmware levels.
2. To understand the preferred topology and be aware of the general and specific considerations for Linux, see Chapter 4, “Planning considerations” on page 39.
3. If necessary, install the supported driver for the NIC adapter and after configuring the IP address, verify that you can ping each port of the IBM Storwize storage system.
4. Configure the host, volumes, and host mapping, as described in 7.2, “Configuring initiators for iSCSI” on page 87.
5. YaST provides an appropriate method to configure iSCSI services and the initiator. To configure the iSCSI service in SSIC, complete the following steps:

```
# rcopen-iscsi start
# rcmultipathd start
# chkconfig open-iscsi on
# chkconfig multipathd on
```

6. Set the iSCSI qualified name (IQN) for the Linux host.

Edit the `/etc/iscsi/initiatorname.iscsi` file to specify the IQN. In this case, the name is the following one:

```
InitiatorName=iqn.2015-06.site:01:4c599d67b183
```

This information is necessary to configure the host object in the IBM Storwize storage system. If the initiator name is changed, the iSCSI service must be restarted by running the following command:

```
# rcopen-iscsi restart
```

Considerations for multipathing

SUSE Linux Server provides its own multipath support, which means that it is not necessary to install a specific device driver. The Device Mapper Multipath module provides the multipathing capability for Linux and is installed by default. Up to eight paths to each device are supported. The multipath-tools package handles the automatic path discovery and grouping. It automatically and periodically tests the path.

The file to configure each specific iSCSI storage device is `/etc/multipath.conf`. This file does not exist by default, so it must be created or copied from the following sample file:

```
/usr/share/doc/packages/multipath-tools/multipath.conf.synthetic
```

Each change in the `multipath.conf` file is not automatically applied. To apply them, complete the following steps:

1. To stop the multipathd service, run the `rcopen-iscsi restart` command.
2. To clear old multipath bindings, run the `multipath -F to` command.
3. To new multipath bindings, run the `multipath -v2 -l` command.

For more information about the content that should be in this file, see 4.3.2, “Planning for Linux” on page 50.

7.4.2 Ethernet network configuration

Example 7-3 shows the base network configuration that is available for this example. In this case, eth0 is used for management traffic and eth1 is used for iSCSI traffic. Later in this section, an example of a configuration is shown for a third Ethernet NIC (not shown in this example yet) that is configured for iSCSI redundancy.

Example 7-3 Basic network configuration for the lab

```
# ifconfig -a
eth0      Link encap:Ethernet  HWaddr 00:0C:29:92:C5:E6
          inet addr:9.113.57.250  Bcast:9.113.57.255  Mask:255.255.254.0
          inet6 addr: fe80::20c:29ff:fe92:c5e6/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:11004731 errors:0 dropped:1025 overruns:0 frame:0
          TX packets:33191 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:1143978545 (1090.9 Mb)  TX bytes:9811803 (9.3 Mb)

eth1      Link encap:Ethernet  HWaddr 00:0C:29:92:C5:F0
          inet addr:192.168.41.121  Bcast:192.168.41.255  Mask:255.255.255.0
          inet6 addr: fe80::20c:29ff:fe92:c5f0/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:9000  Metric:1
          RX packets:2643600 errors:0 dropped:895 overruns:0 frame:0
          TX packets:53446 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:538086563 (513.1 Mb)  TX bytes:4812532 (4.5 Mb)

lo        Link encap:Local Loopback
          inet addr:127.0.0.1  Mask:255.0.0.0
          inet6 addr: ::1/128 Scope:Host
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
          RX packets:3865 errors:0 dropped:0 overruns:0 frame:0
          TX packets:3865 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:274048 (267.6 Kb)  TX bytes:274048 (267.6 Kb)
```

7.4.3 Discovering and logging in to the iSCSI targets

Now, the IP address of only one of the two Ethernet NICs is configured by using the following command:

```
ifconfig eth1 192.168.41.121/24
```

After you map the IBM Storwize storage volume to the host, the following procedure must be done. This section describes each step for discovering the iSCSI targets by using YaST and the CLI.

Note: At the moment of discovering the iSCSI target devices, when you specify one target IP address, the iSCSI initiator discovers all the targets in the canister that correspond to that IP address. This discovery includes even the targets that are not reachable.

Using YaST

To use YaST, complete the following steps:

1. Run the `yast iscsi-client` command. Figure 7-12 shows the initial menu.

```
YaST2 - iscsi-client @ sles11sp2-vm1

iSCSI Initiator Overview
Service—Connected Targets—Discovered Targets

[Service Start]
[ (x) When Booting ]
[ ( ) Manually ]

Initiator Name      Offload Card
iqn.2015-06.site:01: default (Softwareâ†‘)

iSNS Address        iSNS Port
XXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXX

[Help] [Cancel] [ OK ]

F1 Help F9 Cancel F10 OK
```

Figure 7-12 iSCSI initial menu

2. Select Discovered targets. Figure 7-13 shows the menu that opens.

```
YaST2 - iscsi-client @ sles11sp2-vm1

iSCSI Initiator Overview
Service—Connected Targets—Discovered Targets

[Interface | Portal Address | Target Name | Connected]

[Discovery] [Log In] [Delete]

[Help] [Cancel] [ OK ]

F1 Help F5 Delete F9 Cancel F10 OK
```

Figure 7-13 Discovered targets menu

3. Select Discovery and the next panel opens, as shown in Figure 7-14. In this case, the IP addresses for the IBM Storwize V7000 canisters are 192.168.41.150 and 192.168.41.151. If CHAP is configured (it is not configured in this example), the data is specified in this menu.

```

YaST2 - iscsi-client @ sles11sp2-vm1

iSCSI Initiator Discovery

IP Address      Port
192.168.41.150  v 3260^

[ ] No Authentication
Incoming Authentication
Username          Password
[ ] Outgoing Authentication
Username          Password

[Help]          [Back]          [Cancel]          [Next]

F1 Help  F8 Back  F9 Cancel  F10 Next

```

Figure 7-14 Target Information

Figure 7-15 shows the final view after both iSCSI IP addresses of the storage are configured.

```

YaST2 - iscsi-client @ sles11sp2-vm1

iSCSI Initiator Overview
Service—Connected Targets—Discovered Targets—

Interface  Portal Address  Target Name
default    192.168.41.150:3260  iqn.1986-03.com.ibm:2145.cluster9.113
default    192.168.41.151:3260  iqn.1986-03.com.ibm:2145.cluster9.113

[Discovery] [Log In] [Delete]

[Help]          [Cancel]          [ OK ]

F1 Help  F5 Delete  F9 Cancel  F10 OK

```

Figure 7-15 Target information

4. Log in to each of the targets.

Note: The IBM Storwize storage system supports only one iSCSI session between an initiator and a target. Ensure that you do not attempt to connect to the same target more than once.

- a. Select **Log in**. Three options are shown in the Startup box menu:
 - onboot
With this option, iSCSI targets are connected during boot, that is, when root is on iSCSI. As such, it is evaluated from the initrd.
 - manual
With this option, iSCSI targets are not connected by default. The user must do it manually.
 - automatic
With this option, iSCSI targets are connected when the iSCSI service itself starts.
- b. Select Next, as shown in Figure 7-16, and repeat the same procedure for each target.

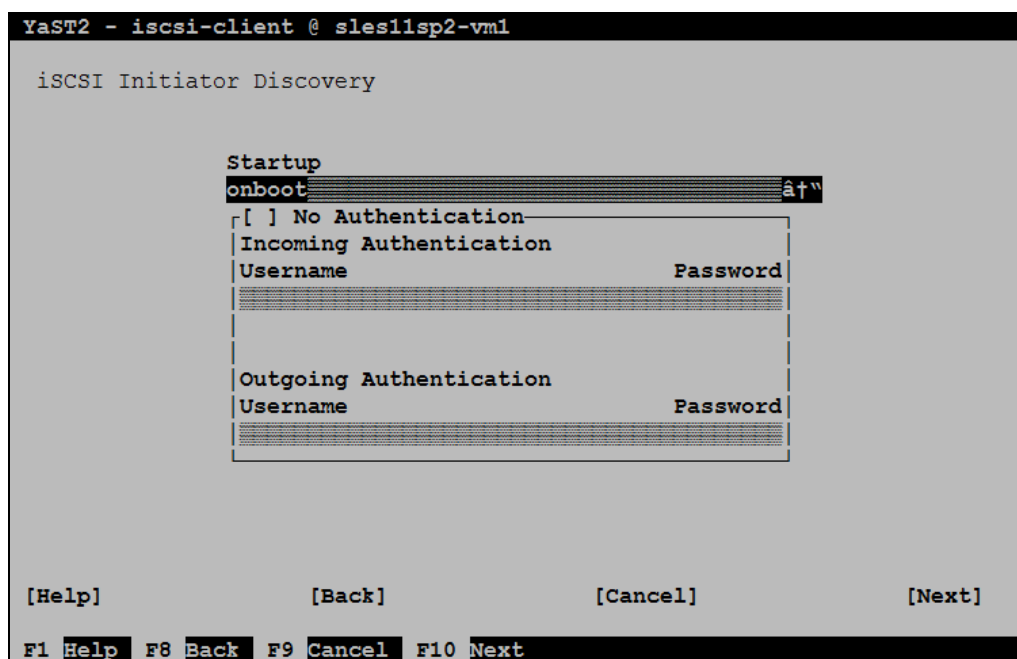


Figure 7-16 Login targets

Using the CLI

To discover the targets by using the CLI, complete the following steps:

1. Run the following command:

```
# iscsiadm -m discovery -t sendtargets -p x.x.x.x
```

Where *x.x.x.x* is the IP address of a node Ethernet port on the SAN Volume Controller clustered system, as shown in Example 7-4.

Example 7-4 Discovering the targets by using the CLI

```
# iscsiadm -m discovery -t sendtargets -p 192.168.41.150:3260
192.168.41.150:3260,1 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node1
# iscsiadm -m discovery -t sendtargets -p 192.168.41.151:3260
192.168.41.151:3260,1 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node2
# iscsiadm -m discoverydb
192.168.41.151:3260 via sendtargets
192.168.41.150:3260 via sendtarget
```

2. Log in to each of the targets.

Note: The IBM Storwize storage system supports only one iSCSI session between an initiator and a target. Ensure that you do not attempt to connect to the same target more than once.

To log in to every target that is already discovered, run the command that is shown in Example 7-5.

Example 7-5 Logging in to the targets by using the CLI

```
# iscsiadm -m node --targetname
iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node1 --portal 192.168.41.150
--login

# iscsiadm -m node --targetname
iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node2 --portal 192.168.41.151
--login
```

Using an iSNS server

The targets can also be discovered by using an iSNS server. For more information, see 7.7.2, “Configuring the iSNS server address on an IBM Storwize storage system” on page 144.

In this section, it is assumed that an iSNS server is already configured. In this case, you need only the IP address and the port of that server, which you can discover by using YaST and the CLI. This example shows how to do the discovery by using iSNS server and the CLI:

1. Add the following lines to the `/etc/iscsi/iscsid.conf` file:

```
isns.address = 192.168.41.227
isns.port = 3205
```

In this case, the IP address of the iSNS server is 192.168.41.227:3205.

2. Discover the targets. Example 7-6 shows how to discover the targets by using iSNS with the CLI.

Example 7-6 Discovering the targets by using an iSNS server

```
# iscsiadm --mode discovery --type isns --portal 192.168.41.227:3205
192.168.41.150:3260,1 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node1
192.168.41.151:3260,1 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node2
```

7.4.4 Understanding iSCSI sessions for software-based initiators

Example 7-7 shows the number of sessions that are already logged in.

Example 7-7 Current sessions

```
# iscsiadm -m session
tcp: [1] 192.168.41.150:3260,2 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node1
tcp: [2] 192.168.41.151:3260,2 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node2
```

In this case, there are two sessions, as shown in Figure 7-17.

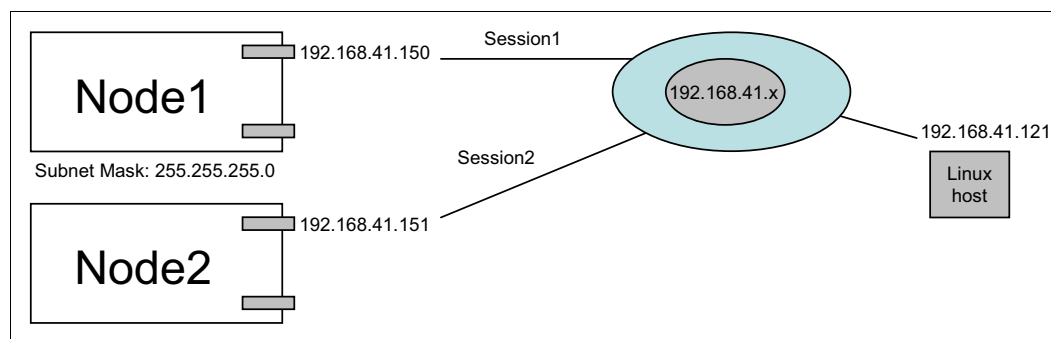


Figure 7-17 iSCSI configuration with two sessions and one NIC

Because software-based initiators are used in this example, by default only one IQN is created. If hardware-based initiators were used, there would be one IQN per physical Ethernet interface.

In Figure 7-17, the relevant thing is that there is redundancy on the IBM Storwize side but not on the host side. This configuration means that if the current NIC goes down, the host loses access to the iSCSI volumes.

To obtain multipathing on the host side, both NICs must be configured.

Note: When using iSCSI software-based initiators, you cannot have more than one IQN per host. Although it is possible to create an additional IQN in Linux by creating another interface by running `iscsiadm`, it is not possible to bind the IP address to the specific IQN.

Adding the second Ethernet NIC

For this example, two other Ethernet ports were configured for iSCSI on the IBM Storwize storage system with IP addresses 192.168.42.150 and 192.168.42.151.

Example 7-8 shows how to add the second Ethernet NIC, discover the new targets, list them, and log in to each one.

Example 7-8 Adding the second Ethernet NIC

```
# ifconfig eth2 192.168.42.121/24
# iscsiadm -m discovery -t sendtargets -p 192.168.42.150:3260
192.168.42.150:3260,1 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node1
# iscsiadm -m discovery -t sendtargets -p 192.168.42.151:3260
192.168.42.151:3260,1 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node2
# iscsiadm -m discoverydb
192.168.41.151:3260 via sendtargets
192.168.42.151:3260 via sendtargets
192.168.42.150:3260 via sendtargets
192.168.41.150:3260 via sendtargets
# iscsiadm -m node --targetname iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node1
--portal 192.168.42.150 --login

# iscsiadm -m node --targetname iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node2
--portal 192.168.42.151 --login
```

Listing the sessions with the new Ethernet NIC

Example 7-9 shows the number of sessions after you create the interface. With this new configuration, there is redundancy if one of the two NICs goes down.

Example 7-9 Listing the number of sessions

```
# iscsiadm -m session
tcp: [1] 192.168.41.150:3260,2 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node1
tcp: [2] 192.168.41.151:3260,2 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node2
tcp: [3] 192.168.42.150:3260,3 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node1
tcp: [4] 192.168.42.151:3260,3 iqn.1986-03.com.ibm:2145.cluster9.113.57.226.node2
```

Figure 7-18 on page 105 shows the current topology now that the following tasks are complete:

1. Start the open-iscsi and multipathd services.
2. Configure the IPs for both Ethernet interfaces.
3. Add the host in the IBM Storwize V7000 storage system with the specific IQN.
4. Discover targets.
5. Log in to the targets.

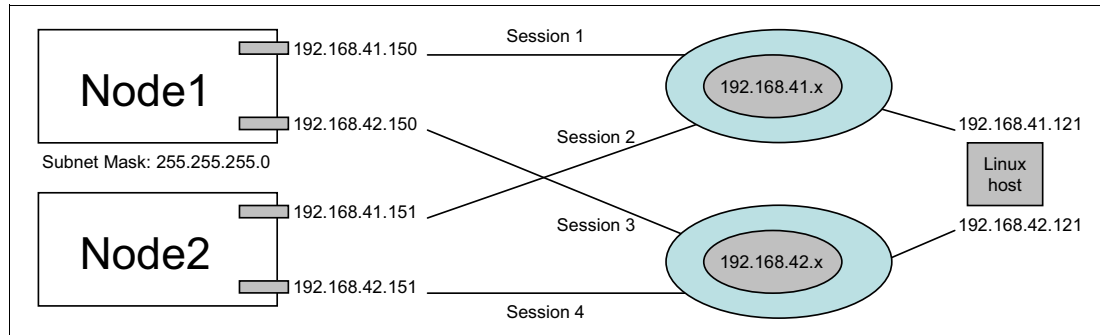


Figure 7-18 Current topology with four sessions logged

Using network bonding

The second method for having network redundancy, even if one of the NICs goes down, is to use network bonding. Network bonding is an operating system feature that allows several Ethernet devices to be aggregated to a single bonding device. The default configuration of the bonding device supports the topology that is described in Figure 7-19, in which one NIC is active waiting and the other is waiting for the primary to fail. This mode is known as *active-backup*.

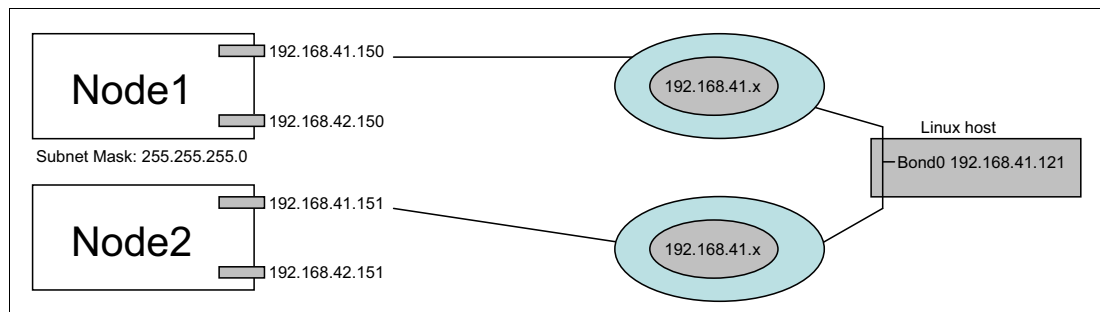


Figure 7-19 Network bonding example topology

For more information about network bonding, see [Network Device Bonding](#).

7.4.5 Working with the IBM Storwize storage volume

To discover the volume, run the following command:

```
rescan-scsi-bus.sh
```

Example 7-10 shows the output from the **fdisk -l** command. For each path, there is one `/dev/sdbX` disk. In this case, `/dev/sdb`, `/dev/sdc`, `/dev/sdd`, and `/dev/sde`. Additionally, there is a `/dev/mapper/mpathvi` device, which is the device that is created automatically by the DM-Multipath driver. It can provide failover in an active/passive configuration.

*Example 7-10 The **fdisk -l** output example after you configure four sessions for a 10.7 GB volume*

```
Disk /dev/sdb: 10.7 GB, 10737418240 bytes
64 heads, 32 sectors/track, 10240 cylinders, total 20971520 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 32768 bytes / 67108864 bytes
Disk identifier: 0x00000000
```

Disk /dev/sdb doesn't contain a valid partition table

Disk /dev/mapper/mpathvi: 10.7 GB, 10737418240 bytes
64 heads, 32 sectors/track, 10240 cylinders, total 20971520 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 32768 bytes / 67108864 bytes
Disk identifier: 0x00000000

Disk /dev/mapper/mpathvi doesn't contain a valid partition table

Disk /dev/sdc: 10.7 GB, 10737418240 bytes
64 heads, 32 sectors/track, 10240 cylinders, total 20971520 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 32768 bytes / 67108864 bytes
Disk identifier: 0x00000000

Disk /dev/sdc doesn't contain a valid partition table

Disk /dev/sdd: 10.7 GB, 10737418240 bytes
64 heads, 32 sectors/track, 10240 cylinders, total 20971520 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 32768 bytes / 67108864 bytes
Disk identifier: 0x00000000

Disk /dev/sdd doesn't contain a valid partition table

Disk /dev/sde: 10.7 GB, 10737418240 bytes
64 heads, 32 sectors/track, 10240 cylinders, total 20971520 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 32768 bytes / 67108864 bytes
Disk identifier: 0x00000000

Disk /dev/sde doesn't contain a valid partition table

From this point, it is possible to use the discovered volume with Logical Volume Manager (LVM) or a traditional extended partition by running the **kpartx** command. This book focuses only on LVM.

Using Logical Volume Manager on the multipath (DM MPIO) device

To start using the volume with LVM, the first thing to do is to initialize a block device to be used as a physical volume, which is analogous to formatting a file system. Example 7-11 shows a basic configuration example of LVM.

Example 7-11 Basic configuration of Logical Volume Manager

```
# pvcreate /dev/mapper/mpathvi  
Physical volume "/dev/mapper/mpathvi" successfully created  
  
# vgcreate vgtest /dev/mapper/mpathvi  
Volume group "vgtest" successfully created
```



```
# lvcreate -l 20 -n test_vol1 vgtest
Logical volume "test_vol1" created

# mkfs.ext4 /dev/vgtest/test_vol1
mke2fs 1.41.9 (22-Aug-2009)
Filesystem label=
OS type: Linux
Block size=1024 (log=0)
Fragment size=1024 (log=0)
20480 inodes, 81920 blocks
4096 blocks (5.00%) reserved for the super user
First data block=1
Maximum filesystem blocks=67371008
10 block groups
8192 blocks per group, 8192 fragments per group
2048 inodes per group
Superblock backups stored on blocks:
    8193, 24577, 40961, 57345, 73729
Writing inode tables: done
Creating journal (4096 blocks): done
Writing superblocks and filesystem accounting information: done
This filesystem will be automatically checked every 31 mounts or
180 days, whichever comes first. Use tune2fs -c or -i to override.

# mount /dev/vgtest/test_vol1 /mnt

# df -h /mnt
Filesystem                                Size  Used Avail Use% Mounted on
/dev/mapper/vgtest-test_vol1             78M   5.6M   68M   8% /mnt
```

Verifying multipathing

The **-ll** option for the **multipath** command is used to display all the necessary information about the current multipath configuration.

Example 7-12 shows the output of the **multipath -ll** command. In this case, it shows the mpathvi device and its specific UID, which is the same view that is seen from the IBM Storwize storage perspective. Additionally it shows the configured multipath policies (round-robin and queue_if_no_path) and two paths for each IBM Storwize canister. The one that is active is from the active canister.

Example 7-12 Output of the multipath -ll command

```
# multipath -ll
mpathvi (36005076400af0001a800000000000066) dm-0 IBM,2145
size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
|+- policy='round-robin 0' prio=50 status=active
| | 6:0:0:0 sde 8:64 active ready running
| | 9:0:0:0 sdh 8:112 active ready running
|+- policy='round-robin 0' prio=10 status=enabled
| | 4:0:0:0 sdc 8:32 active ready running
| | 7:0:0:0 sdf 8:80 active ready running
```

Based on Example 7-12, sde and sdh are paths that correspond to the active canister and sdc and sdf correspond to the passive canister.

Forcing a multipath failover

Example 7-13 shows an example of what the **multipath -ll** command shows after one of the Ethernet NIC goes down.

Example 7-13 Failover example

```
# ifconfig eth1 down
# multipath -ll
mpathv (36005076400af0001a800000000000066) dm-0 IBM,2145
size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
|+- policy='round-robin 0' prio=0 status=enabled
| | 6:0:0:0 sde 8:64 failed faulty running
| | 4:0:0:0 sdc 8:32 failed faulty running
|+- policy='round-robin 0' prio=50 status=active
| | 9:0:0:0 sdh 8:112 active ready running
|+- policy='round-robin 0' prio=10 status=enabled
| | 7:0:0:0 sdf 8:80 active ready running
```

After failover, the only active path is sdh because it corresponds to the active IBM Storwize canister.

7.5 Configuring iSCSI for Windows 2012

This section describes how to configure an IBM Storwize storage system with Microsoft iSCSI initiator software.

7.5.1 Prerequisites

To configure the Windows host, complete the following tasks:

1. Ensure that your host server is using the latest firmware levels.
2. To understand the preferred topology and be aware of the general and specific considerations for Windows hosts, see Chapter 4, “Planning considerations” on page 39.
3. If necessary, install the supported driver for the NIC adapter. After you configure the IP address, verify that you can ping each port of the IBM Storwize storage system.
4. Configure the host, volumes, and host mapping, as described in 7.2, “Configuring initiators for iSCSI” on page 87.
5. Microsoft iSCSI Initiator is installed natively on Windows Server 2012 and no installation is required. However, make sure that the Microsoft iSCSI Initiator Service is running and the startup type for the service is set to automatic in the Windows services applet.
6. Set the IQN for the Microsoft Windows host. To do this task, edit the initiator name in the configuration tab for the iSCSI initiator, as shown here. In this example, the default initiator name is used.

```
iqn.1991-05.com.microsoft:win-itel2q9on01.iscsi.appcluster.com
```

This information is necessary to configure the host object in the IBM Storwize storage system, as shown in Figure 7-20.

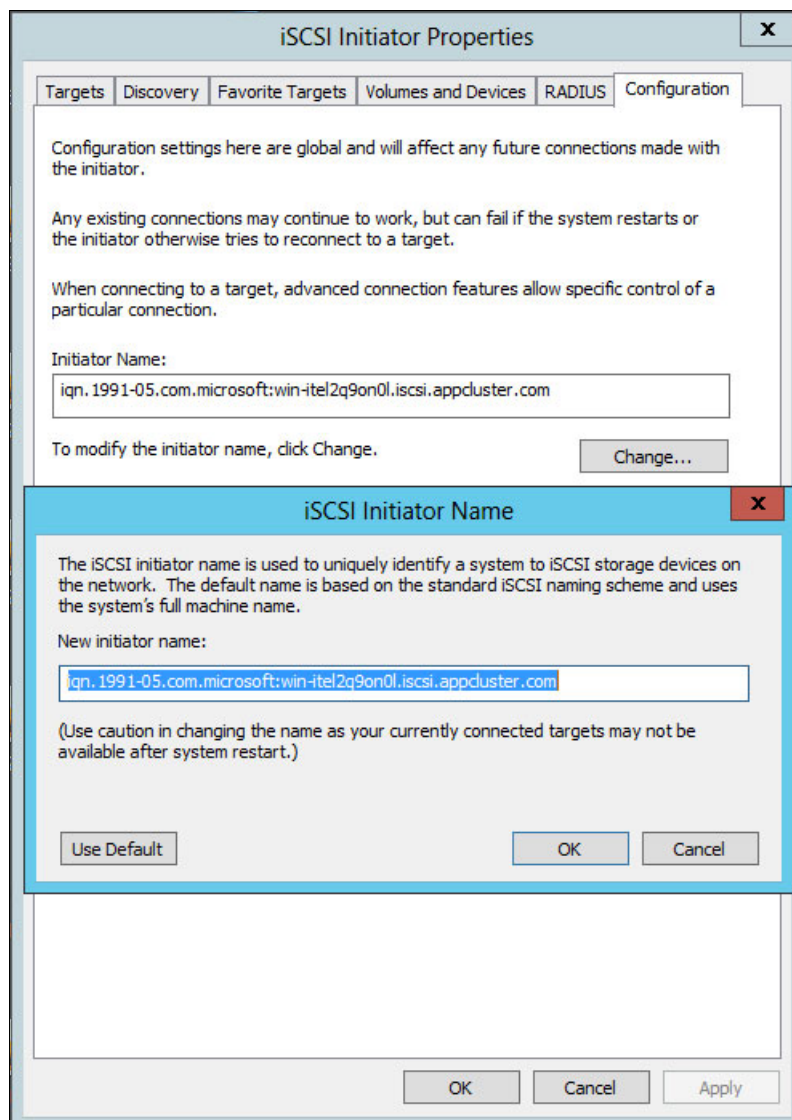


Figure 7-20 iSCSI initiator properties

Considerations for multipathing

Microsoft provides its MPIO multipathing solution for multipath management in the Microsoft Windows operating system. Therefore, it is not essential to install a third-party multipathing driver. IBM Storwize storage systems support the generic Microsoft device-specific module (DSM) that is available in Windows 2012 for iSCSI. MPIO is an optional feature in Windows Server 2012, and is not installed by default.

To install MPIO on your server, complete the following steps:

1. Open **Server Manager** and in the server manager tree, click **Features**.
2. In the Features area, click **Add Features**.
3. In the Add Features wizard, in the Select Features window, select the **Multipath I/O** check box, and then click **Next**.

4. In the Confirm Installation Selections window, click **Install**.
5. After the installation is completed, the system must be restarted.

When MPIO is installed, the Microsoft DSM is also installed, as is the MPIO control panel. The control panel can be used to do the following tasks:

- ▶ Configure MPIO functions.
- ▶ Create MPIO configuration reports.

Note: At the time of writing, the IBM Subsystem Device Driver Device Specific Module (SDD DSM) is not supported for iSCSI host attachment with IBM Storwize storage systems on the Microsoft Windows platform.

7.5.2 Ethernet network configuration on Windows hosts

Figure 7-21 on page 111 shows basic network configuration for iSCSI host attachment on Windows hosts. In this configuration, the Ethernet card Ethernet 3 is used for management traffic, and the other two cards are used for iSCSI traffic. Two cards are used to configure a typical multipathing configuration where they form a session with IBM Storwize target ports. The number of Ethernet ports can be increased as needed. However, IBM Storwize storage systems support a maximum of up to eight sessions per initiator port per I/O group.

```

C:\Users\administrator.ISCSI>ipconfig

Windows IP Configuration


Ethernet adapter Ethernet 3:

    Connection-specific DNS Suffix  . : 
    IPv6 Address. . . . . : 2002:976:2c08:202:ecce:93cf:1542:2775
    IPv6 Address. . . . . : 2002:976:2c08:205:ecce:93cf:1542:2775
    IPv6 Address. . . . . : 2002:976:2c08:207:ecce:93cf:1542:2775
    Link-local IPv6 Address . . . . . : fe80::ecce:93cf:1542:2775%18
    IPv4 Address. . . . . : 9.113.57.227
    Subnet Mask . . . . . : 255.255.254.0
    Default Gateway . . . . . : fe80::224:50ff:fe59:3800%18
                                9.113.56.1


Ethernet adapter Ethernet 2:

    Connection-specific DNS Suffix  . : 
    IPv6 Address. . . . . : 2002:976:2c08:202:49aa:4a4b:7eba:1eb5
    IPv6 Address. . . . . : 2002:976:2c08:205:49aa:4a4b:7eba:1eb5
    IPv6 Address. . . . . : 2002:976:2c08:207:49aa:4a4b:7eba:1eb5
    IPv6 Address. . . . . : 2002:976:2c08:202:f7c7:8bf9:4515:cb3e
    IPv6 Address. . . . . : 2002:976:2c08:202:fa6d:5c82:65f5:81c9
    IPv6 Address. . . . . : 2002:976:2c08:205:9a58:bb56:556b:e6f3
    IPv6 Address. . . . . : 2002:976:2c08:207:84d:c840:8698:a9de
    Link-local IPv6 Address . . . . . : fe80::49aa:4a4b:7eba:1eb5%13
    IPv4 Address. . . . . : 192.168.41.115
    Subnet Mask . . . . . : 255.255.255.0
    IPv4 Address. . . . . : 192.168.41.116
    Subnet Mask . . . . . : 255.255.255.0
    IPv4 Address. . . . . : 192.168.41.200
    Subnet Mask . . . . . : 255.255.255.0
    IPv4 Address. . . . . : 192.168.41.201
    Subnet Mask . . . . . : 255.255.255.0
    Default Gateway . . . . . : fe80::224:50ff:fe59:3800%13


Ethernet adapter Ethernet:

    Connection-specific DNS Suffix  . : 
    IPv6 Address. . . . . : 2002:976:2c08:202:a88f:cfdd:9854:cc03
    IPv6 Address. . . . . : 2002:976:2c08:205:a88f:cfdd:9854:cc03
    IPv6 Address. . . . . : 2002:976:2c08:207:a88f:cfdd:9854:cc03
    Link-local IPv6 Address . . . . . : fe80::a88f:cfdd:9854:cc03%12
    IPv4 Address. . . . . : 192.168.41.114
    Subnet Mask . . . . . : 255.255.255.0
    Default Gateway . . . . . : fe80::224:50ff:fe59:3800%12
                                192.168.41.1

```

Figure 7-21 Basic lab network configuration that is shown in Microsoft Windows

7.5.3 iSCSI target discovery for Windows hosts

After you configure IPs on the Windows hosts and make sure that the IBM Storwize IPs are reachable from the host by pinging them, you can do iSCSI target discovery.

There are two ways of discovering iSCSI targets from Windows hosts:

- SendTargets discovery
- iSNS discovery

SendTargets discovery

For this type of discovery, you need the IP addresses of one or more of your system node's Ethernet ports.

To perform discovery with the Microsoft iSCSI Initiator, complete the following steps:

1. Open the iSCSI initiator software. Figure 7-22 shows the initial window after the software opens.

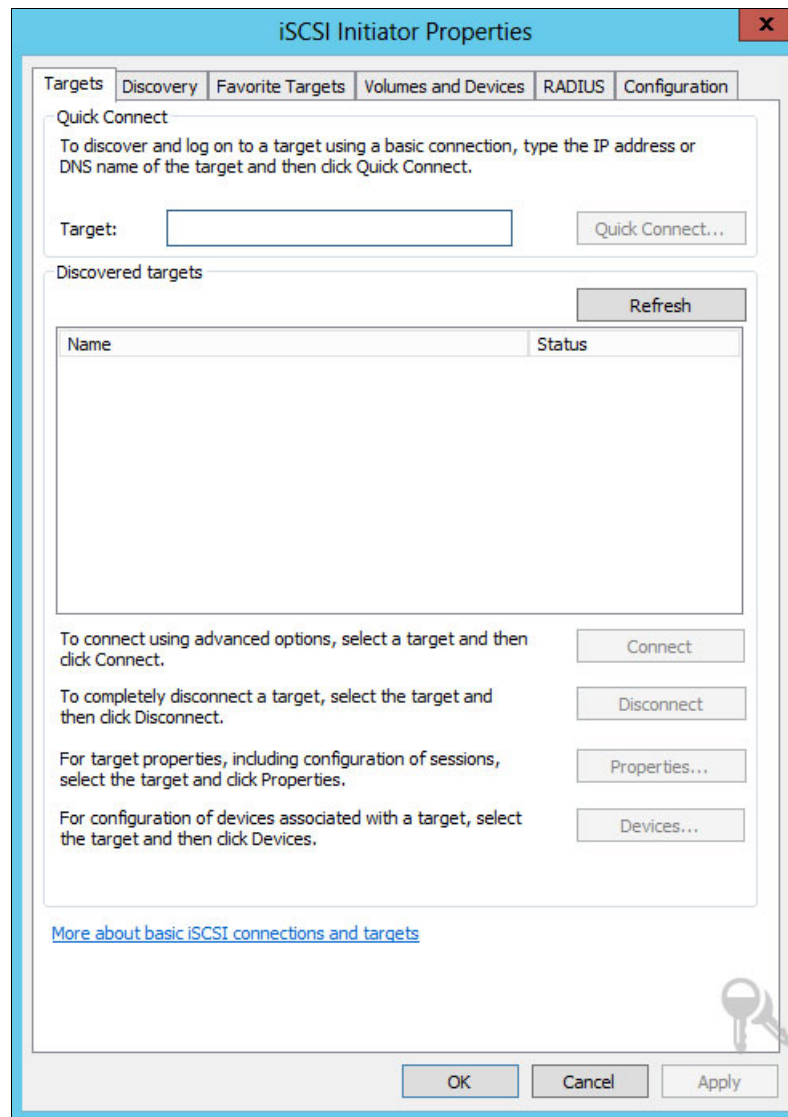


Figure 7-22 iSCSI Initiator Properties

2. Click the **Discovery** tab and then **Discover Portal**.

3. Enter the IP address of the target's Ethernet port, as shown in Figure 7-23, and click **OK**.

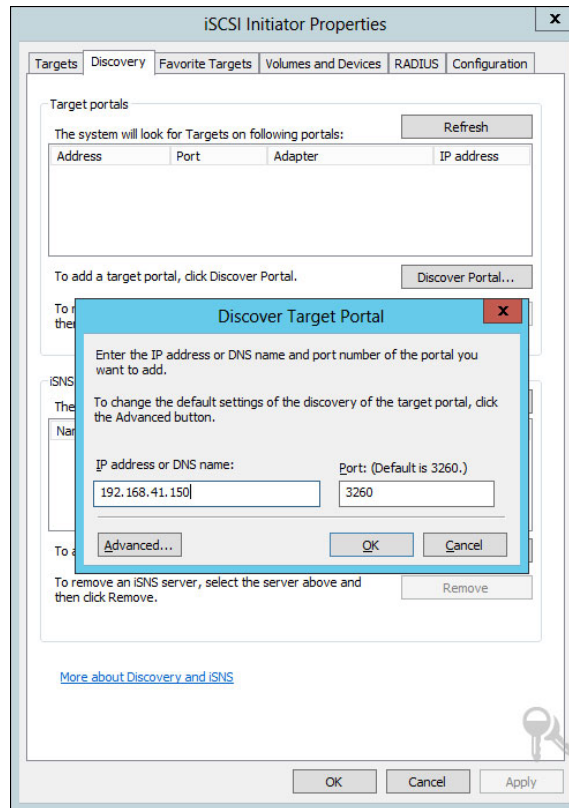


Figure 7-23 iSCSI discovery portal

4. Repeat the same process for all the target IPs with which you want to make sessions.

5. After you discover all the target IPs, you see the targets that are listed under the Discover targets in the Targets tab, as shown in Figure 7-24.

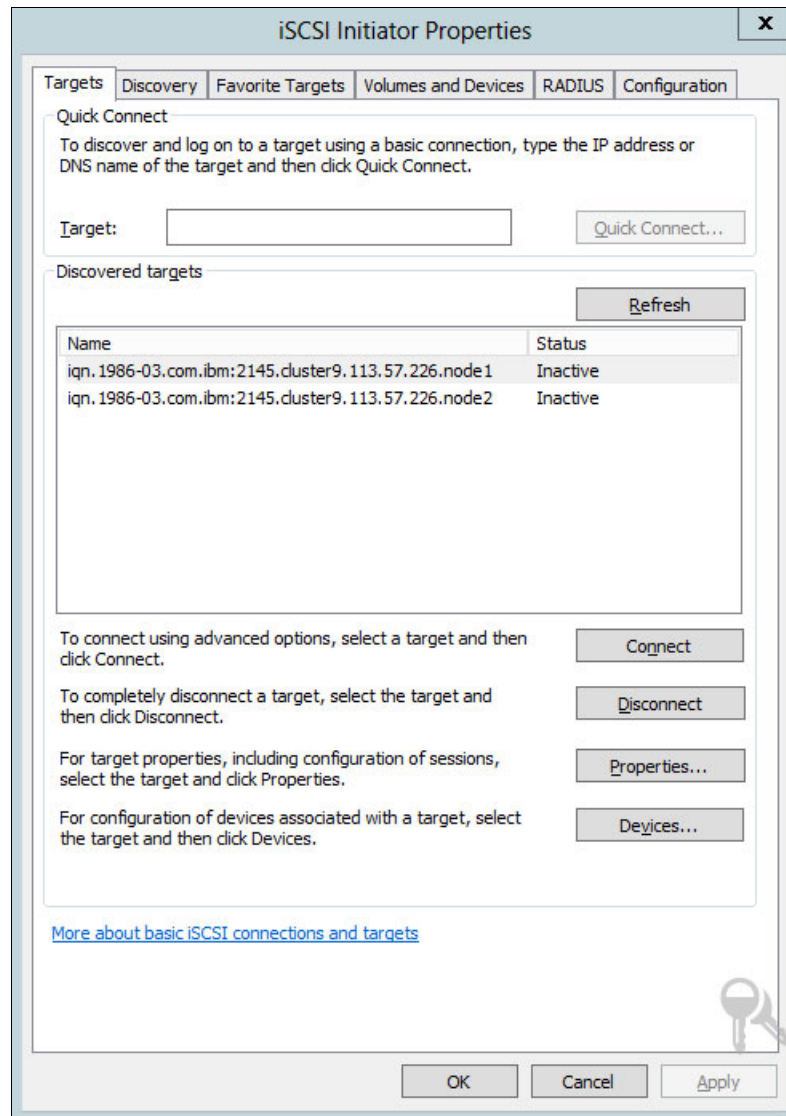


Figure 7-24 Discovered targets

6. To connect to the targets, select each target and click **Connect**. A new window opens, as shown in Figure 7-25.

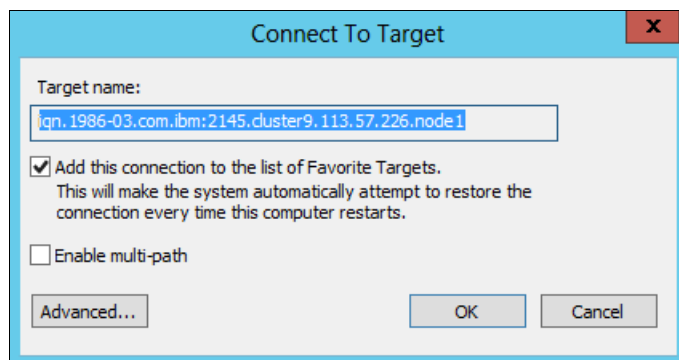


Figure 7-25 Connect to Target

7. To enable multipathing, select the **Enable multi-path** check box.
8. If you intend to configure CHAP authentication or enable header, data digests, or both, click **Advanced**:
 - a. A new window opens, as shown in Figure 7-26. You can select the appropriate check boxes to configure digests and CHAP authentication.

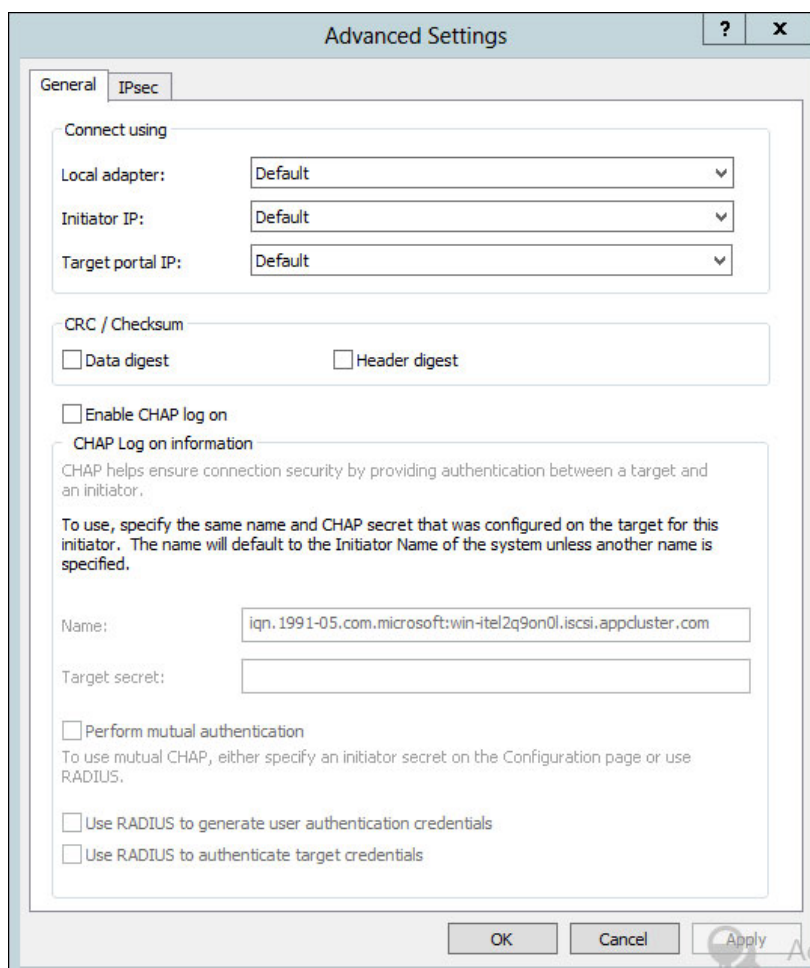


Figure 7-26 Advanced connection settings

- b. Click **OK**.
9. Click **OK** in the Connect to target window to establish a connection.
10. After the connection is established, you can see that both targets are connected, as shown in Figure 7-27.

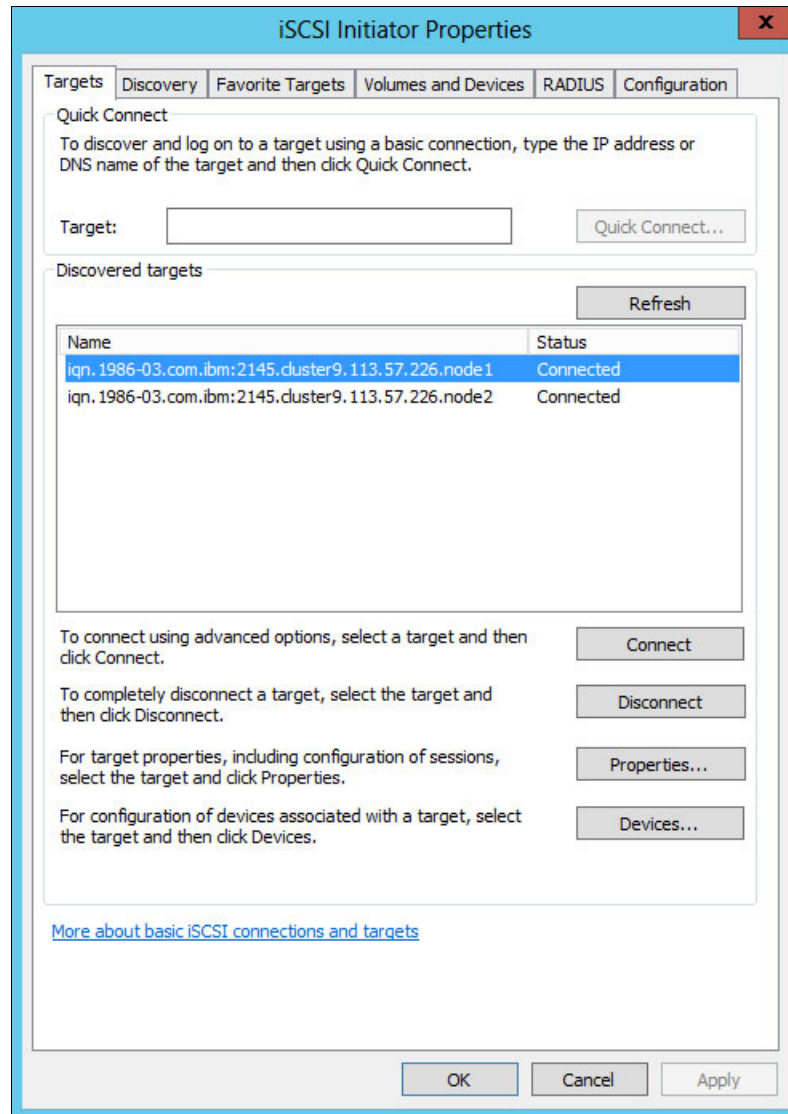


Figure 7-27 Logged in targets

iSNS-based discovery

Another method of discovering target is by using an iSNS server. For information about configuring an iSNS server, see 7.7, “iSNS server configuration” on page 143.

In this section, it is assumed that an iSNS server is already configured. You need the IP address of the iSNS server to do discovery.

To perform iSNS-based discovery of targets by using the Microsoft iSCSI initiator, complete the following steps:

1. In the Discovery tab of the iSCSI initiator software, click **Add Server** in the iSNS servers pane, as shown in Figure 7-28.

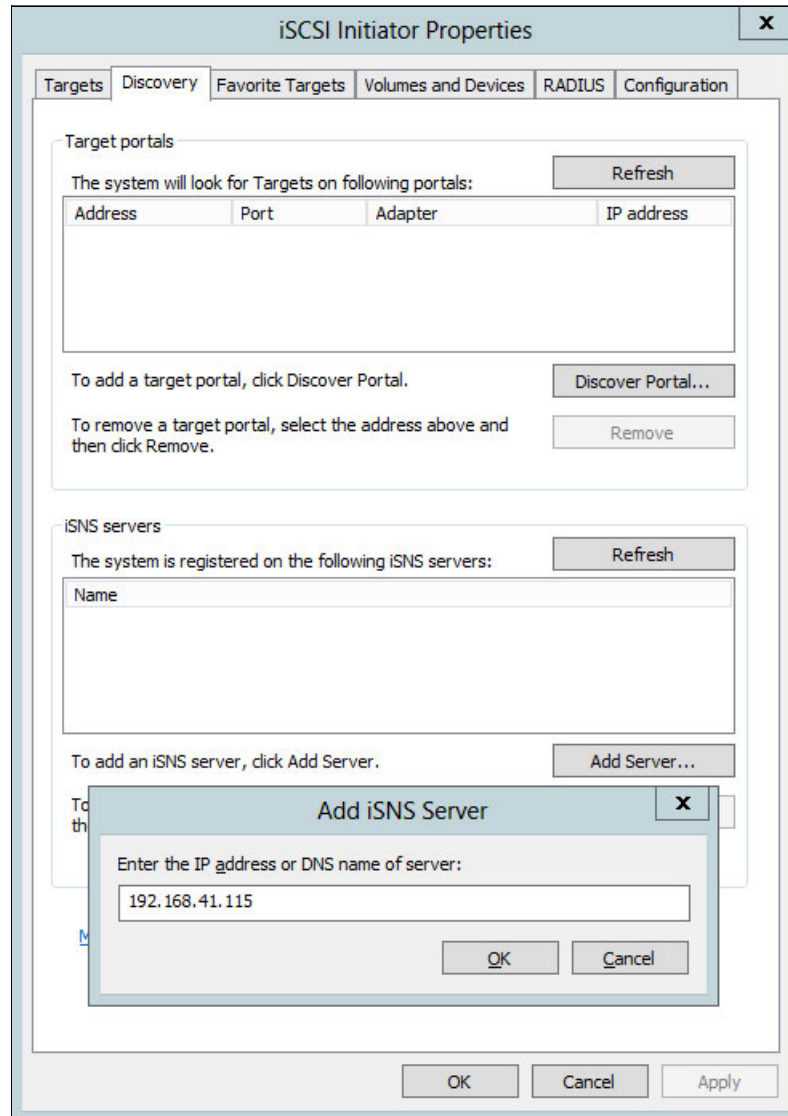


Figure 7-28 Adding an iSNS server IP

2. Click **OK**. The system performs discovery through the iSNS server for targets and automatically connects to the targets. The connected targets are displayed in the Targets tab.

Viewing and managing the discovered disks

After iSCSI targets are discovered by using Microsoft iSCSI initiator, you can use Windows disk management in Computer Management to manage the volumes that are mapped to the host from the IBM Storwize storage systems. Windows should perform an automatic rescan of newly added volumes and they should be visible in the disk management utility. However, if the volumes are not visible, you can do a manual rescan by right-clicking **Disk Management** and selecting **Rescan Disks**, as shown in Figure 7-29.

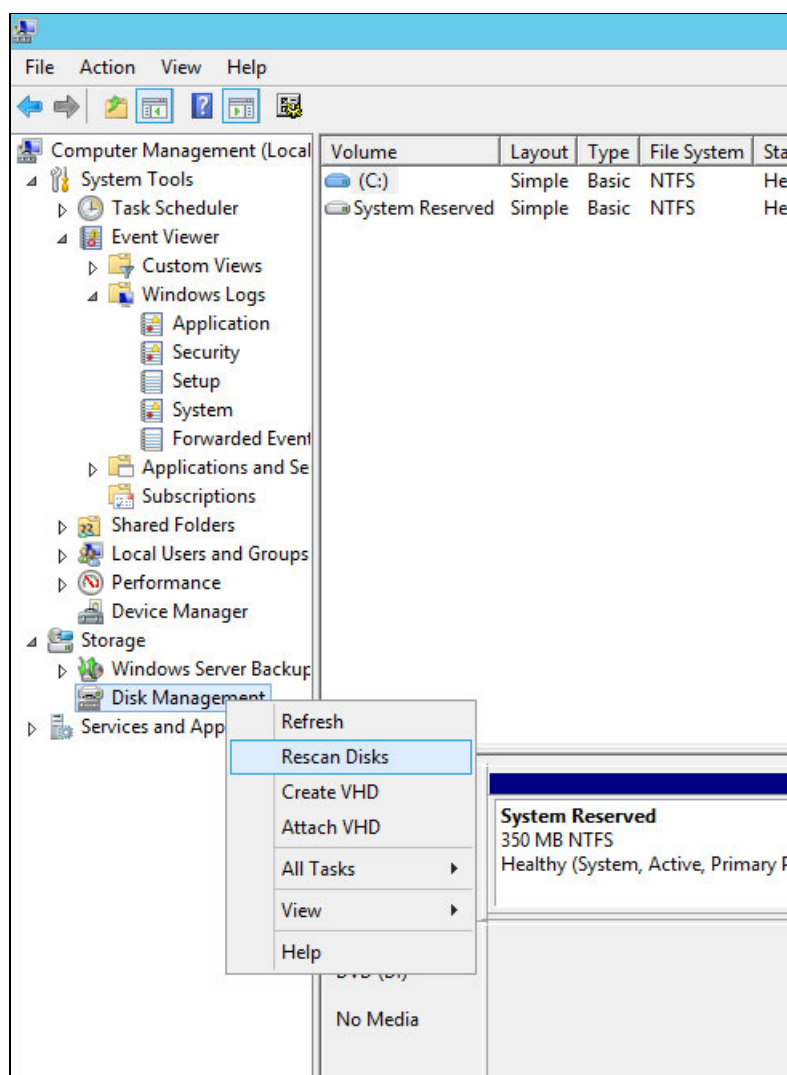


Figure 7-29 Rescanning for disks in the Microsoft Windows operating system

After the volumes are discovered, they can be initialized, formatted, and assigned a drive letter with the disk management utility. You can alternatively use the Microsoft Windows diskpart utility from the command prompt to manage the discovered volumes.

Verifying and managing MPIO for discovered volumes

Microsoft iSCSI MPIO DSM supports a set of load-balance policies that determine how I/O is allocated among the different sessions. Multipathing for devices can be verified and managed by clicking **Devices** → **MPIO**, as shown in Figure 7-30. The figure shows an example in which both paths are up.

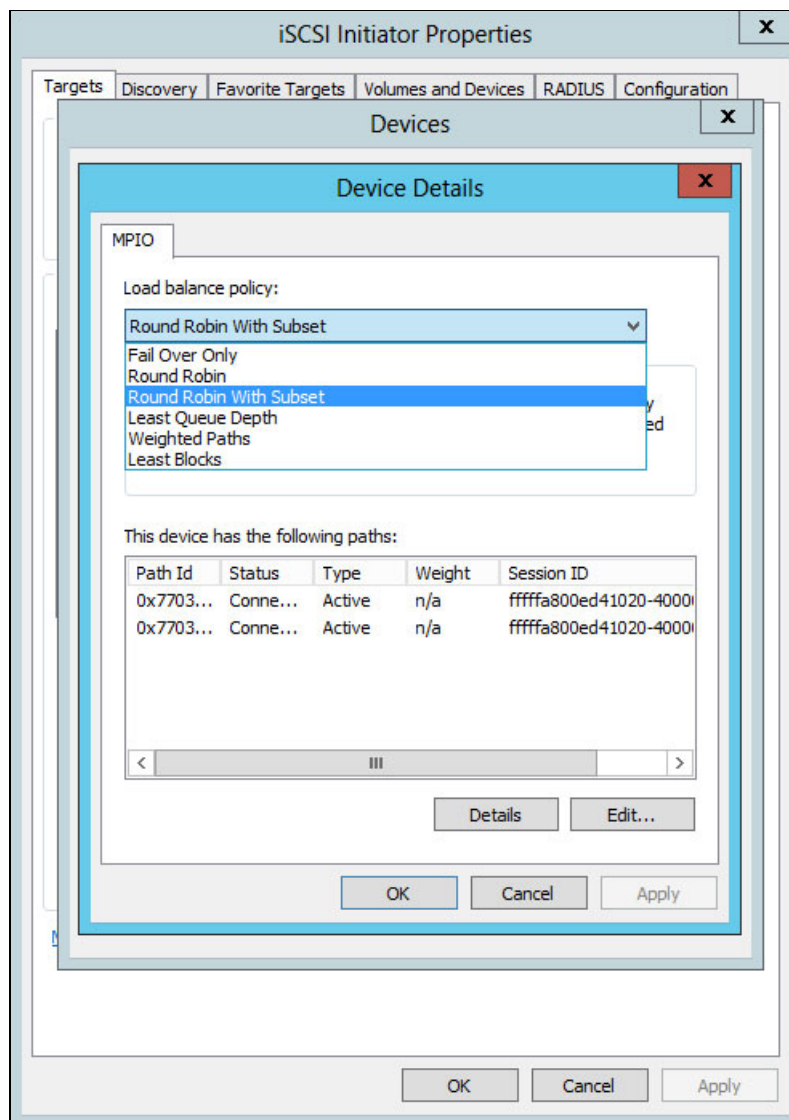


Figure 7-30 Multipath management

Because the IBM Storwize storage system is an active/active array, round-robin is the default policy that is applied. However, different multiple load-balancing options are available. Usage of load-balancing policies might affect performance. It is important to understand the policies thoroughly before you use them. For more information about MPIO load-balancing policies, see [MPIO Policies](#).

Forcing multipath failover

Figure 7-31 shows an example of what the MPIO window looks like when one of the NICs on the host is down.

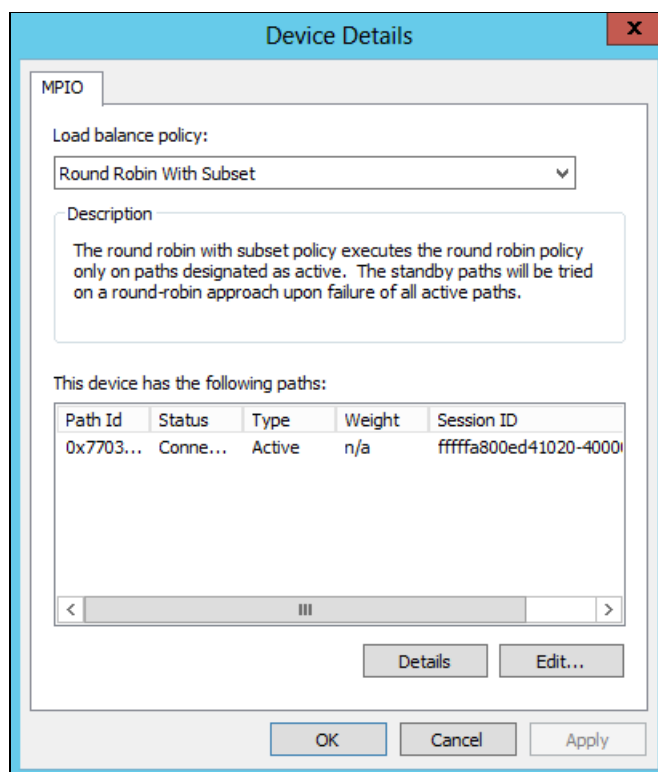


Figure 7-31 Multipath failover

IBM Storwize storage systems support IP target failover. However, to guard against host port failures, use multipathing on the host.

7.6 Configuring iSCSI for VMware ESXi hosts

This section describes the configuration of VMware ESXi hosts for iSCSI by using the software initiator.

Prerequisites

To configure the VMware (ESX) host, complete the following tasks:

- ▶ Be sure that your host server is using the latest firmware levels.
- ▶ To understand the preferred topology and be aware of the general and specific considerations regarding ESX hosts, read Chapter 4, “Planning considerations” on page 39.
- ▶ If necessary, install the supported driver for the NIC adapter. After configuring the IP address, verify that you can ping each port of the IBM Storwize storage system.

- Configure the host, volumes, and host mapping, as described in 7.2, “Configuring initiators for iSCSI” on page 87.
- VMware iSCSI Initiator is installed natively on ESX Server and no installation is required. However, you must add the software initiator to the Storage Adapters category. Complete the following steps:
 - a. Click **Configuration** → **Storage Adapters** on the vSphere Client.
 - b. At the upper right corner, click **Add**.
 - c. Select **Add Software iSCSI Adapter** and click **OK** (Figure 7-32).

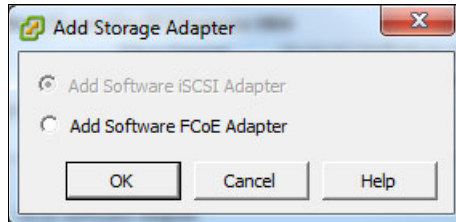


Figure 7-32 Add Storage Adapter

A new iSCSI Software Adapter is created, as shown in Figure 7-33.

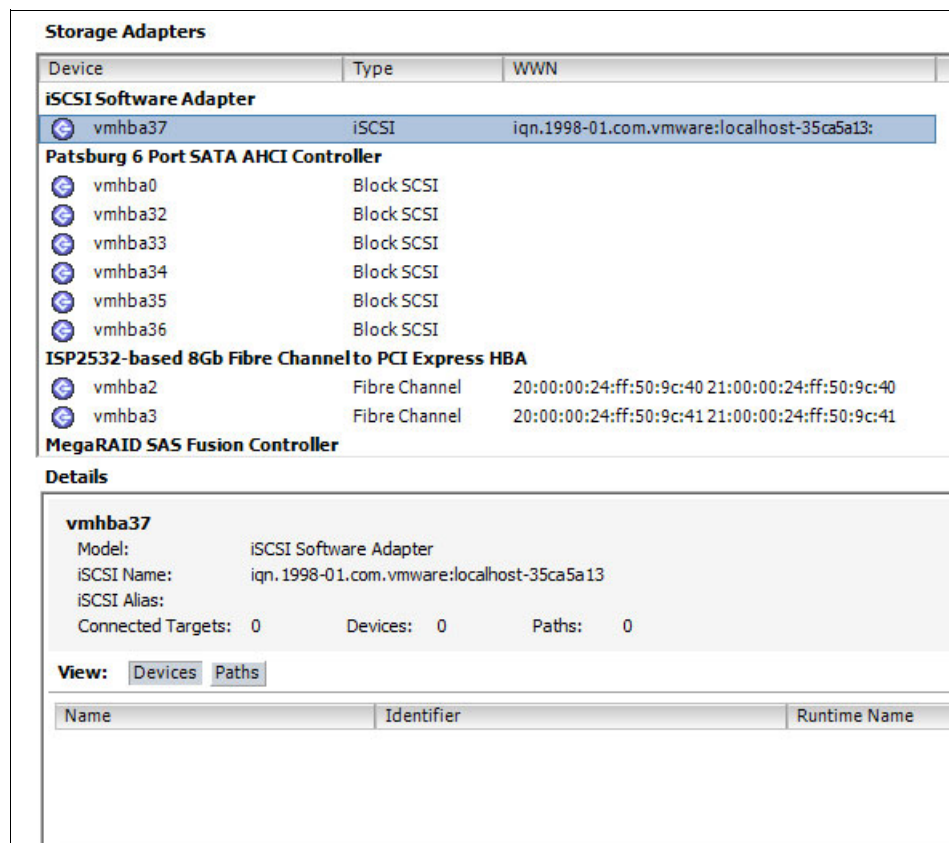


Figure 7-33 Storage adapters

You can optionally set the IQN for the ESX host. To change the IQN, complete the following steps:

1. On the vSphere client, click **Configuration** → **Storage adapters**.
2. Right-click **iSCSI Software Adapter** and click **Properties**.
3. On the General tab, click **Configure** and enter the IQN into the iSCSI name box (Figure 7-34).

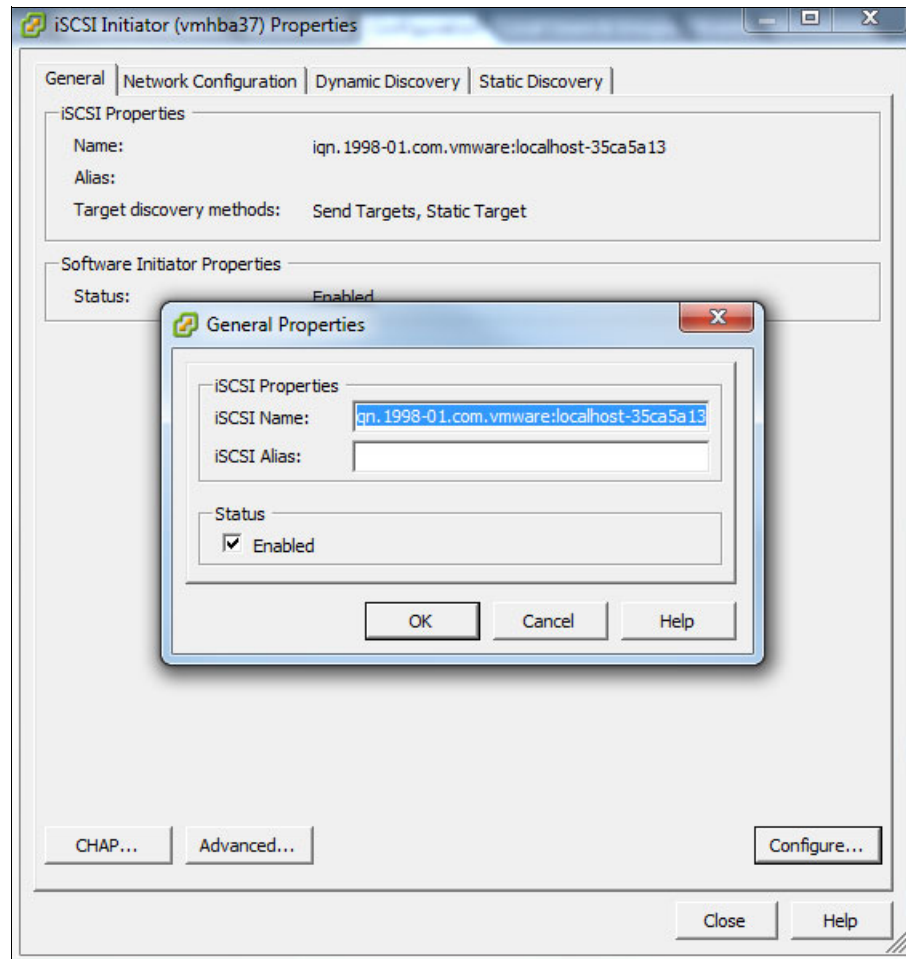


Figure 7-34 General Properties

4. Optional: Assign an alias to the adapter by entering the necessary value into the iSCSI Alias field.

ESXi terminologies to understand before you proceed with the configuration

Be sure to understand the following terms before you begin the configuration:

- ▶ Virtual switch (vSwitch)

A vSwitch is a software program that helps the virtual machines (VMs) communicate among themselves. It contains the VMkernel, which helps the VMware to access the iSCSI I/O and a service console, which is used to perform the management activities.

- ▶ VMkernel

The VMkernel is the operating system of ESX or ESXi servers. All the hardware, including processors, memory, NIC, and host bus adapters (HBAs), is controlled by VMkernels.

- ▶ VM console

A VM console can be described as a specialized network interface that enables the management of the ESX host.

- ▶ VM port groups

These groups help connect all the guest operating systems to the network.

7.6.1 Configuring the Ethernet network on the VMware host

ESXi iSCSI initiators use vSwitches and VMkernel interfaces to drive iSCSI I/O. Therefore, you must configure networking on the ESXi host before the iSCSI targets can be discovered. This section describes the configuration of the network on the ESXi host that is required for configuring iSCSI.

Configuring a vSwitch

To configure a vSwitch, complete the following steps:

1. Click the **Configuration** tab of the vSphere client.
2. Click **Networking** in the Hardware section.
3. Click **Add networking**.

4. Select **Virtual Machine** and click **Next**, as shown in Figure 7-35.

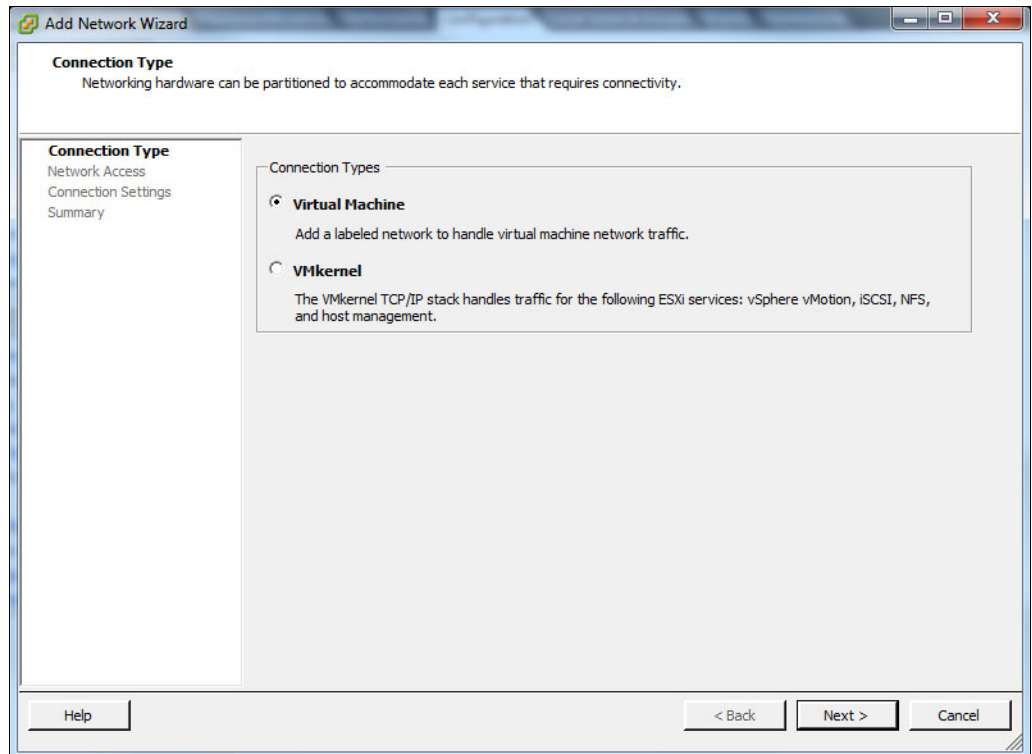


Figure 7-35 Connection types

5. Select **Create a vSphere standard switch**, and click **Next**.

- Under Port Group Properties, enter the network label and optional VLAN ID, if you are using VLANs, for the port group, as shown in Figure 7-36, and click **Next**.

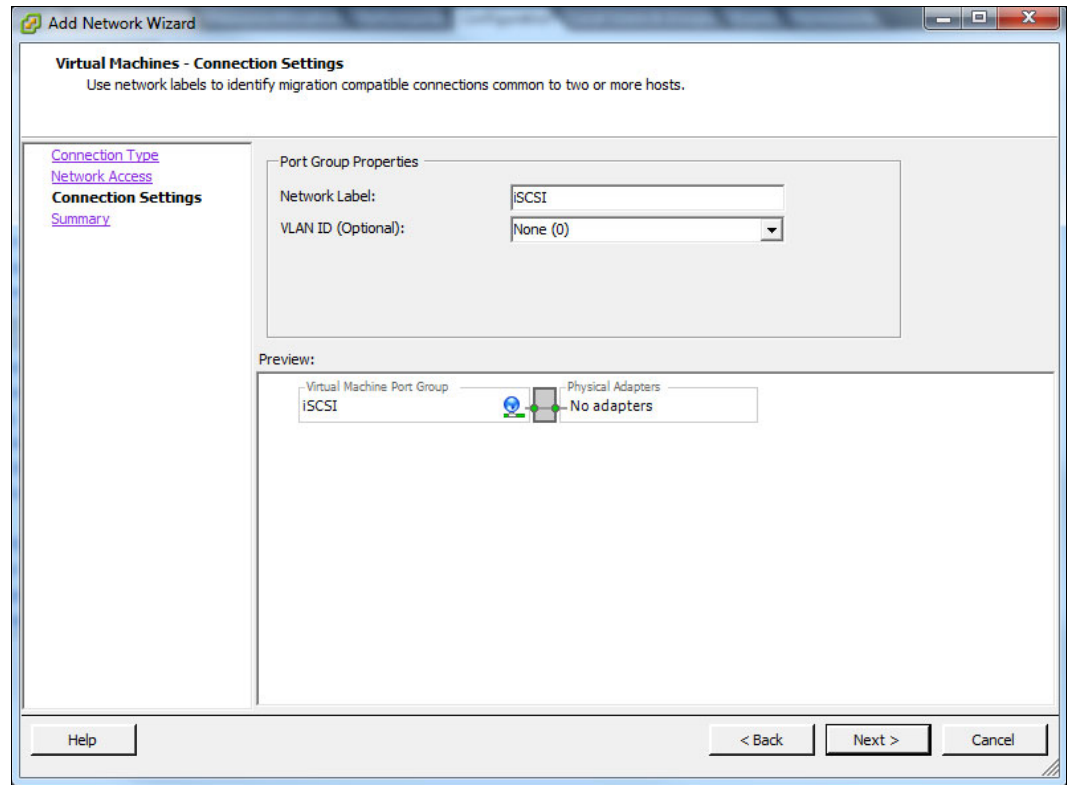


Figure 7-36 Virtual Machines: Connection Settings

- Click **Finish** on the next window to complete the wizard.
- Verify that a new vSwitch appears under networking.

Creating a VMkernel interface

A VMkernel interface must be created in the vSwitch to carry iSCSI I/O. To create a VMkernel interface, complete the following steps:

- Click **Properties** for the vSwitch that is created for iSCSI.
- Click **Add** under vSwitch properties.

3. In the Add Networking wizard, select **VMkernel**, as shown in Figure 7-37.

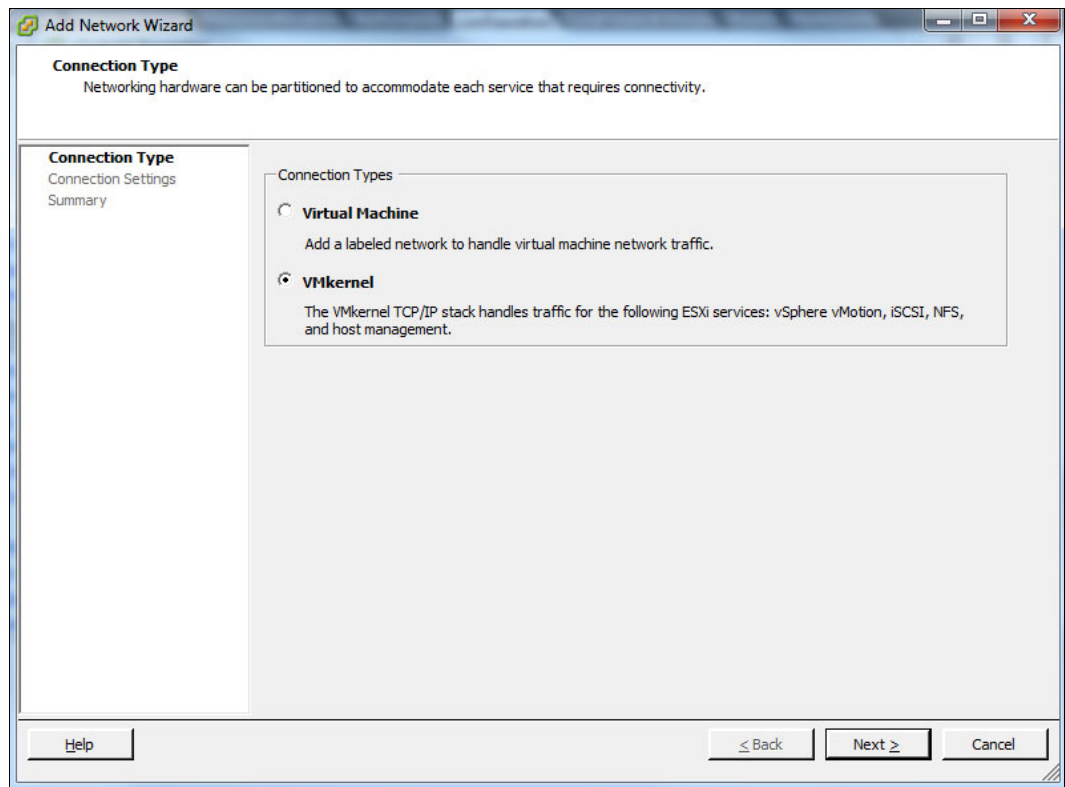


Figure 7-37 Connection type

4. Enter the name for the VMkernel port group and optional VLAN ID if you are using VLAN in your network for iSCSI. If you are using IPv6 for your network, select **IPv6** or **IP and IPv6** and click **Next**. Figure 7-38 shows the configuration window.

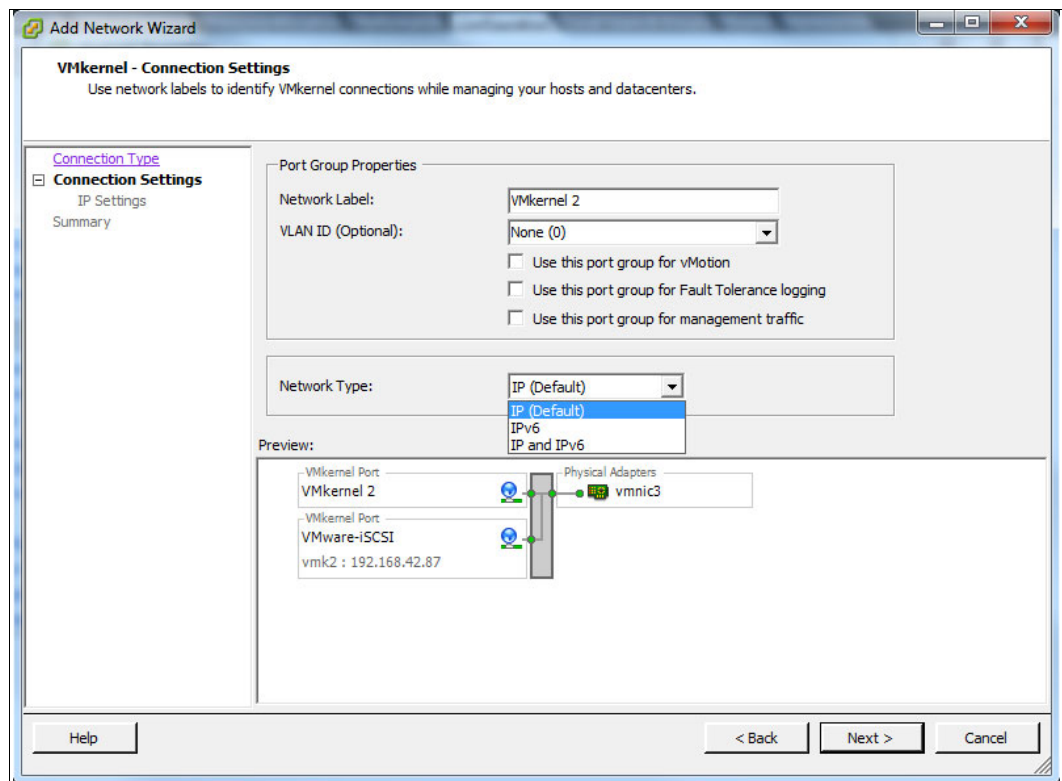


Figure 7-38 VMkernel: Connection Settings

5. If you are using DHCP for your network, select **Obtain IP setting automatically**. Otherwise, select **Use the following IP settings** to configure a static IP for the VMkernel interface on the VMkernel - IP configurations settings window, as shown in Figure 7-39.

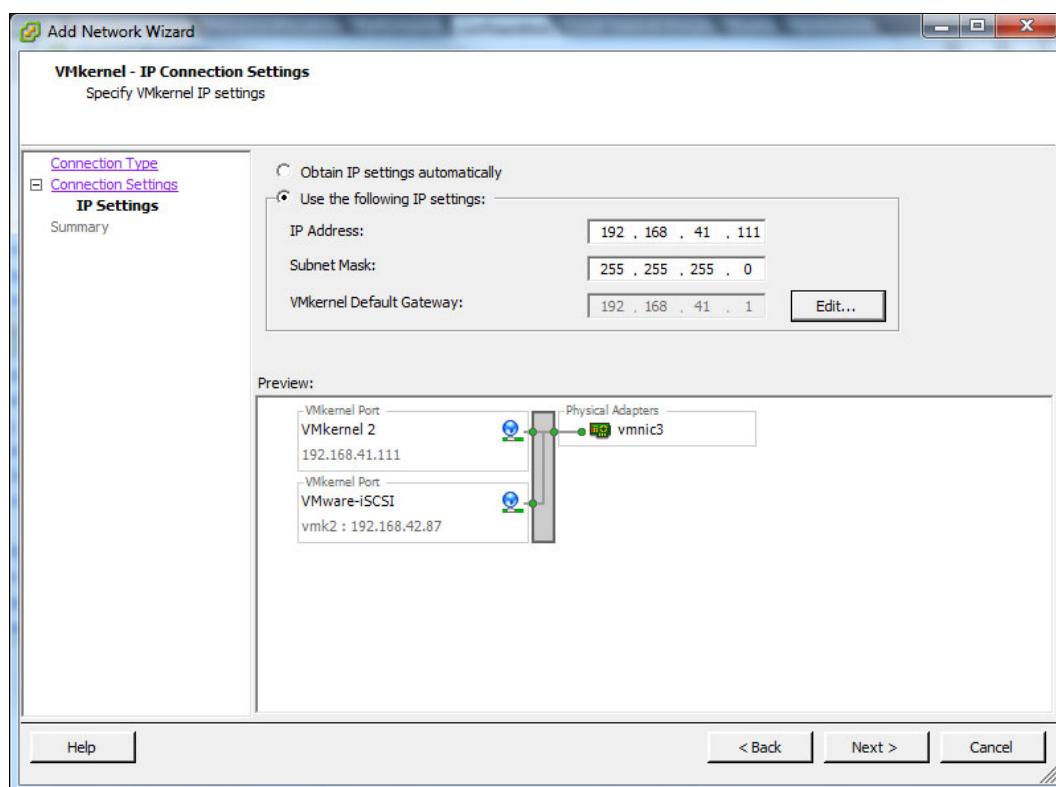


Figure 7-39 VMkernel: IP Connections Settings

6. Click **Finish** on the next window to create the interface.

Discovering targets on ESXi

This section describes how to discover targets on ESXi.

Binding a VMkernel interface to an ESXi iSCSI software adapter

You must bind a VMkernel interface to an ESXi iSCSI software adapter for iSCSI communication to work. To create the bindings, complete the following steps:

1. Click **Configuration** → **Storage Adapter** in the vSphere client.
2. Right-click **Software iSCSI adapter** and click **Properties**.

3. Click **Network Configuration** under Properties and click **Add**, as shown in Figure 7-40.

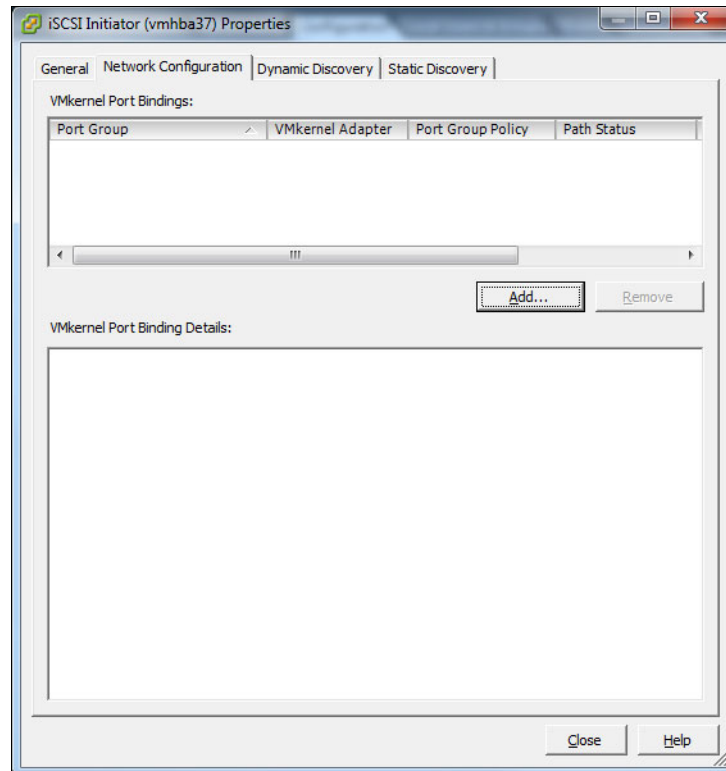


Figure 7-40 iSCSI initiator properties

4. Select the VMkernel interface that you created for iSCSI and click **OK**, as shown in Figure 7-41.

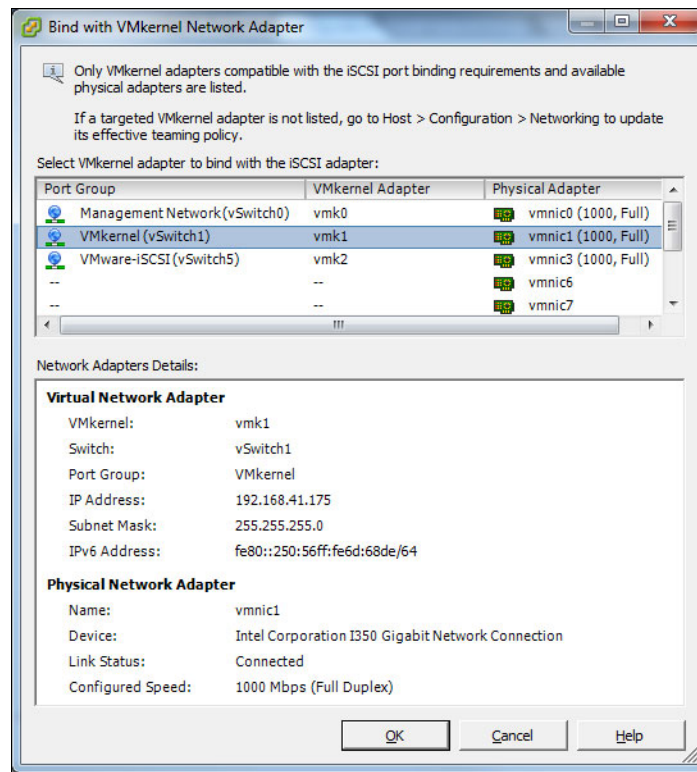


Figure 7-41 Binding with the VMkernel Network Adapter

The VMkernel interface is added under the VMkernel port bindings, as shown in Figure 7-42.

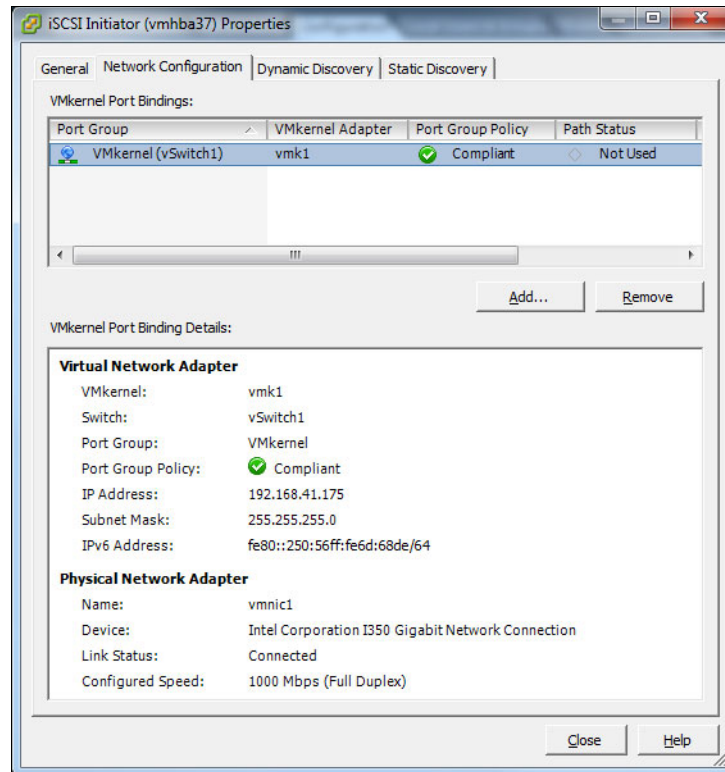


Figure 7-42 VMkernel port bindings

Discovering iSCSI targets

There are two ways to discover targets on ESXi by using software iSCSI:

► Static discovery

In static discovery, the IP is required in addition to the IQNs of all of the targets. The user must connect to all the targets individually by manually adding them. For more information, see “Static discovery” on page 131.

► Dynamic discovery

In dynamic discovery, only the IP address of the target is required. The software performs discovery and automatically connects to all the discovered targets. For more information, see “Dynamic discovery” on page 133.

Static discovery

To configure static discovery in the vSphere client, complete the following steps:

1. Click **Configuration** → **Storage Adapter**.
2. Right-click **Software iSCSI adapter** and click **Properties**.
3. Click **Static discovery** and click **Add**.

4. Enter the IP address and IQN of the target, as shown in Figure 7-43.

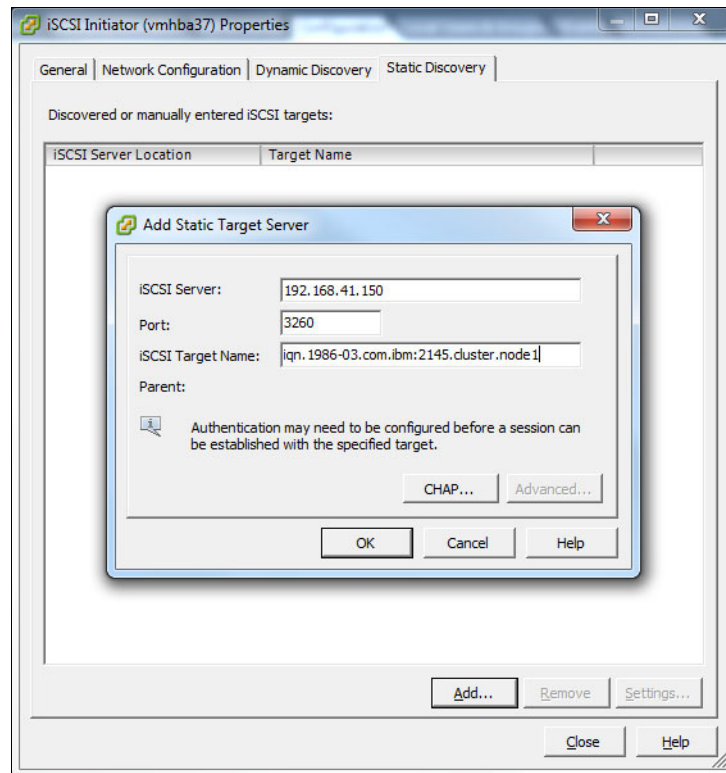


Figure 7-43 Adding a Static Target Server

5. Click **OK**.

The initiator connects to the target and the connected target displays, as shown in Figure 7-44.

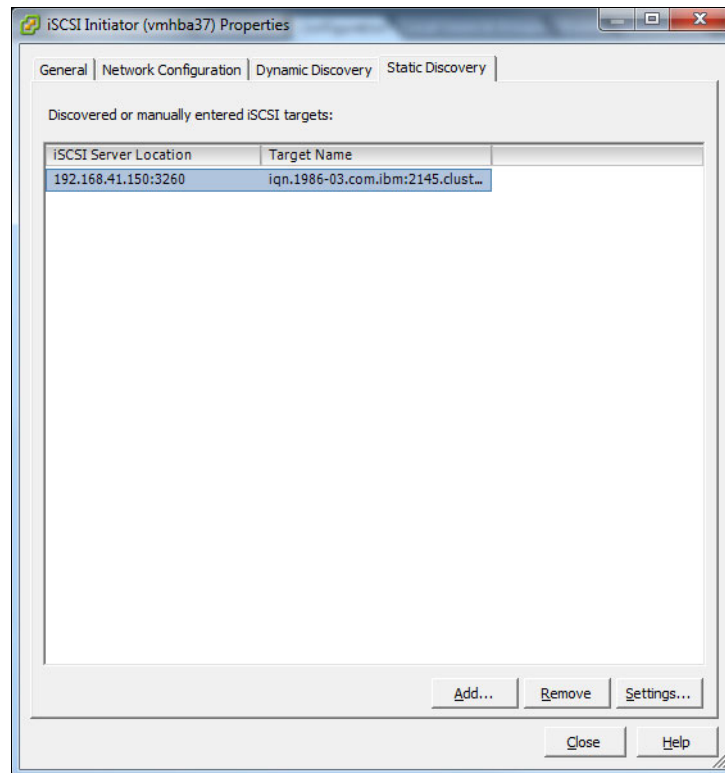


Figure 7-44 Discovered or manually entered targets

Dynamic discovery

To configure dynamic discovery in the vSphere client, complete the following steps:

1. Click **Configuration** → **Storage Adapter** in the vSphere client.
2. Right-click **Software iSCSI adapter** and click **Properties**.
3. Click **Dynamic discovery** and click **Add**.

4. Enter the IP address of the target, as shown in Figure 7-45, and click **OK**.

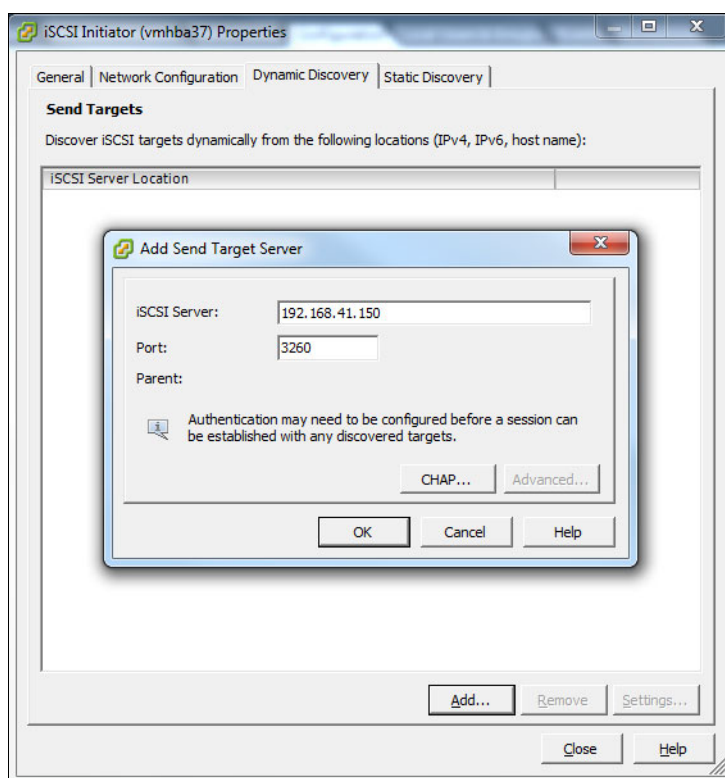


Figure 7-45 iSCSI server location

5. The discovery address is added under SendTargets and Discovered Targets and is shown under Static Discovery.

Note: After discovery, the system prompts you for a rescan of the software adapter. Select **Yes** to perform discovery of the volumes that are mapped to the host.

Setting the CHAP authentication

To set CHAP authentication for iSCSI on ESXi in the vSphere client, complete the following steps:

1. Click **Configuration** and then **Storage Adapter**.
2. Right-click **Software iSCSI adapter** and click **Properties**.
3. Under iSCSI Software adapter properties, click **CHAP**.

4. As shown in Figure 7-46, select **Use CHAP** for “CHAP (target authenticates host)”. Optionally, you can also select **Use CHAP** for “CHAP (host authenticates target)” if you want to use mutual CHAP.

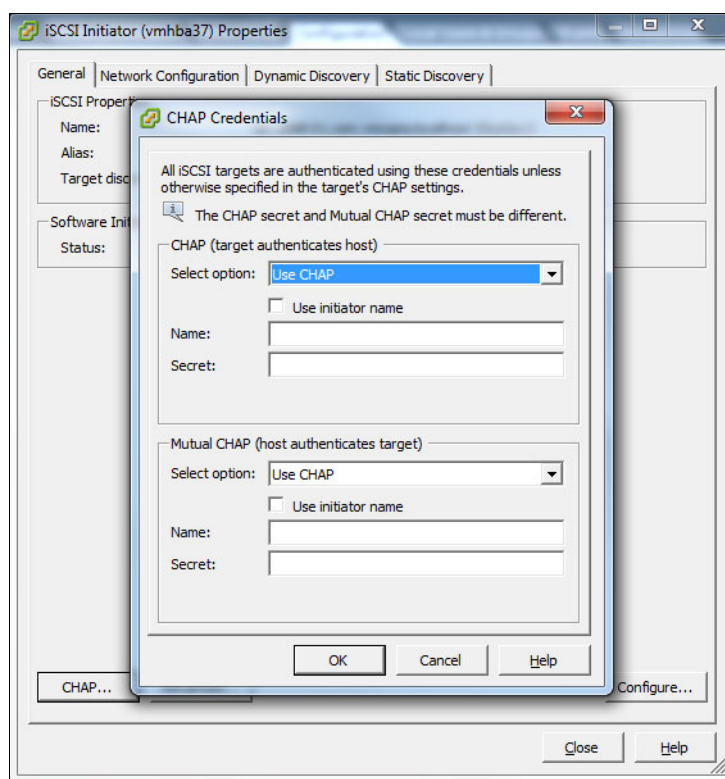


Figure 7-46 Specifying the CHAP credentials

Important: The CHAP name and secret must match the values that are set on the IBM Storwize storage system. If you have used the initiator name (IQN) as the CHAP name on the IBM Storwize storage system, you can select **Use initiator name** for the CHAP setting.

Host multipathing

Multipathing provides the ability to load balance between paths when all paths are present and to handle failures of a path at any point between the server and the storage. The default iSCSI configuration for VMware gives only one path from the initiator to the target. To enable failover and load balancing, port binding must be configured by the administrator to have multiple paths.

Prerequisites for vmknics-based multipathing for software iSCSI

Before you configure vmknics-based multipathing, consider the following things:

1. Two VMkernel port groups are required with an uplink that is connected to them.
2. A VMkernel network adapter should be linked to the software iSCSI adapter. A rediscovery of targets is required to detect multiple paths.

Configuring vmknick-based multipathing

To configure vmknick-based multipathing, complete the following steps:

1. Configure the network as described in 7.5.1, “Prerequisites” on page 108. As part of this process, ensure that you have a vSwitch.

Note: As a preferred practice, have two or more physical adapters for iSCSI on the host machine for multipathing.

2. Configure an additional VMkernel port, which is a prerequisite to configure the port binding, as described in 7.5.1, “Prerequisites” on page 108.7.5.1, “Prerequisites” on page 108.

Note: VMkernel port groups can also be created on different vSwitches. However, if the VMkernel adapters in your configuration are on the same subnet, they can be configured on the same vSwitch.

3. In the previous steps, where you added the network adapters to a vSwitch, network adapters appear as active to each VMkernel port on the vSwitch. Ensure that you override this configuration to ensure that each VMkernel port maps to only one active adapter.
 - a. Under the Port tab of the vSwitch properties, select a VMkernel port and click **Edit**, as shown in Figure 7-47.

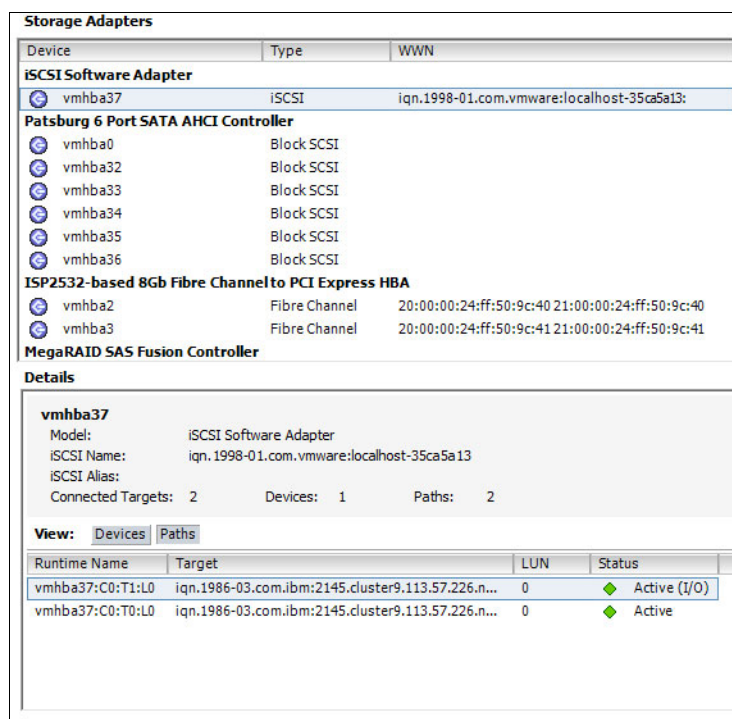


Figure 7-47 Storage Adapters

- b. On the NIC Teaming tab, select **Override switch failover order**, as shown in Figure 7-48.

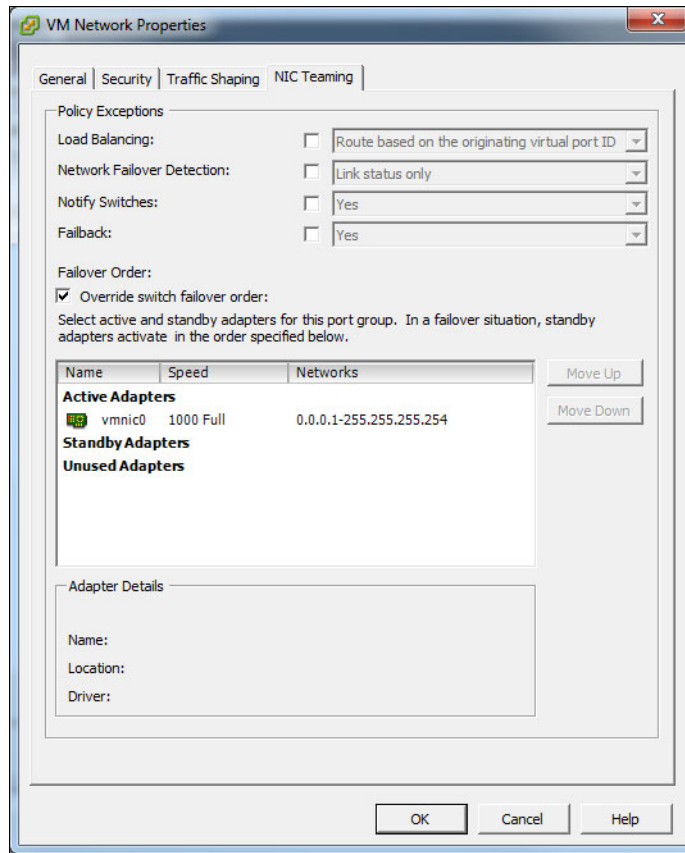


Figure 7-48 VM Network Properties: NIC Teaming

- c. Under Active Adapters, keep only one adapter and then use **Move Down** to shift other adapters under the unused adapters, as shown in Figure 7-49.

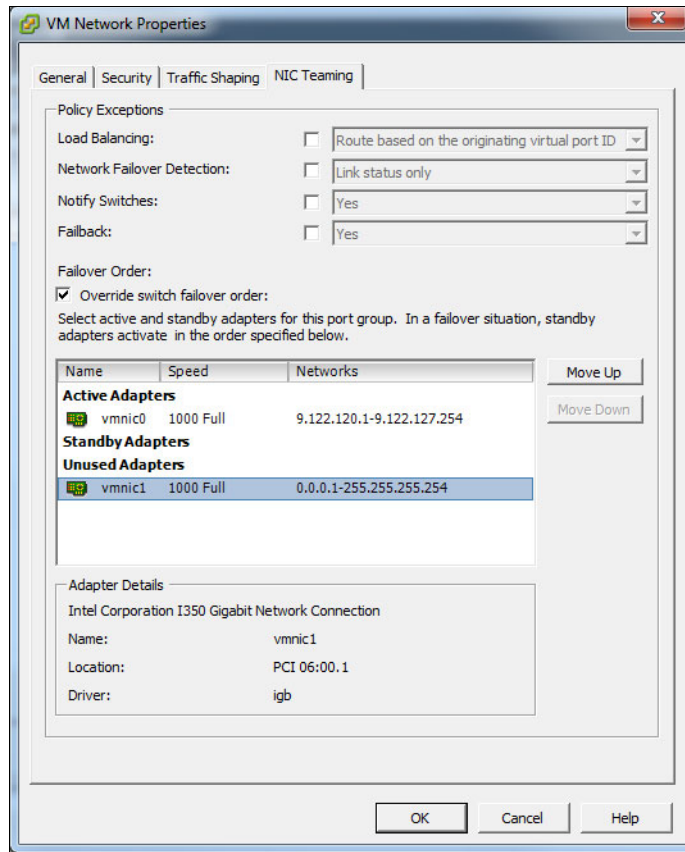


Figure 7-49 VM Network Properties: NIC Teaming

- d. To ensure unique mapping of the port to an adapter, repeat steps a on page 136 to c.
4. Enable the iSCSI Software Adapter, as described in 7.6, "Configuring iSCSI for VMware ESXi hosts" on page 120.

5. Configure the port binding:
 - a. Click the **Configuration** tab and then click **Storage Adapters** → **Properties**.
 - b. Click **Network Configuration** → **ADD** to bind the network adapter to the iSCSI adapter, as shown in Figure 7-50.

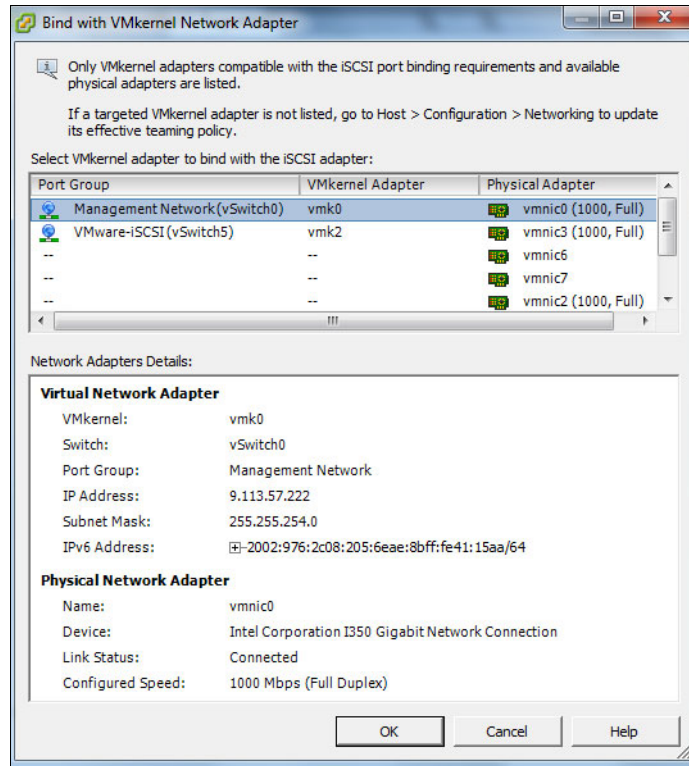


Figure 7-50 Binding with the VMkernel Network Adapter

6. You see a list of all VMkernel adapters that are compatible with the iSCSI port binding requirements. You can select the VMkernel network adapter that you want to bind the iSCSI adapter with and then click **OK**.
7. Repeat step 6 until you bind all the required adapters.

8. When all the VMkernel adapters are bound to the iSCSI software adapters, the adapters are shown, as shown in Figure 7-51.

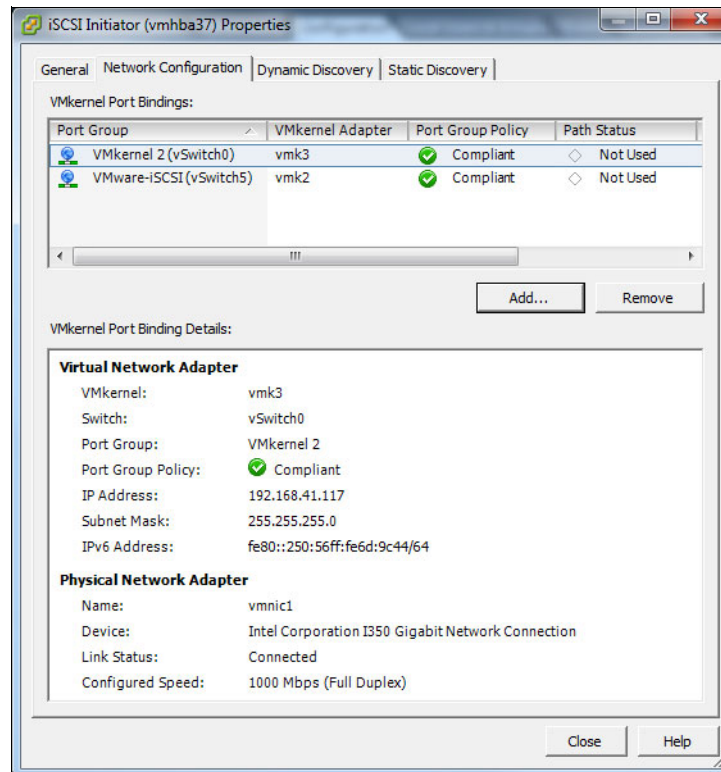


Figure 7-51 Network Configuration

9. Close the Initiator Properties window and click **Rescan** to verify that multiple paths are available for iSCSI LUNs. Figure 7-52 shows a device with multiple paths.

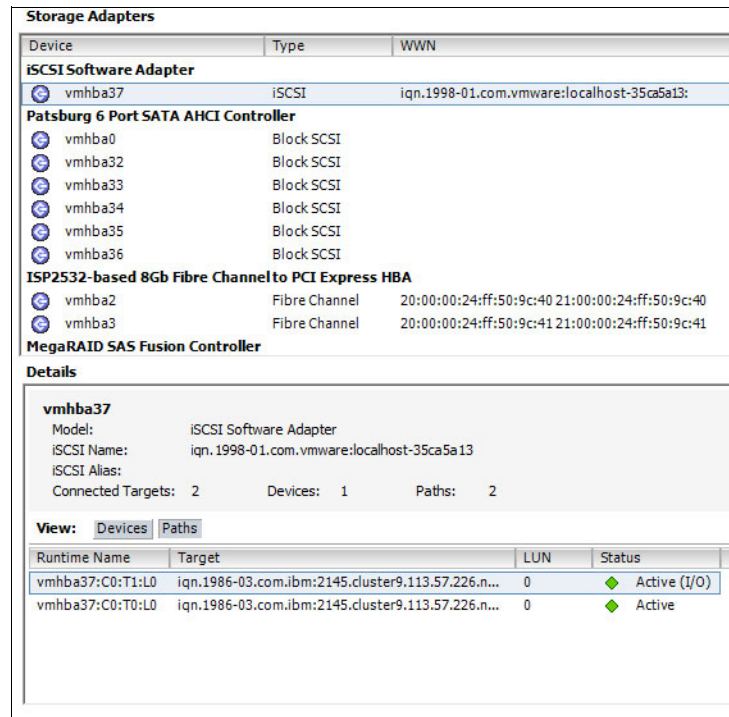


Figure 7-52 Storage Adapters

Choosing a multipathing policy

The following multipath I/O policies can be chosen on ESX or ESXi servers:

- **Fixed:** Always use the preferred path to the disk. If the preferred path is not available, an alternative path to the disk is used. When the preferred path is restored, an automatic failback to the preferred path occurs.
- **Most Recently Used:** Use the most recently used path if the path is available. Whenever a path failure occurs, an alternative path is used. There is no automatic failback to the original path.
- **Round Robin:** Multiple disk paths are used and balanced by using an automatic rotation mechanism.

To change the multipathing policy on ESXi hosts, complete the following steps:

1. In the Storage Adapters section in the vSphere client, select the software adapter, right-click the device or LUN, and click **Manage Paths**, as shown in Figure 7-53.

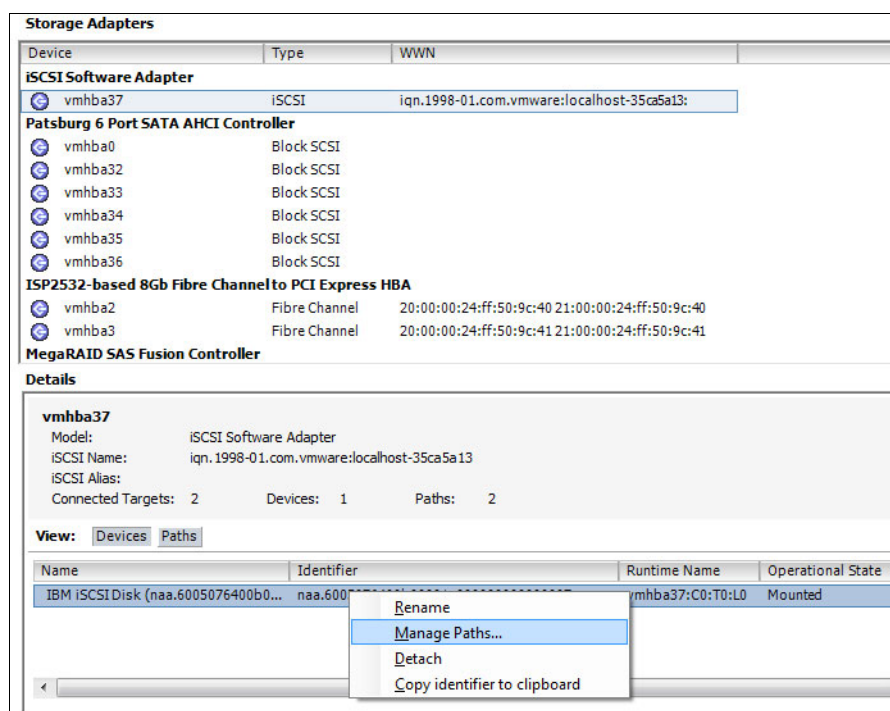


Figure 7-53 Manage Paths

2. Change the policy by using the drop-down menu for Path Selection under Policy, as shown in Figure 7-54.

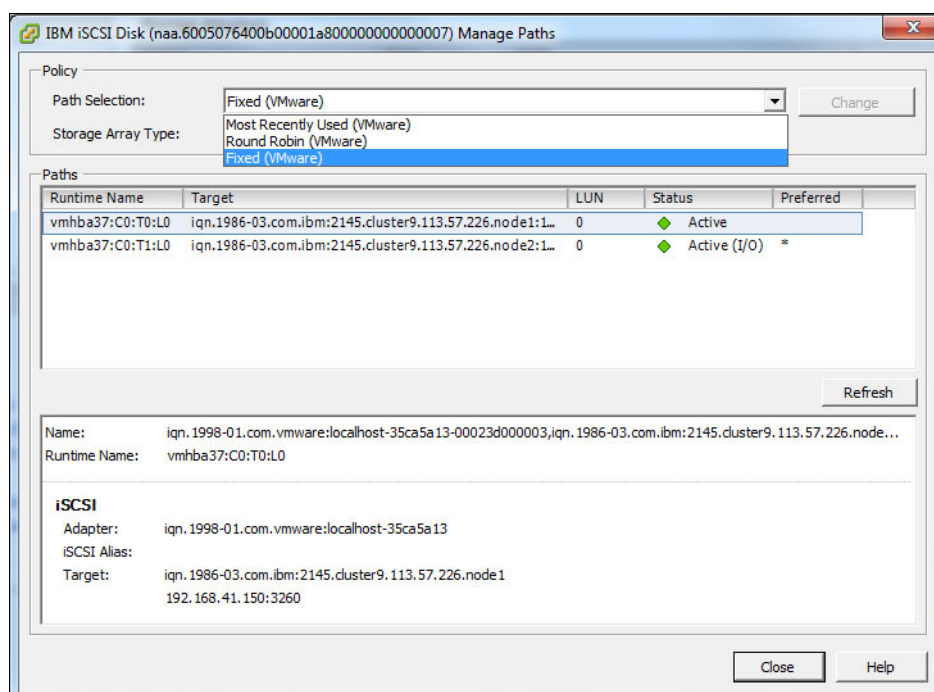


Figure 7-54 Path Selection

An IBM Storwize storage system is an active/active array, so round-robin is the preferred policy. For more information about all the multipathing policies and to select a suitable policy for your environment, see your VMware documentation.

7.7 iSNS server configuration

iSNS is a protocol that allows automatic discovery, management, and configuration of iSCSI devices. iSNS is like a DNS for iSCSI devices.

7.7.1 Enabling iSNS server in Windows 2012

Microsoft Windows Server 2012 includes an iSNS as a standard feature. To install it, complete the following steps:

1. Start Server Manager.
2. Click **Manage** → **Add Roles and Features**.
3. Select **Internet Storage Name Server** and click **Install**, as shown in Figure 7-55.

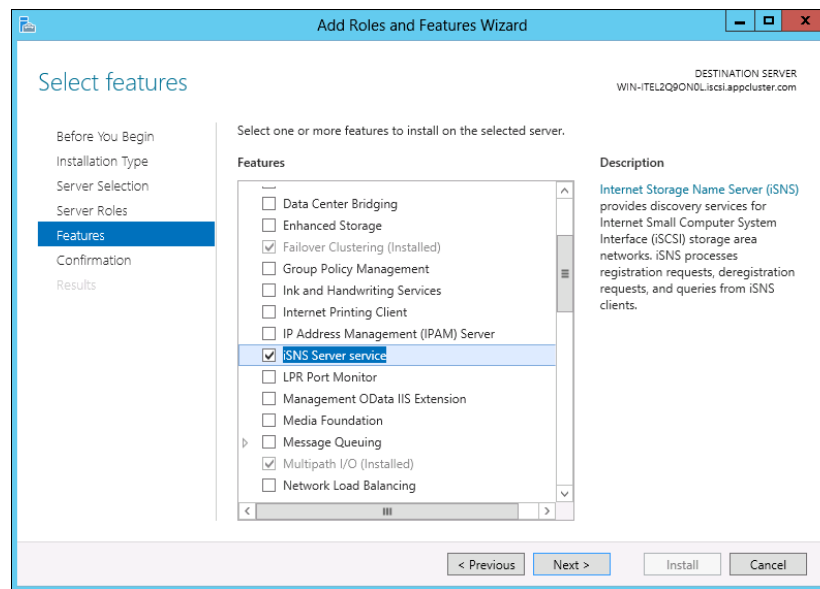


Figure 7-55 Adding the iSNS Server service feature

4. After installation, restart the iSNS server system.

7.7.2 Configuring the iSNS server address on an IBM Storwize storage system

This section shows how to set up an IBM Storwize storage system for iSNS server with the CLI and GUI.

Configuring the iSNS server address with the CLI

To configure the iSNS server address, complete the following steps:

1. To specify an IPv4 address for the iSNS, run the following command:

```
chsystem -isnsip iSNS_server_address
```

Where *iSNS_server_address* is the IP address of the iSCSI storage name service in IPv4 format.

2. To specify an IPv6 address for the iSNS, run the following command:

```
chsystem -isnip_6 ipv6_sns_server_address
```

Where *ipv6_sns_server_address* is the IP address of the iSCSI storage name service in IPv6 format.

Configuring the iSNS server address with the GUI

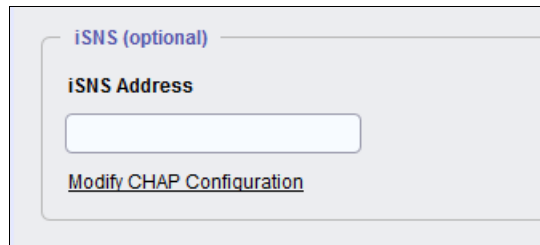
To configure the iSNS server address, complete the following steps:

1. Log in to the IBM Storwize console.
2. Click **Settings** → **Network** → **iSCSI**, as shown in Figure 7-56.



Figure 7-56 iSCSI Configuration

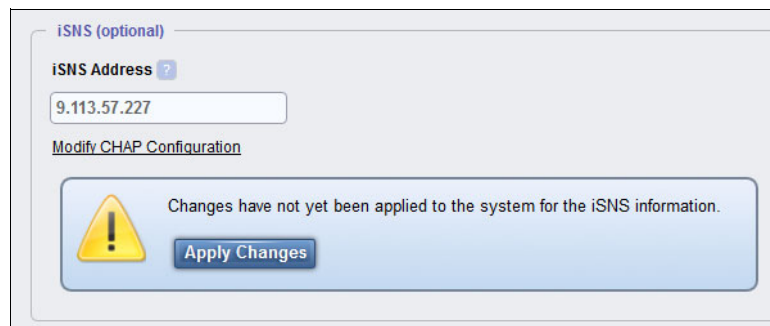
3. On the iSCSI configuration window, add the iSNS address, as shown in Figure 7-57.



The image shows a configuration window titled "iSNS (optional)". Inside, there is a label "iSNS Address" followed by an empty text input field. Below the input field is a link that says "Modify CHAP Configuration".

Figure 7-57 Adding the iSNS address

4. Click **Apply Changes**, as shown in Figure 7-58.



The image shows the same configuration window as Figure 7-57, but now the "iSNS Address" field contains the value "9.113.57.227". Below the input field is a link that says "Modify CHAP Configuration". At the bottom of the window, there is a yellow warning triangle icon and a message that says "Changes have not yet been applied to the system for the iSNS information." Below this message is a button labeled "Apply Changes".

Figure 7-58 Applying changes for the iSNS address configuration

7.7.3 Configuring the iSCSI initiator with iSNS server details

To use the discovery function, the iSCSI initiator must connect to the iSNS server. To set up the iSCSI initiator to connect to the iSNS server, complete the following steps:

1. From Control Panel, click **Administrative Tools** → **iSCSI Initiator** to open the iSCSI Initiator Properties window, as shown in Figure 7-59.

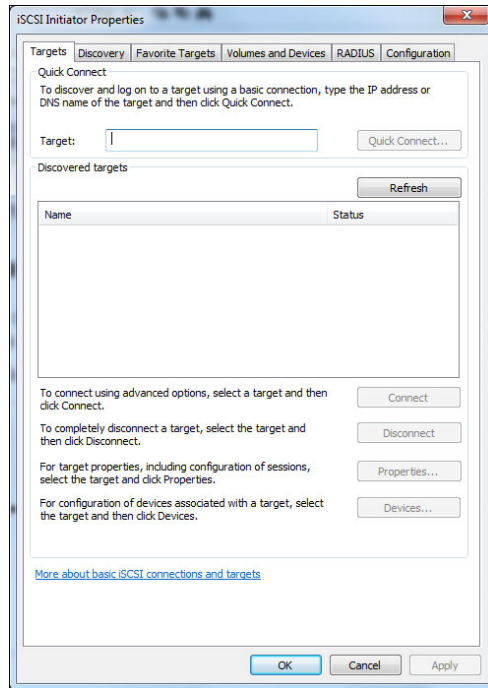


Figure 7-59 iSCSI Initiator Properties

2. Click the **Discovery** tab, as shown in Figure 7-60.

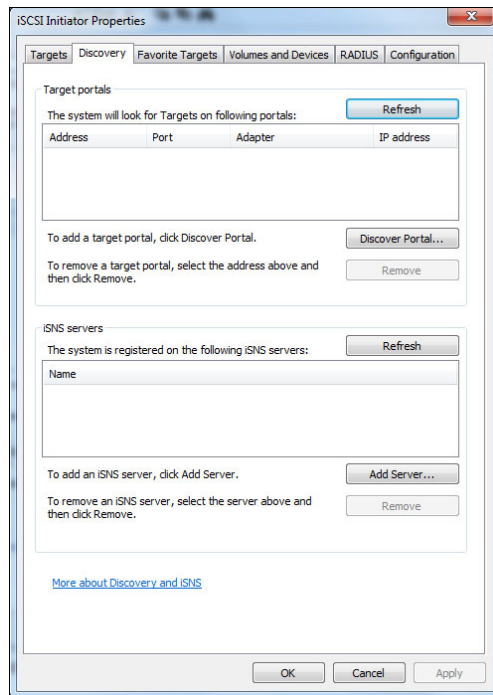


Figure 7-60 iSCSI Initiator Discovery tab

3. Click **Add Server**, as shown in Figure 7-61.

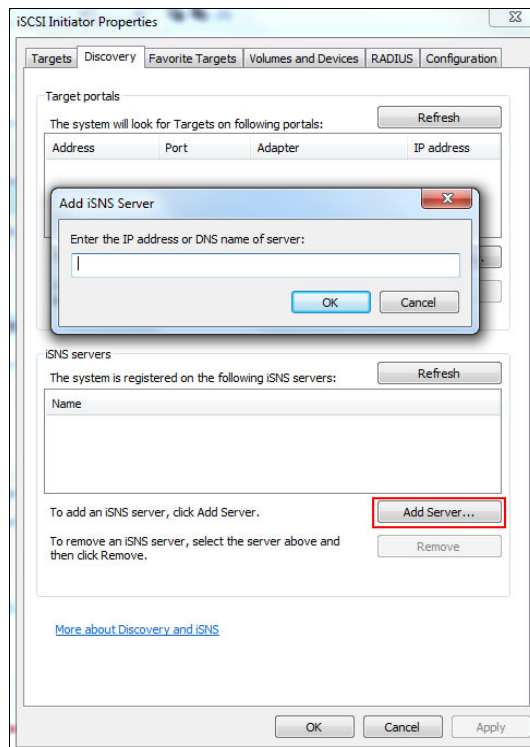


Figure 7-61 Adding the iSNS Server address

7.8 Configuring Priority Flow Control for the IBM Storwize storage system

With PFC, it is possible to prioritize iSCSI traffic over other data traffic in the network by setting bandwidth limits to each type of traffic. Its configuration is defined on the Ethernet switch and can be seen (but not modified) by the IBM Storwize storage system.

This section does not provide detailed information about how to configure PFC from an Ethernet switch perspective. Instead, it focuses on showing how to make a basic configuration on the Ethernet switch and based on that configuration, shows how to verify the flow control configuration on the IBM Storwize storage system. For more information about PFC, see “Priority Flow Control (IEEE standard 802.1 Qbb)” on page 15.

7.8.1 Requirements for PFC

PFC has the following requirements:

- ▶ PFC for iSCSI is supported only on Brocade VDX 10-Gigabit Ethernet (GbE) switches.
- ▶ PFC is supported only with ports that have 10 Gb or higher bandwidth.
- ▶ PFC cannot be manually turned on or off from the system. PFC is automatically turned on if it is configured on the switch (network).
- ▶ PFC is enabled only for those IP addresses that have VLAN tagging that is enabled on the system.

7.8.2 Configuring Priority Flow Control on Brocade VDX

For this example, two ports on the Brocade Ethernet switch are connected (9 and 10).

Example 7-14 shows a basic configuration that is done on the Brocade Ethernet switch to use as a simple reference. In this case, iSCSI traffic is assigned to COS 5, in priority-group-table 3 with 40 percent of bandwidth priority.

Example 7-14 Basic configuration on the Brocade Ethernet switch

```
sw0(config)# conf ter
sw0(config)# protocol lldp
sw0(config-lldp)# iscsi-priority 5
sw0(config-lldp)# exit
sw0(config)# interface TenGigabitEthernet 1/0/9-10
sw0(config-if-te-1/0/9-10)# lldp iscsi-priority 5
sw0(config-if-te-1/0/9-10)# exit
sw0(config)# cee-map default
sw0(config-cee-map-default)# priority-group-table 1 weight 30 pfc on
sw0(config-cee-map-default)# priority-group-table 2 weight 30 pfc off
sw0(config-cee-map-default)# priority-group-table 3 weight 40 pfc on
sw0(config-cee-map-default)# priority-table 2 2 2 1 2 3 2 15.0
sw0(config-lldp)# protocol lldp
sw0(config-lldp)# advertise dcbx-tlv
sw0(config-lldp)# advertise dcbx-iscsi-app-tlv
sw0(config-lldp)# exit
sw0(config)# exit
sw0# show running-config cee-map
cee-map default
```

```

precedence 1
priority-group-table 1 weight 30 pfc on
priority-group-table 15.0 pfc off
priority-group-table 15.1 pfc off
priority-group-table 15.2 pfc off
priority-group-table 15.3 pfc off
priority-group-table 15.4 pfc off
priority-group-table 15.5 pfc off
priority-group-table 15.6 pfc off
priority-group-table 15.7 pfc off
priority-group-table 2 weight 30 pfc off
priority-group-table 3 weight 40 pfc on
priority-table 2 2 2 1 2 3 2 15.0
remap fabric-priority priority 0
remap lossless-priority priority 0n

```

7.8.3 Verifying Priority Flow Control from the IBM Storwize storage system

From the IBM Storwize perspective, to start working with PFC, it is necessary to configure an IP address and VLAN tagging on each Ethernet port that is dedicated for iSCSI. To enable VLAN tagging, complete the following steps:

1. Click **Settings** → **Network** → **Ethernet Ports**, as shown in Figure 7-62.

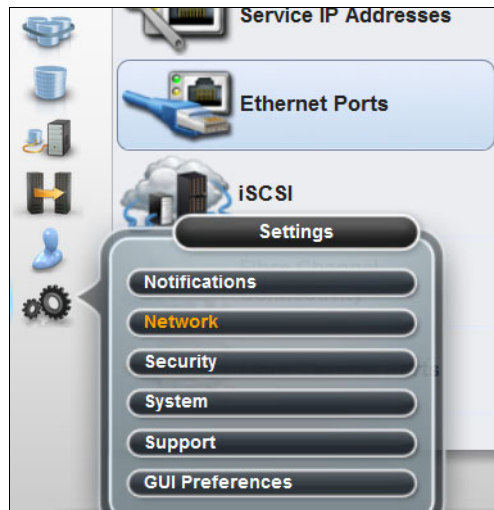


Figure 7-62 Step one to configure VLANs on an IBM Storwize storage system

2. Select the 10-Gbps port that is connected to the Brocade Switch for using iSCSI traffic, right-click, and select **Modify VLAN**, as shown in Figure 7-63.

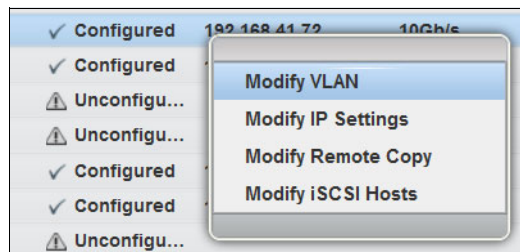


Figure 7-63 Modifying the VLAN

3. Select **Enable** and enter a VLAN tag, as shown in Figure 7-64. The same VLAN tag must be configured in the Brocade switch. In this example, it is 101.

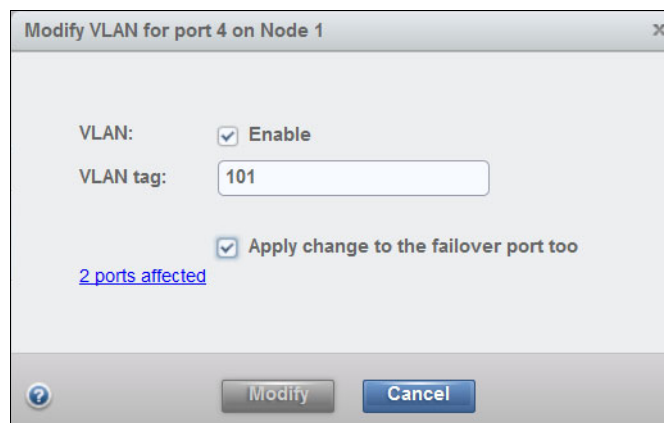


Figure 7-64 VLAN tag configuration

In the IBM Storwize storage system, the following values for each 10 Gbps Ethernet port that is dedicated for iSCSI should automatically change from No to Yes (Lossless) in the Priority Flow Control for IPv4 Column, as shown Figure 7-65.

Speed	Priority Flow Control for IPv4
1Gb/s	
1Gb/s	No
1Gb/s	No
1Gb/s	
1Gb/s	No
1Gb/s	No
10Gb/s	Yes (Lossless)
10Gb/s	Yes (Lossless)

Figure 7-65 Priority Flow Control enabled

The Brocade Ethernet Switch and the IBM Storwize storage system use the DCBX protocol for exchanging the configuration settings. Because they use it, it is possible to see the configured bandwidth allocation for each type of traffic from the IBM Storwize storage system by using the CLI and running **lsportip <Ethernet port>**. In this case, all the information is displayed for port 4, as shown in Example 7-15.

Example 7-15 The *lsportip* 4 output

```
IBM_IBM Storwize:Cluster_9.113.57.226:superuser>lsportip 4
id 4
node_id 1
node_name node1
IP_address 192.168.41.72
mask 255.255.255.0
gateway 192.168.41.1
IP_address_6
prefix_6
gateway_6
MAC 40:f2:e9:e0:01:af
```

```
duplex Full
state configured
speed 10Gb/s
failover no
mtu 1500
link_state active
host yes
remote_copy 0
host_6
remote_copy_6 0
remote_copy_status
remote_copy_status_6
vlan 101
vlan_6
adapter_location 2
adapter_port_id 1
dcbx_state enabled
lossless_iscsi on
lossless_iscsi6
iscsi_priority_tag 5
fcoe_priority_tag 3
pfc_enabled_tags 3:5
pfc_disabled_tags 0:1:2:4:6:7
priority_group_0
priority_group_1 3
priority_group_2
priority_group_3 5
priority_group_4
priority_group_5
priority_group_6
priority_group_7
bandwidth_allocation 0:30:30:40:0:0:0:0
```

The relevant part to mention in Example 7-15 on page 150 is that iSCSI traffic was configured with COS 5 (iscsi_priority_tag 5). COS 5 was assigned to Priority Group 3 and this group has 40 percent of bandwidth priority configured (bandwidth_allocation 0:30:30:40:0:0:0:0). This last value is explained in Figure 7-66.

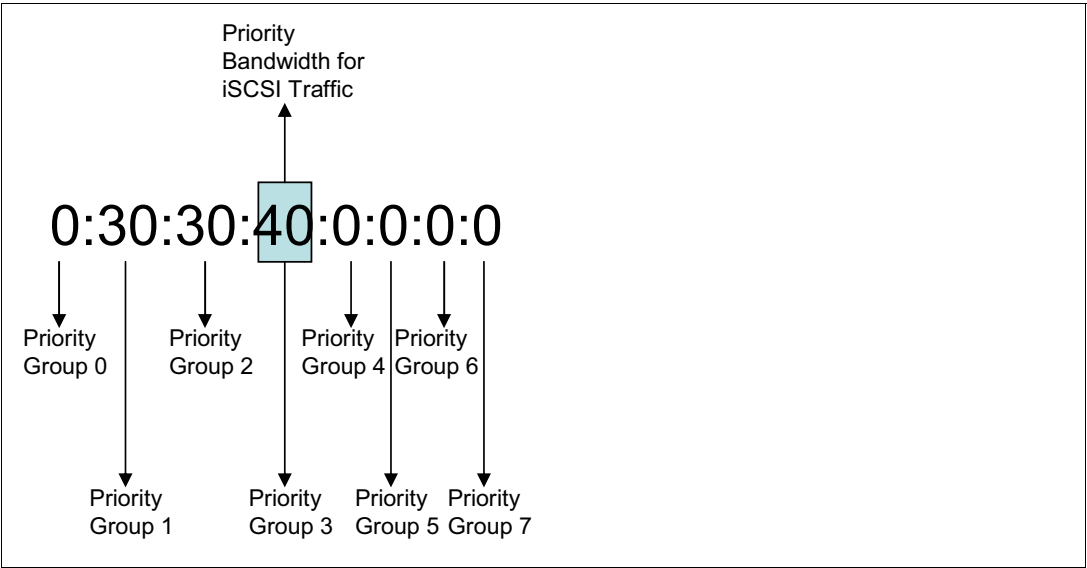


Figure 7-66 Configured priority bandwidth for iSCSI traffic for this example

PFC works by dividing the link into eight different virtual lanes (0 - 7). The ETS protocol is used to allocate each lane a specific link bandwidth for a particular traffic type. Then, each lane can be assigned to a specific priority group (also 0 - 7).

In the **bandwidth_allocation** parameter, each priority group is separated by the character “:”. In this example, the fourth field, which corresponds to priority group 3, has the iSCSI traffic that is assigned and with a 40 percent of bandwidth priority configured.

7.9 Configuring the iSCSI host for the HyperSwap cluster

This section provides information about HyperSwap configuration, iSCSI host site assignment, HyperSwap volume creation, and online conversion of iSCSI host-attached basic volume to HyperSwap volume.

7.9.1 What HyperSwap is

The HyperSwap high availability feature in the IBM SAN Volume Controller software enables continuous data availability if there is a hardware failure, power failure, connectivity failure, or disaster, such as fire or flooding. HyperSwap volumes are highly available and are accessible through two sites at up to 300 km (186.4 miles) apart. A fully independent copy of the data is maintained at each site. At the time of writing to the volume, the data is written on both the sites in synchronous fashion, which provides independent, synchronized copies of HyperSwapped volume at both the sites.

The HyperSwap function builds on two existing technologies in the product:

1. The Non-disruptive Volume Move (NDVM) function that was introduced in Version 6.4 of the SAN Volume Controller software and extended for iSCSI volumes in Version 7.7.1
2. Remote Copy features that include Metro Mirror, Global Mirror, and Global Mirror with Change Volumes

In the typical SAN Volume Controller cluster implementations, the volumes are accessed by a host through a caching I/O group only. In a HyperSwap feature, each volume must be accessed by the hosts that use all the I/O Groups in the clustered system. Therefore, this design needs multi-I/O group access to present a volume from multiple I/O groups. IBM Spectrum Virtualize software supports multi-I/O group volume access of iSCSI-attached volumes from Version 7.7.1. This enables creation and conversion of an iSCSI host-attached normal volume to HyperSwap volumes.

A new Metro Mirror capability, the active-active Metro Mirror, is used to maintain a fully independent copy of the data at each site. When data is written by hosts at either site, both copies are synchronously updated before the write operation is completed. The HyperSwap function automatically optimizes itself to minimize the data that is transmitted between sites and to minimize host read and write latency.

The HyperSwap function works with the standard multipathing drivers that are available on a wide variety of host types, with no additional required host support to access the highly available volume. Where multipathing drivers support Asymmetric Logical Unit Access (ALUA), the storage system informs the multipathing driver of the nodes that are closest to it, and the nodes to use to minimize I/O latency.

7.9.2 Host site assignment

Before you create or convert the volume to a HyperSwap volume for an iSCSI-attached host, you must set up the site attribute for the host. The site attribute corresponds to a physical location that houses the physical objects of the system. In a client installation, the site attribute might correspond to a separate office, a different data center building, or different rooms or racked areas of a single data center that was planned for internal redundancy.

Sites 1, 2, and 3 can be renamed from the default name by using the **chsite** command, as shown in Example 7-16.

Example 7-16 The chsite command example

```
IBM_2145:Redbooks_cluster1:superuser>chsite -name RedbookCluster_site1 1
IBM_2145:Redbooks_cluster1:superuser>chsite -name RedbookCluster_site2 2
IBM_2145:Redbooks_cluster1:superuser>chsite -name RedbookCluster_site3 3
```

To list the sites, you can use the **lssite** command, as shown in Example 7-17.

Example 7-17 The lssite command example

```
IBM_2145:Redbooks_cluster1:superuser>lssite
id site_name
1 RedbookCluster_site1
2 RedbookCluster_site2
3 RedbookCluster_site3
```

Example 7-18 The `chhost` command example

You can also specify the site attribute when the host is created by using the **mkhost** command, as shown in Example 7-19.

Example 7-19 The mkhost command example

You can check the current site definition for the host objects by using the **lshost** command, as shown in Example 7-20.

Example 7-20 The lshost command view

[illegible]

This section describes the creation of HyperSwap volumes, and the online conversion of iSCSI-attached basic volumes to HyperSwap volumes. After you have an assigned site for the iSCSI host, you can create the HyperSwap volumes. You can convert the existing basic volumes to HyperSwap volumes in nondisruptive way after the host site assignment is done.

Creating the HyperSwap volume

After you configure the HyperSwap topology and host site assignments, complete one of the following actions to create the HyperSwap volume:

- If you are using the management GUI, use the Create Volumes wizard to create the HyperSwap volumes. Click **Volumes** → **Volumes** → **Create Volumes** and select **HyperSwap** to create HA volumes, as described in Figure 7-67.

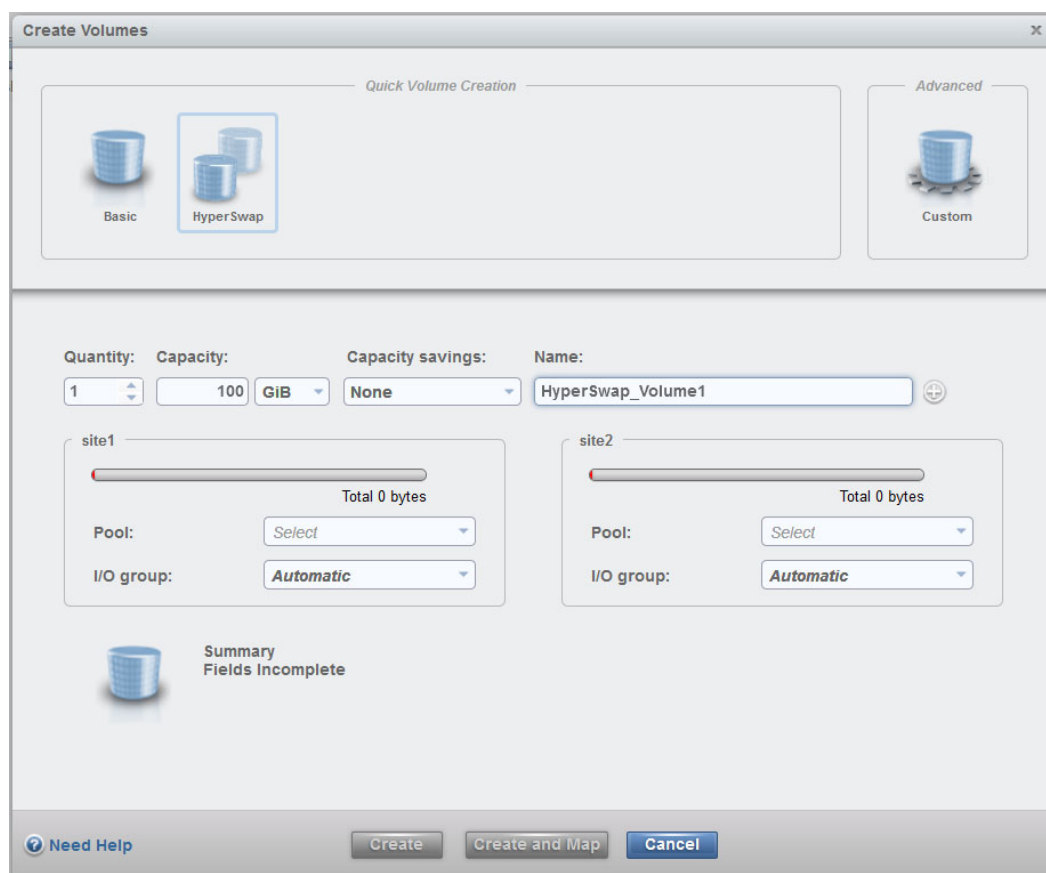


Figure 7-67 HyperSwap Volume creation

- If you are using the CLI, use the **mkvolume** command to create a HyperSwap volume. Example 7-21 shows a HyperSwap volume that is created by specifying two storage pools in independent sites.

Example 7-21 The **mkvolume** command example

```
IBM_2145:Redbooks_cluster1:superuser> mkvolume -size 100 -pool  
site1pool:site2pool  
Volume, id [0], successfully created.
```

Converting to a HyperSwap volume

This subsection provides information about the online conversion of basic iSCSI host-attached volumes to HyperSwap volumes after you upgrade to SAN Volume Controller Version 7.7.1 or later. You may convert the basic volume by adding another copy of the volume at a second site. Before you start the conversion process, check for the system topology to ensure that the system is in HyperSwap mode. Also, check for other prerequisites to create the HyperSwap volume at IBM Knowledge Center.

To convert a basic iSCSI host-attached volume to a HyperSwap volume, use one of the following options:

- To change a basic volume to a HyperSwap volume by using the management GUI, click **Volumes**. Right-click the basic volume and select **Add Volume Copy**. Select the pool at the different site that contains the existing copy, as shown in Figure 7-68. Click **Add**.

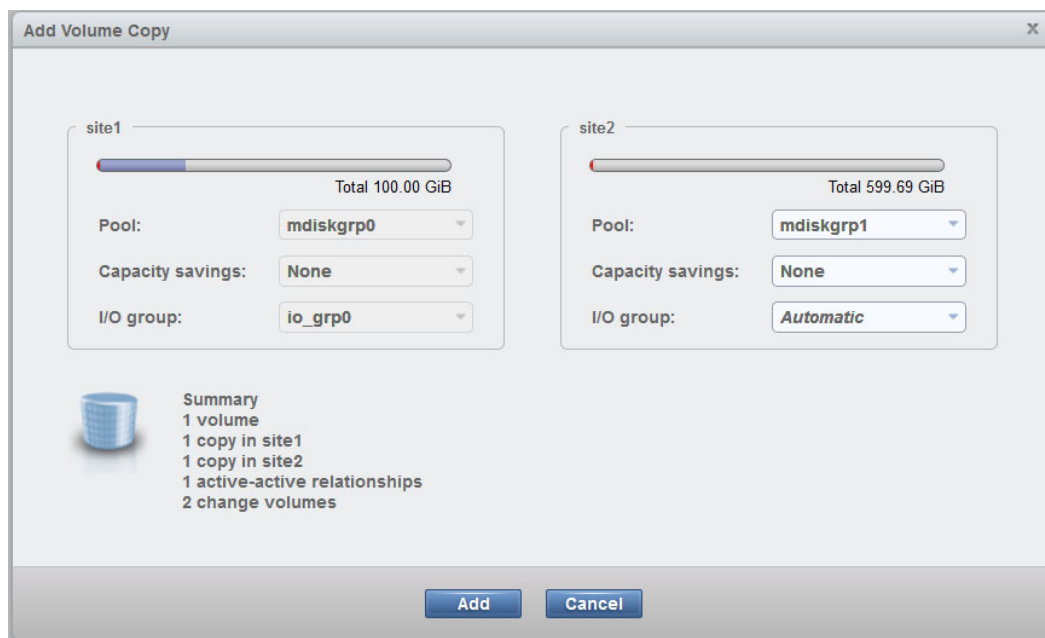


Figure 7-68 Converting a basic volume to a HyperSwap volume


This process creates three VDisks and the active-active relationship across the volume copies to make it a HyperSwapped volume. After the synchronization of the volume copy is complete, your basic volume is converted to a HyperSwap volume.

- The **addvolumecopy** command adds a copy to a HyperSwap volume, which changes a basic volume into a HyperSwap volume with copies on two separate sites. This command internally creates a volume copy on another site and an active-active relation is created between these volume copies. After the relationship is created, the initial data copy operation starts. Upon completion of initial data copy operation, the relationship shows the status as `consistent_synchronized`, and the conversion process is complete.

Example 7-22 shows the CLI example for converting a basic volume to a HyperSwap volume.

Example 7-22 The addvolumecopy command

```
IBM_2145:Redbooks_cluster1:superuser>
IBM_2145:Redbooks_cluster1:superuser> svctask addvolumecopy -pool 1 0
IBM_2145:Redbooks_cluster1:superuser>
```



IBM Spectrum Virtualize and IBM Storwize performance monitoring

This chapter provides a brief overview of the performance monitoring capabilities of the IBM Spectrum Virtualize and IBM Storwize storage systems.

However, it is beyond the scope of this book to provide an in-depth understanding of performance statistics, or explain how to interpret them. For a more comprehensive look at the performance of these systems, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

For IBM storage systems, the official IBM tool for the collection of performance statistics and to supply performance reporting is IBM Spectrum Control (formerly IBM Tivoli Storage Productivity Center). For more information, see 8.3, “Performance data collection with IBM tools” on page 166.

This chapter describes the following topics:

- ▶ 8.1, “Manually gathering performance statistics” on page 158
- ▶ 8.2, “Real-time performance monitoring” on page 159
- ▶ 8.3, “Performance data collection with IBM tools” on page 166

8.1 Manually gathering performance statistics

The IBM Spectrum Virtualize or Storwize storage system is constantly collecting performance statistics. The **lssystem** command shows the **statistics_status**. The default statistics frequency is 5 minutes, which you can adjust by using the **startstats -interval <minutes>** command.

The statistics data is collected in XML files, with a new file created at the end of each sampling period. Each node of a SAN Volume Controller system and each canister in a Storwize system keeps the most recent 16 files of each type. When the 17th file is created, the oldest file is overwritten.

This design provides statistics for the most recent 80-minute period if the default 5-minute sampling interval is used. You can define the sampling interval by running the **startstats -interval <minutes>** command to collect statistics at different intervals. The system supports user-defined sampling intervals of 1 - 60 minutes. Running the **startstats** command resets the statistics timer and gives it a new interval at which to sample.

8.1.1 Statistics file naming

The files that are generated are written to the `/dumps/iostats/` directory. The file name is in the following format:

- ▶ `Nm_stats_<node_serial_number>_<date>_<time>` for managed disk (MDisk) statistics
- ▶ `Nv_stats_<node_serial_number>_<date>_<time>` for virtual disk (VDisk) statistics
- ▶ `Nn_stats_<node_serial_number>_<date>_<time>` for node (or canister) statistics
- ▶ `Nd_stats_<node_serial_number>_<date>_<time>` for disk drive statistics

The `<node_serial_number>` is of the node or canister on which the statistics were collected. The date is in the format `<yymmdd>` and the time is in the format `<hhmmss>`. The following example shows an MDisk statistics file name:

`Nm_stats_7836640-2_140901_164012`

The **lsdumps -prefix /dumps/iostats** command lists the statistics file names, as shown in Example 8-1. The output is truncated and shows only part of the available statistics.

Example 8-1 The lsdump command output

```
IBM_IBM Storwize:Cluster_9.113.57.226:superuser>lsdumps -prefix /dumps/iostats
id  filename
0   Nd_stats_03C4003-2_150821_114831
1   Nm_stats_03C4003-2_150821_114831
2   Nn_stats_03C4003-2_150821_114831
3   Nv_stats_03C4003-2_150821_114831
4   Nd_stats_03C4003-2_150821_114931
5   Nv_stats_03C4003-2_150821_114931
6   Nn_stats_03C4003-2_150821_114931
7   Nm_stats_03C4003-2_150821_114931
8   Nn_stats_03C4003-2_150821_115031
9   Nv_stats_03C4003-2_150821_115031
10  Nd_stats_03C4003-2_150821_115031
```

To retrieve all the statistics files of a system, copy the files from the non-configuration nodes onto the configuration node by using the following command:

```
cpdumps -prefix /dumps/iostats <non_config node id>
```

Then, download the performance statistics files from the configuration node to a local drive on your workstation by using Secure Copy Protocol (SCP). On a Windows workstation, you can use the **pscp.exe** command (included with PuTTY), as shown in the following example:

```
C:\Program Files\PuTTY>pscp -unsafe -load ITS0_IBM  
Storwizeadmin@10.18.229.81:/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.

If you do not use IBM Spectrum Control (or third-party software), you must retrieve and parse these XML files to analyze the long-term statistics. The counters on the files are posted as absolute values. Therefore, the application that processes the performance statistics must compare two samples to calculate the differences from the two files. For the detailed description of the XML file contents, see [IBM Knowledge Center](#).

8.2 Real-time performance monitoring

Real-time performance statistics provide short-term status information for the IBM Storwize storage system. 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 processor use and the bandwidth of volumes, interfaces, and MDisk. Each graph displays the current bandwidth in either megabytes per second (MBps) or I/O 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 configuration 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:

- ▶ Central processing unit (CPU) uses
- ▶ Port use and I/O rates
- ▶ Volume and MDisk I/O rates
- ▶ Bandwidth
- ▶ Latency

Real-time statistics are always collected and cannot be stopped.

8.2.1 Real-time performance monitoring with the CLI

The following commands are available for monitoring the statistics through the CLI:

► **lssystemstats**

Run the **lssystemstats** command to display the most recent values of all of the node or node canister statistics in a clustered system. This command can also be used to display a history of values. You can filter the output to display values only for a subset of available statistics. An example **lssystemstats** CLI output is shown in Figure 8-1 on page 163.

► **lnodestats** and **lnodecanisterstats**

Run the **lnodestats** (IBM Spectrum Virtualize system) or **lnodecanisterstats** (IBM Storwize system) command to display the most recent values of statistics for all of the nodes or node canisters and display all statistics for a particular node canister. Additionally, you can use this command to display a history of values for a subset of available statistics. You can filter the output to display values only from certain nodes or a subset of available statistics.

Example 8-2 shows a **lssystemstats** command output.

Example 8-2 Example lssystemstats command output

```
IBM_2145:Redbooks_cluster1:superuser>lssystemstats
stat_name      stat_current  stat_peak  stat_peak_time
compression_cpu_pc 0           0          161117063047
cpu_pc         1           2          161117063022
fc_mb          0           0          161117063047
fc_io          5061        5354       161117062937
sas_mb         0           0          161117063047
sas_io         0           0          161117063047
iscsi_mb       0           0          161117063047
iscsi_io       0           0          161117063047
write_cache_pc 0           0          161117063047
total_cache_pc 0           0          161117063047
vdisk_mb       0           0          161117063047
vdisk_io       0           0          161117063047
vdisk_ms       0           0          161117063047
mdisk_mb       0           0          161117063047
mdisk_io       0           0          161117063047
mdisk_ms       0           0          161117063047
drive_mb       0           0          161117063047
drive_io       0           0          161117063047
drive_ms       0           0          161117063047
vdisk_r_mb     0           0          161117063047
vdisk_r_io     0           0          161117063047
vdisk_r_ms     0           0          161117063047
vdisk_w_mb     0           0          161117063047
vdisk_w_io     0           0          161117063047
vdisk_w_ms     0           0          161117063047
mdisk_r_mb     0           0          161117063047
mdisk_r_io     0           0          161117063047
mdisk_r_ms     0           0          161117063047
mdisk_w_mb     0           0          161117063047
mdisk_w_io     0           0          161117063047
mdisk_w_ms     0           0          161117063047
drive_r_mb     0           0          161117063047
```

drive_r_io	0	0	161117063047
drive_r_ms	0	0	161117063047
drive_w_mb	0	0	161117063047
drive_w_io	0	0	161117063047
drive_w_ms	0	0	161117063047
iplink_mb	0	0	161117063047
iplink_io	0	0	161117063047
iplink_comp_mb	0	0	161117063047
cloud_up_mb	0	0	161117063047
cloud_up_ms	0	0	161117063047
cloud_down_mb	0	0	161117063047
cloud_down_ms	0	0	161117063047

All three commands list the same set of statistics, but either represent all nodes in the cluster or a particular node (or node canister). The values for these statistics are calculated from the node statistics values in the following way:

Bandwidth	Sum of the bandwidth of all nodes.
Latency	Average latency for the cluster, which is calculated by using data from the whole cluster and not an average of the single node values.
IOPS	Total IOPS of all nodes.
CPU percentage	Average CPU percentage of all nodes.

Table 8-1 has a brief description of each of the statistics that are presented by the **lssystemstats**, **lsnodestats**, and **lsnodecanisterstat** commands.

Table 8-1 Field name descriptions for lssystemstats, lsnodestats, and lsnodecanisterstats statistics

Field name	Unit	Description
compression_cpu_pc	Percentage	Compression CPU use
cpu_pc	Percentage	Use of node CPUs
fc_mb	MBps	Fibre Channel (FC) bandwidth
fc_io	IOPS	FC throughput
sas_mb	MBps	SAS bandwidth
sas_io	IOPS	SAS throughput
iscsi_mb	MBps	iSCSI bandwidth
iscsi_io	IOPS	iSCSI throughput
write_cache_pc	Percentage	Write cache fullness (updated every ten seconds)
total_cache_pc	Percentage	Total cache fullness (updated every ten seconds)
vdisk_mb	MBps	Total VDisk bandwidth
vdisk_io	IOPS	Total VDisk throughput
vdisk_ms	Milliseconds (ms)	Average VDisk latency
mdisk_mb	MBps	MDisk (SAN and RAID) bandwidth
mdisk_io	IOPS	MDisk (SAN and RAID) throughput
mdisk_ms	Milliseconds	Average MDisk latency

Field name	Unit	Description
drive_mb	MBps	Drive bandwidth
drive_io	IOPS	Drive throughput
drive_ms	Milliseconds	Average drive latency
vdisk_w_mb	MBps	VDisk write bandwidth
vdisk_w_io	IOPS	VDisk write throughput
vdisk_w_ms	Milliseconds	Average VDisk write latency
mdisk_w_mb	MBps	MDisk (SAN and RAID) write bandwidth
mdisk_w_io	IOPS	MDisk (SAN and RAID) write throughput
mdisk_w_ms	Milliseconds	Average MDisk write latency
drive_w_mb	MBps	Drive write bandwidth
drive_w_io	IOPS	Drive write throughput
drive_w_ms	Milliseconds	Average drive write latency
vdisk_r_mb	MBps	VDisk read bandwidth
vdisk_r_io	IOPS	VDisk reach throughput
vdisk_r_ms	Milliseconds	Average VDisk read latency
vdisk_w_mb	MBps	VDisk write bandwidth
vdisk_w_io	IOPS	VDisk write throughput
vdisk_w_ms	Milliseconds	Average VDisk write latency
mdisk_r_mb	MBps	MDisk (SAN and RAID) read bandwidth
mdisk_r_io	IOPS	MDisk (SAN and RAID) read throughput
mdisk_r_ms	Milliseconds	Average MDisk read latency
mdisk_w_mb	MBps	MDisk (SAN and RAID) write bandwidth
mdisk_w_io	IOPS	MDisk (SAN and RAID) write throughput
mdisk_w_ms	Milliseconds	Average MDisk write latency
drive_r_mb	MBps	Drive read bandwidth
drive_r_io	IOPS	Drive read throughput
drive_r_ms	Milliseconds	Average drive read latency
drive_w_mb	MBps	Drive write bandwidth
drive_w_io	IOPS	Drive write throughput
drive_w_ms	Milliseconds	Average drive read latency
drive_w_mb	MBps	Drive write bandwidth
drive_w_io	IOPS	Drive write throughput
drive_w_ms	Milliseconds	Average drive write latency
power_w	Watts	Power consumption

Field name	Unit	Description
temp_c	Celsius	Ambient temperature in Celsius degrees
temp_f	Fahrenheit	Ambient temperature in Fahrenheit degrees
iplink_mb	MBps	Internet Protocol (IP) link Bandwidth
iplink_io	IOPS	IP link Throughput
iplink_comp_mb	MBps	IP link compressed Throughput
cloud_up_mb	MBps	Average transfer for cloud upload operations
cloud_up_ms	Milliseconds	Average response time for cloud upload requests
cloud_down_mb	MBps	Average transfer for cloud download operations
cloud_down_ms	Milliseconds	Average response time for cloud download requests

8.2.2 Real-time performance monitoring with the GUI

The real-time statistics are also available from the IBM Storwize GUI. To view them, click **Monitoring** → **Performance**, as shown in Figure 8-1.

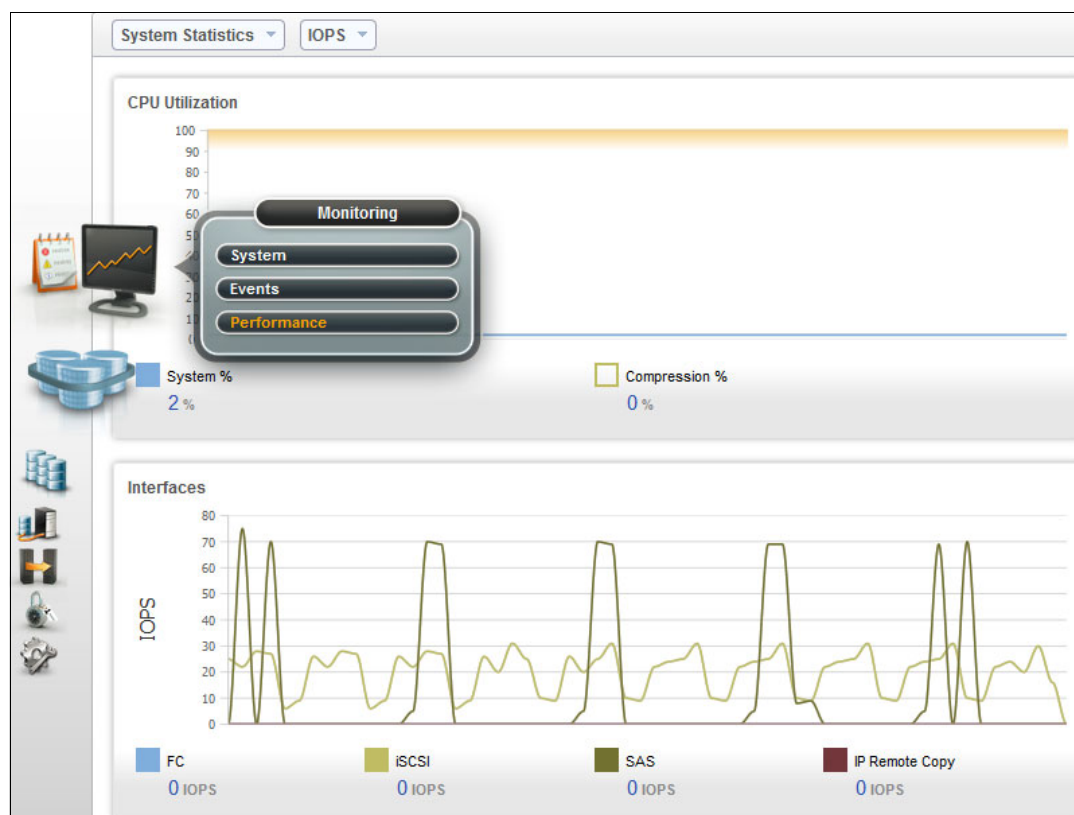


Figure 8-1 IBM Storwize Monitoring menu

The performance monitoring console is divided into four sections that provide use views for the following resources:

- ▶ CPU Use
 - CPU usage (percentage) for general tasks
 - Shows the CPU (percentage) usage for compression (when enabled)
- ▶ Volumes. This shows the overall volume statistics:
 - Read
 - Write
 - Read latency
 - Write latency
- ▶ Interfaces. This shows the overall statistics for each of the available interfaces:
 - FC
 - iSCSI
 - Serial-attached SCSI (SAS)
 - IP Remote Copy
 - IP Remote Copy Compressed
- ▶ MDisk. This shows the following overall statistics for the MDisks:
 - Read
 - Write
 - Read latency
 - Write latency

Figure 8-2 shows real-time performance graphs.



Figure 8-2 Real-time performance graphs

Each graph represents 5 minutes of collected statistics and provides a means of assessing the overall performance of your system. You can select to view performance statistics for each of the available nodes or canisters of the system, as shown in Figure 8-3.

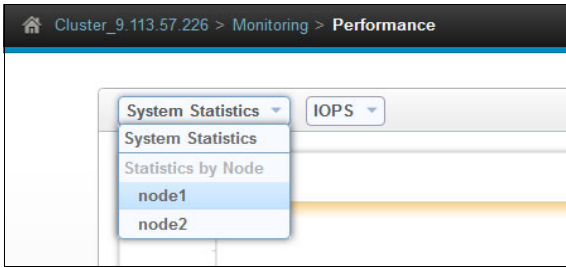


Figure 8-3 Selecting a system node (canister) for System Statistics

It is also possible to change the metric between MBps or IOPS, as shown in Figure 8-4.

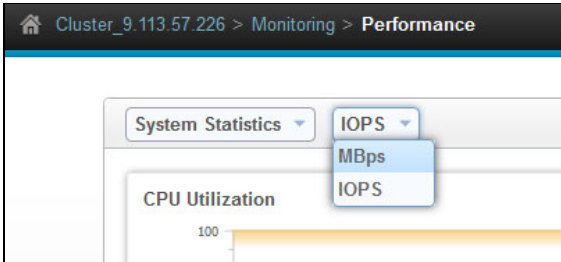


Figure 8-4 Changing a metric to MBps

On any of these views, you can select any point with your cursor to determine the exact 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 8-5.

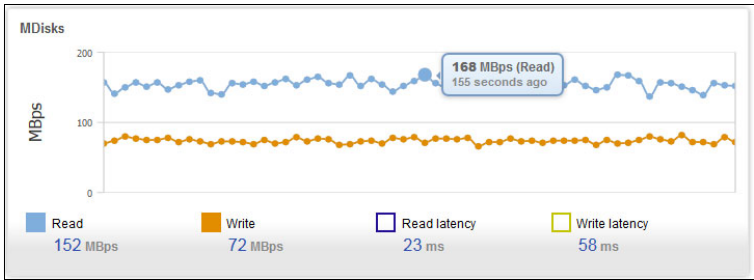


Figure 8-5 Detailed resource use information in a graph

There are various values for each of these resources that you can view by selecting the check box next to a value. For example, for the MDisk view that is shown in Figure 8-6, the four available fields are selected.

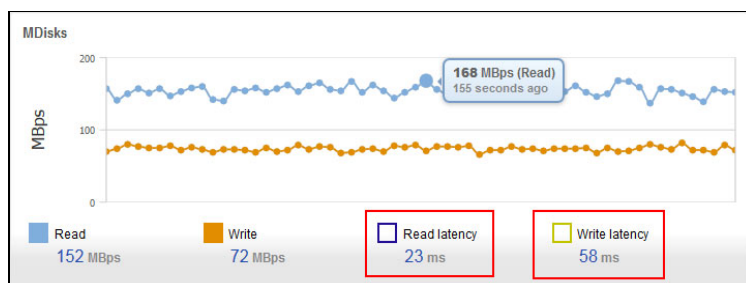


Figure 8-6 Detailed resource use information in the graph

The following detailed resource use information is available in the graph:

- ▶ Read
- ▶ Write
- ▶ Read latency
- ▶ Write latency

8.3 Performance data collection with IBM tools

As explained in 8.1, “Manually gathering performance statistics” on page 158, you can obtain performance reports on an IBM Storwize storage system in standard XML format. However, using .xml files is often an impractical and complex method to analyze the IBM Storwize performance statistics. IBM provides several tools that can be used to view performance information.

8.3.1 IBM Spectrum Control

IBM Spectrum Control is the IBM tool to collect and analyze performance statistics.

For more information about using IBM Spectrum Control to monitor your storage subsystem, see the following resources:

- ▶ *IBM Spectrum Family: IBM Spectrum Control Standard Edition*, SG24-8321
- ▶ [IBM Knowledge Center](#)

8.3.2 IBM Spectrum Control Storage Insights

IBM Spectrum Control Storage Insights is a SaaS offering with its core running over IBM SoftLayer®. IBM Spectrum Control Storage Insight® capabilities can enhance your storage performance as you lower the cost of storage. This solution is cloud-based, so you can get actionable insights in minutes with deeper insights delivered over time as intelligence about your environment builds in the system. The IT workload for maintaining a storage management infrastructure disappears, which enables you to focus on implementing the insights to optimize your storage environment.

The solution is oriented to small and medium businesses that want to avoid the expensive and time-consuming deployment of an on-premises solution and enable less experienced staff to manage storage environment more efficiently by delivering different insights as simple as looking at the many available dashboards. Large organizations can also deploy IBM Spectrum Control Storage Insights to gain visibility of small storage environments, even if they are already using on-premises solutions such as IBM Virtual Storage Center to manage their core storage systems.

For more information about IBM Spectrum Control Storage Insights to view data for your storage subsystem, see the following information:

- ▶ *Regain Control of your Environment with IBM Storage Insights*, REDP-5231
- ▶ [IBM Spectrum Control Storage Insights](#)



IBM Spectrum Virtualize and IBM Storwize storage systems on the OpenStack platform

This chapter introduces OpenStack and its components. It describes details about how to configure a Cinder driver to integrate an IBM Spectrum Virtualize or IBM Storwize storage system as block storage for OpenStack deployments.

This chapter describes the following topics:

- ▶ 9.1, “Introduction to OpenStack components” on page 170
- ▶ 9.2, “Integrating the Cinder driver with IBM Spectrum Virtualize and IBM Storwize storage systems” on page 171

9.1 Introduction to OpenStack components

OpenStack is a no cost and open source software platform for cloud computing. Started by Rackspace Hosting and NASA in 2010, it is now a global collaboration of more than 500 companies. The initiative is managed by the OpenStack Foundation, a non-profit corporate entity that was established in September 2012 to promote OpenStack software and its community. IBM contributes to OpenStack as one of the platinum members.

Different cloud solutions for various different service needs are adopted by different kind of organizations, ranging from small startups to established organizations. As a cloud service platform, OpenStack helps reduce major challenges that are faced by clients with their existing dedicated IT infrastructure and also helps every organization to become more agile and flexible. For more information about other OpenStack storage and cloud deployment options from IBM, see *IBM Private, Public, and Hybrid Cloud Storage Solutions*, REDP-4873.

The mission of the OpenStack initiative is to create a ubiquitous open source cloud computing platform that is simple to implement and massively scalable. OpenStack enables users to place requests for required resources through a self-service portal and use those resources in real time as needed.

OpenStack has a modular architecture with various code names for its components that provide different services. The following core components or services are the minimum components that are required to run an OpenStack system:

Nova	Provides the compute / virtual machine (VM) management services to provision and manage VMs for OpenStack.
Keystone	Provides authentication, Token, Catalog, and Policy services for OpenStack.
Glance	Provides services to store VM images and maintain a catalog of available images.
Neutron	Provides network services for device interfaces in OpenStack.
Cinder	Provides block storage services for use with OpenStack compute instances.
Swift	Provide object storage services that allow users to store much data efficiently and safely.

For more information about OpenStack components, see the [OpenStack documentation](#).

Cinder manages the creation, attachment, and detachment of the block devices to servers. A server can access block storage through different protocols, such as Fibre Channel (FC), Fibre Channel over Ethernet (FCoE), or iSCSI from storage systems that are controlled by Cinder. IBM provides Cinder drivers for IBM Spectrum Virtualize and IBM Storwize systems for FC and iSCSI. For the list of Cinder features that are supported by these drivers, see the [Cinder support matrix](#).

Section 9.2, “Integrating the Cinder driver with IBM Spectrum Virtualize and IBM Storwize storage systems” on page 171 explains how to integrate an IBM Storwize storage system with OpenStack Cinder.

9.2 Integrating the Cinder driver with IBM Spectrum Virtualize and IBM Storwize storage systems

Cinder enables access to persistent block storage for compute instances. The Cinder driver manages the creation, attachment, and detachment of volumes to the compute resources. It is responsible for the control path only. The I/O to the devices runs directly from the compute resources and is mounted over a controlled protocol (for example, iSCSI). Users can request storage resources by using an API. The communication from Cinder to the storage and the Nova compute nodes are part of management or control I/O. Figure 9-1 indicates the data flow from Cinder to Nova and storage controllers.

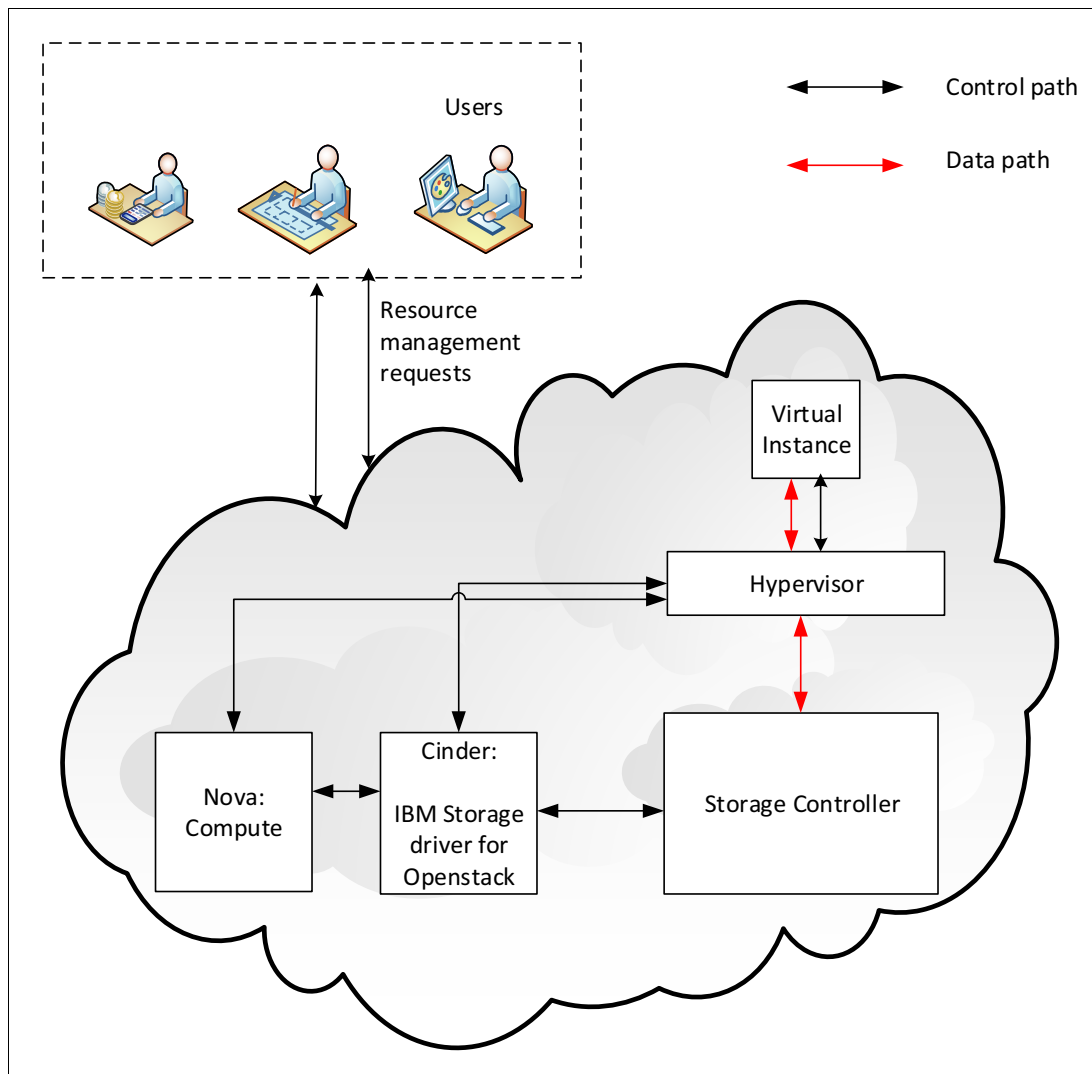


Figure 9-1 Cinder communication with Nova, compute, and storage controllers

Note: This publication focuses on the Newton release of Open Stack. For other versions, see the [OpenStack documentation](#).

Integration of storage systems requires an OpenStack Block Storage driver on the OpenStack Cinder nodes. For IBM Spectrum Virtualize and IBM Storwize systems, that is the *IBM Storwize family and SAN Volume Controller Driver for OpenStack*. The driver is an IBM proprietary solution that supports OpenStack block storage on top of the OpenStack and Cinder open source technologies.

The Cinder driver configuration file provides details about storage controllers that are available for the compute instances. The Cinder configuration file is by default created in `/etc/cinder/cinder.conf`. The minimum required parameters for IBM Storwize storage system to Cinder integration are listed in Example 9-1.

Example 9-1 Cinder configuration with minimum configuration

```
volume_driver = cinder.volume.drivers.ibm.storwize_svc.storwize_svc_iscsi.StorwizeSVCISCSIDriver
san_ip = 9.113.57.226
san_login = superuser
san_password = passw0rd
storwize_svc_volpool_name = mdiskgrp0
storwize_svc_iscsi_chap_enabled=True
volume_backend_name = svc1
```

Table 9-1 lists the minimum required parameters and some important optional parameters.

Table 9-1 Cinder configuration file parameters and descriptions

Configuration parameters or flags	Description
volume_driver	Tells the volume driver which IBM Storwize family and SAN Volume Controller driver to use.
san_ip	IBM Spectrum Virtualize or IBM Storwize system management address (or host name).
san_login	Management login user name.
san_password	Management login password.
san_private_key	Management login SSH private key ^a .
storwize_svc_volpool_name	Name of the pool where to create volumes.
volume_backend_name	Name that is used by Cinder to identify this storage system among multiple storage back-ends. ^b Optional, no default.
storwize_svc_vol_iogrp	ID of the IO group for new volumes. Optional, default value is 0.
storwize_svc_connection_protocol	Connection protocol to use. Optional, currently supports iSCSI or FC, default value is iSCSI.
storwize_svc_iscsi_chap_enabled	Enable CHAP authentication for iSCSI connections. Optional, default value is True.

a. Authentication requires either a password or an SSH private key. One must be specified. If both are specified, the driver uses only the SSH private key.

b. Multiple back-end storage systems can serve the same OpenStack Compute configuration. At volume creation, the Cinder scheduler uses filters to decide in which back end the volume is created. For more information, see the [OpenStack documentation](#).

By default, the driver creates non-compressed, thin-provisioned volumes with a grain size of 256 KB and auto-expansion enabled. This behavior can be modified by optional parameters in the Cinder configuration file. For these and other extra parameters that can be used, see [IBM Storwize Family and SAN Volume Controller Driver Options in Cinder](#).

This website explains also how to integrate IBM Spectrum Virtualize or IBM Storwize remote mirroring (called *back-end storage replication* in Cinder terminology).

After editing the following Cinder configuration file, you must restart the Cinder service:

```
/root/scripts/restartCinder.sh
```

9.2.1 Volume creation and host attachment with OpenStack

After the minimum configurations are done along with required network connections, it is possible to create and manage volumes from Cinder and Nova nodes. Example 9-2 shows how to create the volume from the Cinder node and list the volume to view the status of the volume that is created. The syntax for command for the **cinder create** command is:

```
cinder create --name <volume_name> <size_in_Gb>
```

Example 9-2 Cinder create and list example

```
$ cinder create --name volume1 1
```

Property	Value
attachments	[]
availability_zone	nova
bootable	false
consistencygroup_id	None
created_at	2015-10-30T08:45:39.000000
description	None
encrypted	False
id	714fa399-eb23-4fd0-bc31-8267cc058cc5
metadata	{}
migration_status	None
multiattach	False
name	volume1
os-vol-host-attr:host	dig_openstack@SVCdriver#svcdriver
os-vol-mig-status-attr:migstat	None
os-vol-mig-status-attr:name_id	None
os-vol-tenant-attr:tenant_id	1f288d01f0a64d0fb4140e65e6409ee6
os-volume-replication:driver_data	None
os-volume-replication:extended_status	None
replication_status	disabled
size	1
snapshot_id	None
source_volid	None
status	creating
updated_at	2015-10-30T08:45:39.000000
user_id	6220597255ff46229d80ceb64fac8985
volume_type	svcdriver

```
$ cinder list
```

ID	Status	Display Name	Size	Volume Type	Bootable	Attached to
714fa399-eb23-4fd0-bc31-8267cc058cc5	available	Volume1	1	svcdriver	false	

9.2.2 Volume attachment from Nova

After the volumes are created from the Cinder driver, volumes attachment must be done from the Nova component. Example 9-3 shows how to attach the volume to the host with the parameters that are mentioned in the Cinder configuration file. The syntax of the **Nova volume attach** command is:

```
nova volume-attach <INSTANCE_ID> <VOLUME_ID> [<DEVICE>]
```

The optional parameter *<DEVICE>* is either a device file for the target system (for example, */dev/vdb*) or the default value *auto*. For more information, see [Attach a Volume to an Instance](#).

Example 9-3 Volume attachment from Nova

```
nova volume-attach e10a856d-a66f-486c-95a6-3e5f9688b7f0
714fa399-eb23-4fd0-bc31-8267cc058cc5
```

Property	Value
device	/dev/vdb
id	714fa399-eb23-4fd0-bc31-8267cc058cc5
serverId	e10a856d-a66f-486c-95a6-3e5f9688b7f0
volumeId	714fa399-eb23-4fd0-bc31-8267cc058cc5

For all the OpenStack configurations that are done from OpenStack control nodes, you can use the audit logs from the IBM Storwize storage system to identify the command that was triggered from OpenStack, as shown in Example 9-4.

Example 9-4 Audit logs from an IBM Storwize storage system indicating commands that are triggered from OpenStack

399	151027093451	openstack	9.122.121.65	0	114	svctask mkvdisk -name volume-c74daee9-1045-42fb-b88d-6fada39e82f1
-iogrp 0	-mdiskgrp	mdiskgrp0	-size 1	-unit gb	-autoexpand	-grainsize 256 -rsize 2% -warning 0% -easytier on
400	151027093832	openstack	9.122.121.65	0	115	svctask mkvdisk -name volume-936e737d-f23b-482b-b8f6-0c9cb4f9e847
-iogrp 0	-mdiskgrp	mdiskgrp0	-size 1	-unit gb	-autoexpand	-grainsize 256 -rsize 2% -warning 0% -easytier on
401	151027094527	openstack	9.122.121.65	0	116	svctask mkvdisk -name volume-289696ce-f44e-4ac5-8483-bd9f7d9c2532
-iogrp 0	-mdiskgrp	mdiskgrp0	-size 1	-unit gb	-autoexpand	-grainsize 256 -rsize 2% -warning 0% -easytier on
402	151027094840	openstack	9.122.121.65	0	6	svctask mkhost -name dig_openstack-74177470 -force -iscsiname
iqn.1994-05.com.redhat:8d096bbc0c7						
403	151027094840	openstack	9.122.121.65	0		svctask chhost -chapsecret AtgnZEYY7n5Ha3d 6
404	151027094841	openstack	9.122.121.65	0		svctask mkvdiskhostmap -host dig_openstack-74177470 -scsi 0 116



Troubleshooting

This chapter introduces the various interfaces that are available for troubleshooting on the IBM Storwize storage system. It also describes some procedures to determine the cause of problems when you are configuring or using Internet Small Computer System Interface (iSCSI) on an IBM SAN Volume Controller storage system and an IBM Storwize storage system.

It provides examples that explain how to use the management GUI, Service Assistant GUI, command-line interface (CLI), Service CLI, and other diagnostic and monitoring tools. It also includes descriptions of advanced techniques to retrieve logs that are not normally accessible through the management GUI.

This chapter describes the following topics:

- ▶ 10.1, “Storage tools on an IBM Storwize storage system” on page 176
- ▶ 10.2, “Storage logs that are used for analysis” on page 191
- ▶ 10.3, “Different IP addresses on the IBM Storwize storage system” on page 196
- ▶ 10.4, “Problem determination” on page 198

10.1 Storage tools on an IBM Storwize storage system

You can select from many user management interfaces to check the status of the IBM Storwize storage system.

10.1.1 Management GUI

To access the management GUI, open a web browser and go to the cluster IP address:

`https://<Cluster IP>`

Figure 10-1 shows the login window of the management GUI.



Figure 10-1 Management GUI login window

Node properties

It is possible to check some basic properties of a node by hovering your cursor over the GUI image at the rear view of the IBM Storwize canister. This method is a quick way to check which node is the configuration node.

Figure 10-2 shows the node properties that are displayed when you hover your cursor over the node image in the management GUI.

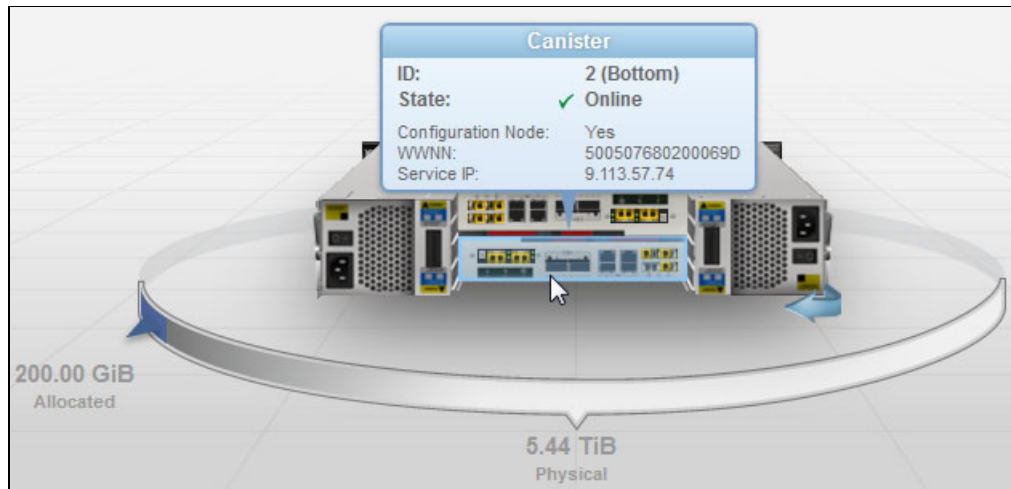


Figure 10-2 Node status and properties in the management GUI

Health Status bar

The Health Status bar provides the initial indication of SAN Volume Controller or IBM Storwize storage system status from the management GUI. The status can be one of the following colors:

- ▶ Green: Healthy
- ▶ Amber: Degraded
- ▶ Red: Unhealthy

Figure 10-3 demonstrates the error alerting and health status in the management GUI.

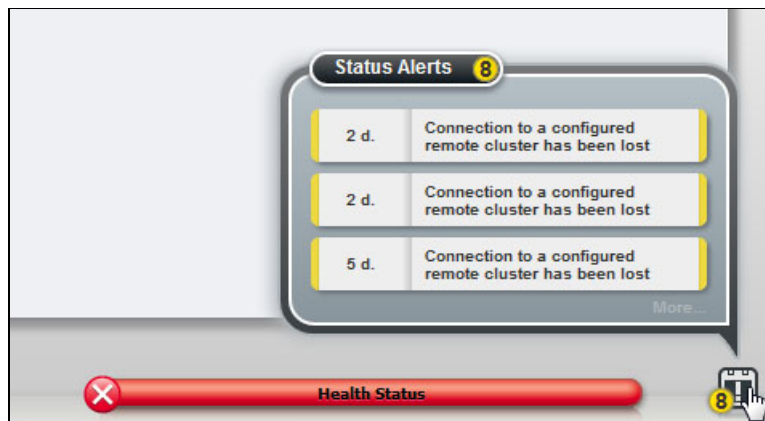


Figure 10-3 Health status in the management GUI

Event Log Viewer

To view errors, click **Monitoring** → **Event Log**.

Figure 10-4 shows the Event Log Viewer from the management GUI.

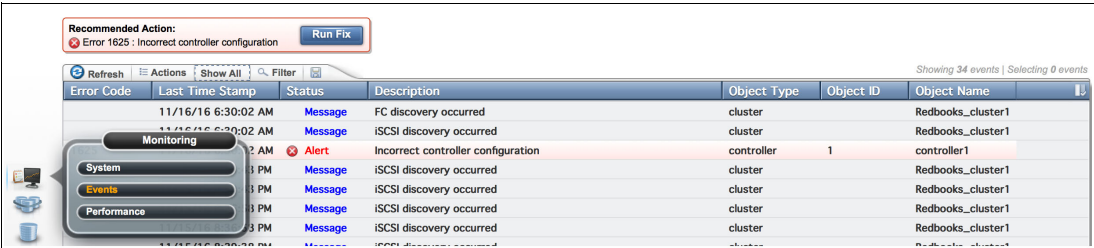


Figure 10-4 Event Log Viewer in the management GUI

Figure 10-5 explains the entries in the Status column.

Status	Description
Alert	This event requires attention. Follow the fix procedure or service action to resolve.
Alert	This event has already been fixed.
Expired	This event no longer represents a concern.
Monitoring	This event is not yet of concern.
Message	This event provides useful information about system activity.

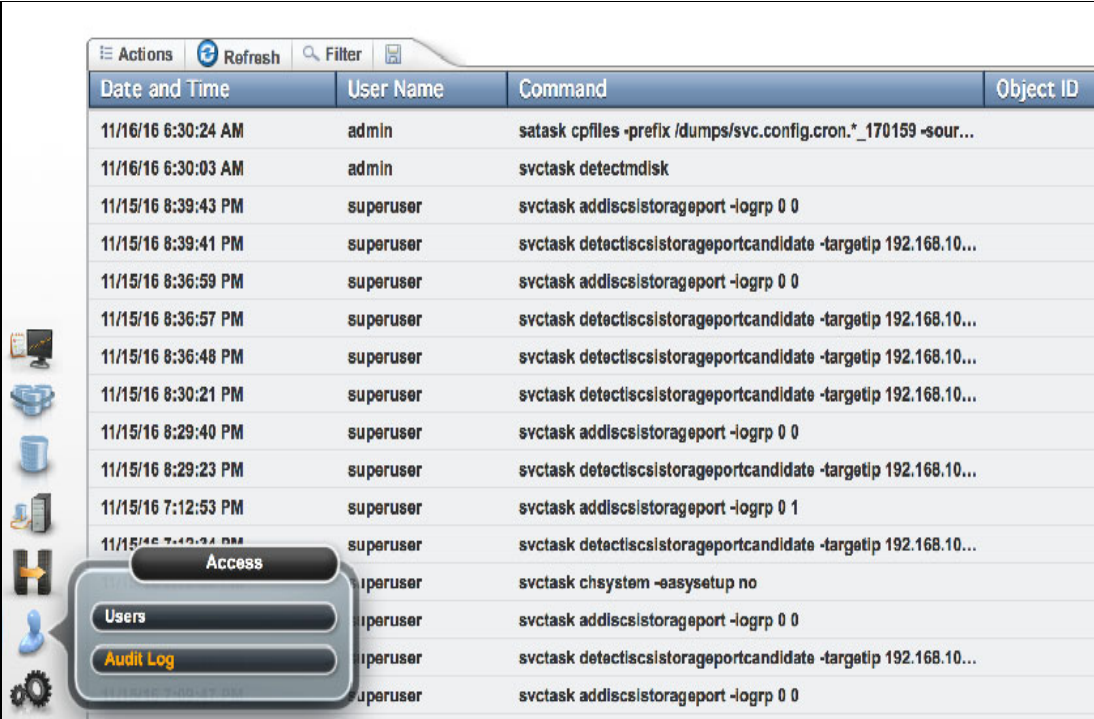
Figure 10-5 GUI event viewer status

If unfixed errors are detected, a separate Recommended Actions window opens above the event list, as shown in Figure 10-4. Click **Run Fix** to run the automatic fix procedure in the recommended order.

Audit Log

To view the audit log, click **Access** → **Audit Log**.

Figure 10-6 shows the Audit Log viewer from the management GUI.



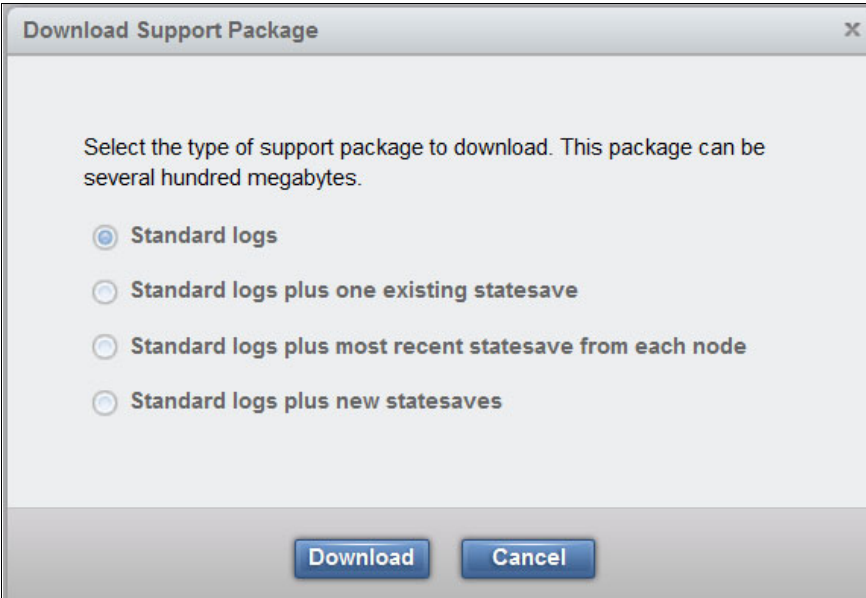
Date and Time	User Name	Command	Object ID
11/16/16 6:30:24 AM	admin	satask cpfiles -prefix /dumps/svc.config.cron.*_170159 -sour...	
11/16/16 6:30:03 AM	admin	svctask detectmdisk	
11/15/16 8:39:43 PM	superuser	svctask adddiscsistorageport -logrp 0 0	
11/15/16 8:39:41 PM	superuser	svctask detectdiscsistorageportcandidate -targetip 192.168.10...	
11/15/16 8:36:59 PM	superuser	svctask adddiscsistorageport -logrp 0 0	
11/15/16 8:36:57 PM	superuser	svctask detectdiscsistorageportcandidate -targetip 192.168.10...	
11/15/16 8:36:48 PM	superuser	svctask detectdiscsistorageportcandidate -targetip 192.168.10...	
11/15/16 8:30:21 PM	superuser	svctask detectdiscsistorageportcandidate -targetip 192.168.10...	
11/15/16 8:29:40 PM	superuser	svctask adddiscsistorageport -logrp 0 0	
11/15/16 8:29:23 PM	superuser	svctask detectdiscsistorageportcandidate -targetip 192.168.10...	
11/15/16 7:12:53 PM	superuser	svctask adddiscsistorageport -logrp 0 1	
11/15/16 7:12:34 PM	superuser	svctask detectdiscsistorageportcandidate -targetip 192.168.10...	
	superuser	svctask chsystem -easyssetup no	
	superuser	svctask adddiscsistorageport -logrp 0 0	
	superuser	svctask detectdiscsistorageportcandidate -targetip 192.168.10...	
	superuser	svctask adddiscsistorageport -logrp 0 0	

Figure 10-6 Audit Log viewer in the management GUI

Downloading a support package

You can capture a data collection bundle from the management GUI.

Figure 10-7 shows the four support package options.



Download Support Package

Select the type of support package to download. This package can be several hundred megabytes.

☒ Standard logs

☐ Standard logs plus one existing statesave

☐ Standard logs plus most recent statesave from each node

☐ Standard logs plus new statesaves

Download Cancel

Figure 10-7 Download Support Package from the management GUI

The type of support package that is required for problem analysis might vary depending on the nature of the fault. IBM Support Center published a guide to help determine the type of support package to collect for your support request, which is found at [What Data Should You Collect for a Problem on SVC or Storwize Systems](#).

If you have any questions regarding which support package to download, contact IBM Support Center.

Showing a full log listing

It is sometimes more relevant to download individual files from the SAN Volume Controller or IBM Storwize cluster. The management GUI Download Support Package menu also includes an option to select individual files from the configuration node.

Figure 10-8 shows the file list that is displayed when you select the **Show full log listing** option from the Download Support Package window.

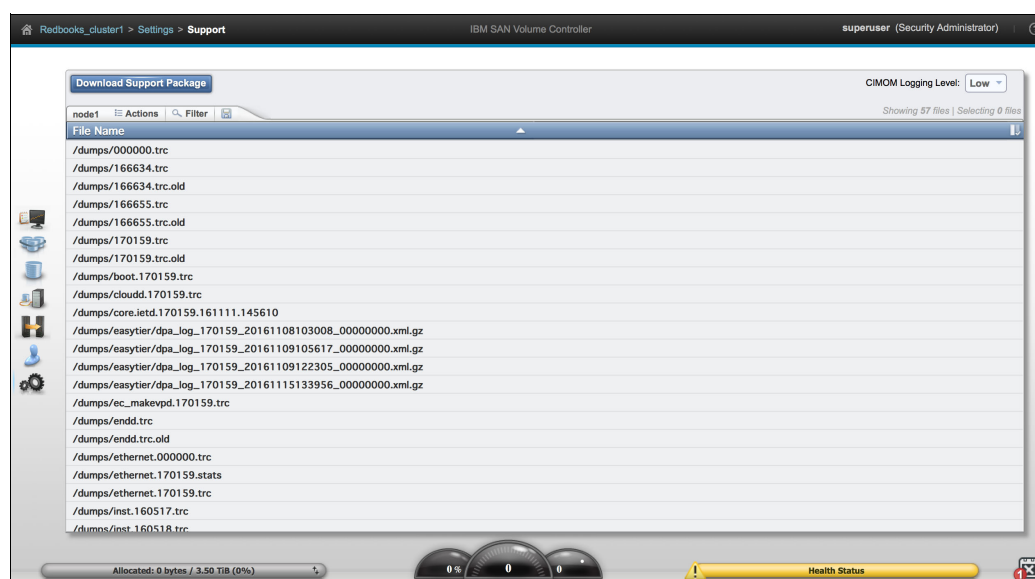


Figure 10-8 Showing the full log listing option in the management GUI

Setting notifications

Notifications are selected from the Settings menu in the management GUI. There are three menu options in the Notifications window:

- ▶ Email. This setting enables event notifications to be forwarded by email by using Simple Mail Transfer Protocol (SMTP). Call Home can also be enabled so that critical faults generate a problem management record (PMR) that is then sent directly to the appropriate IBM Support Center.
- ▶ SNMP. If Simple Network Management Protocol (SNMP) is enabled, an SNMP trap is sent to an SNMP manager when a new event is generated.
- ▶ Syslog. Log messages can be forwarded on an IP network by using the syslog protocol.

Figure 10-9 shows the Notifications window in the management GUI.

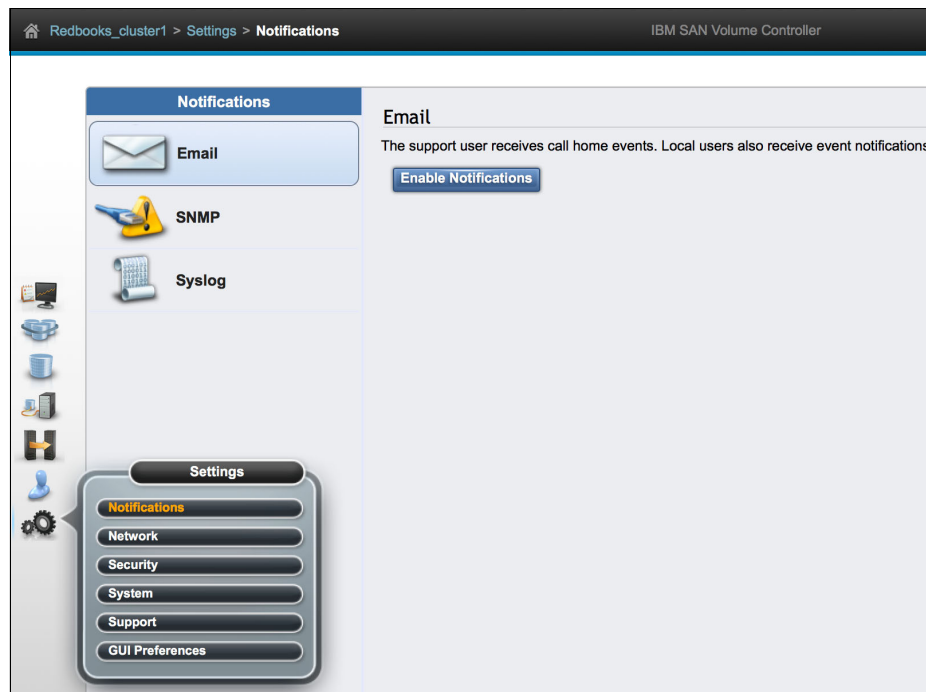


Figure 10-9 Notification window in the management GUI

10.1.2 Service Assistant GUI

The Service Assistant GUI allows the user to access low-level diagnostic and administrative functions by using a direct connection to a selected node instead of only the configuration node by using the cluster IP address.

To access this tool, open a web browser and go to the following URL:

`https://<Service IP>/service`

Figure 10-10 shows the home menu in the Service Assistant GUI.

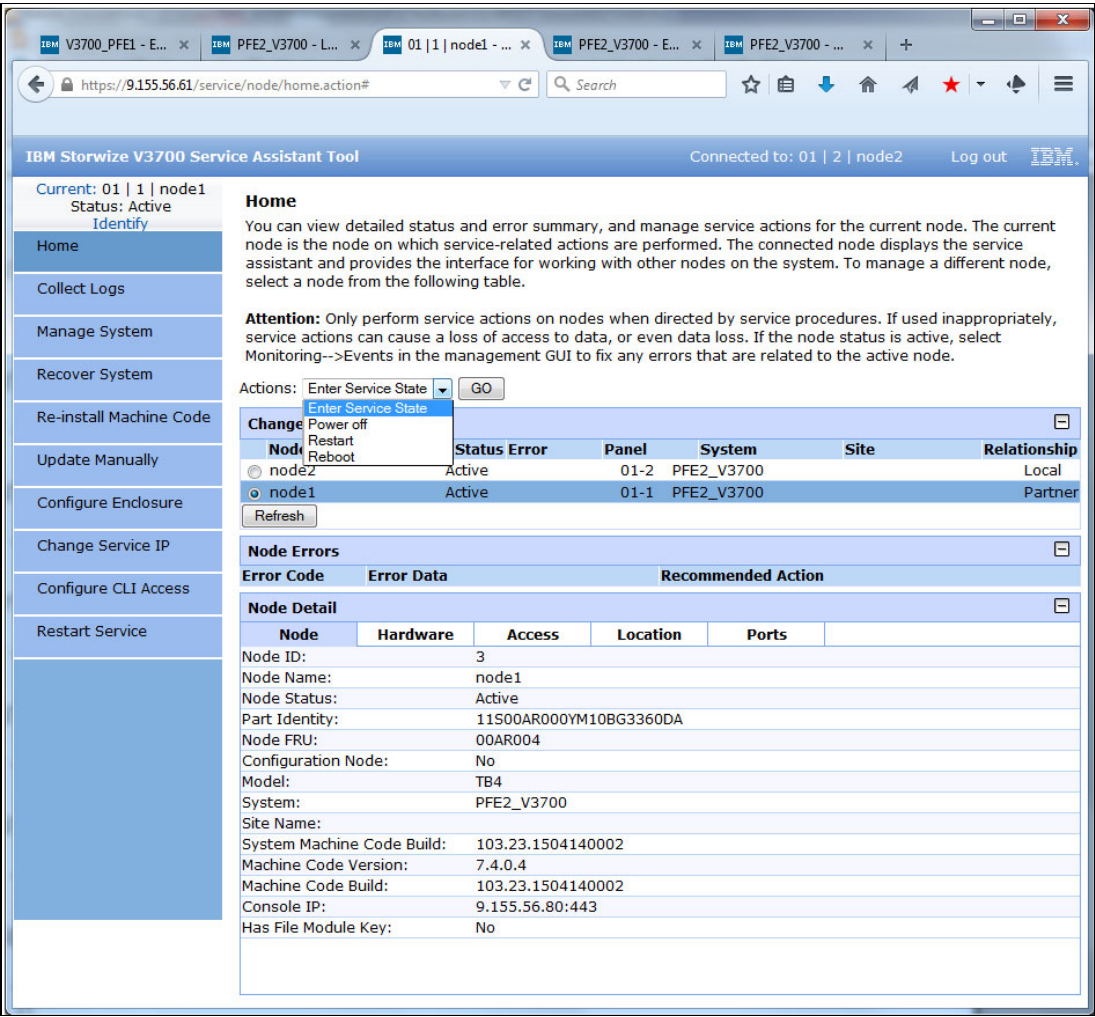


Figure 10-10 Service Assistant GUI

The following diagnostic options might be useful when you are working on iSCSI-related problems in the Service Assistant GUI:

- ▶ Download support package
- ▶ Download individual files from SAN Volume Controller / IBM Storwize node
- ▶ Node Reset
- ▶ Change Service IP address

10.1.3 Command-line interface

The CLI provides the most comprehensive selection of commands for managing the SAN Volume Controller and IBM Storwize cluster. The full list of commands, including syntax, is available at [IBM Knowledge Center](#).

iSCSI CLI configuration commands

These commands can be used to set or change Ethernet port and iSCSI configuration settings.

The **cfgportip** command

The **cfgportip** command is used to assign an IP address to each node Ethernet port for iSCSI input/output (I/O).

Figure 10-11 shows the syntax options for the **cfgportip** command.

For standard port configuration:

```
>>- cfgportip -- -- -node --+- node_name -+-- ----->
                               '- node_id ---'

>--+- -ip -- ipv4addr -- -mask -- subnet_mask -- -gw -- ipv4gw ---+-->
      '- -ip_6 -- ipv6addr -- prefix_6 -- prefix -- -gw_6 -- ipv6gw -'

>--+-----+-- +-----+-- ----->
      '- -failover -'      +- -host --+- yes -+---+
                           |           '- no --' |
                           +- -host_6 --+- yes -+--'
                           |           '- no --' |

>--+-----+-- +-----+-- ----->
      +- -remotecopy -- remote_copy_port_group_id ---+
      '- -remotecopy_6 -- remote_copy_port_group_id -'

>--+-----+-- +-----+-- ----->
      +- -vlan-- vlan_id_ip4-+      +- -vlan_6-- vlanid_ip6-+
      '- -novlan-----'      '- -novlan_6-----'

>--+-----+-- +-----+-- --port_id-----><
      +- -storage--+yes-+---+      '- -force-'
      |           '-no--' |
      +- -storage_6--+yes-+--
      |           '-no--' |
      |-----|
```

For maximum transmission unit (MTU):

```
>>- cfgportip - +- -mtu ---mtu---+-----+-- -port_id-><
                '- -defaultmtu---' '- -iogrp--+ io_grp_id---+-'
                                   '- io_grp_name-'
```

Figure 10-11 The **cfgportip** command syntax

For more information, see 7.1, “Configuring the IBM Storwize storage system for iSCSI” on page 84.

The **rmportip** command

The **rmportip** command is used to remove an iSCSI IP address from a node Ethernet port.

Figure 10-12 shows the syntax options for the **rmportip** command.

```
>>- rmportip -- --+-----+-- -- -ip_6 ----->
               '- -failover -'

>-- -node --+ node_name +- --port_id-- -----><
               '- node_id ---'
```

Figure 10-12 The **rmportip** command syntax

The chsystem command

The **chsystem** command is relevant in iSCSI configuration. It is used for the following tasks:

- ▶ Changing the cluster IP address
- ▶ Configuring the NTP server
- ▶ Specifying an iSNS server
- ▶ Setting the authentication method that is used for iSCSI
- ▶ Setting or clearing the Challenge Handshake Authentication Protocol (CHAP) secret

The mkhost command

The **mkhost** command allows the user to define the host with a connectivity protocol, including the following options:

- ▶ **-saswpn** for SAS attached host
- ▶ **-fcwpn** for Fibre Channel (FC) attached host
- ▶ **-iscsiname** for iSCSI attached host

For iSCSI attached hosts, the **mkhost** command can include one or more IQNs:

```
mkhost -iscsiname iqn.localhost.hostid.7f000001 -name newhost
mkhost -iscsiname iqn.localhost.hostid.7f000001 iqn.localhost.hostid.7f000002
-name newhost2
```

When the iSCSI host is created by using the **mkhost** command with the **iscsiname** parameter, the host is initially configured with the authentication method as none, and no CHAP secret is set. To set a CHAP secret for authenticating the iSCSI host with the SAN Volume Controller storage system, use the **chhost** command with the **chapsecret** parameter.

The addhostport command

The **addhostport** command allows the user to change the original host by defining by additional host ports. For an iSCSI host, these additional host ports are represented as IQNs:

```
addhostport -iscsiname iqn.localhost.hostid.7f000002 newhost
```

The chhost command

The **chhost** command allows the user to change certain properties of a host definition, such as name or type. In addition, it allows the CHAP secret that is used to authenticate the host for iSCSI I/O to be set.

Figure 10-13 shows the syntax for the **chhost** command.

```
>>- chhost -- +-----+>
              '- -type --+- hpux -----+- -'
                  +- tpgs-----+
                  +- generic -----+
                  +- openvms -----+
                  +- adminlun -----+
                  '- hide_secondary -'

>--+-----+>
  '- -mask - port_login_mask -'

>--+-----+>
  '- -name -- new_name_arg -'

>--+-----+>
  +- -chapsecret -- chap_secret +-
  '- -nochapsecret -----'

>--+-----+--+ host_name +-----><
  +- -site --+- site_name +--+ '- host_id ---'
  |           '- site_id ---' |
  '- -nosite -----'
```

Figure 10-13 The **chhost** command syntax

The CHAP authentication can also be defined by running the **chsystem** command.

The rmhost command

The **rmhost** command is used to delete a host object.

The mkvdiskhostmap command

The **mkvdiskhostmap** command is used to create a mapping between a volume and a host, which makes the volume accessible for I/O operations to the specified host.

iSCSI status and diagnostic CLI commands

These CLI commands might be useful for checking the network and status during iSCSI problem determination.

The ping command

This command changed in Version 7.5 and later. In Version 7.4.x and earlier, the **ping** command enabled only a single parameter to specify the target IP address. For Version 7.5 and later, it is necessary to select the IBM Storwize source IP address in addition to the destination address. This requirement adds diagnostic value for checking network connectivity in a redundant topology where multiple Ethernet ports are configured for iSCSI on the same node.

Figure 10-14 shows the syntax for the **ping** command in Version 7.5 and later.

```
>>- ping ----->
>--+ -srcip4 --source_ipv4_address destination_ipv4_address-+->
'- -srcip6 --source_ipv6_address destination_ipv6_address-'
```

Figure 10-14 Syntax of the **ping** command (Version 7.5)

Important: The **ping** command can be issued only against a source IP address that exists on the node in which you run the command.

The **lspportip** command

The **lspportip** command is used to list the iSCSI IP addresses that are assigned for each port on each node in the clustered system.

Figure 10-15 shows the **lspportip** command syntax.

```
>>- lspportip -- --+-----+----->
'- -filtervalue -- attribute=value -'

>--+-----+-----+----->
'- -filtervalue? -' '- -nohdr -'

>--+-----+-----+----->
'- -delim -- delimiter -' '- ethernet_port_id -'
```

Figure 10-15 The **lspportip** command syntax

The **lspportip** command without any parameters provides a concise summary of port settings and statuses for all Ethernet ports.

Example 10-1 shows an example of the output (columns are truncated for clarity).

Example 10-1 Example **lspportip** output

```
IBM_IBM Storwize:v7k_r51:superuser>lspportip
id name IP_address mask state speed failover link_state host
1 node1 unconfigured 1Gb/s no active
1 node1 unconfigured 1Gb/s yes active
2 node1 192.168.51.20 255.255.255.0 configured 1Gb/s no active yes
2 node1 configured 1Gb/s yes active
3 node1 unconfigured no inactive
3 node1 unconfigured yes inactive
4 node1 unconfigured no inactive
4 node1 unconfigured yes inactive
1 node2 unconfigured 1Gb/s no active
1 node2 unconfigured 1Gb/s yes active
2 node2 192.168.51.19 255.255.255.0 configured 1Gb/s no active yes
2 node2 configured 1Gb/s yes active
3 node2 unconfigured no inactive
3 node2 unconfigured yes inactive
4 node2 unconfigured no inactive
4 node2 unconfigured yes inactive
```

Example 10-2 shows an example of the **lspportip** command with an optional parameter (**-filtervalue state=configured**) that displays only the ports that are configured for iSCSI with their IP failover partner ports (the columns are truncated for clarity).

Example 10-2 Example lspportip output with filter

```
IBM_IBM Storwize:v7k_r51:superuser>lspportip -filtervalue state=configured
```

id	name	IP_address	mask	state	speed	failover	link_state	host
2	node1	192.168.51.20	255.255.255.0	configured	1Gb/s	no	active	yes
2	node1			configured	1Gb/s	yes	active	
2	node2	192.168.51.19	255.255.255.0	configured	1Gb/s	no	active	yes
2	node2			configured	1Gb/s	yes	active	

Tip: It is also possible to use a filter value of **host=yes** to exclude the IP failover ports:

```
IBM_IBM Storwize:v7k_r51:superuser>lspportip -filtervalue host=yes
```

The **lspportip** command with **ethernet_port_id** included as a parameter provides a detailed view of the specified port, as shown in Example 10-3.

Example 10-3 Example lspportip command output with ethernet_port_id included

```
IBM_IBM Storwize:v7k_r51:superuser>lspportip 2
id 2
node_id 1
node_name node1
IP_address 192.168.51.20
mask 255.255.255.0
gateway 192.168.51.1
IP_address_6
prefix_6
gateway_6
MAC 6c:ae:8b:7e:73:28
duplex Full
state configured
speed 1Gb/s
failover no
mtu 1500
link_state active
host yes
remote_copy 0
host_6
remote_copy_6 0
remote_copy_status
remote_copy_status_6
vlan
vlan_6
adapter_location 0
adapter_port_id 2
```

The lsroute command

This command displays the IP routing table. The table provides details of the gateway that is used for IP traffic to a range of IP addresses for each Ethernet port. This information can be used to diagnose configuration node accessibility problems. The **lsroute** command is equivalent to the Linux **route** command.

There are no parameters for the **lsroute** command.

Example 10-4 shows an example of this command.

Example 10-4 Example lsroute command output

IBM_IBM Storwize:v7k_r51:superuser>lsroute							
Kernel IP routing table							
Destination	Gateway	Genmask	Flags	Metric	Ref	Use	Iface
192.168.50.0	0.0.0.0	255.255.254.0	U	0	0	0	eth0
169.254.0.0	0.0.0.0	255.255.0.0	U	1003	0	0	eth0
0.0.0.0	192.168.50.1	0.0.0.0	UG	0	0	0	eth0

The lspportfc command

The **lspportfc** command is used to show the status and properties of all FC ports on each node in a clustered system. It also displays information about the 10-Gb Ethernet ports that might be relevant for iSCSI problem determination.

The **lspportfc** command provides a concise summary, but providing the object ID as a parameter returns detailed status information.

The lsiscsiauth command

The **lsiscsiauth** command lists the CHAP secret that is configured for authenticating an entity to the IBM Storwize clustered system.

This command is available in Version 7.5 and later.

The lshost command

The **lshost** command is used to generate a list with concise information about all the hosts that are visible to the clustered system and detailed information about a single host.

The lsvdisk command

The **lsvdisk** command is used to check the status of the volume. It returns either a concise list or a detailed view of volumes that are recognized by the clustered system, depending whether the VDisk ID is included as a parameter.

The lsvdiskhostmap command

The **lsvdiskhostmap** command shows which volumes are mapped to which hosts. The volume is visible only to hosts that are included in the mapping list.

The lsdependentvdisks command

The **lsdependentvdisks** command lists any volumes that depend on a single node. Hosts that are defined with a connection to just a single node or when the redundant link is down appear in the list.

Data collection CLI commands

These CLI commands can be used for capturing logs that might be relevant for diagnosing iSCSI problems.

You can use the Management GUI to generate and download a Support Data file with or without memory dumps. This information might be requested by IBM when a support call is logged.

The CLI can be used for generating a specific log.

The lseventlog command

The **lseventlog** command is used to display a concise view of the system event log, or a detailed view of one entry from the log.

Figure 10-16 shows the syntax for the **lseventlog** command.

```
>>- lseventlog --+-----+----->
               '- -filtervalue -- attribute_value -'

>--+-----+--+-----+----->
   '- -filtervalue? -' '- -alert --+-yes+-'
                               '-no--'

>--+-----+--+-----+----->
   '- -message --+-yes+-' '- -monitoring --+-yes+-'
               '-no--'                               '-no--'

>--+-----+--+-----+----->
   '- -expired --+-yes+-' '- -fixed --+-yes+-'
               '-no--'                               '-no--'

>--+-----+--+-----+----->
   '- -config --+-yes+-' '- -count -- entry_limit-'
               '-no--'

>--+-----+--+-----+-----><
   '- -order --+-date-----+-' '-sequence_number-'
               '-severity-'
```

Figure 10-16 The lseventlog command syntax

The catauditlog command

The **catauditlog** command displays the in-memory contents of the audit log, which provides a recent history of commands that were run either from the management GUI or CLI.

10.1.4 Service CLI

The Service CLI is intended for use by advanced users who are confident with using the CLI. It is accessed from an SSH session as with the standard CLI, but using the service IP address rather than cluster IP.

The Service CLI is used to manage a node canister in a control enclosure by using the task commands and information commands. An example where this might be useful is when checking error logs, trace files, and status from the non-configuration node. Most actions can also be run from the Service Assistant GUI.

10.1.5 USB

The USB port is used for the initial configuration of the IBM Storwize cluster and data collection when the other GUI and CLI interfaces are not available. It can also be used for running the System Administration tasks such as superuser password reset, node restart, and changing the Service IP address.

If you insert an empty USB key, a text file that is named `satask_result.html` is generated. It contains basic configuration and status information.

If you insert a USB key that contains a file that is named `satask.txt`, the node tries to run the command that is specified in the file.

10.1.6 Visual indicators (Ethernet port LED status)

Figure 10-17 shows the LED status indicators for the 1 Gb Ethernet ports.

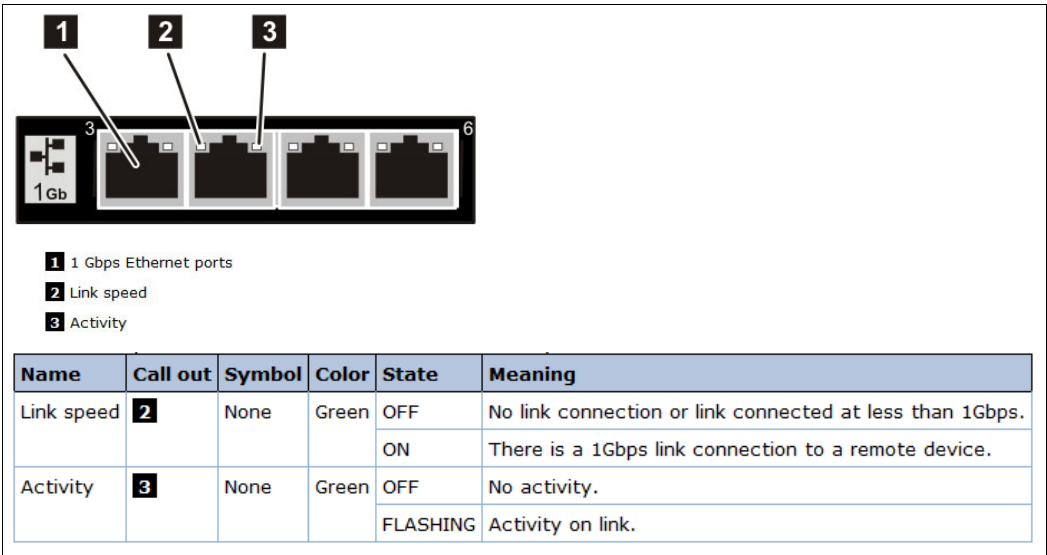


Figure 10-17 1 Gb Ethernet ports

Figure 10-18 shows the LED status indicators for the 10 Gb Ethernet ports.

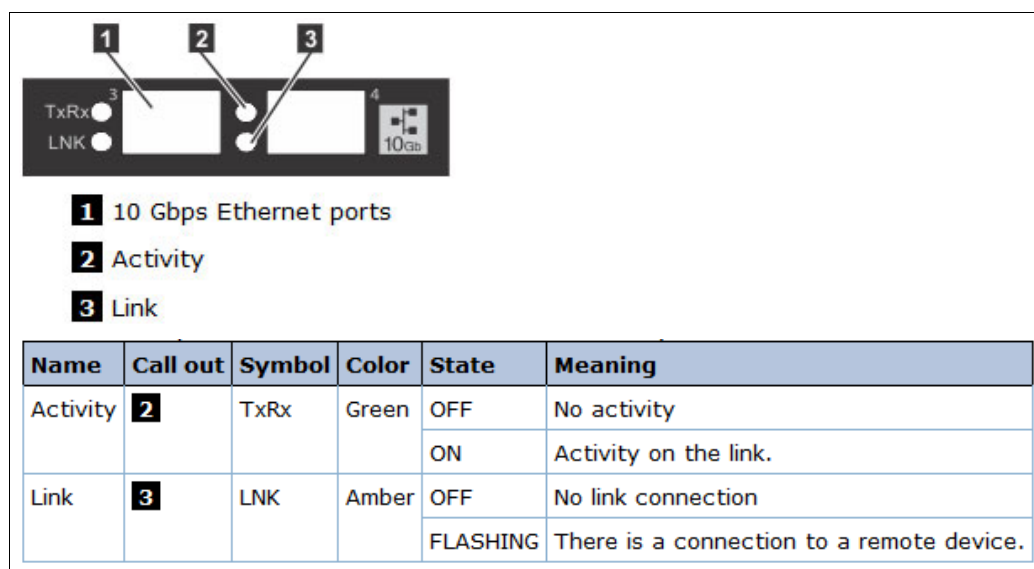


Figure 10-18 10 Gb Ethernet ports

10.2 Storage logs that are used for analysis

There are a number of logs that are available on the SAN Volume Controller and IBM Storwize storage system that can be useful for analyzing iSCSI problems.

10.2.1 Support Package on the IBM Storwize cluster

There are four Support Package options that are available from the management GUI that generate a bundle of logs that are collated from all nodes in the IBM Storwize cluster:

- ▶ Option 1 (Standard Logs) is the basic data collection bundle. It includes the event log, trace files, configuration backup file, performance statistics, and various other logs from all nodes in the cluster. The file size is relatively small.
- ▶ Option 2 includes the Standard Logs together with the most recent dump file (if available) from the configuration node only. This option might be required for analysis following a software error (node assert). The file size varies depending on whether a dump file is present.
- ▶ Option 3 includes the Standard Logs together with the most recent dump file for all nodes. This option might be required for analysis following a software error (node assert). The file size might be larger depending on how many nodes in the cluster contain a dump file.
- ▶ Option 4 generates a new statesave on each node in the cluster and includes them in the support package. The file size always is relatively large.

The Support Package can also be generated from the Service Assistant GUI, but includes logs only from a single node.

The Support Package is downloaded to the management workstation. A snap file is created and compressed with a .tgz extension. The serial number, node ID, and time stamp are included in the file name:

snap.78G01RN-2.151024.133346.tgz

The .tgz is a compressed tar archive that can be extracted on a Linux host by using the following command:

```
tar -zxvf <filename.tgz>
```

There are also a number of Windows utilities that are available to extract and view the contents of the .tgz archive file.

10.2.2 Event log on the IBM Storwize cluster

Events that are detected are saved in an event log. When an entry is made in this event log, the condition is analyzed. If any service activity is required, a notification is sent, if notifications are set up.

The following methods are used to notify you and IBM Support Center of a new event:

- ▶ If SNMP is enabled, an SNMP trap is sent to an SNMP manager that is configured by the customer.
- ▶ If enabled, log messages can be forwarded on an IP network by using the syslog protocol.
- ▶ If enabled, event notifications can be forwarded by email by using SMTP.
- ▶ Call Home can be enabled so that critical faults generate a PMR that is then sent directly to the appropriate IBM Support Center by using email.

Events are classified as either alerts or messages:

- ▶ An alert is logged when the event requires some action. Some alerts have an associated error code that defines the service action that is required. The service actions are automated through the fix procedures. If the alert does not have an error code, the alert represents an unexpected change in state. This situation must be investigated to see whether it is expected or represents a failure. Investigate an alert and resolve it as soon as it is reported.
- ▶ A message is logged when a change that is expected is reported, for example, an IBM FlashCopy operation completes.

Fields in the event log

The event log contains fields with information that are used for diagnosing IBM Storwize errors, as shown in Table 10-1.

Table 10-1 Fields in the event log

Data field	Description
Error code	Indicates that the event represents an error in the system that can be fixed by following the fix procedure or service action that is identified by the error code. Not all events have an error code. Different events have the same error code if the same service action is required for each one.
Sequence number	Identifies the event within the system.
Event count	The number of events that are coalesced into this event log record.
Object type	The object type to which the event relates.
Object ID	Uniquely identifies the object within the system to which the event relates.
Object name	The name of the object in the system to which the event relates.

Data field	Description
Copy ID	If the object is a volume and the event refers to a specific copy of the volume, this field is the number of the copy to which the event relates.
Reporting node ID	Typically identifies the node that is responsible for the object to which the event relates. For events that relate to nodes, it identifies the node that logged the event, which can be different from the node that is identified by the object ID.
Reporting node name	Typically identifies the node that contains the object to which the event relates. For events that relate to nodes, it identifies the node that logged the event, which can be different from the node that is identified by the object name.
Fixed	Where an alert is shown for an error or warning condition, it indicates that the user marked the event as fixed,
First time stamp	The time when this error event was reported. If events of a similar type are being coalesced together so that one event log record represents more than one event, this field is the time the first error event was logged.
Last time stamp	The time when the last instance of this error event was recorded into this event log record.
Root sequence number	If set, it is the sequence number of an event that represents an error that probably caused this event to be reported. Resolve the root event first.
Sense data	Additional data that gives the details of the condition that caused the event to be logged.

10.2.3 Audit log on the IBM Storwize cluster

The audit log provides a history of actions that are submitted through the management GUI or the CLI. It can be used to monitor administrative activity on the IBM Storwize storage system, and is an important tool in problem determination to verify what changes were made.

The audit log entries include the following information:

- ▶ The time and date when the action or command was submitted on the system
- ▶ The name of the user who performed the action or command
- ▶ The IP address of the system where the action or command was submitted
- ▶ The parameters that were submitted with the command
- ▶ The results of the command or action
- ▶ The sequence number and the object identifier that is associated with the command or action

10.2.4 Ethernet logs and statistics on IBM Storwize nodes

Each IBM Storwize node contains two Ethernet log files in the /dumps folder.

On node 1 of the IBM Storwize enclosure, the serial number is 78G01RN:

```
/dumps/ethernet.78G01RN-1.stats
/dumps/ethernet.78G01RN-1.trc
```

On node 2 of the IBM Storwize enclosure, the serial number is 78G01RN:

```
/dumps/ethernet.78G01RN-2.stats
/dumps/ethernet.78G01RN-2.trc
```

These logs are included within the support package. They can also be downloaded as individual files from the management GUI (click **Settings** → **Support** → **Show full log listing**) or by using the Service Assistant GUI (Collect Logs). Using the Service Assistant GUI allows selection of either node, but the management GUI can see only the files that are present on the configuration node.

There is an entry that is generated in the .trc and .stats files every time that an interface state changes, which allows the user to check the history of IP addresses that are assigned to each Ethernet port. It also provides an overview of traffic on each port, MTU size, and error counts, as shown in Example 10-5.

Example 10-5 Example log entry

```
==== START: ec_ifdown_ioport at Wed Oct 7 13:53:35 IST 2015
parms: -i -p 2 -a 192.168.51.20 -m 255.255.255.0 -g 192.168.51.1 -v 0 DO_NOT_FAILOVER_IP:
killall -9 arping with arguments -c 8 -U -I eth1 192.168.51.20
ip -4 route flush table eth1_2
ip -4 rule del from 192.168.51.20 table eth1_2
ip -4 rule del to 192.168.51.20 table eth1_2
ip -4 addr del 192.168.51.20/32 dev eth1 label eth1:2 brd +
status: ifconfig
eth0      Link encap:Ethernet  HWaddr 6C:AE:8B:7E:5D:05
          inet addr:192.168.51.19 Bcast:192.168.51.255 Mask:255.255.254.0
          inet6 addr: 2002:976:2c08:207:6eae:8bff:fe7e:5d05/64 Scope:Global
          inet6 addr: 2002:976:2c08:202:6eae:8bff:fe7e:5d05/64 Scope:Global
          inet6 addr: 2002:976:2c08:205:6eae:8bff:fe7e:5d05/64 Scope:Global
          inet6 addr: fe80::6eae:8bff:fe7e:5d05/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:19820741 errors:2 dropped:0 overruns:0 frame:2
          TX packets:91091 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:1443796538 (1.3 GiB)  TX bytes:12428798 (11.8 MiB)
          Interrupt:16 Memory:fba00000-fba20000

eth0:10   Link encap:Ethernet  HWaddr 6C:AE:8B:7E:5D:05
          inet addr:9.113.57.74 Bcast:9.113.57.255 Mask:255.255.255.255
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          Interrupt:16 Memory:fba00000-fba20000

eth0:20   Link encap:Ethernet  HWaddr 6C:AE:8B:7E:5D:05
          inet addr:9.113.57.228 Bcast:9.113.57.255 Mask:255.255.255.255
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          Interrupt:16 Memory:fba00000-fba20000

eth1      Link encap:Ethernet  HWaddr 6C:AE:8B:7E:5D:04
          inet6 addr: 2002:976:2c08:202:6eae:8bff:fe7e:5d04/64 Scope:Global
          inet6 addr: 2002:976:2c08:207:6eae:8bff:fe7e:5d04/64 Scope:Global
          inet6 addr: 2002:976:2c08:205:6eae:8bff:fe7e:5d04/64 Scope:Global
          inet6 addr: fe80::6eae:8bff:fe7e:5d04/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:241147567 errors:253 dropped:0 overruns:0 frame:253
          TX packets:440390841 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:16117707881 (15.0 GiB)  TX bytes:655754179350 (610.7 GiB)
          Interrupt:16 Memory:fbb00000-fbb20000

eth1:1    Link encap:Ethernet  HWaddr 6C:AE:8B:7E:5D:04
          inet addr:192.168.51.19 Bcast:0.0.0.0 Mask:255.255.255.255
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          Interrupt:16 Memory:fbb00000-fbb20000
```



```

eth2      Link encap:Ethernet  HWaddr 00:90:FA:07:01:E6
          UP BROADCAST MULTICAST  MTU:1500  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:0 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:0 (0.0 b)  TX bytes:0 (0.0 b)

eth3      Link encap:Ethernet  HWaddr 00:90:FA:07:01:E8
          UP BROADCAST MULTICAST  MTU:1500  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:0 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:0 (0.0 b)  TX bytes:0 (0.0 b)

lo        Link encap:Local Loopback
          inet addr:127.0.0.1  Mask:255.0.0.0
          inet6 addr: ::1/128 Scope:Host
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
          RX packets:71680 errors:0 dropped:0 overruns:0 frame:0
          TX packets:71680 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:7195465 (6.8 MiB)  TX bytes:7195465 (6.8 MiB)

```

This example shows that on 7 October at 13:53, the Ethernet port 1 on node 2 had three IPv4 addresses assigned:

- ▶ 192.168.51.19 (iSCSI IP)
- ▶ 9.113.57.74 (service IP)
- ▶ 9.113.57.228 (cluster IP)

On Ethernet port 2, there is only one IPv4 address assigned: 192.168.51.19 (iSCSI IP).

10.2.5 iSCSI logs on IBM Storwize nodes

If an iSNS server is included in the configuration, then it is possible to view and manage all iSCSI devices through the iSNS client interface. All initiator and target devices register with an iSNS database and exchange properties during a query.

The list of active iSCSI sessions is also included in a file that is named `proc_net_iet_session`, which is present in the `/dumps/syslog` folder on each node. The file name includes a suffix of a serial number and node ID. For Node 2 on enclosure serial number 78G01RN, this name is `/dumps/syslogs/proc_net_iet_session.78G01RN-2`.

Example 10-6 shows a sample `proc_net_iet_session` file.

Example 10-6 Sample `proc_net_iet_session` file

```

tid:1 name:iqn.1986-03.com.ibm:2145.v7kr51.node1
login_id:2 sid:281475047817728 name:iqn.1994-05.com.redhat:rhel-r51u28 port:0
cid:0 local_ip:192.168.51.20 remote_ip:192.168.51.28 state:active hd:none dd:none

```

These logs include iSCSI host and network configuration properties that are essential for host-side problem determination. They contain initiator and target IP addresses and IQNs for all iSCSI sessions.

With current code releases at the time of writing, complete logs are not included in the Support Package bundle and they are not available for download as individual files from either the management GUI or Service Assistant GUI.

Here is a workaround procedure for retrieving the `proc_net_iet_session` files from both nodes:

1. Obtain a list of nodes in the cluster and identify which node is the configuration node from the management GUI or CLI.
2. Enter the following command to copy all `proc_net_iet_session` files onto the configuration node. This command must be repeated for every non-configuration node in the cluster.

```
satask cpfiles -prefix "/dumps/syslogs/proc_net*" -source <source_panel_name>
```
3. Use the management GUI to generate a basic Support Package (option 1). This compressed (.tgz) file contains a valid `proc_net_iet_sessions` file from each node.

10.3 Different IP addresses on the IBM Storwize storage system

Figure 10-19 shows the different IP addresses that may be set on a 2-node IBM Storwize cluster where both internal Ethernet ports are configured for iSCSI.

The primary cluster IP address that is used for accessing the management GUI and CLI is available only on the configuration node. In any IBM Storwize cluster, there always is only one node that is assigned as the configuration node. Any node in the IBM Storwize cluster can take over the configuration node role when the current configuration node is either restarted or goes offline.

Each IBM Storwize cluster has the primary cluster IP address on Ethernet port 1 of the configuration node. It is possible to assign a secondary cluster IP address that is available only through Ethernet port 2 of the configuration node. The example in Figure 10-19 shows two cluster IP addresses that are assigned, with the configuration node being Node 1.

Node 1 (config node)		Node 2	
E/net Port 1	IP Addresses	E/net Port 1	IP Addresses
	Cluster: 10.1.10.11 Service: 10.1.10.13 iSCSI: 10.1.10.20		Cluster: n/a Service: 10.1.10.14 iSCSI: 10.1.10.22
E/net Port 2	IP Addresses	E/net Port 2	IP Addresses
	Cluster: 10.1.10.12 Service: n/a iSCSI: 10.1.10.21		Cluster: n/a Service: n/a iSCSI: 10.1.10.23

Figure 10-19 IP addresses on a two-node IBM Storwize cluster with iSCSI

Each node is also assigned a unique service IP address that is available only through Ethernet port 1. This address is used for accessing the Service Assistant GUI and service CLI.

The example in Figure 10-19 on page 196 also shows both internal Ethernet ports on each node that is configured for iSCSI. It is also possible to assign another IP address for replication, although this should not use an Ethernet port that already is configured for iSCSI host access.

10.3.1 Path failover mechanisms in iSCSI on an IBM Storwize storage system

There are two mechanisms that are used in an IBM Storwize storage system for providing redundant paths if a link or component failure occurs:

- Multipath driver** The multipath driver on the host reroutes the I/Os across available paths where available.
- IP failover** Ethernet ports that are configured for iSCSI behave as a redundant pair on partner nodes within an IO group. This mechanism takes effect only when a node is offline, as is the case during a concurrent code update (CCU). A link failure does not trigger an iSCSI IP failover. This feature cannot be disabled or changed.

10.3.2 iSCSI IP failover

Figure 10-20 demonstrates the iSCSI IP address failover if an IBM Storwize node is taken offline. Node 2 inherited the iSCSI IP addresses from each corresponding partner node port in addition to its original iSCSI IP address. The IQN is also transferred across so that the identity of the iSCSI targets remains unchanged to any hosts that access this IBM Storwize unit.

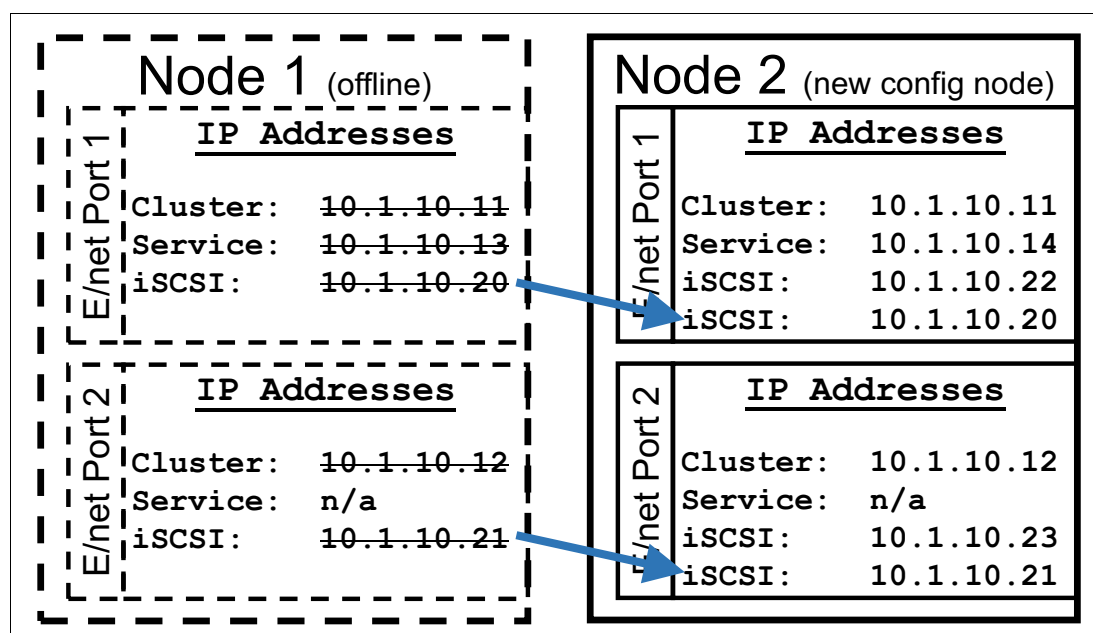


Figure 10-20 iSCSI IP failover

Node 1 was previously the configuration node, so now Node 2 also took over the configuration node role, together with the two cluster IP addresses. If this was a larger cluster with four, six, or eight nodes, then the configuration role does not necessarily switch to the partner in the I/O group.

This example shows only two Ethernet ports per IBM Storwize node. The same IP failover applies to any additional configured Ethernet ports on host interface cards (HICs) in the IBM Storwize storage system.

The iSCSI IP address failover takes place only when a node is offline or no longer a member of the IBM Storwize cluster. When Node 1 rejoins the cluster, the iSCSI IP addresses returns to their original configuration, which makes maintenance activities such as a machine code update almost transparent to the iSCSI connected host, even without a multipath driver.

It is important to understand that this process might be going on in the background while checking for certain iSCSI problems, perhaps where multipathing did not behave as expected when a single IBM Storwize node restarted.

10.4 Problem determination

There are multiple ways of gathering evidence to help determine the underlying cause of failure. There are also various fault symptoms, different network topologies, and host or multipath configurations. This section covers some of the common issues that can be seen with typical iSCSI configurations on IBM Storwize storage systems.

This section is divided into seven subsections that should cover many cases where troubleshooting is required. It is not intended to be a step-by-step guide. In situations where a problem develops, the user already should have a reasonable idea of which subsections might be relevant. For new installations where the iSCSI connectivity never worked, it might be worth reviewing all the subsections.

This section assumes that the IBM Storwize storage system is in an optimal state, but iSCSI hosts do not detect the IBM Storwize storage targets. Directed maintenance procedures should first be run from the management GUI to identify and fix all unfixed errors before proceeding.

If possible, ensure that all hosts, switches, and IBM Storwize storage systems have synchronized clocks, preferably by using NTP. It makes log analysis easier if adjustment for time differences is not required.

Here are the subsections:

- ▶ Section 10.4.1, “Problem determination: Obtaining a basic configuration overview” on page 199
Always maintain an up-to-date log of the system configuration. For iSCSI problem determination, it is important to understand the network topology together with IP addresses of all management, service, host, and IBM Storwize iSCSI Ethernet ports. In many cases where this is available, it is possible to spot quickly suspect areas based on where problems are detected and where problems are not seen.
- ▶ Section 10.4.2, “Problem determination: Checking the network configuration” on page 201
The iSCSI protocol depends on a reliable network transport layer. If there are any defects in the network, then there also are iSCSI problems.

- ▶ Section 10.4.3, “Problem determination: Checking the IBM Storwize configuration” on page 204

In cases where a host loses access to iSCSI target or the target cannot be discovered, it is worth checking the volume status to ensure that they are online and mapped correctly.

- ▶ Section 10.4.4, “Problem determination: Checking authentication” on page 204

If CHAP authentication is enabled, then the target must authenticate the initiator with a predefined secret key before granting access.

- ▶ Section 10.4.5, “Problem determination: Checking active sessions from the IBM Storwize storage system” on page 205

This subsection covers the scenarios where iSCSI target LUNs are accessible to the hosts but not discovered on all expected paths. It is necessary to identify the problematic paths.

- ▶ Section 10.4.6, “Problem determination: Checking a host configuration” on page 206

In a redundant topology, the multi-path driver on the host reroutes I/Os if a path failure occurs. There are different host logs that are available to view on each operating system. It is worth confirming that configuration rules, dependencies, and prerequisites are met.

- ▶ Section 10.4.7, “Problem determination: Checking for performance problems” on page 207

There is a separate chapter on performance that provides various suggestions for optimizing or tuning the IBM Storwize storage system in an iSCSI environment. This subsection covers some of the known defects that might contribute to performance degradation.

10.4.1 Problem determination: Obtaining a basic configuration overview

It is normally necessary to have a clear understanding of the network topology together with host attachment details to interpret whether the iSCSI status is as expected. There are some CLI commands that can provide a general overview of the setup from the IBM Storwize storage system:

- ▶ Check node IDs and identify the configuration node by running the following command on an IBM Storwize model:

```
lsnodecanister
```

Here is the equivalent command on the SAN Volume Controller storage system:

```
lsnode
```

These commands provide a view of the nodes in the cluster, together with the panel IDs that are required in the next command.

- ▶ Check the configuration of each node by running the following command:

```
sainfo lsservicestatus <panel_id>
```

This command should be run for each node. The key configuration information to note includes the information that is shown in Example 10-7.

Example 10-7 Key configuration information

```
panel_name
node_id
node_name
cluster_port
cluster_ip
cluster_gw
```

```
cluster_mask
cluster_ip_6
cluster_gw_6
cluster_mask_6
config_node
service_IP_address
service_gateway
service_subnet_mask
service_IP_address_6
service_gateway_6
service_subnet_mask_6
node_code_version
```

The fields most relevant for network and iSCSI problem determination are:

- ▶ ethernet_ports
- ▶ ethernet_port_id
- ▶ port_status
- ▶ port_speed

Section 10.2.5, “iSCSI logs on IBM Storwize nodes” on page 195 describes a procedure for obtaining more iSCSI host configuration and network properties from the `proc_net_iet_sessions` files. Example 10-8 shows a sample `proc_net_iet_session` file from node 1.

Example 10-8 Sample `proc_net_iet_session` file

```
tid:1 name:iqn.1986-03.com.ibm:2145.v7kr51.node1
login_id:2 sid:281475047817728 name:iqn.1994-05.com.redhat:rhel-r51u28 port:0
cid:0 local_ip:192.168.51.20 remote_ip:192.168.51.28 state:active hd:none dd:none
```

Example 10-9 shows the same information from the partner node in the I/O Group, node 2.

Example 10-9 Sample `proc_net_iet_session` file from the partner node

```
tid:1 name:iqn.1986-03.com.ibm:2145.v7kr51.node2
login_id:1 sid:281475031040512 name:iqn.1994-05.com.redhat:rhel-r51u28 port:0
cid:0 local_ip:192.168.51.19 remote_ip:192.168.51.28 state:active hd:none dd:none
```

These files provide a useful and detailed view of the current iSCSI topology as seen from the IBM Storwize storage system. The sample files show the following information:

- ▶ There is a single Red Hat Enterprise Linux server that is attached with iSCSI connectivity.
- ▶ It is using a single NIC for this iSCSI connection.
- ▶ The IP address of the host NIC is 192.168.51.28.
- ▶ The initiator IQN is `iqn.1994-05.com.redhat:rhel-r51u28`.
- ▶ There is a single iSCSI session to node 1 through the interface IP address 192.168.51.20.
- ▶ The target IQN for the node 1 session is `1986-03.com.ibm:2145.v7kr51.node1`.
- ▶ There is a single iSCSI session to node 2 through the interface IP address 192.168.51.19.
- ▶ The target IQN for the node 2 session is `1986-03.com.ibm:2145.v7kr51.node2`.
- ▶ The initiator IQN shows that the host is using a software initiator.

The **lspointip** output shows which IBM Storwize ports are used for these iSCSI sessions, as shown in Example 10-10.

Example 10-10 Example *lspointip* output that shows ports for iSCSI sessions

IBM_IBM Storwize:v7k_r51:superuser>lspointip -filtervalue host=yes														
id	node_id	node_name	IP_address	mask	gateway	IP_address_6	prefix_6	gateway_6	MAC	duplex	state	speed		
failover	link_state	host	remote_copy	host_6	remote_copy_6	remote_copy_status	remote_copy_status_6	vlan	vlan_6	adapter_location				
adapter_port_id														
2 1	node1		192.168.51.20	255.255.255.0	192.168.51.1				6c:ae:8b:7e:73:28	Full	configured	1Gb/s	no	
active	yes	0		0					0	2				
2 2	node2		192.168.51.19	255.255.255.0	192.168.51.1				6c:ae:8b:7e:5d:04	Full	configured	1Gb/s	no	
active	yes	0		0					0	2				

This information should be sufficient to produce an accurate topology diagram.

Figure 10-21 shows the basic iSCSI topology diagram that can be compiled with just data that is extracted from the `proc_net_uet_session` files from each node and the **lspointip** output.

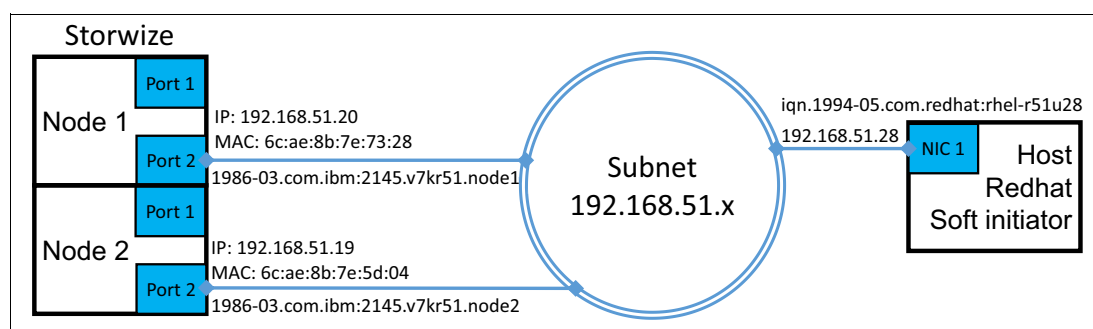


Figure 10-21 Basic iSCSI topology diagram

If there are problems attaching to the Ethernet hosts, the fault might be related to the network, the IBM Storwize storage system, or the host. There are logical steps that can be followed to determine the underlying cause of failure. It is often possible to narrow down the areas likely to be contributing to the problem by understanding the iSCSI network topology and clearly defining the problem symptoms.

10.4.2 Problem determination: Checking the network configuration

The iSCSI protocol allows SCSI commands to be transported over IP networks. Therefore, the first basic test in any fault scenario where hosts cannot access the storage should involve testing whether there is a fault in the TCP/IP link between the host and storage port.

Checking the end-to-end network connection

The following command can be used to check the IP addresses of IBM Storwize (iSCSI target) ports:

```
lspointip
```

The **ping** command can then be run either from the IBM Storwize storage system or the affected host. With IBM Storwize V7.5 or later, the IP addresses of both the iSCSI initiator and target are required as parameters, as shown in Example 10-11.

Example 10-11 Example *ping* command

```
IBM_IBM Storwize:V3700_PFE1:superuser>ping -srcip4 206.30.15.40 206.30.15.1
PING 206.30.15.1 (206.30.15.1) from 206.30.15.40 : 56(84) bytes of data.
64 bytes from 206.30.15.1: icmp_seq=1 ttl=255 time=2.16 ms
```

```
64 bytes from 206.30.15.1: icmp_seq=2 ttl=255 time=0.331 ms
64 bytes from 206.30.15.1: icmp_seq=3 ttl=255 time=0.347 ms
64 bytes from 206.30.15.1: icmp_seq=4 ttl=255 time=0.372 ms
64 bytes from 206.30.15.1: icmp_seq=5 ttl=255 time=0.385 ms

--- 206.30.15.1 ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4004ms
rtt min/avg/max/mdev = 0.331/0.719/2.163/0.722 ms
```

If the IBM Storwize iSCSI port IP address does not respond to **ping** either to or from the host through any of the configured paths, then it is necessary to take a closer look at the network. The network path consists of multiple components. In the minimum configuration, these components include a network interface card on the host, the Ethernet port on the IBM Storwize storage system, the Ethernet switches, and the cables that connect them. If the links are 10 Gbps, there also are small form-factor (SFF) transceivers and fiber cables to consider.

Checking the status of the Ethernet port on the IBM Storwize storage system

The **lsportip** output and Ethernet port LED status shows the Ethernet port link state, including the connection speed.

If a 1-Gbps link is down, connecting at a lower speed, or generating errors then, complete the following tasks:

- ▶ Check or replace the cable.
- ▶ Confirm that correct grade CAT5 or CAT6 cable is used.
- ▶ Check the switch logs for errors.
- ▶ If available, try an alternative port on the IBM Storwize storage system.
- ▶ If available, try connecting to an alternative switch port.

If a 10-Gbps link is down, connecting at a lower speed, or generating errors, complete the following tasks:

- ▶ Check or replace the fiber cable.
- ▶ Confirm that OM2 or OM3 multimode fiber cable is used within maximum distance limits.
- ▶ Check the switch logs for errors.
- ▶ If available, try an alternative port on the IBM Storwize HIC.
- ▶ If available, try a different 10-Gbps SFF in the same IBM Storwize HIC port.
- ▶ If available, try connecting to an alternative switch port.
- ▶ Replace the part. The replacement 10-Gbps SFF FRU part number is 31P1549.

Checking the firewall and router settings

TCP port 3260 must be unblocked in the firewall for iSCSI communication.

Checking the network status from the host

There are many tools that are available on the host to examine the status of the network and analyze traffic. Analyzing the output of these tools is outside the scope of this publication. However, sometimes an IBM Support representative might request a network trace capture while a problem is being re-created. The IBM Storwize storage system logs and dump files provide a good indication of what the cluster is doing, and these tools show what is being sent in the network.

There are rare cases where a packet trace is required. The IBM Support representative advises you whether this is necessary to assist with an investigation on a particular case. They also provide specific instructions. There are similar packet trace utilities that are available on Windows, Linux, and VMware ESX and ESXi.

Wireshark

Wireshark is a packet capture and trace utility for Windows. You can download it from the [Wireshark website](#).

Wireshark is a GUI-based tool. The user selects the connection (source and target IP addresses) that requires analysis. By default, Wireshark captures all the traffic on the selected interface with data, and when the capture is stopped, the captured data can be saved in a trace file.

Trace files are likely to be large. A 1-GB file that is split into manageable chunks of around 200 MB is usually sufficient to cover the period of failure if the trace is started immediately before you re-create the fault.

In the Wireshark capture options, the following settings might be required:

1. In the Capture section:
 - Clear **Limit each packet to X bytes** (you want a full capture of packets).
 - Change Buffer Size to 2 MB.
2. In the Capture File(s) section:
 - Select **Use multiple files**.
 - Select **Next file every** and set 200 MB.
 - Select **Ring buffer with** and set five files.
3. In the Display Options section, clear all the options.
4. In the Name Resolution section, clear all the options.

Stop the packet capture as soon as the issue is re-created. Upload the file to IBM with an Option 4 Support Package (with statesaves).

The tcpdump and tcpdump-uw tools on ESX/ESXi

For detailed documentation about this utility, see this [VMware Knowledge base article](#).

To list the VMkernel interfaces that are used for iSCSI, run the following command:

```
# esxcfg-vmknics -l
```

Then, to generate up to ten 100-MB trace files and keep on wrapping around until the problem can be re-created, run the following command:

```
# tcpdump-uw -i vmk0 -s 1514 -C 100M -W 10 -w /var/tmp/test.pcap
```

This command assumes an MTU size of 1500. Using parameters **-s 9014 -B 9** enables the capture of Jumbo frames. The test.pcap file is in a PCAP format that can be analyzed with Wireshark.

The tcpdump tool on Linux

To capture packets without full data but enough data to recognize iSCSI Headers (smaller capture size), run the following command:

```
tcpdump -i <ethX> -s 200 -w <capture file name>
```

To capture packets along with data (you can select the large capture size), run the following command:

```
tcpdump -i <ethX> -s 65536 -w <capture file name>
```

To generate up to ten 100-MB trace files and keep on wrapping around until the problem can be re-created, run the following command:

```
tcpdump -i eth0 -s 200 -w test.pcap -C 100 -W 10
```

Checking the IP address, subnet mask, and gateway settings

Make sure that there are no errors in the network definitions on the host and IBM Storwize network ports.

Checking the VLAN settings

If applicable, for virtual local area network (VLAN) configuration guidelines for an iSCSI environment, see [IBM Knowledge Center](#).

10.4.3 Problem determination: Checking the IBM Storwize configuration

After you confirm that the IBM Storwize target responds to **ping** from the host, and the host still cannot discover the LUNs, complete the following tasks:

- ▶ Verify that the volume is online by running the following command:

```
lsvdisk
```
- ▶ Verify that the host is defined correctly on the IBM Storwize storage system by running the following command:

```
lshost
```
- ▶ Verify that the volume is mapped correctly by running the following command:

```
lsvdiskhostmap
```

10.4.4 Problem determination: Checking authentication

It is possible to use the CLI to configure the CHAP to authenticate the IBM Storwize cluster to the iSCSI-attached hosts. After the CHAP is set for the cluster, all attached hosts must be configured to authenticate this way.

To help in problem determination, this step can be delayed until after the first one or two hosts are configured and their connectivity is tested without authentication.

To set the authentication method for the iSCSI communications of the cluster, run the following command:

```
svctask chcluster -iscsiauthmethod chap -chapsecret chap_secret
```

Where **chap** sets the authentication method for the iSCSI communications of the cluster and **chap_secret** sets the CHAP secret to be used to authenticate the cluster through iSCSI. This parameter is required if the **iscsiauthmethod chap** parameter is specified. The specified CHAP secret cannot begin or end with a space.

To clear any previously set CHAP secret for iSCSI authentication, run the following command:

```
svctask chcluster -nochapsecret
```

The **nochapsecret** parameter is not allowed if the **chapsecret** parameter is specified.

The **lsiscsiauth** command lists the CHAP secret that is configured for authenticating an entity to the IBM Storwize cluster. The command also displays the configured iSCSI authentication method. For example, run the following CLI command:

```
svcinfo lsiscsiauth
```

10.4.5 Problem determination: Checking active sessions from the IBM Storwize storage system

If LUNs are discovered successfully but are not visible through all the expected paths from the host, check the host status by running the IBM Storwize **lshost** command. Check the following things in the output:

- ▶ The **port_count** field in the **lshost** output shows the number of IQNs that are associated with the iSCSI host. Each IQN is listed with a separate **iscsi_name**.
- ▶ **lshost** shows a **host_status** value of one of the following values:
 - **online**: The host has full connectivity. A host that uses iSCSI connectivity is online if it has an iSCSI session with each I/O group with which the host has volume mappings.
 - **offline**: The host has no connectivity. There are no iSCSI sessions with any I/O group with which the host has volume mappings. This might be because the host is powered off. This also means that if an iSCSI host is logged in to only I/O groups for which it is not configured, the associated host object status is offline.
 - **degraded**: An iSCSI host that has no mapped volumes is degraded if it is logged in to some, but not all, of the I/O groups to which it belongs. It is also degraded if the host has multiple initiators (shown in **iscsi_name**) and one of them is offline.
- ▶ **lshost** displays the number of iSCSI sessions from the host IQN in the **node_logged_in_count** field.
- ▶ **lshost** shows a **state** value for the **iscsi_name**, which can be one of the following values:
 - **active**: Each I/O group with volume mappings has at least one associated iSCSI session for the specified **iscsiname**.
 - **inactive**: The host has no volume mappings but at least one iSCSI session for the specified **iscsiname** is present.
 - **offline**: One or more I/O Groups with volume mappings for the defined host do not have an associated iSCSI session.

The lshost output with IP failover: The **lshost** output remains unchanged when a node is offline because the Ethernet port on the partner node in the cluster pair automatically adopts the missing iSCSI IP addresses. Therefore, the iSCSI sessions that are previously associated with node 1 port 1 continue through node 2 port 1 while node 1 is not an online member of the IBM Storwize cluster.

In FC or SAS configurations, the **host_status** value in **lshost** normally changes to **degraded** whenever paths to one node are unavailable.

10.4.6 Problem determination: Checking a host configuration

There are some known issues that cause a host or multiple hosts to be unable to discover the IBM Storwize target.

Checking for a changed node name or cluster name

The IQN for each node or node canister is generated by using the clustered system and node or node canister names. Therefore, changing either name also changes the IQN of all of the nodes or node canisters in the clustered system and might require reconfiguration of all iSCSI-attached hosts.

This is an important consideration. Any IBM Storwize configuration with iSCSI-attached hosts should not have any action involving a change of cluster or node name that is performed on them, unless this is a planned activity. This is normally nondisruptive with FC- or SAS-attached hosts, but causes an impact with iSCSI.

Note: Replacing a node canister does not change the node name if the correct replacement procedure is followed. Only the node ID might change.

Checking for incorrect characters in the IQN definition

It is not uncommon to find the IQN definition in the host target list with incorrect characters that do not match the actual IQN. The numeral 1 and lowercase L character (l) are easy to mix up if they are entered manually. It is worth verifying that the IQN is defined correctly at this stage of problem determination.

Checking for a supported host configuration

IBM tests certain combinations of host OS, adapter, firmware levels, multi-path drivers, and SAN boot and clustering software. The approved list is published for each IBM Storwize code release through the [IBM System Storage Interoperation Center \(SSIC\)](#).

Additional interoperability matrix documents are published at the following locations:

- ▶ [V7.8.x Supported Hardware List, Device Drive, Firmware, and Recommended Software Levels for SAN Volume Controller](#)
- ▶ [V7.7.x Supported Hardware List, Device Driver, Firmware, and Recommended Software Levels for SAN Volume Controller](#)
- ▶ [V7.6.x Supported Hardware List, Device Driver, Firmware, and Recommended Software Levels for SAN Volume Controller](#)
- ▶ [V7.5.x Supported Hardware List, Device Driver, Firmware, and Recommended Software Levels for SAN Volume Controller](#)

An RPQ or SCORE request must be submitted for any configurations that are not included in these documents.

Checking the session limits per host

Consider the following session limit information:

- ▶ An IBM Storwize storage system has a limit of four sessions per node from each initiator IQN.
- ▶ There is a limit of one session per IBM Storwize node for any node in a host HA cluster configuration.

Checking for failed paths from the host

The next step is to identify failed paths from the host:

- ▶ Use multi-path driver commands to determine the path status. These commands are described for different operating systems in Chapter 7, “Configuring the IBM Storwize storage system and hosts for iSCSI” on page 83.
- ▶ To determine which paths are expected and which are missing, see the topology in 10.4.1, “Problem determination: Obtaining a basic configuration overview” on page 199.
- ▶ To check whether path errors are detected by the multipath driver, see 10.4.2, “Problem determination: Checking the network configuration” on page 201.

10.4.7 Problem determination: Checking for performance problems

Some of the attributes and host parameters that might affect iSCSI performance include the following things:

- ▶ Transmission Control Protocol (TCP) Delayed ACK
- ▶ Ethernet Jumbo frame
- ▶ Network bottleneck or over subscription
- ▶ iSCSI session login balance
- ▶ Priority Flow Control (PFC)
- ▶ Network errors

TCP Delayed ACK

TCP Delayed ACK is a technique that is used to improve network performance. It reduces protocol processing impact by combining several ACKs into a single response. The receiver usually returns only an ACK for every second segment.

A similar technique for reducing protocol processing impact is also used by the sender. The Nagle algorithm improves network efficiency by buffering smaller segments until an ACK is received for the previously transmitted segments.

There are certain workload patterns that can cause the combined effect of both these techniques to have a negative impact on iSCSI performance. Therefore, sometimes disabling the Delayed ACK on the host can improve iSCSI read I/O performance on the IBM Storwize storage system.

If there is evidence of degraded iSCSI read performance, update to the latest machine code. There are many enhancements that are introduced in Version 7.1.

If problems persist, consider disabling TCP Delayed ACK on the host if this option is available on the operating system. If the host has iSCSI offload network adapters with hardware initiators, the TCP Delayed ACK must be disabled from the adapter interface rather than OS. It might not be possible to disable this feature on some cards.

To disable Delayed ACK when you use software initiators, see the following knowledge base articles:

- ▶ For Windows hosts, see [Slow performance occurs when you copy data to a TCP server by using a Windows Sockets API program.](#)
- ▶ For VMware hosts, see [ESX/ESXi hosts might experience read or write performance issues with certain storage arrays.](#)

There is no option to disable TCP Delayed ACK in most standard Linux versions or in Emulex OCe10102 and OCe11102 adapters.

Ethernet Jumbo frames

The most commonly used size for Jumbo frames is 9000 bytes. You can enable Jumbo frames on the IBM Storwize storage system by running **cfgportip** with the **-mtu** parameter to specify the maximum transmission unit (MTU). To set an MTU of 9000 on port 1 in I/O group 0, use the following syntax:

```
cfgportip -mtu 9000 -iogrp 0 1
```

The setting can be changed back to the default value of 1500 bytes by running the following command:

```
cfgportip -defaultmtu -iogrp 0 1
```

Jumbo frames provide a performance benefit only if the network supports the larger MTU size from end to end. It is possible to test whether a ping packet can be delivered without fragmentation.

On a Windows host, the command syntax is as follows:

```
ping -t <iscsi target ip> -S <iscsi initiator ip> -f -l <new mtu size - packet overhead (usually 36)>
```

The following command can be used to check whether a 9000-byte MTU is set correctly on a Windows system:

```
ping -t -S 192.168.1.117 192.168.1.217 -f -l 8964
```

If successful, the reply should a result that is similar to the following example:

```
192.168.1.217: bytes=8964 time=1ms TTL=52
```

If Jumbo frames are not enabled on the host running the **ping**, the reply shows the following information:

```
ping: sendto: Message too long
```

If Jumbo frames are not enabled at the destination or a switch in between, the reply shows the following information:

```
Request timeout for icmp_seq 0
```

On a Linux host, the equivalent command syntax is as follows:

```
ping -l <source iscsi initiator ip> -s <new mtu size> -M do <iscsi target ip>
```

On VMware ESXi, the command syntax is as follows:

```
ping <iscsi target ip> -I <source iscsi initiator ip> -s <new mtu size - 28> -d
```

Network bottleneck or oversubscription

The IBM Storwize iSCSI solution design should avoid any bottlenecks and oversubscription. The network must be balanced to avoid any packet drops, which can affect storage performance.

It is also not unusual for a workload to change over time. Ideally, the solution planner factors in sufficient capacity and bandwidth to cater for future expansion. The system administrator should monitor activity patterns to identify new bottlenecks before they cause impact.

A support package includes cluster performance statistics for a period of around 4 hours with a 15-minute sampling interval. In many cases, the sampling interval must be reduced to the minimum of 5 minutes for a more useful set of statistics. However, this configuration reduces the coverage period to just over an hour before starting the support package collection. To change the sampling interval of the system, use the following command:

```
svctask startstats -interval interval_in_minutes
```

The acceptable range of this command is 1 - 60 minutes with intervals of 1 minute.

The IBM Storwize management GUI includes a basic performance monitoring function that can be selected from the Monitoring menu. This function can be combined with network data from the switches and server performance statistics from the hosts to provide a system-wide view. The Linux and AIX operating systems have the **iostat** tool, and Windows has `perfmon.msc /s`. In addition, tools such as IBM Spectrum Control can be used to provide more comprehensive performance management.

Use dedicated NICs on servers for host traffic to reduce network congestion and latency.

iSCSI session login balance

For the best performance, set up a single iSCSI session between an initiator and the target iSCSI port. Therefore, if there are two ports that are configured on a host, each must set up a single iSCSI session per iSCSI port on any SAN Volume Controller node. It does not help to configure multiple iSCSI sessions from an initiator port to multiple iSCSI target ports on the same node on the SAN Volume Controller storage system.

While you are configuring iSCSi ports on the SAN Volume Controller storage system, it is preferable to configure each iSCSI port on the same SAN Volume Controller node on a different subnet or VLAN.

For maximum availability, identical Ethernet port IDs on each node of a 2-node cluster *must* be configured on the same subnet or VLAN. In case the same iSCSI host connects to more than one I/O group, it is preferable to extend the same logic to the additional I/O groups.

The maximum number of sessions per node is limited to four per host. This limitation can result in the fifth session being closed. As a result, there is no control over which four sessions are used. They might not represent an even balance for optimal performance or redundancy. If there are more than four IP addresses that are configured on a SAN Volume Controller storage system and these addresses are configured on a mix of 1 Gb and 10-Gb ports, iSCSI sessions must be configured carefully to achieve the correct connectivity between the initiator and target ports.

Priority Flow Control

Use the **lspportip** command with a port ID number to show the detailed output. This command includes the following two fields in the output:

- ▶ `lossless_iscsi`
- ▶ `lossless_iscsi6`

If PFC was enabled on the switch, one or both of these fields should display a value of on to indicate that PFC is also enabled on the IBM Storwize storage system. If both values remain off, it might be due to the following reasons:

- ▶ The VLAN is not set for that IP. Check the following things:
 - For IP address type IPv4, check the `vlan` field in the `lspointip` output. It should not be blank.
 - For IP address type IPv6, check the `vlan_6` field in the `lspointip` output. It should not be blank.
 - If the `vlan` and `vlan_6` fields are blank, see 7.1.2, “Setting optional iSCSI settings on IBM Storwize storage systems” on page 85 for more information about configuring VLAN for iSCSI.
- ▶ The host flag is not set for that IP. Check the following things:
 - For IP address type IPv4, check the `host` field in the `lspointip` output. It should be yes.
 - For IP address type IPv6, check the `host_6` field in the `lspointip` output. It should be yes.
 - If the `host` and `host_6` fields are both set to no, set the host flag for the IP type by running the `cfgpointip` command.
- ▶ PFC is not set correctly on the switch.

If the VLAN is correctly set, and the host flag is also set, but the `lossless_iscsi` or `lossless_iscsi6` field is still showing as off, some switch settings might be missing or incorrect. Do the following tasks to verify the switch configuration:

- ▶ Verify that the priority tag is set for iSCSI traffic.
- ▶ Verify that PFC is enabled for the priority tag that is assigned to iSCSI CoS.
- ▶ Verify that DCBx is enabled on the switch.
- ▶ Consult the documentation for enabling PFC on your specific switch.
- ▶ Consult the documentation for enabling PFC on Red Hat Enterprise Linux (RHEL) and Windows hosts specific to your configuration.

Network errors

The packet retransmission count should be less than 0.2% to provide good performance. Even a relatively low level of dropped packets can have a significant impact on iSCSI I/O performance.

Using IOMeter to simulate a workload

Sometimes, it is necessary to generate a workload during initial installation, which allows multi-path failover to be tested, helps identify potential bottlenecks, simulates expected user workload patterns, and stress tests the entire solution to expose any marginal areas before commencing live operation.

IOMeter is a stress tool that can be used with Windows and Linux hosts. For more information about IOMeter, see the [IOMeter website](#).

Note: Select settings for IOMeter carefully. Ideally, the workload should simulate the expected user load and application activity. As described in “TCP Delayed ACK” on page 207, it is possible to induce degraded iSCSI performance with certain workload patterns.

iSCSI virtualization

This part describes Internet Small Computer System Interface (iSCSI) virtualization and how to plan, configure, secure, and troubleshoot connections.

This part describes the following topics:

- ▶ Chapter 11, “iSCSI virtualization overview” on page 213
- ▶ Chapter 12, “External virtualization of IBM Storwize storage systems” on page 227
- ▶ Chapter 13, “Virtualization of IBM Spectrum Accelerate storage systems” on page 239
- ▶ Chapter 14, “External virtualization of Dell Equallogic PS Series” on page 265
- ▶ Chapter 15, “Configuration and administration of iSCSI” on page 297
- ▶ Chapter 16, “Troubleshooting iSCSi virtualization” on page 319



iSCSI virtualization overview

This chapter provides an overview of how to virtualize back-end storage controllers behind SAN Volume Controller or IBM Storwize systems that are connected over Internet Small Computer System Interface (iSCSI).

This chapter uses the term iSCSI controllers to refer to back-end storage systems that are connected over iSCSI. The term Fibre Channel (FC) controller is used to refer to back-end storage systems that are connected over FC.

This chapter details the planning considerations that are needed when virtualizing an iSCSI controller. It starts with describing the fundamental aspects that differentiate FC controllers from iSCSI controllers, and how to model connectivity to iSCSI controllers. The connectivity options, security aspects, and configuration limits are described, and followed by detailed steps to virtualize iSCSI controllers.

This chapter contains the following topics:

- ▶ 11.1, “Planning considerations for iSCSI virtualization” on page 214
- ▶ 11.2, “iSCSI external virtualization steps” on page 220

11.1 Planning considerations for iSCSI virtualization

The iSCSI protocol differs from FC in many ways, and virtualizing iSCSI controllers requires a different set of considerations. This chapter starts with recognizing these differences, presents a way to model connectivity to iSCSI controllers, and lays out various considerations when you plan to virtualize storage controllers over iSCSI.

11.1.1 Fibre Channel versus iSCSI virtualization

There are some basic differences when using an FC fabric versus an Ethernet network, so the process of discovery and session establishment is different for iSCSI controllers.

Discovery and session establishment in Fibre Channel

Discovery in a switched FC fabric always happens through a name server. One of the switches in the fabric acts as a name server. When new entities (servers, storage, or virtualization appliances) are connected to the switch over FC, they automatically log in to the name server. When FC controllers join the fabric, they register the target capability. When an initiator logs in to the fabric, it queries the name server for targets on the fabric. It uses the discovery response to log in to the list of discovered targets. Therefore, discovery and session is an automated process in the case of FC controllers. The only prerequisite is to ensure that the initiators and targets are part of the same zone.

Discovery and session establishment in iSCSI

Discovery process refers to several things:

- Discovery controllers that are available in the network

The usage of name servers is optional for iSCSI. iSCSI initiators can be configured to directly connect to a target iSCSI controller.

Although name servers are used and supported by iSCSI, this function is not embedded in the Ethernet switches as with FC. The name servers must be configured as separate entities on the same Ethernet network as the iSCSI initiators and targets. An iSCSI target should explicitly register with an iSCSI Name Server (iSNS). The IP address of the name server must be explicitly configured on the target iSCSI controller. To do discovery of targets, the initiator must be explicitly configured with the name server address before you do discovery. Therefore, the discovery process in iSCSI is not automated as with FC.

After you connect, the initiator automatically sends a **sendtarget** request to the target controller to get a list of the IP addresses and IQN names. This action does not require any manual actions.

- Getting a list of the capabilities of the target

The iSCSI initiator and target exchange capabilities are part of the operational negotiation phase of iSCSI Login process. This process does not require any manual actions.

- Getting a list of available LUNs from the target

This is a SCSI workflow, and independent of the transport protocol that is used. All LUNs that are mapped to the initiator by using the target-specific LUN mapping rules are reported to the initiator by using SCSI commands.

Establishing sessions to iSCSI controllers from SAN Volume Controller or IBM Storwize storage systems

With FC connectivity, after an initiator discovers the target through the fabric name server, it automatically attempts to log in and establish sessions to the target ports through all zoned ports that are connected to the fabric.

While using iSCSI connectivity for connecting to back-end iSCSI controllers, connections must be set up manually.

Figure 11-1 depicts the overall differences while discovering and establishing sessions to iSCSI controllers versus FC controllers.

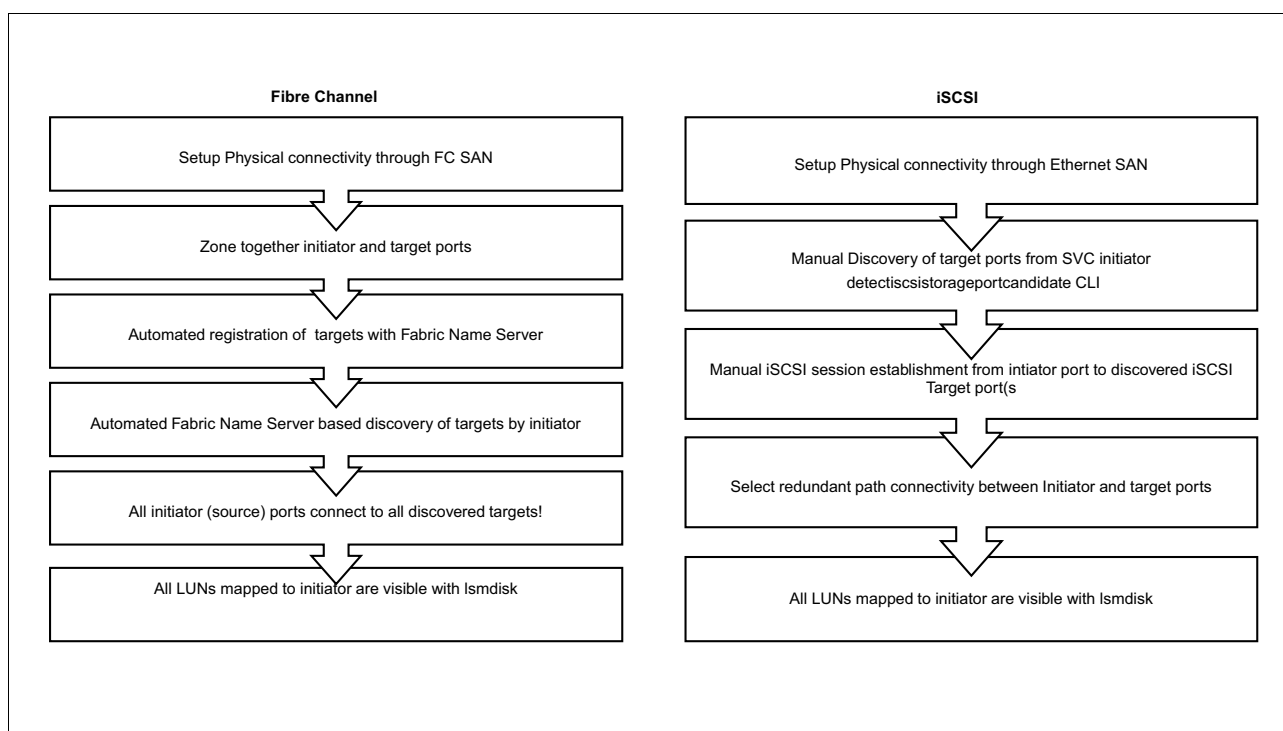


Figure 11-1 Differences while discovering and establishing sessions to iSCSI controllers versus Fibre Channel controllers

11.1.2 Storage port configuration model

This section describes some of the basic concepts that you must know to virtualize iSCSI controllers behind SAN Volume Controller or IBM Storwize.

Source Ethernet ports

Whenever a storage controller is virtualized behind SAN Volume Controller or IBM Storwize for high availability, it is expected that every node in the initiator system has at least one path to the target controller. More paths between each initiator node and the target controller help increase the available bandwidth and enable more redundancy for the initiator node/target connectivity.

The first step in the iSCSI controller configuration is to select the Ethernet port on the nodes of the initiator system. To ensure symmetric connectivity from every node, the configuration CLI for iSCSI virtualization refers to the source Ethernet port, which is the port number of the Ethernet port on every node in the system, as shown in Figure 11-2.

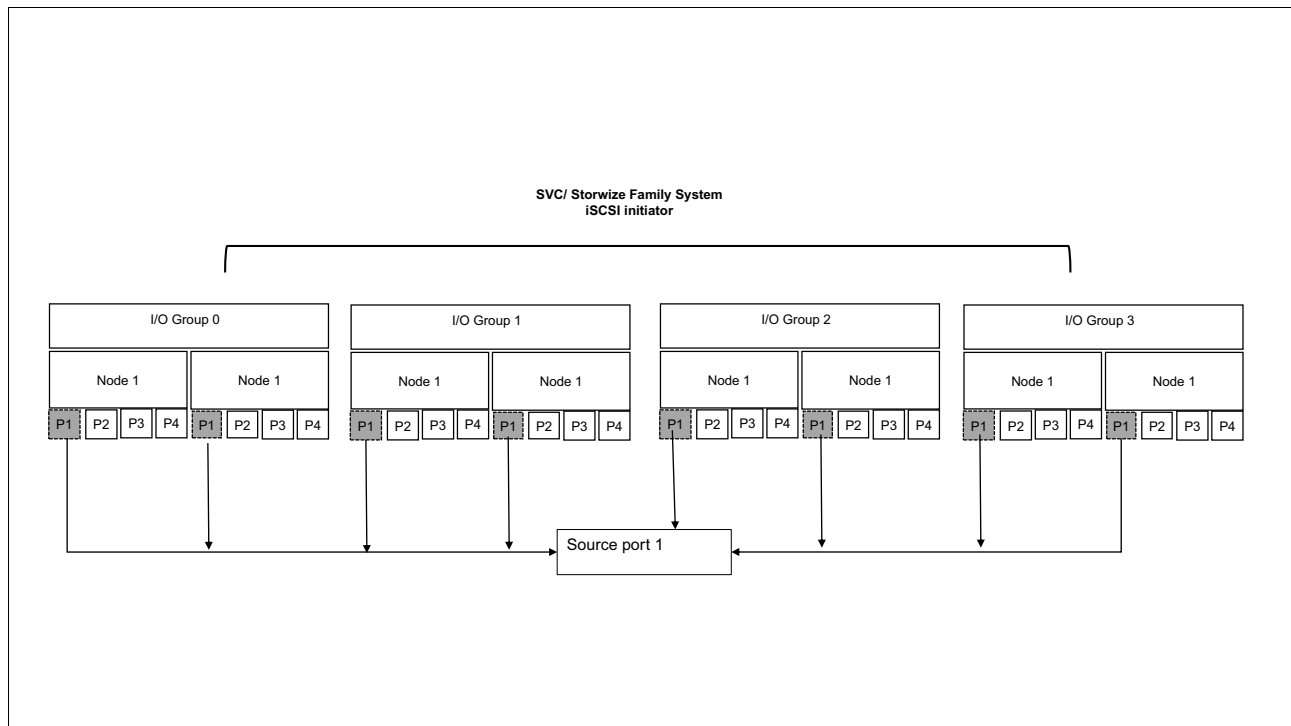


Figure 11-2 SAN Volume Controller and IBM Storwize family iSCSI initiator

Storage ports

A *storage port* is a target controller iSCSI endpoint. It is characterized as a target controller IQN, which is accessible through an IP address on an Ethernet port of the target controller. The concept is fundamental to the configuration interface for iSCSI virtualization. While discovering target controllers and establishing connectivity between initiator and target, the sessions are established between source Ethernet ports (as described in “Source Ethernet ports” on page 215) and target storage ports (as described in this section).

The total candidate storage ports on a controller can be calculated by using the following equation:

Number of candidate storage ports = (number of target IQNs) x (number of configured Ethernet ports)

Here are two examples:

► Example 1

An iSCSI controller has a single IQN per controller node and multiple Ethernet ports that are configured with IP addresses for iSCSI connectivity. All the LUNs from the controller are accessible through the same target IQN. If there are four IP addresses that are configured, then the target controller has four candidate storage ports.

► Example 2

A controller has four configured Ethernet ports for an iSCSI controller, but associates a single IQN per exported LUN. If there are four LUNs/ IQNs, and all IQNs are accessible through all Ethernet ports, then the number of candidate storage ports on the controller is 16.

To establish a connection from source Ethernet ports to target storage ports, the initiator must send a discovery request from the source Ethernet ports to a specific storage port. The discovery output is a subset of the overall storage port candidates for the target controller.

Selecting the I/O group versus system-wide connectivity

There are iSCSI controllers that associate a different target IQN with every LUN that is exported, which requires a separate session between the initiator nodes and target nodes for every LUN that is mapped to the initiator by the target.

Because there are limits on the total number of sessions that the initiators and targets support, you are not required to have connectivity to an iSCSI target controller from every node in the initiator system. You may, when working with controllers that associate an IQN per LUN, to configure connectivity to a target node from nodes of a single IO group only so that you can reduce the session sprawl for a target controller while ensuring redundant connectivity through nodes of a single IO group.

Before you decide whether to use I/O group versus system-wide connectivity to a target controller, see the target controller documentation to understand how many IQNs are exported by the controller and how many LUNs can be accessed through the target. Also, you must consider the maximum number of sessions that are allowed by the initiator (SAN Volume Controller or IBM Storwize storage system) and target controller.

By default, system-wide connectivity is assumed.

11.1.3 Controller considerations

All iSCSI controllers provide access to storage as SCSI LUNS. There are various different implementations of iSCSI controllers, which might require different configuration steps to virtualize the controller behind SAN Volume Controller or IBM Storwize storage systems.

Here are the primary external characteristics that differ across controllers:

► Number of target IQNs that are reported to initiators

An iSCSI initiator sends a discovery request to an iSCSI target by sending a **sendtarget** request. The target responds with a list of IQN and IP addresses that are configured on the controller to which the initiator is granted access.

Some controllers export only a single target IQN through multiple Ethernet ports. All the mapped LUNs from the controller are accessible to the initiator if the initiator logs in to the target IQN.

There are other controllers that export multiple target IQNs. Each IQN maps to a different LUN.

- Number of IP addresses that are visible to initiators

All iSCSI controllers enable connectivity through multiple Ethernet ports to enable redundant connections and to increase the throughput of the data transfer. When an initiator must connect to a target, it must specify a target IP address. Although most controllers enable setting up connections from initiator ports to target ports through the different target IP addresses, some controllers recommend establishing connections to a virtual IP address to enable the controller to load balance connections across the different Ethernet ports.

Depending on how many IQNs are available from a target and how many target IP addresses initiators can establish connections to, the configuration model from initiator to target can be chosen to provide sufficient redundancy and maximum throughput.

Having an IQN per LUN leads to many sessions being established, especially if there are many LUNs that are available from the storage controller. There are limits on the number of sessions that can be established from every node and the number of controllers that can be virtualized behind SAN Volume Controller or IBM Storwize storage systems.

Traditionally, every back-end controller (independent of the type of connectivity that is used) must have at least one session from every initiator node. To avoid session sprawl, for iSCSI controllers that map a target IQN for every LUN that is exported, there is an option that is available to establish sessions from a single IO group only. Also, if the target controller recommends connecting through only a single virtual IP of the target, the session creation process should consider distributing the sessions across the available initiator ports for maximum throughput.

Some controllers map a single target IQN for a group of clustered iSCSI controllers. In such a case, it is not possible to distinguish individual target controllers within the clustered system. Based on the number of Ethernet ports that are available on initiator nodes, the number of sessions can be configured to connect across Ethernet ports of multiple target controllers in the clustered system.

11.1.4 Stretched cluster and HyperSwap topology

iSCSI back-end controllers are supported in both stretched cluster and HyperSwap system topologies.

In a stretched system configuration, each node on the system can be on a different site. If one site experiences a failure, the other site can continue to operate without disruption. You can create an IBM HyperSwap topology system configuration where each I/O group in the system is physically on a different site. When used with active-active relationships to create HyperSwap volumes, these configurations can be used to maintain access to data on the system when power failures or site-wide outages occur.

In both HyperSwap and stretched configurations, each site is defined as an independent failure domain. If one site experiences a failure, then 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 automatically breaks a tie in case of a link failure between the two main sites.

Controllers that are connected through iSCSI can be placed in a system in stretched topology. When a controller is placed in a system with a stretched or HyperSwap topology, it should be visible and have connectivity to nodes only at a particular site, regardless of whether I/O Group or clusterwide connectivity was set while establishing sessions by using the **addiscsistorageport** CLI.

To enable connectivity only to nodes of a particular site, you can use an option site parameter of **addiscsistorageport**. When you run **addiscsistorageport**, you can specify **site_name** or **site_id** from the **lssite** CLI. The **lsscsistorage** command also shows which site (if any) was used to establish connectivity.

Use a third site to house a quorum disk or IP quorum application. Quorum disks cannot be on iSCSI-attached storage systems; therefore, iSCSI storage cannot be configured on a third site.

11.1.5 Security

The security features for connectivity to back-end controllers can be categorized in to two areas:

- Authorization

A target controller might require that any initiator node or nodes must present authentication credentials with every discovery or normal iSCSI session. To enable this function, one-way CHAP authentication is supported by most iSCSI controllers. The CLIs for discovery and session establishment can specify the user name and CHAP secret. These credentials, if specified, are stored persistently and used every time that a session is reestablished from initiator source ports to target ports.

- Access control

Access control is enabled at each target controller. A target controller creates host objects and maps volumes to hosts. Some controllers present multiple volumes that are mapped to a host through a single target IQN, and other controllers might present a unique IQN for each LUN. In the latter case, a discovery request from the initiator system to the target lists multiple target IQNs, and a separate session that is initiated by **addiscsistorageport** must be established for every discovered target IQN. In the former, a single invocation of **addiscsistorageport** is sufficient to provide the initiator nodes access to all volumes that are exported by the target controller to the host.

11.1.6 Limits and considerations

Quorum disks cannot be placed on iSCSI-attached storage systems. As a result, the following limits should be considered:

- iSCSI storage cannot be configured on a third site when using stretched or HyperSwap topology. Using IP quorum is preferred as a tie breaker in a stretched or HyperSwap topology.
- An iSCSI MDisk should not be selected when choosing a quorum disk.

SAN Volume Controller iSCSI initiators show only the first 64 discovered IQNs that are sent by the target controller in response to a discovery request from the initiator.

Connections can be established only from an initiator source to the first 64 IQNs for any back-end iSCSI controller.

A maximum of 256 sessions can be established from each initiator node to one or more target storage ports.

When a discovery request is sent from initiator source ports to target storage ports, the target sends a list of IQNs. In controllers that export a unique IQN per LUN, the list might depend on the number of configured LUNs in the target controller. Because each target IQN requires one or more redundant sessions to be established from all nodes in an initiator IO group / cluster, this situation can lead to session sprawl both at the initiator and target. A SAN Volume Controller connectivity limit of 64 target IQNs per controller helps manage this sprawl.

11.2 iSCSI external virtualization steps

This section describes the steps that are required to configure an IBM Storwize storage system as an iSCSI initiator to connect to an IBM Storwize or non IBM Storwize iSCSI back-end controller.

11.2.1 Port selection

List the Ethernet ports on the initiator Storwize nodes, which act as the initiator, and the Ethernet ports on the iSCSI controller, which serve as the back end or target. Ports with the same capability or speed should be chosen to avoid bottlenecks on the I/O path during data migration and virtualization. The ports listing can be seen on the initiator nodes by using the **svctaks lsportip** command. See the documentation for different target controllers to see how to view the Ethernet ports and speeds.

Note: The port speed for IBM Storwize ports is visible only when the port is cabled and the link is active.

11.2.2 Source port configuration

The Ethernet ports on the initiator SAN Volume Controller system can be configured for host attachment, back-end storage attachment, or IP replication. Back-end storage connectivity is turned off by default for SAN Volume Controller Ethernet ports. Configure the initiator ports by using the GUI or CLI (**svctask cfgportip**) and set the **storage/storage_6** flag to yes to enable the ports to be used for back-end storage connectivity.

Example 11-1 shows an example for enabling a source port on the initiator system for back-end storage controller connectivity by using the IPv4 address that is configured on the port.

Example 11-1 Source port configuration

```
IBM_2145:Redbooks_cluster1:superuser>cfgportip -node node1 -ip 192.168.104.109  
-mask 255.255.0.0 -gw 192.168.100.1 -storage yes 3
```

11.2.3 Target port configuration

Configure the selected ports on the target controller for host connectivity. For specific steps to configure ports on different back-end controllers that can be virtualized by the IBM Storwize initiator system, see the following chapters.

11.2.4 Host mapping and authentication settings on target controllers

You must create host mappings on the target system so that the target system recognizes the initiator system and presents volumes to it to be virtualized. For a target IBM Storwize controller, this process involves creating a host object, associating the IQN for initiator system's nodes with that host object, and mapping volumes to that host object (these are the volumes that the initiator system virtualizes). See specific target controller documentation for host attachment guides. The IQN of a node in an initiator Storwize system can be found by using the **svcinfo lsnodecanister** command.

Optionally, if the target controller requires authentication during discovery or login from the initiator, the user name and CHAP secret must be configured on the target system. For more information about how to accomplish this task on different controllers, see the following chapters.

11.2.5 Understanding the storage port model for a back-end controller

Different controllers expose a different number of IQNs and target port IP addresses. As a result, there might be a different number of storage ports, as described in “Storage ports” on page 216. The number of storage ports depends on the controller type. To identify which storage port model suits the back-end controller that is being virtualized, see 11.1.3, “Controller considerations” on page 217. This section also enables you to understand the discovery output and how sessions can be set up between the initiator source ports and target storage ports by running multiple iterations of discovery and configuration.

You should decide which iSCSI sessions to establish during the planning phase because this decision constrains physical aspects of your configuration, such as the number of Ethernet switches and physical ports that are required.

11.2.6 Discovering storage ports from the initiator

After all the initiator source ports or target storage ports are configured, you must discover the target storage ports by using the initiator source ports. This discovery can be achieved either by using the system management GUI or by using the **svctask discoveriscsistorageportcandidate** command. The command specifies the source ports' number and the target port IP to be discovered.

Optional parameters that can be specified include the user name and password for controllers that require CHAP authentication during discovery. If the user name is not specified and CHAP is specified, the initiator node IQN is used as the default user name.

Discovery can be done in two modes:

- ▶ **Cluster wide:** Discovery is done from the numbers source port on all nodes in the initiator system to a target port IP. This is the default mode for discovery, unless specified otherwise.
- ▶ **I/O Group wide:** A discovery request to target controller storage ports can be sent only from a selected source port of a specified I/O group in the initiator system.

The SAN Volume Controller iSCSI initiator sends an iSCSI **sendtargets** discovery request to the back-end controller target port IP that is specified. If sessions must be established to multiple target ports, discovery request must be sent to every target port after discovery, and then you must do session establishment from one port. A new discovery request purges the old discovery output.

Example 11-2 shows an example of triggering a cluster-wide discovery from source port 3 of the initiator system. The discovery request is sent to the controller with target port 192.168.104.190. The target controller requires authentication, so **chapsecret** is specified. Because the user name (which is optional) is not specified, the node IQN is used to set the user name while sending the discovery request.

Example 11-2 Triggering a cluster-wide discovery

```
IBM_2145:Redbooks_cluster1:superuser>detectiscsistorageportcandidate -srcportid 3
-targetip 192.168.104.190 -chapsecret secret1
```

Example 11-3 shows an example of triggering I/O group-specific discovery from source port 3 of the initiator system. The example is similar to Example 11-2 except that the discovery request is initiated only through the nodes of I/O group 2, assuming that the initiator system has an I/O group with I/O group ID 2. Also, in this example, the target controller does not require any authentication for discovery, so the user name and CHAP secret are not required.

Example 11-3 Triggering an I/O group discovery

```
IBM_2145:Redbooks_cluster1:superuser>detectiscsistorageportcandidate -iogrp 2
-srcportid 3 -targetip 192.168.104.198
```

11.2.7 Viewing the discovery results

The output of the successful discovery can be seen in the management GUI or by running **svcinfo lsiscsistorageportcandidate**. For each target storage port to which the discovery request is sent, the discovery output shows multiple rows of output, one per discovered target IQN.

Example 11-4 shows the result of the successful discovery. The discovered controller has only a single target IQN.

Example 11-4 Discovery results

```
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageportcandidate
id src_port_id target_ipv4      target_ipv6 target_iscsiname iogroup_list
configured status site_id site_name
0  3           192.168.104.190
iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1 1:1:--:-- no full
```

The first column in the discovery output shows the discovery row ID. This row ID is used in the next step while setting up sessions to the discovered target controller. The source port ID and target IP that are specified in the discovery request are shown next. The IQN of the discovered controller is shown under the **target_iscsiname** column.

The **iogroup_list** shows a colon-separated list of I/O groups with target ports.

The **configured** column indicates whether the discovered target IQN has any established sessions with source ports or target ports. The values are yes and no (no by default). This is the result of a previous discovery and session establishment.

The **status** column indicates whether the discovery was successful.

The results from a discovery request are valid only until the following actions occur:

- ▶ A new discovery is issued, which purges the previous discovery results before sending a new discovery request.
- ▶ T2/T3/T4 cluster recovery is done, which purges the previous discovery results.

In general, it is preferable to run discovery immediately before adding a session.

11.2.8 Adding sessions to discovered storage ports

The next step after viewing discovery results is to establish sessions between initiator source ports and the target storage ports, which are described in 11.2.7, “Viewing the discovery results” on page 222. This task can be done from the management GUI by selecting the discovered storage ports, or by running the **svctask addiscsistorageport** command and referencing the row number that is discovered by the **svcinfo lsiscsistorageportcandidate** command.

If the discovery output shows multiple IQNs corresponding to multiple LUNS on the target controller, run **addiscsistorageport** once for every discovered storage port, referencing the row number of the discovered storage port from the output of the **svcinfo lsiscsistorageportcandidate** command.

There are several optional parameters that can be specified while adding sessions to target storage ports. A user name and CHAP secret can be specified if the back-end controller being virtualized requires CHAP authentication for session establishment. If a user name is not specified and CHAP is specified, the initiator node IQN is used by default.

Sessions can be configured in two modes:

- ▶ Cluster-wide: Sessions can be configured from the numbers source port on all nodes in the initiator system to the specified target storage port (referencing the discovery output row number). This is the default mode for session establishment, unless specified otherwise.
- ▶ IO Group-wide: Sessions to a target controller storage port can be established only from the selected source port of a specified I/O group in the initiator system. This mode is the recommended mode when virtualizing controllers that expose every LUN as a unique IQN. Specifying a single I/O group for target controller connectivity helps ensure that a single target controller does not consume many of the overall, allowed back-end session count.

Another option that can be specified when establishing sessions from initiator source ports to target storage ports is the site. This option is useful when an initiator system is configured in a stretched cluster or HyperSwap configuration. In such configurations, a controller should be visible only to nodes in a particular site. The allowed options for storage controller connectivity are site1 and site 2. iSCSI controllers are not supported in site3, and are used for tie-breaking in split-brain scenarios.

Example 11-5 shows how to establish connectivity to the target controller that was discovered in Example 11-4 on page 222.

Example 11-5 Establishing connectivity to the target controller

```
IBM_2145:Redbooks_cluster1:superuser>addiscsistorageport -chapsecret secret1 0
```

The **addiscsistorageport** command can specify a site name or site ID in case the system is configured with multiple sites, as in the case of a stretched cluster or HyperSwap topology.

Example 11-6 shows how to establish connectivity to the target controller that was discovered in Example 11-4 on page 222 through nodes only in site 1.

Example 11-6 Establishing connectivity to the target controller

```
IBM_2145:Redbooks_cluster1:superuser>addiscsistorageport -site 1 -chapsecret
secret1 0
```

11.2.9 Viewing established sessions to storage ports

After sessions are established to the storage ports of a back-end controller, sessions to one or more storage ports can be viewed either in the management GUI or by using the **lsiscsistorageport** command. When you use the command, there are two views: a concise view and a detailed view.

The concise view provides a consolidated listing of all sessions from the initiator system to all back-end target iSCSI controllers. Each row of output refers to connectivity from initiator system nodes to a single target storage port. You can view this output by running the **svctask addiscsistorageport** command. Each invocation of the command results in a separate row of output, as shown by the output of the **lsiscsistorageport** command.

The detailed view can be used to see the connectivity status from every initiator node port to the target storage ports for a selected row of the output of **lsiscsistorageport**.

Example 11-7 shows the concise **lsiscsistorageport** view on the initiator system in the example that is illustrated by Example 11-1 on page 220 through Example 11-4 on page 222. The steps in the preceding examples are repeated to add connectivity to node2 of the target controller through source ports 3 and 4.

Example 11-7 Output of the lsiscsistorageport command

```
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageport
id src_port_id target_ipv4 target_ipv6 target_iscsiname
controller_id iogroup_list status site_id site_name
1 3 192.168.104.190 iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1
4 1:1:-:- full
2 3 192.168.104.192 iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node2
4 1:1:-:- full
3 4 192.168.104.191 iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1
4 1:1:-:- full
4 4 192.168.104.193 iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node2
4 1:1:-:- full
```

The first column specifies the row ID view and denotes the sessions that are established from the specified initiator node ports to a back-end controller target iSCSI qualified name (IQN) through a target Internet Protocol (IP) address. The value is 0 - 1024.

The column **src_port_id** is the source port identifier for the node Ethernet port number that is displayed in the **lsportip** output.

The columns **target_ipv4** / **target_ipv6** indicate the IPv4/ IPv6 address of the iSCSI back-end controller target port to which connectivity is established.

The **target_iscsiname** indicates the IQN of the iSCSI back-end controller to which the connectivity is established.

The `controller_id` indicates the controller ID that is displayed in the `lscontroller` output.

The `iogroup_list` indicates the colon-separated list of I/O groups through which connectivity is established to the target port. The values are 0 and 1:

- ▶ 0 indicates that the I/O group is available in the system, but discovery is either not triggered through the I/O group or discovery through the I/O group fails.
- ▶ 1 indicates that the I/O group is present and discovery is successful through the I/O group.
- ▶ - indicates that the I/O group is not valid or is not present in the system.

Status indicates the connectivity status from all nodes in the system to the target port. Here are the values:

- ▶ Full: If you specify a single I/O group by using the `addiscsistorageport` command and you establish the session from all nodes in the specified I/O group, the status is `full`.
- ▶ Partial: If you specify a single I/O group by using the `addiscsistorageport` command and you establish the session from a single node in the specified I/O group, the status is `partial`.
- ▶ None: If you specify a single I/O group by using the `addiscsistorageport` command and you do not establish the session from any node in the specified I/O group, the status is `none`.

The `site_id` and `site_name` parameters indicate the site ID / site name (if the nodes being discovered belong to a site). These parameters apply to stretched and HyperSwap systems.

Example 11-8 provides a sample output of the detailed view of the `lsiscsistorageport` command. The example shows an example for connectivity that is established to a back-end IBM Storwize controller. However, the output is similar to connectivity to other target controllers.

Example 11-8 Output of the `lsiscsistorageport` command

```
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageport 1
id 1
src_port_id 3
target_ipv4 192.168.104.190
target_ipv6
target_iscsiname iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1
controller_id 4
iogroup_list 1:1::-
status full
site_id
site_name
node_id 1
node_name node1
src_ipv4 192.168.104.199
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node1
connected yes
node_id 2
node_name node2
src_ipv4 192.168.104.197
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node2
connected yes
node_id 3
```

```
node_name node3
src_ipv4 192.168.104.198
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node3
connected yes
node_id 4
node_name node4
src_ipv4 192.168.104.196
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node4
connected yes
```

The detailed view enumerates all the fields that are visible in the concise view. While each row in the detailed view sums up the connectivity status across all nodes in an I/O group or all I/O groups of the initiator system, the detailed view provides a connectivity view from each initiator node part of the site/iogroup/cluster that is indicated in the concise view output. It helps to view and isolate connectivity issues to an initiator node and source port.

The `node_name` / `node_id` indicates the node name and node ID of the initiator node through which a session is established to a storage port that is identified by `target_ipv4` / `target_ipv6` and `target_iscsi_name`. The `node_name` and `node_id` are as referenced in other clustered system CLIs.

`src_ipv4` / `src_ipv6` refers to the IPv4/ IPv6 address of the source port that is referred to by `src_port_id` on the initiator node that is referenced by `node_name` / `node_id`.

`src_iscsiname` refer to the IQN of the source node that is referenced by `node_name` / `node_id`.

`connected` indicates whether the connection is successfully established from a specified source port (`src_port_id`) of an initiator node that is referenced by `node_name` / `node_id`, which has an IP address `src_ipv4` / `src_ipv6` to a target storage port with `target_iscsiname` IQN and IP address `target_ipv4` / `target_ipv6`. The values are yes and no.



External virtualization of IBM Storwize storage systems

This chapter contains a detailed description about using an IBM Storwize system as a back-end storage controller connected through Internet Small Computer System Interface (iSCSI).

In such a configuration, there are two Storwize systems: One that is acting as a back-end storage controller, and another that is virtualizing it. The system that is acting as the back end is an iSCSI target, and the system that is virtualizing it is an iSCSI initiator.

To avoid confusion, this chapter uses the term *target system* to refer to the system that is acting as the back end, and the term *initiator system* to refer to the system that is virtualizing it.

This chapter describes the following topics:

- ▶ 12.1, “Planning considerations” on page 228
- ▶ 12.2, “Target configuration” on page 229
- ▶ 12.3, “Initiator configuration” on page 232
- ▶ 12.4, “Configuration validation” on page 235

12.1 Planning considerations

This section describes the things that you should consider before using an IBM Storwize system as an iSCSI-connected back end.

You should decide which iSCSI sessions to establish during the planning phase because this decision constrains physical aspects of your configuration, such as the number of Ethernet switches and physical ports that are required.

Because each node has a unique IQN in SAN Volume Controller and IBM Storwize systems, it is possible to have sessions from each node of the initiator system to each node of the target system. To achieve maximum availability, each node in the initiator system should establish at least one session with each node in the target system. This is also known as *cluster-wide connectivity*.

Furthermore, for redundancy, each pair of nodes should have two sessions, with each session using a link across a different Ethernet switch. Each SAN Volume Controller or IBM Storwize system has a limit to the maximum number of external virtualization iSCSI sessions. In addition, there is no performance benefit from having more than two sessions between a pair of nodes. Therefore, do not establish more than two sessions between each pair.

To avoid bottlenecks in the paths between the initiator system and the target system, ensure that all of the ports that are involved in a connection are at the same speed. For example, when using a 10 Gbps source port on the initiator system, ensure that the target port is also 10 Gbps, and that the connection goes through a 10 Gbps switch. When establishing cluster-wide connectivity, the initiator system uses the same source port from each initiator-system node to connect to a given target port. Therefore, you should also ensure that each source port with a given index on the initiator system has the same speed.

This chapter uses the configuration that is shown in Figure 12-1 as an example, showing the steps that are required to set up a simple configuration that includes a Storwize target system.

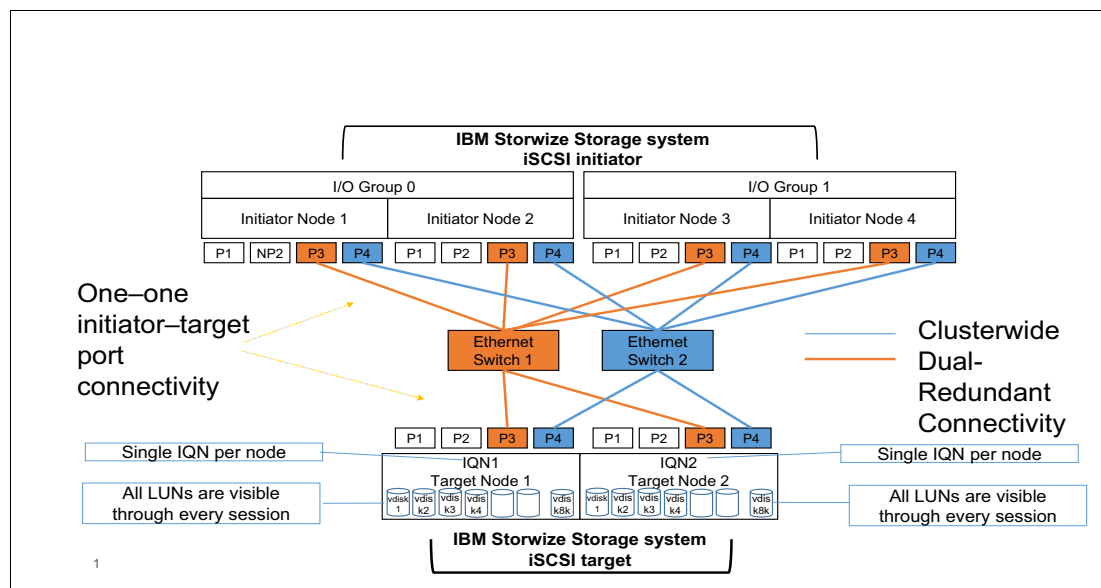


Figure 12-1 An example configuration with an iSCSI-connected Storwize target system

In Figure 12-1 on page 228, each node of the initiator system has two redundant paths to each node of the target system, with one using Ethernet Switch 1 (in orange) and one using Ethernet Switch 2 (in blue). The initiator system has cluster-wide connectivity to each node of the target system for maximum availability. Each source port of the initiator system can have at most one connection to each port of the target system.

12.1.1 Limits and considerations

From the perspective of the target system, the initiator system is an iSCSI-attached host. Therefore, the limits and considerations that apply to the use of a Storwize system as the target system include the iSCSI host-attachment limits and considerations for IBM Storwize systems. Section 4.5, “IBM Storwize family and iSCSI limits” on page 53 gives references to the detailed lists of limits and considerations for SAN Volume Controller and IBM Storwize products. Also, the limits and considerations for iSCSI external virtualization in general still apply. Section 11.1.6, “Limits and considerations” on page 219 gives references to detailed lists of these limits and restrictions.

12.1.2 Performance considerations

To maximize the performance benefit from using jumbo frames, use the largest possible MTU size in any given network configuration. There is no performance benefit from increasing the MTU size unless all the components in the path use the increased MTU size, so the optimal MTU size setting depends on the MTU size settings of the whole network. The largest MTU size that is supported by IBM Storwize systems is 9000. You can change the MTU size that is configured on a port on either the initiator system or the target system by using the `cfgport ip` command. For more information, see “Ethernet Jumbo frames” on page 208.

12.2 Target configuration

This section describes steps that must be carried out on the target system to enable it to be used as a back-end storage controller that is connected by iSCSI. This section assumes that the back-end system already has volumes set up to be virtualized by the initiator system. For more information, see *Implementing the IBM Storwize V5000 Gen2 (including the Storwize V5010, V5020, and V5030)*, SG24-8162.

From the perspective of the target system, the initiator system is a host that is attached by iSCSI. For this reason, configuring the target system in this case is the same as configuring an IBM Storwize system for iSCSI-connected host attachment, but using the initiator system as the “host”. For more information, see Chapter 7, “Configuring the IBM Storwize storage system and hosts for iSCSI” on page 83.

12.2.1 System layer

If the initiator system is a SAN Volume Controller or is in the replication layer, the Storwize target system must be in the storage layer. You can change the system layer by using the `chsystem` command. For more information about system layers, see [IBM Knowledge Center](#).

12.2.2 Host mappings

You must create host mappings so that the target system recognizes the initiator system and presents volumes to it to be virtualized. This task involves creating a host object, associating the initiator system's IQN with that host object, and mapping volumes to that host object (these are the volumes that the initiator system virtualizes).

In Figure 12-1 on page 228, the initiator system has four IBM Storwize node canisters. Because in SAN Volume Controller and IBM Storwize systems each node has its own IQN, the system has four IQNs. The target system should represent the initiator system with one host object, and associate all four IQNs with that host object.

Example 12-1 shows how to set up the host mappings in this example by using the CLI:

1. Create a host object and associate the initiator system's four IQNs with that host object by using **mkhost**.
2. Map a volume to that host object by using **mkvdiskhostmap**.

Example 12-1 Steps that are required to set up host mappings on the target system by using the CLI

```
IBM_Storwize:Redbooks_Backend_cluster:superuser>mkhost -name
Redbooks_initiator_system -iscsiname
iqn.1986-03.com.ibm:2145.redbookscluster1.node1,iqn.1986-03.com.ibm:2145.redbookscluster1.node2,iqn.1986-03.com.ibm:2145.redbookscluster1.node3,iqn.1986-03.com.ibm:2145.redbookscluster1.node4 -iogrp 0
Host, id [0], successfully created
IBM_Storwize:Redbooks_Backend_cluster:superuser>mkvdiskhostmap -host
Redbooks_initiator_system Vdisk_0
Virtual Disk to Host map, id [0], successfully created
```

The **-iogrp 0** option configures the target system to present volumes only from I/O group 0 to the host. In this case, this is the only I/O group in the target system. You can find the initiator system's IQNs by using the **lshost** command on the initiator system. It is also possible to add IQNs to an existing host object by using **addhostport**.

12.2.3 Authentication

SAN Volume Controller and IBM Storwize systems support only one-way (the target authenticates the initiator) CHAP authentication for external virtualization by using iSCSI. Therefore, you should configure only one-way CHAP authentication on the target system, although it is possible to configure two-way CHAP authentication on it (because SAN Volume Controller and IBM Storwize systems support two-way CHAP authentication for host attachment).

To configure one-way CHAP authentication on the target system, set a CHAP secret for the host object that represents the initiator system. This can be done either by using the **chhost** command or by using the iSCSI configuration pane in the GUI. For more information, see 5.2, "Configuring CHAP for an IBM Storwize storage system" on page 57.

Example 12-2 on page 231 demonstrates how to set a CHAP secret for the host object that represents the initiator system in the example that is illustrated by Figure 12-1 on page 228. Set a CHAP secret for the host object that represents the initiator system by using **chhost**. There is no feedback from a successful invocation.

Example 12-2 Configuring a CHAP secret on the target system for one-way (target authenticates initiator) CHAP authentication

```
IBM_Storwize:Redbooks_Backend_cluster:superuser>chhost -chapsecret secret1
Redbooks_initiator_system
```

After a CHAP secret is set, the target system automatically uses it to authenticate the initiator system. Therefore, you must specify the target system's CHAP secret both when you discover its ports and when you establish sessions with it from the initiator system. These procedures are described in 12.3, "Initiator configuration" on page 232.

12.2.4 Port configuration

For the initiator system to establish iSCSI sessions with the target system, the target system's iSCSI ports must have IP addresses. For more information about how to set the target system's iSCSI ports' IP addresses, see 7.1.1, "Setting the IBM Storwize iSCSI IP address" on page 84. While following these instructions, ensure that you note the IP addresses being set and to which ports they belong. You need this information when discovering the ports from the initiator and establishing sessions in 12.3, "Initiator configuration" on page 232.

In the example that is illustrated by Figure 12-1 on page 228, each node in the target system has four Ethernet ports: Two 1 Gbps ports and two 10 Gbps ports. Each connection should be between ports with the same speed to maximize the link performance.

In Figure 12-1 on page 228, the connections are from 10 Gbps ports on the initiator system to the 10 Gbps ports on the target system. These are ports 3 and 4 on each node of the target system. Therefore, you must assign IP addresses to these four 10 Gbps ports. Example 12-3 shows how to do this with the CLI. Use the **cfgportip** command to configure IP addresses for the iSCSI ports.

Remember: There is no feedback from a successful invocation.

Tip: In this example, the ports do not use VLAN tagging. If you require VLAN tagging, for example, to use Priority Flow Control (PFC), you also must use the **-vlan** or **-vlan6** options.

Example 12-3 Configuring IP addresses for the target system's iSCSI ports

```
IBM_Storwize:Redbooks_Backend_cluster:superuser>cfgportip -node node1 -ip
192.168.104.190 -mask 255.255.0.0 -gw 192.168.100.1 3
IBM_Storwize:Redbooks_Backend_cluster:superuser>cfgportip -node node1 -ip
192.168.104.191 -mask 255.255.0.0 -gw 192.168.100.1 4
IBM_Storwize:Redbooks_Backend_cluster:superuser>cfgportip -node node2 -ip
192.168.104.192 -mask 255.255.0.0 -gw 192.168.100.1 3
IBM_Storwize:Redbooks_Backend_cluster:superuser>cfgportip -node node2 -ip
192.168.104.193 -mask 255.255.0.0 -gw 192.168.100.1 4
```

12.3 Initiator configuration

This section describes steps that must be carried out on the initiator system to virtualize LUNs that are presented by an IBM Storwize target system that is connected with iSCSI. Part 3, “iSCSI virtualization” on page 211 contains instructions for configuring a SAN Volume Controller or IBM Storwize initiator system for external virtualization of a storage controller that is connected by using iSCSI. Those instructions are generic regarding the back-end controller that is used as the target system. These instructions are specific for external virtualization of an IBM Storwize target system, making specific reference to the example that is illustrated by Figure 12-1 on page 228.

12.3.1 Establishing connections and sessions

The initiator system can establish sessions with the target system by using either the CLI or the GUI. You should already know which connections you want to establish during the planning phase of an installation to plan the network architecture, which depends on the planned iSCSI connections. Section 12.1, “Planning considerations” on page 228 describes the connections that should be established between a SAN Volume Controller or IBM Storwize initiator system and an IBM Storwize target system.

In the example that is shown in Figure 12-1 on page 228, there are 16 iSCSI sessions to establish (two sessions from each of four nodes in the initiator system to each of two nodes in the target system). These 16 sessions are treated by the target system as four groups of four sessions; each group of four sessions is encapsulated in an iSCSI storage port object.

For an IBM Storwize target system, you should configure cluster-wide connectivity to achieve maximum availability. Therefore, each iSCSI storage port object in the example includes a session from each node of the initiator system, of which there are four. Table 12-1 shows the details of the four iSCSI storage port objects in this example.

Table 12-1 The iSCSI storage port objects in the example that is illustrated by Figure 12-1 on page 228

Initiator port ID	Initiator I/O group	Target node name	Target port ID	Target port IP address
3	All	node1	3	192.168.104.190
4	All	node1	4	192.168.104.191
3	All	node2	3	192.168.104.192
4	All	node2	4	192.168.104.193

You can configure these iSCSI sessions by using either the CLI or the GUI on the initiator system. Example 12-4 on page 233 shows how to configure one of the iSCSI storage port objects from Table 12-1, in particular the one that is described in the first row of the table.

With both the **detectiscsistorageportcandidate** and the **addiscsistorageportcandidate** commands, the **-chapsecret** option specifies the CHAP secret that is used for one-way (the target authenticates the initiator) CHAP authentication. Include this option to use the CHAP secret that is set on the target system in 12.2.3, “Authentication” on page 230. Alternatively, do not include this option if the target system has no CHAP secret. To configure cluster-wide connectivity, do not use the **-iogrp** option with either CLI because doing so configures connectivity to only one I/O group.

Example 12-4 Creating an iSCSI storage port with the CLI

```
IBM_2145:Redbooks_cluster1:superuser>detectiscsistorageportcandidate -srcportid 3 -targetip  
192.168.104.190 -chapsecret secret1  
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageportcandidate  
id src_port_id target_ipv4 target_ipv6 target_iscsiname  
iogroup_list configured status site_id site_name  
0 3 192.168.104.190 iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1  
1:1::- no full  
IBM_2145:Redbooks_cluster1:superuser>addiscsistorageport -chapsecret secret1 0
```

Figure 12-2 and Figure 12-3 on page 234 show the steps that are required on the initiator system's GUI to configure all of the iSCSI storage port objects that are described in Table 12-1 on page 232. Figure 12-2 shows the Add External iSCSI Storage wizard. Access this wizard by clicking **Add External iSCSI Storage** in the upper left of the **Pools** → **External Storage** pane.

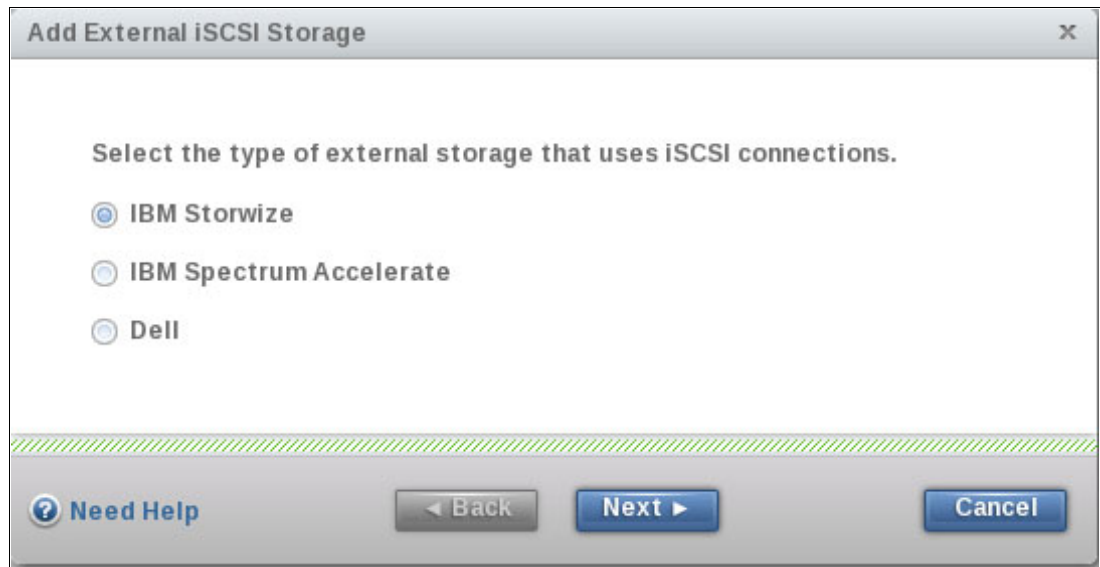


Figure 12-2 The first stage of the Add External iSCSI Storage wizard

In Figure 12-2, click **IBM Storwize** and then **Next** to configure the iSCSI storage port objects to virtualize an IBM Storwize target system.

Figure 12-3 shows the second step of the Add External iSCSI Storage wizard. The details that are entered into the wizard in Figure 12-3 result in the initiator system creating all of the iSCSI storage port objects that are described in Table 12-1 on page 232, which configures all of the iSCSI sessions in the example that is illustrated by Figure 12-1 on page 228.

Figure 12-3 The second step of the Add External iSCSI Storage wizard

The CHAP secret field sets the CHAP secret that is used for one-way (the target authenticates the initiator) CHAP authentication, both when discovering and when establishing sessions with the target system. Enter the CHAP secret that was set on the target system in 12.2.3, “Authentication” on page 230, or leave the field blank if the target system has no CHAP secret configured.

Each Target port on remote storage field corresponds to an iSCSI storage port object that the wizard creates. Each such object has cluster-wide connectivity to maximize availability. The source port on the initiator system for each iSCSI storage port object will be the port that is selected in the Select source port list. The target IP address for each iSCSI storage port object will be the IP address that is entered into the Target port on remote storage field.

To maximize availability and have redundancy in case of a path failure, you should configure two redundant connections to each node of the target system. To enforce this setting, the wizard does not allow you to continue unless all four Target port on remote storage fields are complete.

12.4 Configuration validation

When you complete the necessary steps to configure both the target system and the initiator system, you can verify the configuration by using the CLI on both systems.

You can use the **lsiscsistorageport** command to view information about the configured iSCSI storage port objects on the initiator system. Example 12-5 shows the concise **lsiscsistorageport** view on the initiator system in the example that is illustrated by Figure 12-1 on page 228.

Example 12-5 The lsiscsistorageport concise view

```
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageport
id src_port_id target_ipv4 target_ipv6 target_iscsiname
controller_id iogroup_list status site_id site_name
1 3 192.168.104.190 iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1
4 1:1::- full
2 3 192.168.104.192 iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node2
4 1:1::- full
3 4 192.168.104.191 iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1
4 1:1::- full
4 4 192.168.104.193 iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node2
4 1:1::- full
```

The view in Example 12-5 shows all of the configured iSCSI storage port objects, each of which is characterized by a source port ID, target IP address, and target IQN. The view shows some basic connectivity and configuration information for each object. Each object in this view corresponds to a row in Table 12-1 on page 232. In this example, the entry in the I/O group list field for each object is 1:1::- and the entry in the status field for each object is full, which indicates that the initiator system has good connectivity to the target system.

The entries in the I/O group list fields are colon-separated lists of keys. Each place in the list describes the connectivity status of an I/O group. The 1 key in the first two places indicates that the first two I/O groups are both meant to have connectivity, and actually do have connectivity. The minus (-) key in the final two places indicates that the final two I/O groups are not meant to have connectivity. In this case, this is because the iSCSI storage port object is configured to have cluster-wide connectivity, but the initiator system has only two I/O groups.

A 0 key in any place indicates that some nodes from the associated I/O group are meant to have connectivity but do not. The first place in the list always refers to I/O group 0, the second place to I/O group 1, and so on; this is regardless of which I/O groups are actually present in the system.

The full entries in the status fields indicate that every node that should have connectivity does have connectivity. An entry of partial indicates that only some nodes that are meant to have connectivity do have connectivity. An entry of none indicates that no nodes that are meant to have connectivity do have connectivity.

The **lscsistorageport** command also allows a detailed view that gives more detailed information about a specific iSCSI storage port object. If the concise view shows that the initiator system does not have full connectivity to the target system, you can use the detailed view to see which nodes do not have connectivity.

Example 12-6 shows the **lscsistorageport** detailed view for the first iSCSI storage port object on the initiator system in the example that is illustrated by Figure 12-1 on page 228.

Example 12-6 The lscsistorageport detailed view

```
IBM_2145:Redbooks_cluster1:superuser>lscsistorageport 1
id 1
src_port_id 3
target_ipv4 192.168.104.190
target_ipv6
target_iscsiname iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1
controller_id 4
iogroup_list 1:1::-
status full
site_id
site_name
node_id 1
node_name node1
src_ipv4 192.168.104.199
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node1
connected yes
node_id 2
node_name node2
src_ipv4 192.168.104.197
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node2
connected yes
node_id 3
node_name node3
src_ipv4 192.168.104.198
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node3
connected yes
node_id 4
node_name node4
src_ipv4 192.168.104.196
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node4
connected yes
```

Example 12-6 shows the detailed view for the iSCSI storage port object with ID 1, which is the object that is described in the first row of Table 12-1 on page 232. In addition to the fields in the concise view, the detailed view also gives per-node connectivity information. In this case, the entry in the connected field in each block of the output is yes, indicating that every node in the initiator system has connectivity to the target port.

For a full description of the **lscsistorageport** command, see [IBM Knowledge Center](#).

Example 12-7 The lshost detailed view

The view on the target system that is shown in Example 12-7 provides detailed information about the host object corresponding to the initiator system. The value in the port count field is 4, which is the number of IQNs that are associated with this host object. This value is correct because there are four nodes in the initiator system. The value in the I/O group count field is 1 because only one I/O group from the target system is configured to present volumes to the initiator system.

For each block of the output relating to a specific IQN of the host system, the value in the `nodes_logged_in` field is 4, and the value in the `state` field is `active`. This value indicates that there is good connectivity to each node of the initiator system. The state fields give the number of logins that nodes from the target system have with that IQN from the initiator system.

Chapter 12. External virtualization of IBM Storwize storage systems 237

For a full description of the **lshost** command, see [IBM Knowledge Center](#).

In addition to the information in the **lscsistorageport** and **lshost** views, there is further relevant information in the **lscontroller** and **lportip** views. The **lscontroller** view on the initiator system contains information about the controller object that represents the target system. It is documented in [IBM Knowledge Center](#).

The **lportip** view on either the initiator system or the target system contains information about the IP ports on that system. It is documented in [IBM Knowledge Center](#).



Virtualization of IBM Spectrum Accelerate storage systems

This chapter describes how to virtualize external storage systems from the IBM Spectrum Accelerate™ family as Internet Small Computer System Interface (iSCSI) targets, including IBM XIV® systems, IBM FlashSystem® A9000 systems, and IBM Spectrum Accelerate software-defined storage (SDS) solutions.

The chapter contains the following sections:

- ▶ 13.1, “Planning considerations” on page 240
- ▶ 13.2, “Target configuration” on page 243
- ▶ 13.3, “Initiator configuration” on page 254

For comprehensive descriptions of concepts, architecture, and implementation of these storage systems see the following IBM Redbooks publications:

- ▶ *IBM XIV Storage System Architecture and Implementation*, SG24-7659
- ▶ *IBM FlashSystem A9000 and IBM FlashSystem A9000R Architecture and Implementation*, SG24-8345
- ▶ *IBM Spectrum Accelerate Deployment, Usage, and Maintenance*, SG24-8267

13.1 Planning considerations

IBM Spectrum Accelerate storage systems support horizontal scale-out capability, which means a single system can have 3 - 15 nodes (or modules). The connectivity paradigm for virtualization by a SAN Volume Controller or IBM Storwize system is determined by the characteristics of the XIV, IBM FlashSystem A9000, and IBM Spectrum Accelerate iSCSI targets:

- ▶ Each target port has its own IP address.
- ▶ The system has a single system-wide IQN that handles all connections.
- ▶ All iSCSI LUNs are visible through every iSCSI session.

For redundancy, you must connect at least two target nodes to two or more SAN Volume Controller or IBM Storwize initiator ports through different Ethernet switches. More connections can be configured between a source system and a target, depending on the availability of source ports with storage connectivity enabled. Figure 13-1 shows the connection schema.

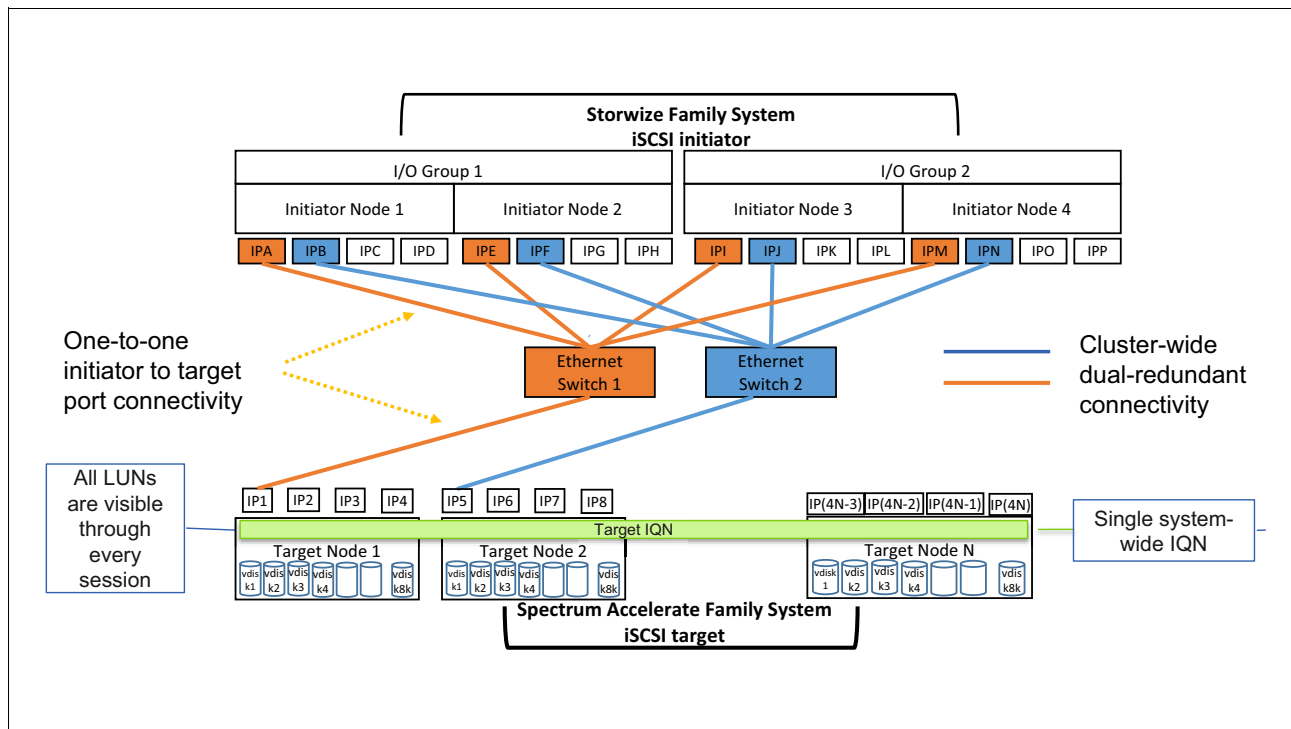


Figure 13-1 Connecting a virtualized XIV, IBM FlashSystem A9000, IBM FlashSystem A9000R, or IBM Spectrum Accelerate iSCSI storage system

In this example, the IBM Storwize system consists of two control enclosures (I/O groups) with two initiator nodes per enclosure. Each node has four Ethernet ports, labeled as IPA (first port of the first node) to IPP (fourth port of the fourth node). Two ports per node are configured as initiator ports through two switches to target ports of the IBM Spectrum Accelerate system. The other two Ethernet ports on each node are not configured.

The first ports (orange) on each initiator and target node are connected through Ethernet switch 1. The second ports (blue) on each initiator and target node are connected through Ethernet switch 2. Because of the system-wide IQN, each Storwize initiator node can access all IBM Spectrum Accelerate LUNs.

When you define initiator port connections, the configuration applies to all the ports on the system. That means in this example, the following items are true:

- ▶ Port IP1 on IBM Spectrum Accelerate node 1 is the target port for IBM Storwize source ports IPA, IPE, IPI, and IPM.
- ▶ Port IP5 on IBM Spectrum Accelerate node 2 is the target port for IBM Storwize source ports IPB, IPF, IPJ, and IPN.

Figure 13-1 on page 240 shows the minimum configuration with redundancy. In this configuration, extra Ethernet ports remain unconfigured but can be connected to increase the throughput. You can connect to as many XIV target ports across nodes equal to SAN Volume Controller or IBM Storwize initiator ports.

Important: The current implementation of iSCSI virtualization supports only one XIV target port per SAN Volume Controller or IBM Storwize initiator port. You must connect different target ports to different initiator ports.

In Figure 13-1 on page 240, the initiator ports IPC, IPD, IPG, IPH, IPK, IPL, IPO, and IPP remain unconfigured. On the target nodes, ports IP2, IP3, IP4, IP6, IP7, and IP8 are also unconfigured. Initiator ports IPC, IPG, IPK, and IPO can be connected through Ethernet switch 1 to any unused target port across the target nodes. When these connections are configured, another path between the source and target nodes is created. Similarly, the initiator ports IPD, IPH, IPL, and IPP can be connected through Ethernet switch 2 to any unused target port across the target nodes.

13.1.1 Limits and considerations for IBM XIV, IBM FlashSystem A9000, and IBM Spectrum Accelerate

For XIV, IBM FlashSystem A9000, and IBM Spectrum Accelerate systems, the number of iSCSI ports per node or module varies by model. For more information, see the following publications:

- ▶ *IBM XIV Storage System Architecture and Implementation*, SG24-7659
- ▶ *IBM FlashSystem A9000 and IBM FlashSystem A9000R Architecture and Implementation*, SG24-8345
- ▶ *IBM Spectrum Accelerate Deployment, Usage, and Maintenance*, SG24-8267

Here are the minimum code versions for iSCSI virtualization:

- ▶ XIV Gen3 firmware 11.2.0.c
- ▶ IBM FlashSystem A9000 firmware 12.0.0
- ▶ IBM Spectrum Accelerate software 11.5.x

XIV, IBM FlashSystem A9000, and IBM Spectrum Accelerate systems do not support IP link aggregation. Ethernet ports cannot be bonded.

The [SAN Volume Controller and IBM Storwize Family iSCSI Storage Attachment Support Matrix](#) provides up-to-date information.

13.1.2 Performance considerations

The default maximum transmission unit (MTU) of these storage systems is 4500 bytes. To match the SAN Volume Controller or IBM Storwize initiators, you should set the MTU to 9000 bytes. That is the upper limit for the MTU on XIV, IBM FlashSystem A9000, and IBM Spectrum Accelerate systems. As a prerequisite, you must configure the whole network to this MTU value if it is supported by your switches and routers. If you want an MTU other than the default, then it must be configured for each target port of the XIV, IBM FlashSystem A9000, or IBM Spectrum Accelerate system.

Tip: If the MTU that is being used by the XIV system is higher than the network can transmit, the frames are discarded. The frames are discarded because the do-not-fragment bit is normally set to on. Use the **ping -l** command to test the specification of the packet payload size from a Windows workstation in the same subnet. A **ping** command normally contains 28 bytes of IP and ICMP headers plus payload. Add the **-f** parameter to prevent packet fragmentation.

For example, the **ping -f -l 8972 10.1.1.1** command sends a 9000-byte frame to the 10.1.1.1 IP address (8972 bytes of payload and 28 bytes of headers). If this command succeeds, then you can configure an MTU of 9000 in the XIV GUI or XCLI.

13.1.3 Migration considerations

If iSCSI hosts are already attached to the XIV system, then their target ports are often configured with a smaller MTU size. Typical values are 1500 or 4500 bytes. Setting the MTU size to 9000 for virtualization and attaching these hosts to the SAN Volume Controller or IBM Storwize system can cause performance and stability issues during the migration because the XIV ports send large frames to the SAN Volume Controller or IBM Storwize nodes and to the hosts.

Therefore, you must consider this issue in your migration plan. Here are two options:

- You use different XIV target ports for SAN Volume Controller or IBM Storwize nodes and for the hosts. The target ports for the SAN Volume Controller or IBM Storwize initiators are reconfigured with an MTU of 9000 bytes (or the best value that the switches and routers support). The target ports for the hosts remain unchanged until hosts do not need these ports anymore.

When all migration tasks are complete, the target ports for the hosts can be reconfigured for usage as extra virtualization ports.

- Alternatively, you can configure the SAN Volume Controller or IBM Storwize nodes to use the smaller MTU size for migration. Therefore, these initiator nodes and the hosts can share target ports.

When all migration tasks are complete, and SAN Volume Controller or IBM Storwize nodes are the only initiators for the XIV ports, you change the MTU size to the optimal value.

13.2 Target configuration

To configure your XIV, IBM FlashSystem A9000, and IBM Spectrum Accelerate system, you must install the Hyper-Scale Manager graphical user interface (GUI) and the XIV command-line interface (XCLI) on your workstation. You log in to the GUI or XCLI and add the iSCSI target system to the inventory. (For more information about how to configure the GUI and XCLI, see the IBM Redbooks publications for your system.)

Note: The examples show an XIV Storage System (machine type 2810), but apply also to IBM FlashSystem A9000 and IBM Spectrum Accelerate systems.

Every XIV system has a unique IQN. The format of the IQN is simple, and includes a fixed text string followed by the last digits of the system's serial number.

Important: Do not attempt to change the IQN. If you need to change the IQN, you must engage IBM Support.

To display the IQN of the XIV Storage System, complete the following steps:

1. From the XIV GUI, click **Systems** → **System Settings** → **System**.
2. The Settings dialog box opens. Select the **Parameters** tab, as shown in Figure 13-2.

If you are displaying multiple XIV systems from the All Systems view, you can right-click an XIV system, and select **Properties** → **Parameters** to get the same information.

The screenshot shows a window titled "XIV IBM-SVC-XIV Settings" with a close button (X) in the top right corner. On the left is a sidebar with five tabs: "General", "Parameters", "Multi-tenancy", "SNMP", and "Misc". The "Parameters" tab is selected. The main area displays several configuration items, each with a label and a corresponding input field or dropdown menu:

- iSCSI Name:** A text field containing "iqn.2005-10.com.xivstorage:041529".
- Time Zone:** A dropdown menu showing "GMT".
- NTP Server:** An empty text field.
- DNS Primary:** An empty text field.
- DNS Secondary:** An empty text field.
- Use IPv6:** A dropdown menu showing "Yes".
- Volume Default SSD Caching:** A dropdown menu.
- Application Administrator Capabilities:** A dropdown menu showing "Basic".
- Interconnect MTU:** A text field containing "9000".

At the bottom of the dialog are two buttons: "Update" and "Cancel".

Figure 13-2 XIV system settings with iSCSI IQN

To show the same information in the XCLI, run the XCLI **config_get** command, as shown in Example 13-1.

Example 13-1 XIV system settings in XCLI

```
XIV IBM-SVC-XIV>>config_get
Name                               Value
dns_primary                        XIV IBM-SVC-XIV
dns_secondary                      Unknown
system_name                        XIV IBM-SVC-XIV
snmp_location                      Unknown
```

snmp_contact	Unknown
snmp_community	XIV
snmp_trap_community	XIV
snmp_type	V2C
snmpv3_user	
snmpv3_encryption_type	AES
snmpv3_encryption_passphrase	****
snmpv3_authentication_type	SHA
snmpv3_authentication_passphrase	****
system_id	41529
machine_type	2810
machine_model	999
machine_serial_number	9041529
machine_unique_id	a9d61f84264c42a8895b3ccaae95fb2e
email_sender_address	
email_reply_to_address	
email_subject_format	{severity}: {description}
iscsi_name	iqn.2005-10.com.xivstorage:041529
ntp_server	
support_center_port_type	Management
isns_server	
ipv6_state	enabled
ipsec_state	disabled
ipsec_track_tunnels	no
impending_power_loss_detection_method	UPS

After installing the XIV management tools and collecting the necessary information, you configure the XIV system in three steps:

1. Configure the iSCSI target ports (13.2.1, “Port configuration” on page 245).
2. Define the SAN Volume Controller or IBM Storwize initiator nodes as hosts (13.2.2, “Host mappings and authentication” on page 247).
3. Map XIV LUNs to these nodes (13.2.3, “Mapping XIV LUNs to the SAN Volume Controller or IBM Storwize system” on page 253).

13.2.1 Port configuration

To set up the iSCSI ports, complete these steps:

1. From the XIV GUI, select **View** → **Host and Clusters** → **iSCSI Connectivity**, and then click **Define IP Interface - iSCSI**.
2. In the window that opens (Figure 13-3), enter the parameters for one iSCSI target port:
 - If you are managing multiple XIV systems from the GUI, select the iSCSI target system from the Systems drop-down menu.
 - In the Name field, enter a text string to identify this port in other GUI panes and XCLI output. In the example, the string describes the port as Module 1, iSCSI port 1.
 - Specify the IP address, the netmask, and the default gateway in the next fields.
 - Set the MTU to 9000, if it is supported by your network.
 - From the drop-down menu **Node** and the **Port Number** switch, select the module and its port to which these settings apply. (The Port Number switch varies, depending on the ports that are available in this module.)
3. Click **Define** to complete the IP interface and iSCSI setup.

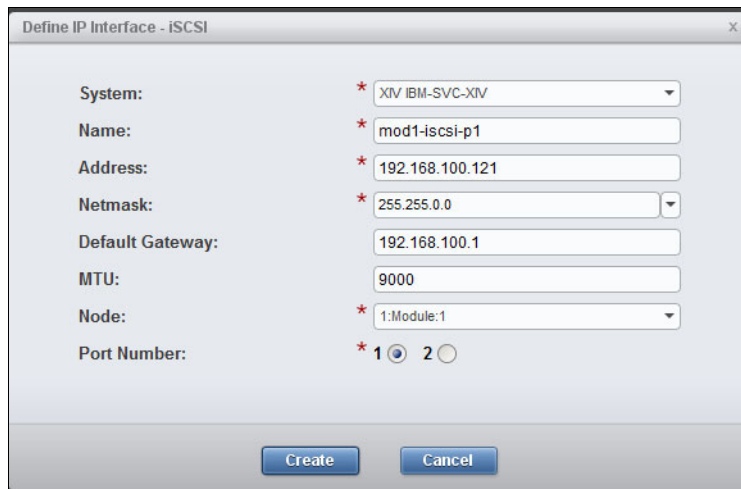


Figure 13-3 Defining an IP interface for iSCSI

To configure iSCSI ports by using the XCLI, enter the **ipinterface_create** command, as shown in Example 13-2.

Example 13-2 Defining an IP interface for iSCSI by using XCLI

```
XIV IBM-SVC-XIV>>ipinterface_create ipinterface="mod1-iscsi-p1"  
address=192.168.100.121 netmask=255.255.0.0 module=1:Module:1 ports="1"  
gateway=192.168.100.1 mtu=9000
```

As explained in 13.1, “Planning considerations” on page 240, you must configure at least two iSCSI target ports in different XIV modules. Therefore, the minimum configuration of iSCSI ports in the XIV GUI looks like Figure 13-4.

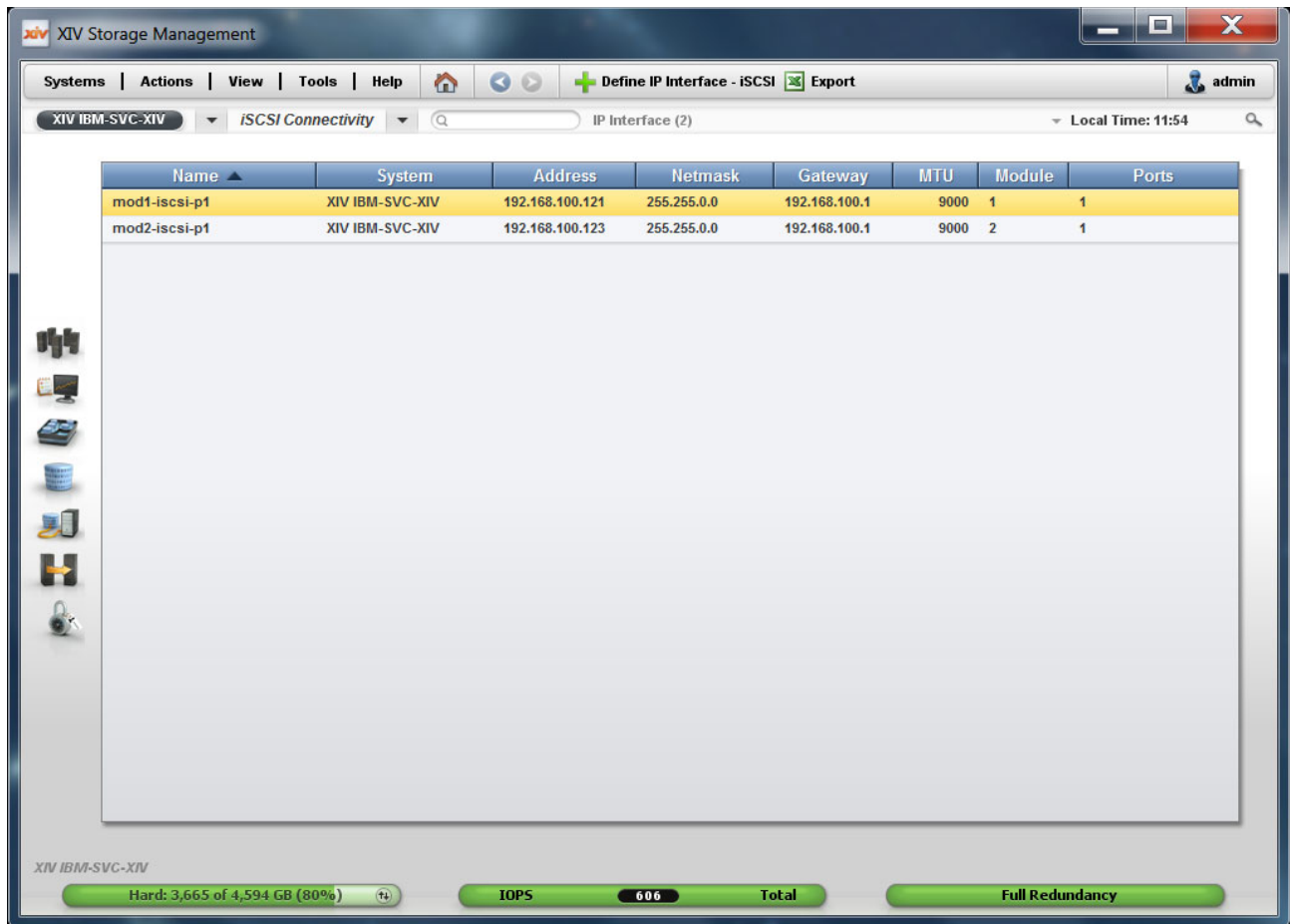


Figure 13-4 Minimum iSCSI port configuration in the XIV GUI

To show the same information in the XCLI, run the XCLI `ipinterface_list` command, as shown in Example 13-3.

Example 13-3 Minimum iSCSI port configuration in XCLI

```
XIV IBM-SVC-XIV>>ipinterface_list
Name          Type          IP Address      Network Mask    Default Gateway  IPv6 Address
IPv6 Gateway  MTU    Module    Ports    IP access group name
management    Management    9.113.57.64    255.255.254.0    9.113.56.1
1500    1:Module:1
interconnect  Interconnect  192.168.64.161  255.255.255.0
9000    1:Module:1
management    Management    9.113.57.65    255.255.254.0    9.113.56.1
1500    1:Module:2
interconnect  Interconnect  192.168.64.162  255.255.255.0
9000    1:Module:2
management    Management    9.113.57.78    255.255.254.0    9.113.56.1
1500    1:Module:3
interconnect  Interconnect  192.168.64.163  255.255.255.0
9000    1:Module:3
```

```
mod1-iscsi-pl  iSCSI      192.168.100.121  255.255.0.0    192.168.100.1
9000    1:Module:1    1
mod2-iscsi-pl  iSCSI      192.168.100.123  255.255.0.0    192.168.100.1
9000    1:Module:2    1
```

The **ipinterface_list** command displays configured network ports only. The output lines that are highlighted in blue correspond to the same two iSCSI ports that are shown in Figure 13-4 on page 246. The other output lines display XIV management and interconnect ports. The rows might be in a different order each time you run this command. To see a complete list of IP interfaces, use the **ipinterface_list_ports** command.

13.2.2 Host mappings and authentication

The XIV, IBM FlashSystem A9000, and IBM Spectrum Accelerate systems support unidirectional iSCSI Challenge Handshake Authentication Protocol (CHAP).

Note: The CHAP configuration is defined on a per-initiator basis. There are no global configurations for CHAP that affect all the initiators that are connected to the system.

For the iSCSI initiator to log in with CHAP, both the `iscsi_chap_name` and the `iscsi_chap_secret` parameters must be set. After both of these parameters are set, the initiator can run an iSCSI login to the IBM XIV Storage System only if the login information is correct.

CHAP name and secret parameter guidelines

The following guidelines apply to the CHAP name and secret parameters:

- ▶ Both the `iscsi_chap_name` and `iscsi_chap_secret` parameters must either be specified or not specified. You cannot specify just one of them.
- ▶ The `iscsi_chap_name` and `iscsi_chap_secret` parameters must be unique. If they are not unique, an error message is displayed. However, the command does not fail.
- ▶ The secret must be 96 - 128 bits. You can use one of the following methods to enter the secret:
 - Base64 requires that 0b is used as a prefix for the entry. Each subsequent character that is entered is treated as a 6-bit equivalent length.
 - Hex requires that 0x is used as a prefix for the entry. Each subsequent character that is entered is treated as a 4-bit equivalent length.
 - String requires that a prefix is not used (it cannot be prefixed with 0b or 0x). Each character that is entered is treated as an 8-bit equivalent length.
- ▶ If the `iscsi_chap_secret` parameter does not conform to the required secret length (96 - 128 bits), the command fails.
- ▶ If you change the `iscsi_chap_name` or `iscsi_chap_secret` parameters, a warning message is displayed. The message says that the changes will be applied the next time that the host is connected.

CHAP can be configured either through the XIV GUI or through XCLI when you add the iSCSI initiators as XIV hosts. The SAN Volume Controller or IBM Storwize nodes that act as iSCSI initiators must be configured as iSCSI hosts in the XIV, IBM FlashSystem A9000, or IBM Spectrum Accelerate system.

Tip: Because all initiator nodes share the XIV LUNs to be virtualized, you configure these nodes as a cluster of hosts.

The section “Configuring SAN Volume Controller or IBM Storwize nodes as XIV hosts with the GUI” describes how to configure the initiator nodes as a host cluster with the XIV GUI. The section “Configuring SAN Volume Controller or IBM Storwize nodes as XIV hosts with the XCLI” on page 252 explains the same procedure with the XCLI. When the initiators are configured as hosts, you can map XIV LUNs to them, as shown in 13.2.3, “Mapping XIV LUNs to the SAN Volume Controller or IBM Storwize system” on page 253.

Configuring SAN Volume Controller or IBM Storwize nodes as XIV hosts with the GUI

To configure the SAN Volume Controller or Storwize nodes as XIV hosts with the GUI, complete these steps:

1. In the XIV Storage System main GUI window, hover your cursor over the **Hosts and Clusters** icon and select **Hosts and Clusters**. Alternatively, you can select **View** → **Hosts and Clusters** → **Hosts and Clusters** from the top menu bar.
2. The Hosts window opens, and shows a list of hosts (if any) that are already defined. To add a cluster, click **Add Cluster**.
3. The **Add Cluster** dialog box opens, as shown in Figure 13-5. Enter a name for the cluster. For the Type, the default option is correct.

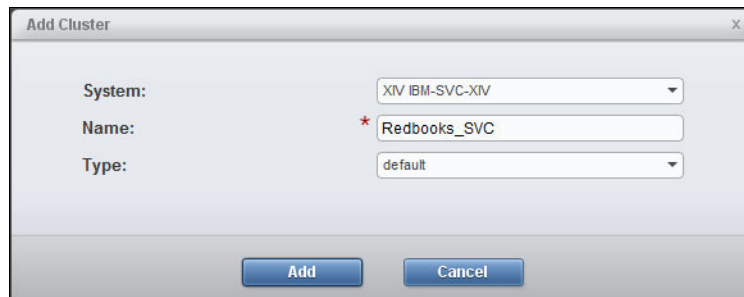
The image shows a screenshot of the 'Add Cluster' dialog box in the XIV Storage System GUI. The dialog box has a title bar with 'Add Cluster' and a close button. It contains three fields: 'System:' with a dropdown menu showing 'XIV IBM-SVC-XIV', 'Name:' with a text input field containing 'Redbooks_SVC' and a red asterisk indicating a required field, and 'Type:' with a dropdown menu showing 'default'. At the bottom of the dialog box are two buttons: 'Add' and 'Cancel'.

Figure 13-5 Add Cluster dialog box

4. To add a node to the cluster, select the cluster in the **Hosts and Clusters** pane, right-click, and select **Add Host** (Figure 13-6).

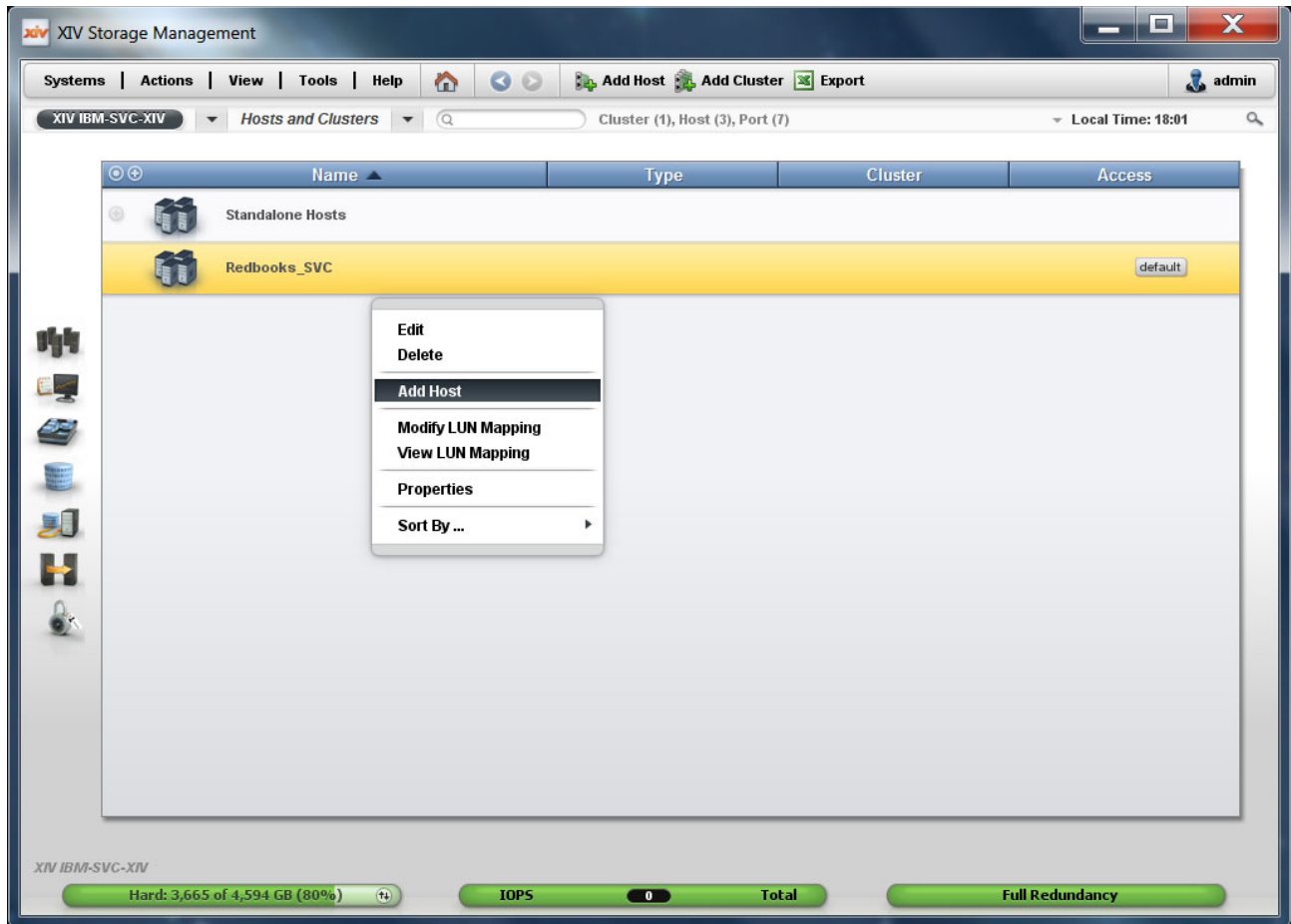
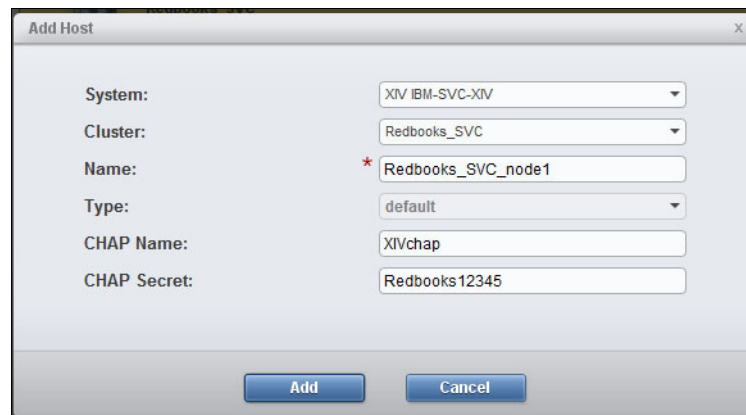


Figure 13-6 Adding a host to the cluster

5. The Add Host dialog box opens, as shown in Figure 13-7. Enter a name that identifies your SAN Volume Controller or IBM Storwize node. For the Type, the default option is correct. Enter the CHAP name and the CHAP secret.

Important: You must use the same CHAP name and CHAP secret for each initiator node because the target discovery runs on all connected SAN Volume Controller or IBM Storwize nodes for the single system-wide target IQN.

The image shows a screenshot of a software window titled "Add Host". It contains several labeled input fields: "System:" with a dropdown menu showing "XIV IBM-SVC-XIV"; "Cluster:" with a dropdown menu showing "Redbooks_SVC"; "Name:" with a text box containing "Redbooks_SVC_node1" and a red asterisk icon to its left; "Type:" with a dropdown menu showing "default"; "CHAP Name:" with a text box containing "XIVchap"; and "CHAP Secret:" with a text box containing "Redbooks12345". At the bottom of the window are two buttons: "Add" and "Cancel".

System:	XIV IBM-SVC-XIV
Cluster:	Redbooks_SVC
Name:	* Redbooks_SVC_node1
Type:	default
CHAP Name:	XIVchap
CHAP Secret:	Redbooks12345

Figure 13-7 Host parameters

6. After adding all SAN Volume Controller or IBM Storwize nodes as hosts to the cluster, the Hosts and Clusters pane looks similar to Figure 13-8.

Host access to XIV LUNs is granted depending on the host adapter ID. For an iSCSI connection, the host adapter ID is the initiator IQN. To add an IQN to a host definition, right-click the host and select **Add Port** from the menu.

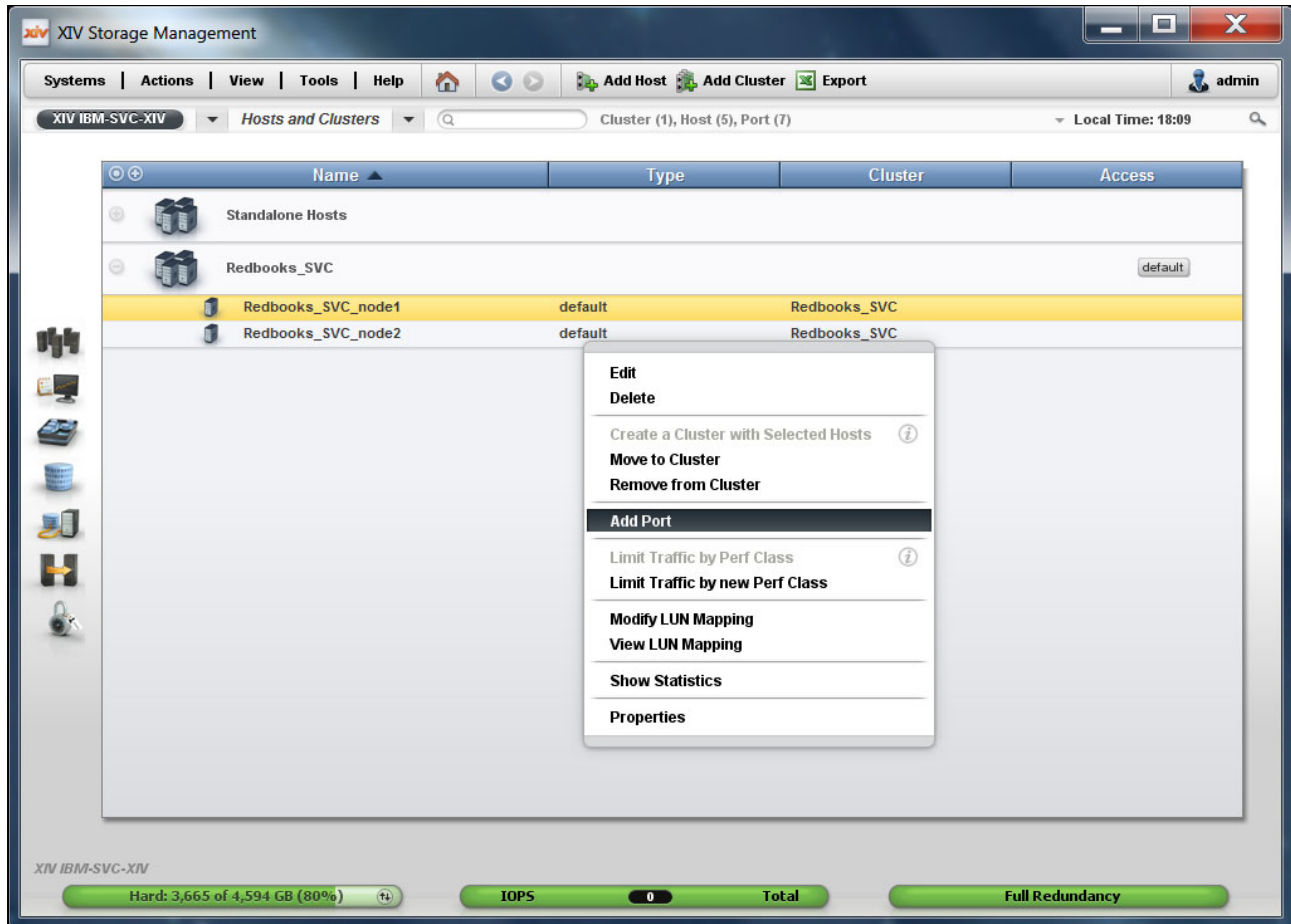


Figure 13-8 Adding a port to a node of the SAN Volume Controller or IBM Storwize cluster

7. The Add Port window opens, as shown in Figure 13-9. Select **iSCSI** as the Port Type and enter the IQN of the node as the iSCSI Name.

Because all Ethernet ports of a SAN Volume Controller or IBM Storwize node use the same IQN, a single iSCSI port definition per host is sufficient, independent of the number of Ethernet ports.

The 'Add Port' window contains the following fields and values:

- System:** XIV IBM-SVC-XIV
- Host Name:** Redbooks_SVC_node1
- Port Type:** iSCSI
- iSCSI Name:** * q3.com.ibm:2145.redbookscluster1.node1

Buttons: Add, Cancel

Figure 13-9 iSCSI initiator port parameters

8. Repeat steps 4 on page 249 to 7 for all initiator nodes in your SAN Volume Controller or IBM Storwize cluster.

Now, map your XIV LUNs to your SAN Volume Controller or IBM Storwize system (see 13.2.3, “Mapping XIV LUNs to the SAN Volume Controller or IBM Storwize system” on page 253). The XIV Host Connectivity GUI pane does not show any connections to the newly configured iSCSI hosts. You see the connections after the iSCSI discovery during the initiator configuration (see 13.3, “Initiator configuration” on page 254).

Configuring SAN Volume Controller or IBM Storwize nodes as XIV hosts with the XCLI

To configure the SAN Volume Controller or IBM Storwize nodes as XIV hosts with the XCLI, complete these steps:

1. Define a cluster by using the **cluster_create** command:
2. Define a host as a member of this cluster by using the **host_define** command:

```
cluster_create cluster=[SVCClusterName]
```

```
host_define host=[SVCNodeName] cluster=[SVCClusterName]
iscsi_chap_name=[SVCChapName] iscsi_chap_secret=[SVCChapSecret]
```

Important: You must use the same CHAP name and CHAP secret for each initiator node because the target discovery runs on all connected SAN Volume Controller or IBM Storwize nodes for the single system-wide target IQN.

3. Add an iSCSI port by using the **host_add_port** command:

```
host_add_port host=[SVCNodeName] iscsi_name=[SVCNodeIQN]
```

Because all Ethernet ports of a SAN Volume Controller or IBM Storwize node use the same IQN, a single port definition per host name is sufficient, independent of the number of Ethernet ports.

4. Repeat steps 2 and 3 for all initiator nodes of your SAN Volume Controller or IBM Storwize cluster.

Now, map XIV LUNs to your SAN Volume Controller or IBM Storwize system (see 13.2.3, “Mapping XIV LUNs to the SAN Volume Controller or IBM Storwize system” on page 253). The output of the XCLI `host_list` command does not show any connections to the newly configured iSCSI hosts. You see the connections after the iSCSI discovery during the initiator configuration (see 13.3, “Initiator configuration” on page 254).

13.2.3 Mapping XIV LUNs to the SAN Volume Controller or IBM Storwize system

The final configuration step is to map XIV LUNs (that becomes MDisks) to the initiator nodes. Using the GUI, complete the following steps:

1. In the Hosts and Clusters configuration window, right-click the cluster of your initiator nodes to which the LUN is to be mapped and select **Modify LUN Mappings** (Figure 13-10).

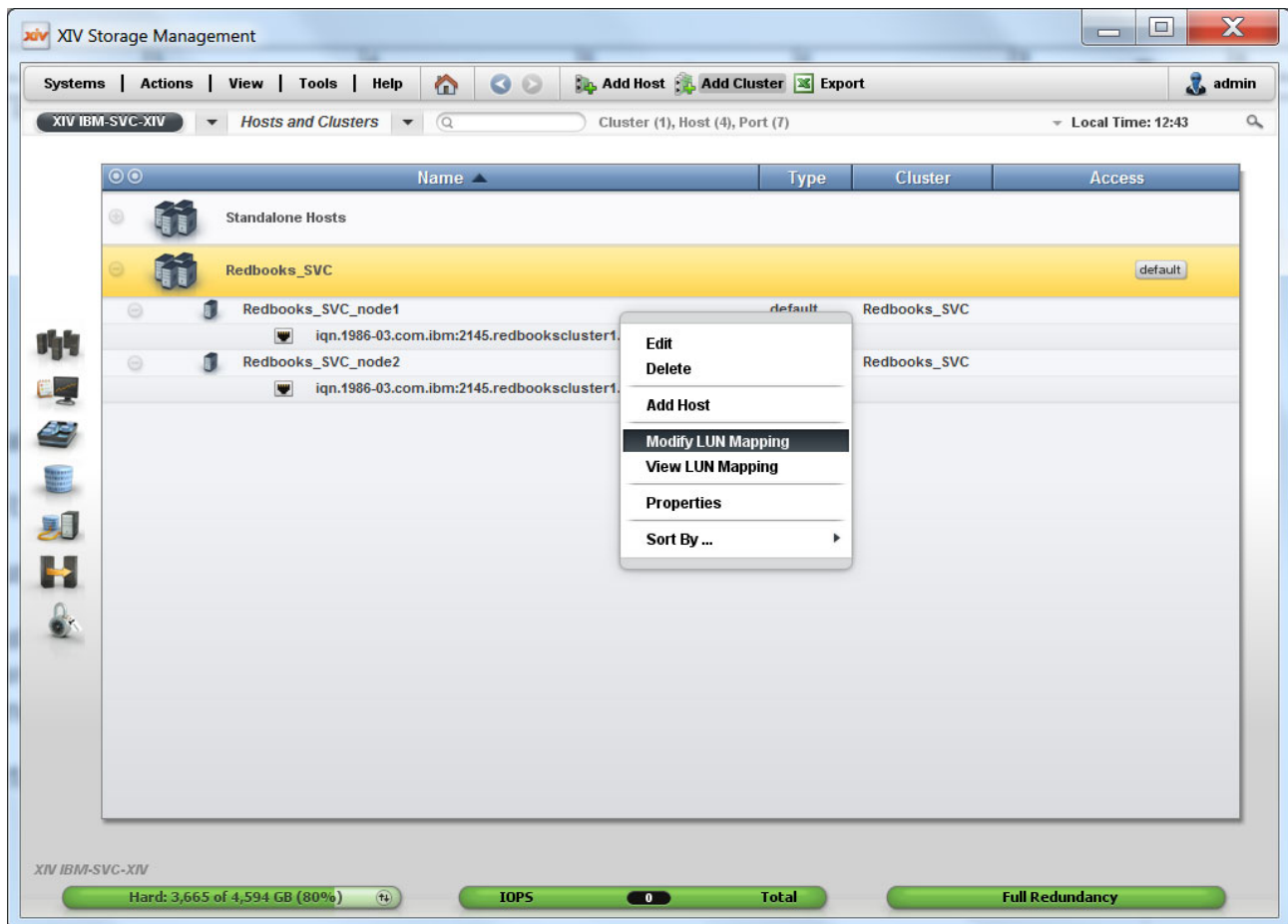


Figure 13-10 Modifying the LUN mapping in the XIV GUI

2. The Volume to LUN Mapping window opens, as shown in Figure 13-11. Select the volume that will become an MDisk from the left pane. The GUI suggests a LUN ID to which to map the volume; there is no need to change it. Click **Map** and the volume is assigned immediately.

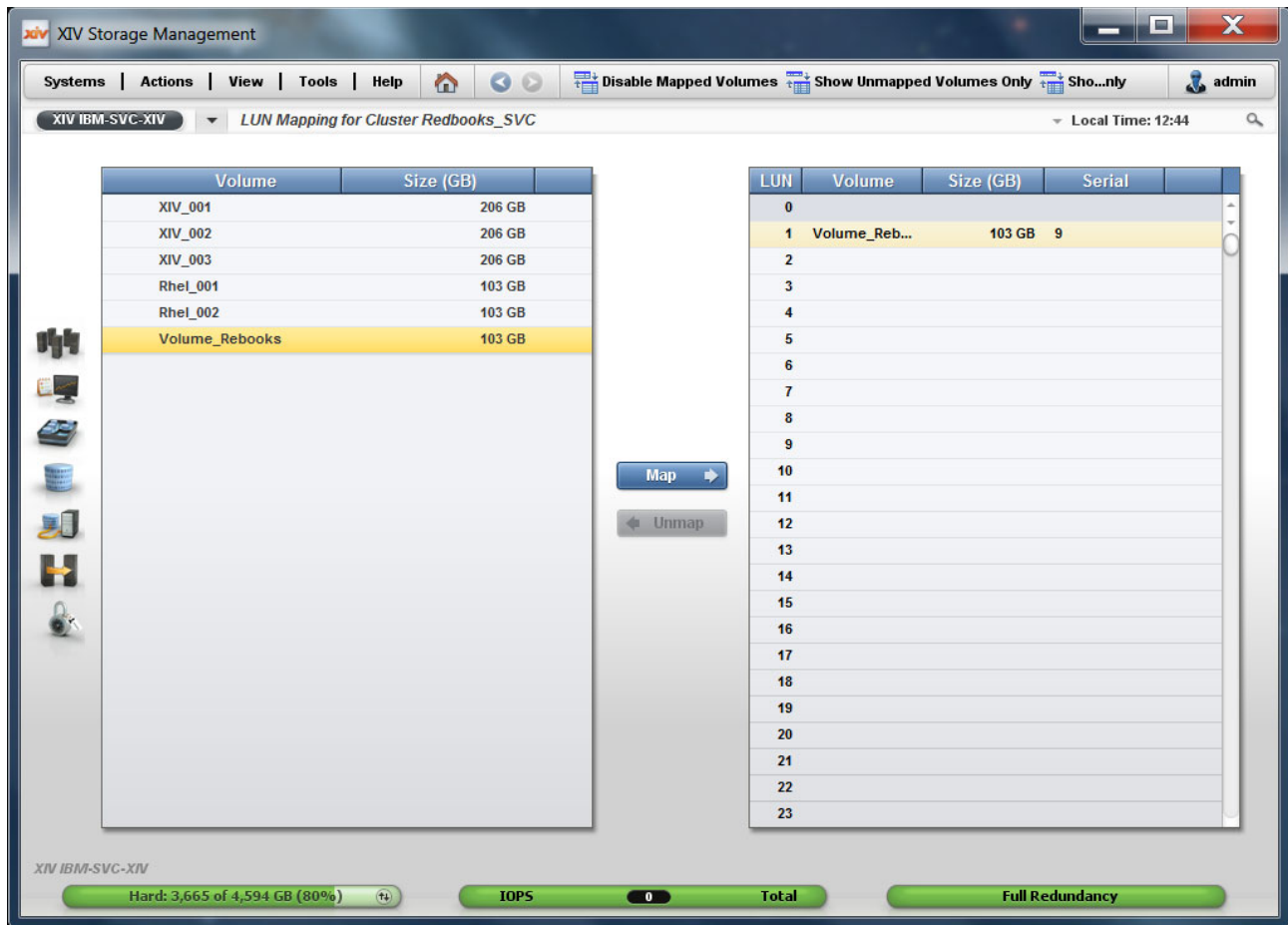


Figure 13-11 Selecting a LUN for mapping in the XIV GUI

Alternatively, you can use the `map_vol` command to map a LUN to the cluster of initiator nodes, as shown in Example 13-4. The LUN ID is a mandatory parameter.

Example 13-4 Mapping an XIV LUN to the cluster iSCSI initiator nodes

```
XIV IBM-SVC-XIV>>map_vol cluster=Redbooks_SVC vol=Volume_Redbooks lun=1
```

13.3 Initiator configuration

After mapping the XIV LUNs (which become MDisks) to the SAN Volume Controller or Storwize cluster, you configure these cluster nodes as initiators. The following sections describe how to accomplish this task:

- ▶ Section 13.3.1, “Overview” on page 255 gives an overview of the whole procedure.
- ▶ Section 13.3.2, “Workflow that uses the CLI” on page 255 shows the detailed configuration steps by using the CLI.

- ▶ Section 13.3.3, “Workflow with GUI” on page 257 shows the configuration by using the GUI.
- ▶ Section 13.3.4, “Configuration validation” on page 261 explains how to validate the configuration before you create SAN Volume Controller or IBM Storwize volumes.

13.3.1 Overview

As a prerequisite, the SAN Volume Controller or IBM Storwize initiator ports must be configured with IPv4 or IPv6 addresses. For more information, see Chapter 11, “iSCSI virtualization overview” on page 213.

The configuration procedure for the SAN Volume Controller or IBM Storwize initiators consists of two phases:

- ▶ Discovery of the targets
- ▶ Session establishment to these targets

You must repeat these two phases for each initiator port. Section 13.3.2, “Workflow that uses the CLI” on page 255 explains these steps by using the CLI. Section 13.3.2, “Workflow that uses the CLI” on page 255 shows the steps by using the GUI.

13.3.2 Workflow that uses the CLI

To configure the XIV, IBM FlashSystem A9000, or IBM Spectrum Accelerate storage system as an external iSCSI controller, complete the following steps:

1. to discover manually the XIV iSCSI target with the provided IP and CHAP secret, run the following command:

```
detectiscsistorageportcandidate -targetip [XIVIP1] -srcportid [SVCSourcePortID]
-username [SVCChapName] -chapsecret [SVCChapSecret]
```

The **[SVCChapName]** and **[SVCChapSecret]** values must match the **[SVCChapName]** and **[SVCChapSecret]** values that are specified for authentication for the XIV system’s host mappings (see 13.2.2, “Host mappings and authentication” on page 247). Otherwise, the SAN Volume Controller or IBM Storwize system cannot detect the XIV target ports.

2. List the status of the most recently discovered target by running the following command:

```
lscsistorageportcandidate
```

3. After the discovery was successful, you can establish sessions from the initiator port to the most recently discovered target by running the following command:

```
addiscsistorageport -username [SVCChapName] -chapsecret [SVCChapSecret] [ID]
```

In this command, **[ID]** is the row ID of the **lscsistorageportcandidate** command in step 2. Again, the **[SVCChapName]** and **[SVCChapSecret]** values are the same as specified for authentication for the XIV system’s host mappings.

4. Verify that new sessions are established by running the following command:

```
lscsistorageport
```

The command lists the status of the initiator and target for each session in separate rows.

Additionally, you can verify the detailed status of initiator node connectivity through the ports to the target by running the following command:

```
lscsistorageport [ID]
```

In this command, **[ID]** is the row ID of the **lscsistorageport** command in step 4.

- Repeat steps 1 on page 255 to 4 on page 255 for each initiator port. Choose a different target IP each time, as described in 13.1, “Planning considerations” on page 240.

Example 13-5 shows steps 1 on page 255 to 4 on page 255 for one initiator port with comments.

Example 13-5 Initiator port configuration

```
# Step 1 on page 255: Discover the XIV iSCSI target 192.168.100.121 from initiator port 3.
IBM_2145:Redbooks_cluster1:superuser>detectiscsistorageportcandidate -targetip 192.168.100.121
-srcportid 3 -username XIVchap -chapsecret Redbooks12345
#
# Step 2 on page 255: List the status of the most recently discovered target.
# The single output row (with ID 0) shows the target from source port 3 in IO group 1.
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageportcandidate
id src_port_id target_ipv4 target_ipv6 target_iscsiname iogroup_list configured status
site_id site_name
0 3 192.168.100.121 iqn.2005-10.com.xivstorage:041529 1:-::- no full
#
# Step 3 on page 255: Establish sessions from the initiator port to the most recently discovered
target.
# Use the row ID 0 from the previous output as parameter in this command.
IBM_2145:Redbooks_cluster1:superuser>addiscsistorageport -username XIVchap -chapsecret
Redbooks12345 0
#
# Step 4 on page 255: Verify that new sessions had been established.
# (Extra output rows for other iSCSI controllers omitted.)
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageport
id src_port_id target_ipv4 target_ipv6 target_iscsiname controller_id
iogroup_list status site_id site_name
4 3 192.168.100.121 iqn.2005-10.com.xivstorage:041529 3
1:-::- full
```

Now, you can configure the XIV LUNs that are mapped to the cluster of initiator nodes (see 13.2.3, “Mapping XIV LUNs to the SAN Volume Controller or IBM Storwize system” on page 253) as managed disks in the SAN Volume Controller or IBM Storwize system. The steps are the same as for external LUNs that are attached through Fibre Channel (FC):

- Verify the controller by using the **lscontroller** command.
- Discover the managed disks by using the **detectmdisk** command.
- Verify the managed disks with the **lsmdisk** command.
- Create a pool by running the **mkmdiskgrp** command if necessary and add the managed disks to the pool by running the **addmdisk** command.

Example 13-6 shows that the XIV LUNs that are mapped in the examples of 13.2.3, “Mapping XIV LUNs to the SAN Volume Controller or IBM Storwize system” on page 253. (Extra output rows for other external controllers are omitted.)

Example 13-6 Configuration of the XIV LUNs as managed disks

```
# Step 1: Verify the controller.
# The XIV system is shown as controller4.
IBM_2145:Redbooks_cluster1:superuser>lscontroller
id controller_name ctrl_s/n vendor_id product_id_low product_id_high site_id site_name
0 controller4 A2390000 IBM 2810XIV- LUN-0
#
# Step 2: Discover the new LUNs.
```

```
# The command completes asynchronously and can take some minutes to complete.
IBM_2145:Redbooks_cluster1:superuser>detectmdisk
#
# Step 3 on page 256: Verify the LUNs.
# The MDisk with ID 7 is presented by controller4 in status unmanaged.
IBM_2145:Redbooks_cluster1:superuser>lsmdisk
id name      status mode      mdisk_grp_id mdisk_grp_name capacity ctrl_LUN_#      controller_name
UID                                                tier          encrypt site_id
site_name distributed dedupe
7 mdisk7 online unmanaged          96.2GB  0000000000000001 controller4
00173800a239000900000000000000000000000000000000000000000000000000 tier_enterprise no
no no
#
# Step 4 on page 256: Create a pool and add the MDisk with ID 7 to this pool.
IBM_2145:Redbooks_cluster1:superuser>mkmdiskgrp -name XIVpool -mdisk 7 -ext 1024
MDisk Group, id [0], successfully created
```

13.3.3 Workflow with GUI

To configure the XIV, IBM FlashSystem A9000, or IBM Spectrum Accelerate storage system as an external iSCSI controller by using the SAN Volume Controller or Storwize GUI, complete the following steps:

1. Click **Pools** → **External Storage**, and then click **Add External iSCSI Storage** (Figure 13-12).

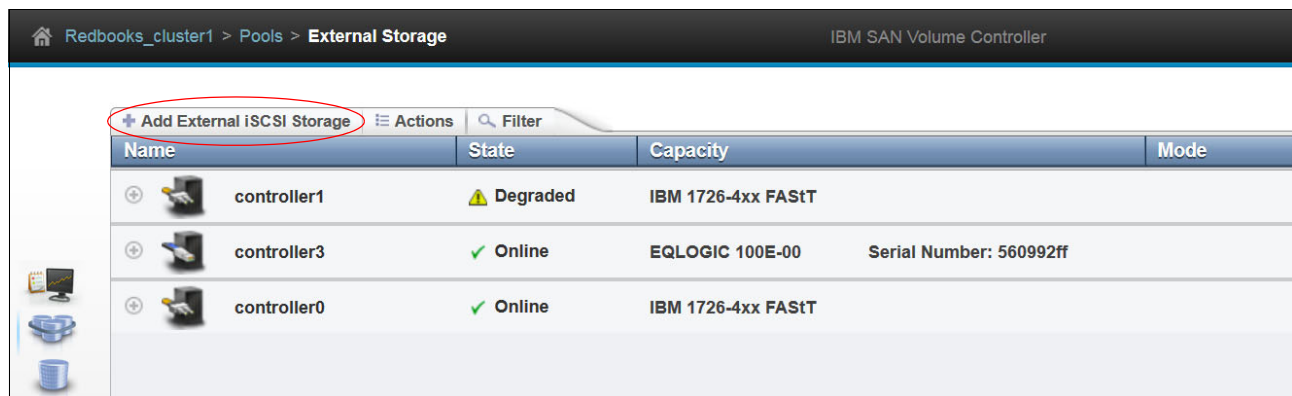


Figure 13-12 External Storage window to add external iSCSI storage

2. The dialog box to select the type of the iSCSI controller opens (Figure 13-13). Select **IBM Spectrum Accelerate** for XIV, IBM FlashSystem A9000, and IBM Spectrum Accelerate systems.

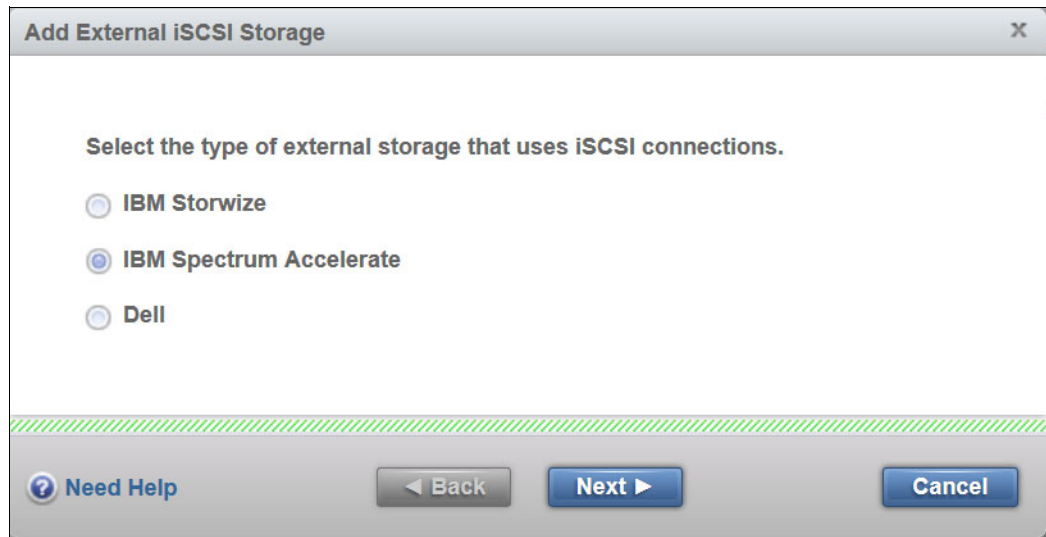


Figure 13-13 Selecting the type of the iSCSI storage controller

3. Click **Next**. The window for the iSCSI controller parameters opens (Figure 13-14). Enter the CHAP name (as User name), the CHAP secret, and at least two initiator port names with their XIV target port addresses.

The screenshot shows a window titled "Add External iSCSI Storage" with a close button (X) in the top right corner. The window contains the following fields and controls:

- User name (optional):** A text input field containing "XIVchap".
- CHAP secret:** A text input field containing "Redbooks12345".
- Source port 1 connections:** A section with a title bar and expand/collapse buttons (+/-). It contains:
 - Select source port 1:** A dropdown menu showing "Port3".
 - Target port on remote storage 1:** A text input field containing "192.168.100.121".
- Source port 2 connections:** A section with a title bar and expand/collapse buttons (+/-). It contains:
 - Select source port 2:** A dropdown menu showing "Port4".
 - Target port on remote storage 2:** A text input field containing "192.168.100.123".

At the bottom of the window, there is a progress bar (partially filled with green) and three buttons: "Need Help" (with a question mark icon), "Back" (with a left arrow icon), and "Next" (with a right arrow icon). A "Cancel" button is also present on the far right.

Figure 13-14 iSCSI controller parameters

- Click **Next**. The summary window opens (Figure 13-15). Verify the parameters and click **Finish**.

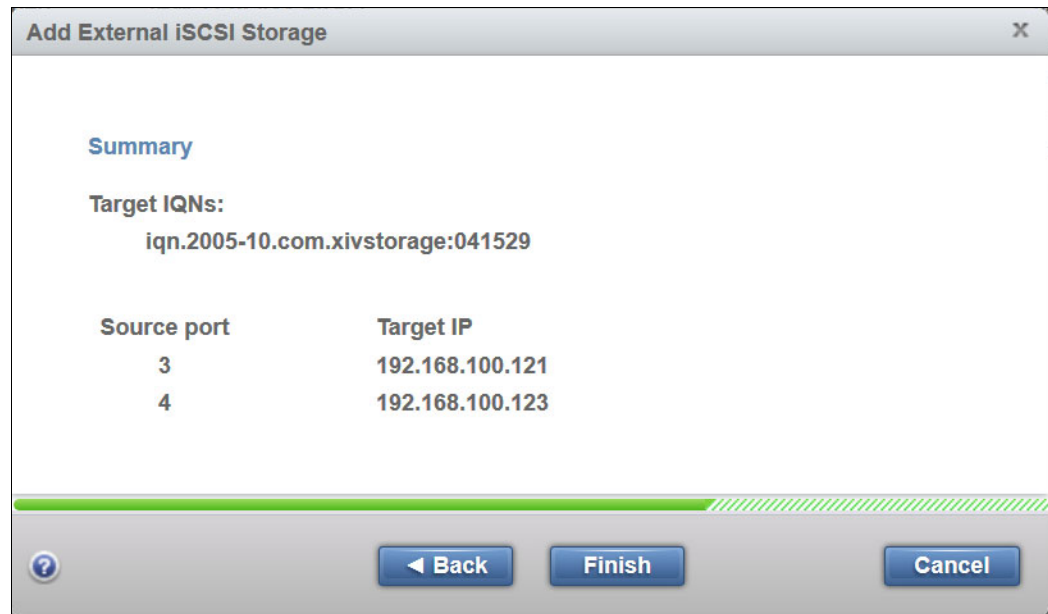


Figure 13-15 iSCSI controller summary

- The External Storage window shows the new controller with its LUNs as MDisks (Figure 13-16).

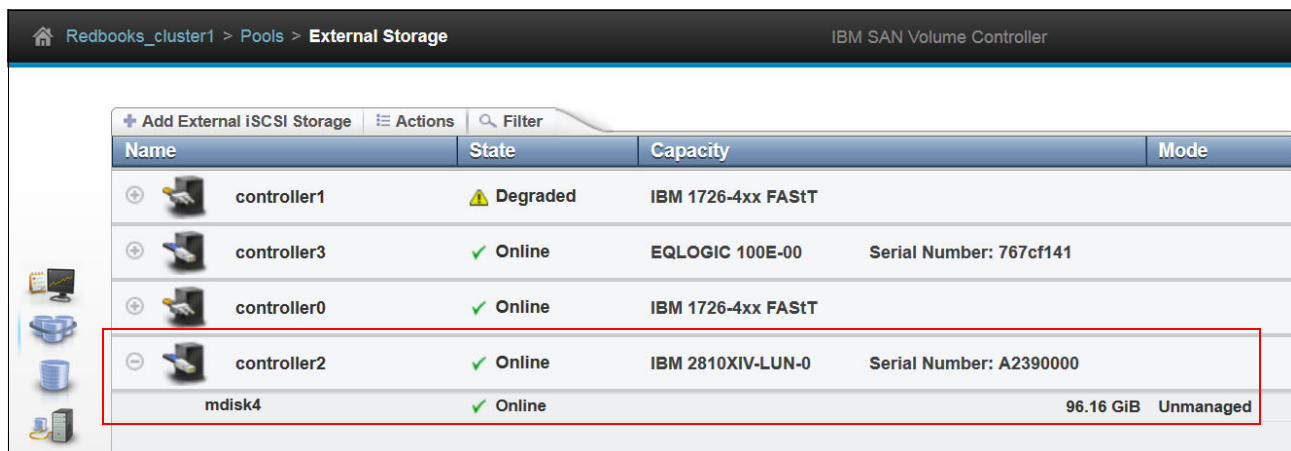


Figure 13-16 New external iSCSI controller with LUN

Now, you can configure the XIV LUNs that are mapped to the cluster of initiator nodes (see 13.2.3, “Mapping XIV LUNs to the SAN Volume Controller or IBM Storwize system” on page 253) as managed disks in the SAN Volume Controller or IBM Storwize system. The steps are the same as for external LUNs that are attached through FC, that is, click **Pools** → **MDisks by Pools**, create a pool if necessary, and add the managed disks to the pool (Figure 13-17).

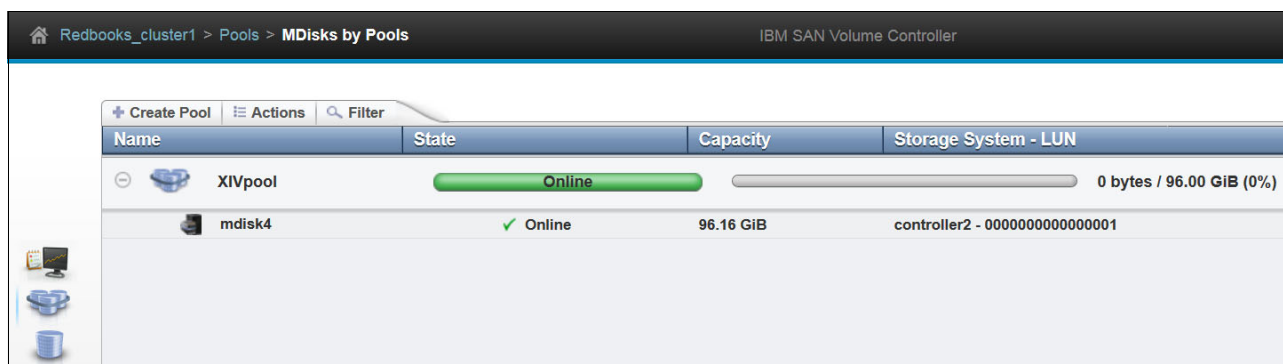


Figure 13-17 MDisks from the XIV that is configured in the new pool

13.3.4 Configuration validation

To validate the configuration from the initiator side, you can use the `lsiscsistorageport` command. Without parameters, the command lists all the established iSCSI sessions. Then, you see the row IDs of this output in `lsiscsiport [ID]` commands to display the details of the XIV sessions. Example 13-7 shows an output for the iSCSI session that is configured in 13.3.2, “Workflow that uses the CLI” on page 255.

Example 13-7 Detailed output of lsiscsiport

```
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageport 1
id 1
src_port_id 3
target_ipv4 192.168.100.121
target_ipv6
target_iscsiname iqn.2005-10.com.xivstorage:041529
controller_id 3
iogroup_list 1:-::-
status full
site_id
site_name
node_id 1
node_name node1
src_ipv4 192.168.104.199
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node1
connected yes
node_id 2
node_name node2
src_ipv4 192.168.104.197
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node2
connected yes
```

The `iogroup_list` field shows a colon-separated list of discovery result codes per I/O group:

- ▶ Value 0 indicates that the I/O group is available in the system, but discovery is either not triggered through the I/O group or discovery through the I/O group failed.
- ▶ Value 1 indicates that the I/O group is present and discovery is successful through the I/O group.
- ▶ Value - (dash) indicates that the I/O group is not valid or is not present in the system.

In this example, the first I/O group discovered successfully the iSCSI target (value 1 in field `iogroup_list` and value full in field `status`), and the fields `connected` show yes for both nodes of the I/O group.

For configuration validation from the target side, you can use XIV GUI window and select **Hosts and Clusters** → **Host Connectivity** (Figure 13-18).

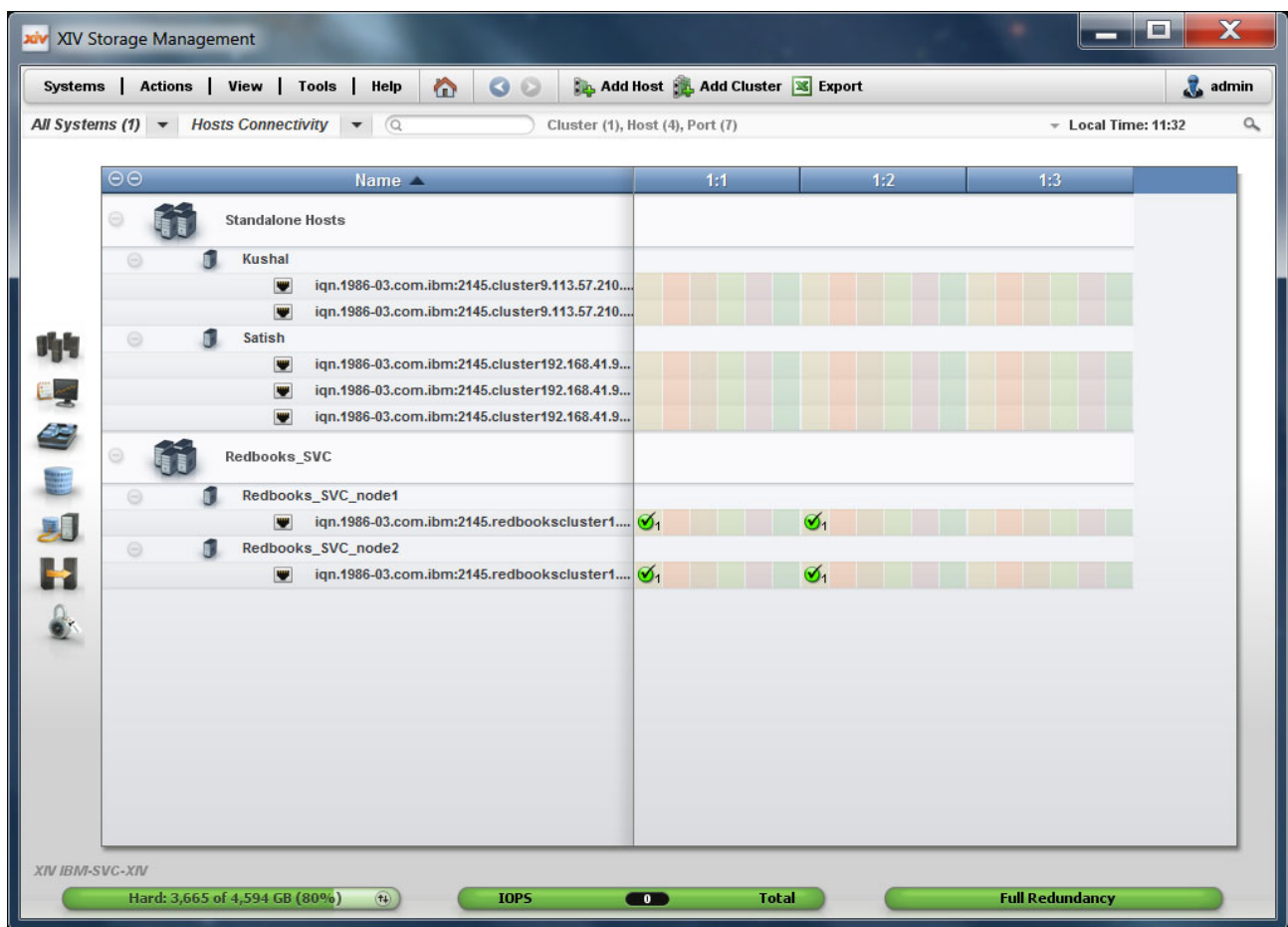


Figure 13-18 XIV host connectivity from two initiator nodes to two target nodes (modules)

The green check marks show that two iSCSI initiators (from the host cluster Redbooks_SVC) successfully established sessions to port 1 in two different target nodes (modules 1:1 and 1:2). You can detect failed connections easily as missing check marks.

Using the XCLI, you can get the same output from the **host_connectivity_list** command, as shown in Example 13-8.

Example 13-8 XCLI output of host_connectivity_list with sessions from two initiator nodes to two target nodes

XIV IBM-SVC-XIV>>host_connectivity_list				
Host		Host Port	Module	Local FC
port	Local iSCSI	port	Type	
Redbooks_SVC_node2		iqn.1986-03.com.ibm:2145.redbookscluster1.node2	1:Module:1	
mod1-iscsi-p1		iSCSI		
Redbooks_SVC_node1		iqn.1986-03.com.ibm:2145.redbookscluster1.node1	1:Module:1	
mod1-iscsi-p1		iSCSI		
Redbooks_SVC_node2		iqn.1986-03.com.ibm:2145.redbookscluster1.node2	1:Module:2	
mod2-iscsi-p1		iSCSI		
Redbooks_SVC_node1		iqn.1986-03.com.ibm:2145.redbookscluster1.node1	1:Module:2	
mod2-iscsi-p1		iSCSI		

For further analysis of connectivity from the XIV side, Table 13-1 lists some useful built-in tools. See the full XCLI command descriptions in the [IBM XIV Storage System documentation](#).

Table 13-1 XIV built-in tools

Tool	Description
host_connectivity_list	Lists FC and iSCSI connectivity to hosts.
ipinterface_list_ports	Lists all Ethernet ports, their configuration, and their status.
ipinterface_run_arp	Prints the ARP database of a specified IP address.
ipinterface_run_traceroute	Tests connectivity to a remote IP address.



External virtualization of Dell Equallogic PS Series

This chapter describes the IBM SAN Volume Controller cluster with support for the Internet Small Computer System Interface (iSCSI)-based external storage, and guidelines to virtualize a Dell Equallogic PS Series storage controller over iSCSI. This chapter also describes the command-line interface (CLI) and graphical user interface (GUI) configuration guidelines for initiator and target controllers, along with migration considerations.

This chapter describes the following topics:

- ▶ 14.1, “Planning considerations” on page 266
- ▶ 14.2, “Target configuration” on page 268
- ▶ 14.3, “Initiator configuration” on page 284

14.1 Planning considerations

One of the first prerequisites to consider is the interconnect for the IP network. Figure 14-1 shows a suggested configuration for the Dell Equallogic PS Series connections in this chapter.

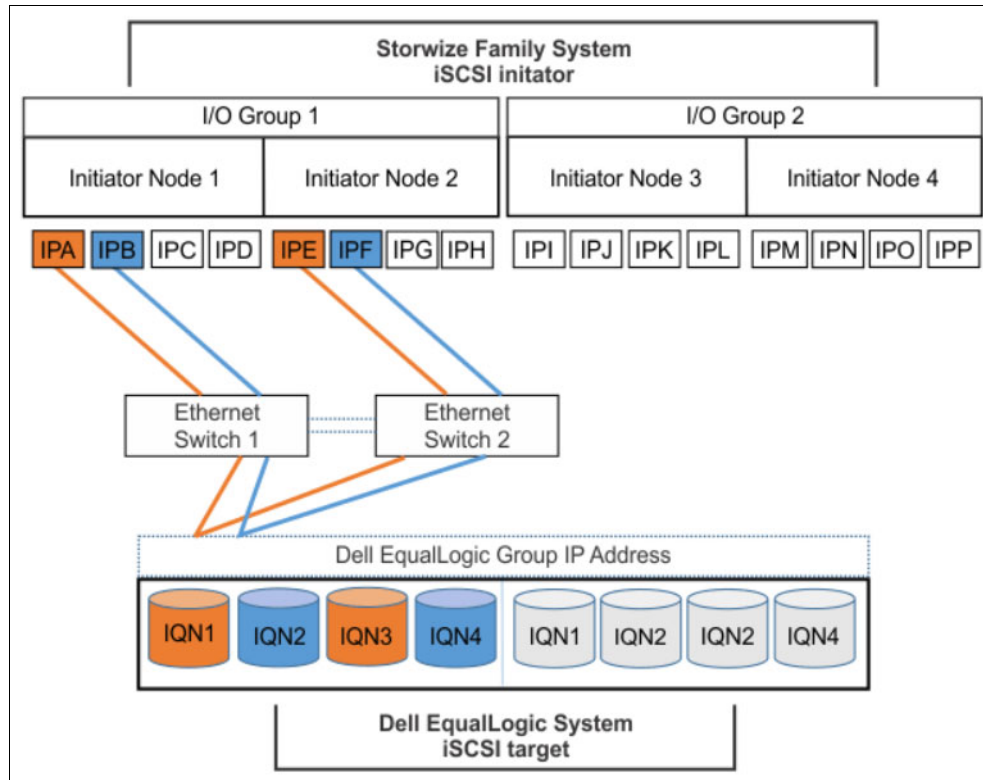


Figure 14-1 Connecting a Dell Equallogic PS Series iSCSI storage

Figure 14-1 provides a detailed description of the fundamentals of how to configure Dell Equallogic PS Series controllers with SAN Volume Controller. The front-end system consists of two SAN Volume Controller I/O groups, with each I/O group possessing two nodes. Each node has a maximum of four initiator ports. The target side that is shown in the figure reflects a Dell Equallogic PS Series controller with a group IP configured, which possess the access to the iSCSI qualified name (IQN) of each node. Ethernet switches are used for the networking that is required for the communication between the initiator and the target.

The ports that are colored in orange and blue signify the logic behind the establishment of the iSCSI sessions between the initiator and the target. IPA (initiator port-node1) on the initiator is connected to the group IP and eventually to IQN-1 through Ethernet switch 1. Similarly, IPB (initiator port-node1) is connected to the group IP and eventually to IQN-2 through the Ethernet switch 2. The same logic is implemented while connecting the initiator ports IPE and IPF to the IQN-1 and IQN-2 by using the respective switches.

Also, while establishing sessions, defining the source ports implies that connections to the target ports will be established by all of the initiator ports on the system. The target IQN, IQN-1, will have sessions with IPA and IPE, which are the initiator ports on the system. The same is true for the target IQN, IQN-2, because they maintain sessions with IPB and IPF.

The ports that stand idle, both on the initiator and the target, can be used for further connections by using the same logic, which helps to increase throughput and the level of session redundancy.

14.1.1 Dell Equallogic PS Series connection considerations

This section describes the considerations when you connect a Dell Equallogic PS Series controller to a SAN Volume Controller or IBM Storwize system. You should consider these points while planning for the Dell Equallogic PS Series external storage for migration or virtualization.

No redundant paths per node

Connecting to the target IQN by using two ports of the same initiator nodes is not allowed. This is a limitation in the current code-line, which results in reduced session redundancy per node. There is I/O group-wide redundancy present because each node has one connection to back-end storage.

Session limits

Although it is possible to establish sessions by using both cluster-wide and I/O group-wide connectivity, it is preferable to have I/O group-wide connectivity because Dell has a limitation on the number of sessions between the initiator and the target system. Currently, 128 sessions to one back-end controller are allowed. Therefore, the recommendation is to exhaust the number of sessions to gain maximum storage access, and not to access the same storage through multiple I/O groups.

LUN mapping limits

Every LUN in the Dell controller has a unique IQN, and the session must be established with each LUN separately, which consumes many sessions from the initiator to the Dell controller. With the current available code version, only 64 LUNs can be virtualized from an external Dell controller. Therefore, it is preferable to have fewer larger LUNs to reduce the iSCSI sessions for the Dell controller if you are planning to connect the Dell controller for virtualization.

14.1.2 Migration considerations

In the Dell Equallogic PS Series implementation design, every volume is associated with one IQN and you must establish the session to the back-end LUN, which creates multiple iSCSI sessions and might reach the limit if not managed properly. The initiator system has a limit of 64 LUNs per external Dell Equallogic PS Series controller to be virtualized behind SAN Volume Controller or an IBM Storwize system. If there are more than 64 LUNs that are mapped to the initiator system from Dell Equallogic PS Series, discovery output is truncated, and only the first 64 LUNs are listed in the `lsiscsistorageportcandidate` output.

At the time of data migration from the internal Dell Equallogic PS Series controller to a SAN Volume Controller or IBM Storwize system, if there are more than 64 LUNs on Dell, consider migrating the LUNs in chunks of 64:

1. Connect the first 64 LUNs to the initiator system and migrate them to the SAN Volume Controller or IBM Storwize system.
2. Remove the target sessions when the migration is complete.
3. Change the access policy on the target Dell controller for those LUNs.

4. Remove the LUN access to the initiator system.
5. Now, redo the discovery and login process to connect another chunk of 64 LUNs to initiator system and perform migration activity.

14.2 Target configuration

To configure your Dell Equallogic PS Series system as a target controller for a SAN Volume Controller or IBM Storwize cluster, you must perform the initial setup activities that are described in the Dell documentation center to start the Dell system. (For more information about how to configure the Dell controller, see the product documentation for your system.) This initial setup includes using the setup utility to configure an array, and creating a PS Series group with the array as the initial group member.

For hardware installation information, see the *PS Series QuickStart* guide in your Dell documentation. The following subsections provide information about setting up Group IP and iSCSI IPs, volume creation, and access policy settings on the Dell Equallogic PS Series storage array:

- ▶ 14.2.1, “Port configuration” on page 268
- ▶ 14.2.2, “Setting up access policies” on page 271
- ▶ 14.2.3, “Creating volumes and applying access policies” on page 276
- ▶ 14.2.4, “Modifying the settings of existing volumes” on page 281

14.2.1 Port configuration

This section provides information about Group IP configuration and iSCSI port configuration for the Dell Equallogic PS Series controller.

Setting up the Group IP

The Group IP address is the network address for the Dell controller group. You should use the Group IP address as the iSCSI discovery address when you connect initiators to iSCSI targets in the group. You also must use the Group IP address to access the group for management purposes, unless you configure a dedicated management network.

Prerequisite: Before changing the Group IP address or the Group name for a system that is already configured, check your Dell documentation for information about the effects of these changes.

Also, for more information, see [Group Network Configuration](#).

While assigning IP addresses to Dell Equallogic PS Series Arrays or Members, you must assign a single Group IP address. Ethernet ports are assigned IP addresses on the active controller in the group, and an optional management IP address.

To configure or change the Group IP, complete the following steps:

1. Open the Dell Equallogic PS Series Group manager by using the management IP address in a supported browser.

2. Select **Group Configuration** → **General**. Specify the IP address and the netmask, as shown in Figure 14-2.

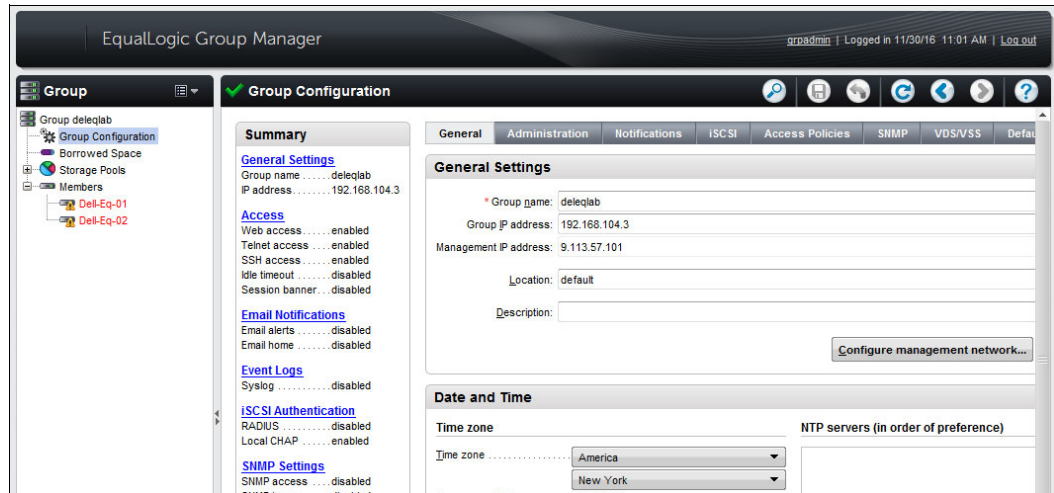


Figure 14-2 Dell Equallogic PS Series Group IP configuration

3. Click **Save all changes**.
4. Refresh the window to view the changed Group IP address of the system.

Network interface configuration

Dell Equallogic PS Series Controllers run in active-standby mode. The active controller is running, and all Ethernet traffic and I/O processing is done by the active controller. The standby controller is running, and replicates cache from the active controller.

If the active controller fails, the standby controller becomes active and serves all Ethernet traffic. The IP addresses that are assigned to the active controller's Ethernet ports are failovers to the standby controller when a failure of the active controller occurs. Because of this mechanism of IP failover, the IP assignment is done for the active controller only, and settings are overridden to the failover controller upon failure of the active controller.

To set up the iSCSI ports, complete the following steps:

1. From the Dell Equallogic PS Series Group Manager console in the browser, click **Member** → **Network**. This window that opens shows the IP configuration of the member, and displays a list of network interfaces, as shown in Figure 14-3.

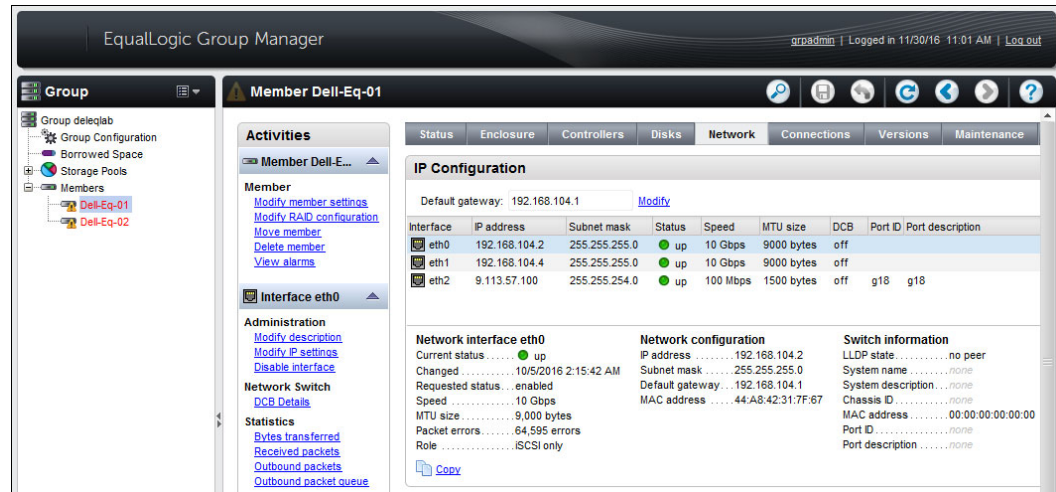


Figure 14-3 Network interface view of Dell Equallogic PS Series

2. Right-click the Ethernet interface and select **Modify IP settings**, as shown in Figure 14-4.

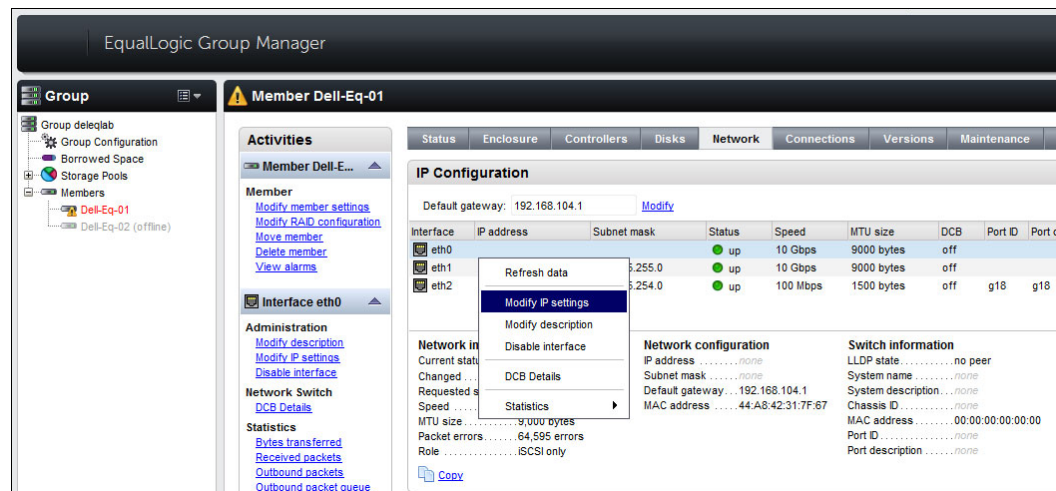


Figure 14-4 Modifying the IP settings of the Dell controller

3. In the dialog box that is shown in Figure 14-5, enter the parameters for one iSCSI target port. Specify the IP address, the Subnet mask, and the Default gateway. Select the **Enable interface** check box.

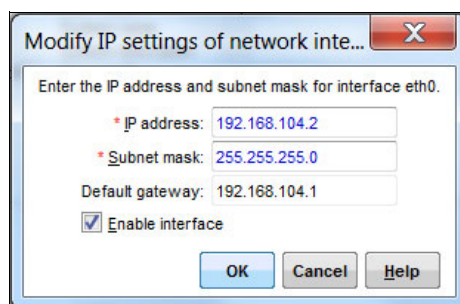


Figure 14-5 Window to assign the IP address of the network interface

4. Click **OK**.
5. Click the interface name and validate the changed settings in the detailed view, as shown in Figure 14-6.

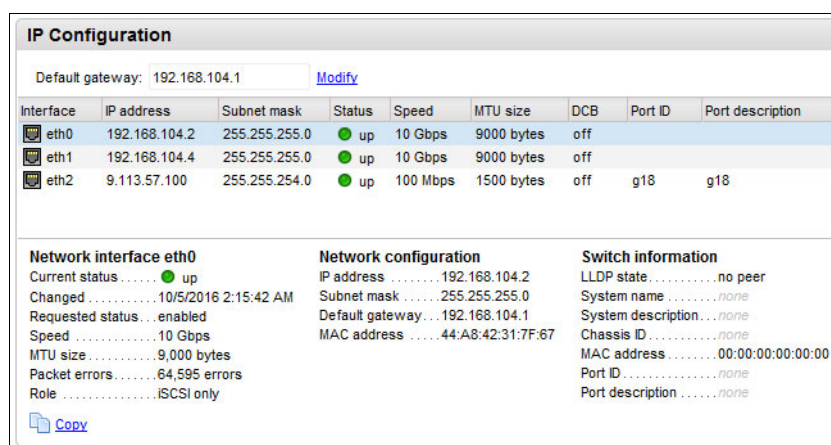


Figure 14-6 Window to view the changed IP addresses of the Dell system

6. Repeat the steps to configure other planned network interfaces.

14.2.2 Setting up access policies

Access control policies are the centralized mechanism for managing access controls for volume access. This access policy mechanism is implemented to ensure that only wanted hosts can access the volumes. For iSCSI connectivity, this access control can be achieved by using either one or a combination of the following methods:

- ▶ The host iSCSI initiator IQN
- ▶ The IP address that the server is using for iSCSI access
- ▶ Challenge Handshake Authentication Protocol (CHAP)

Dell Equallogic PS Series enables you to create the access policies and apply them on the volume that allows access to the defined hosts to the volume. One access policy can be applied to multiple volumes in the system. If you are planning to virtualize LUNs from SAN Volume Controller or IBM Storwize systems, then it is advisable that you use the same access policy for all the LUNs that are mapped to the SAN Volume Controller or IBM Storwize system.

To create access policies on Dell Equallogic PS Series, complete the following steps:

1. Open the Dell Equallogic PS Series Group manager by using the management IP address in a supported browser.
2. Open the Group Configuration window and click the **Access Policies** tab, as shown in Figure 14-7.

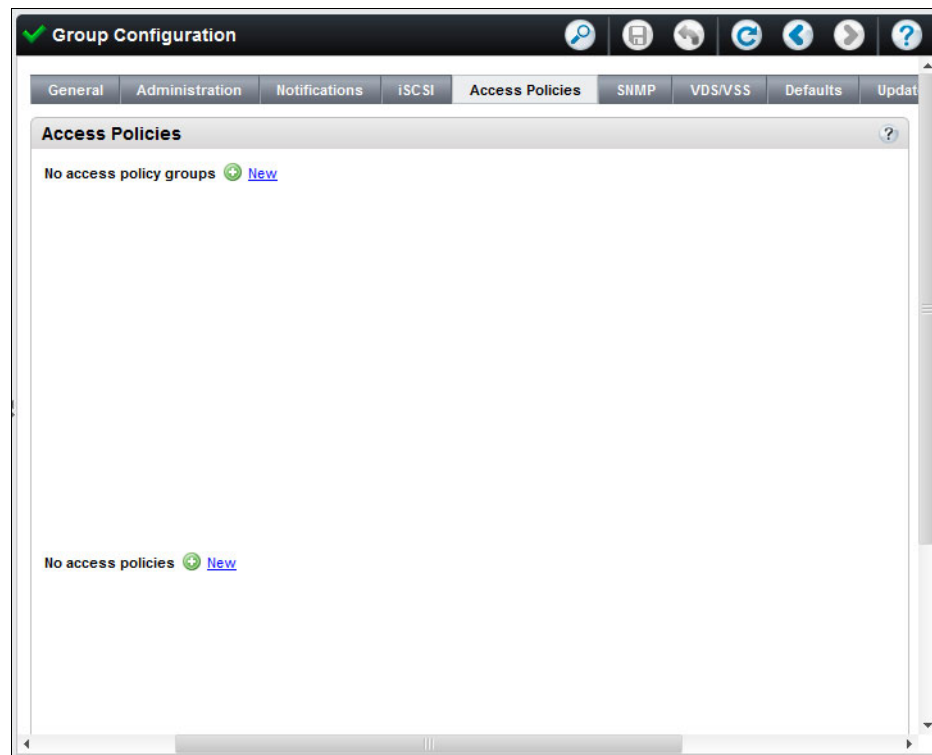


Figure 14-7 Adding an access policy

3. Click the **New Access Policy** dialog box.

4. Provide a policy Name with an optional Description, as shown in Figure 14-8.

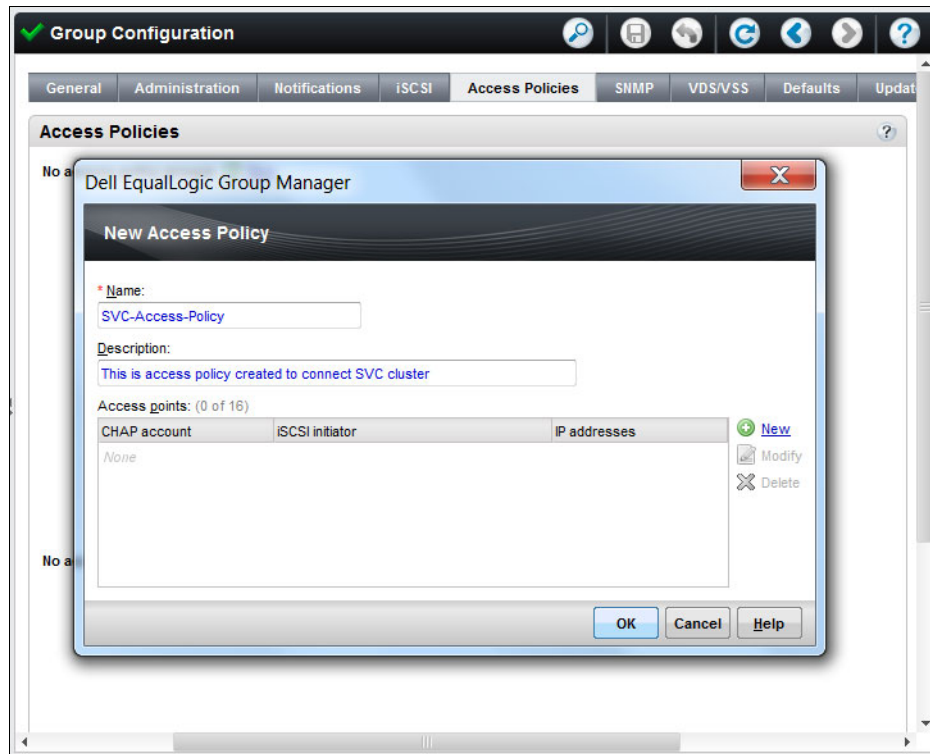


Figure 14-8 Providing an access policy name

5. Select **New** to create an extended access point. Figure 14-9 shows the New Extended Access Point dialog box.

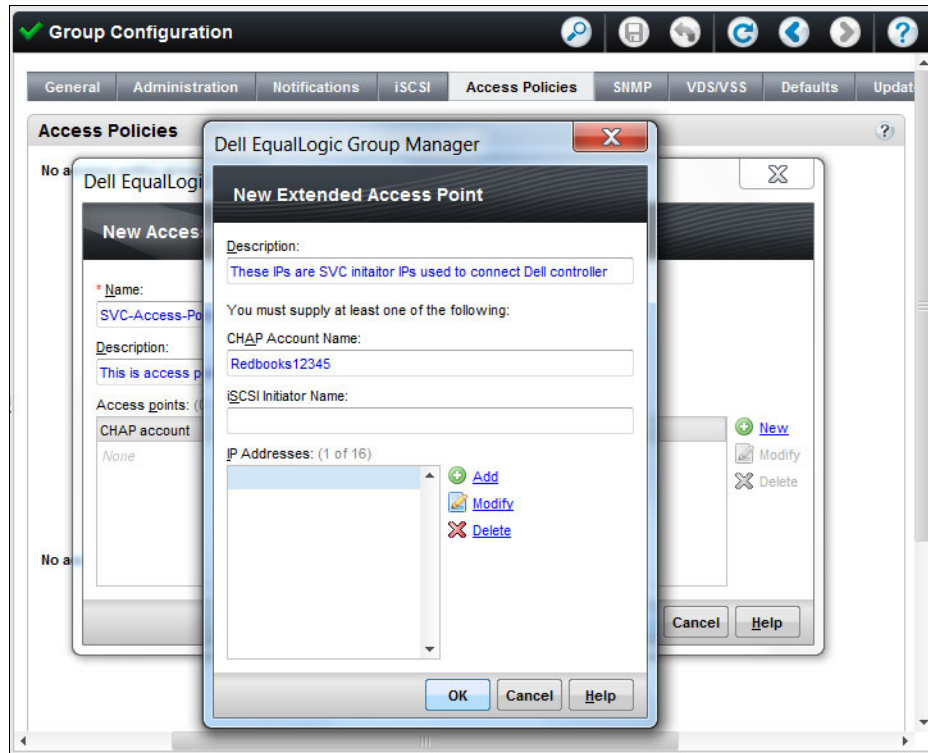


Figure 14-9 Dialog box to add the access policy details

6. Click **Add** to provide the IP addresses of the initiator ports that are intended to access the volume. CHAP authentication is optional and can be used, if required. For connecting a SAN Volume Controller or IBM Storwize controller, it is advised to use an IP address while creating the access policies. Therefore, you must add multiple IP addresses to one extended access point. Repeating this step for all of the initiator IP addresses creates a single extended access policy for all of the initiator IP addresses.

Figure 14-10 shows the added IP addresses of the IBM Storwize initiator.

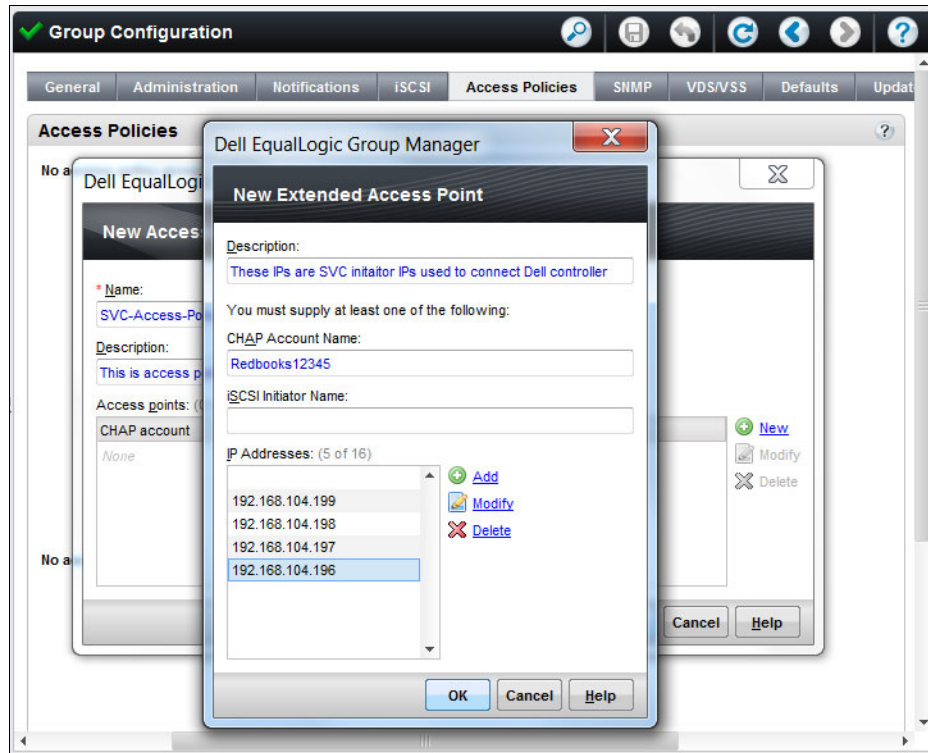


Figure 14-10 Added IP addresses in the access policy

- Click **OK**. You see the list of IP addresses that are added in the extended access point, as shown in Figure 14-11.

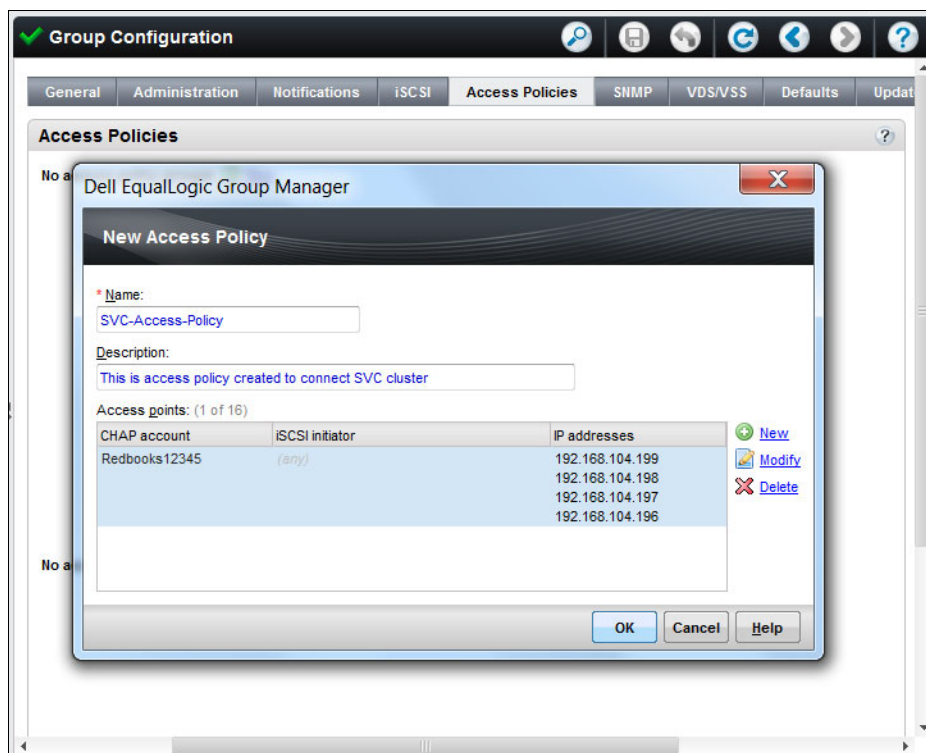


Figure 14-11 Window to review the new access policy details

- Click **OK**. You see the newly created access policy in the Access Policies tab of the Group Configuration window, as shown in Figure 14-12.

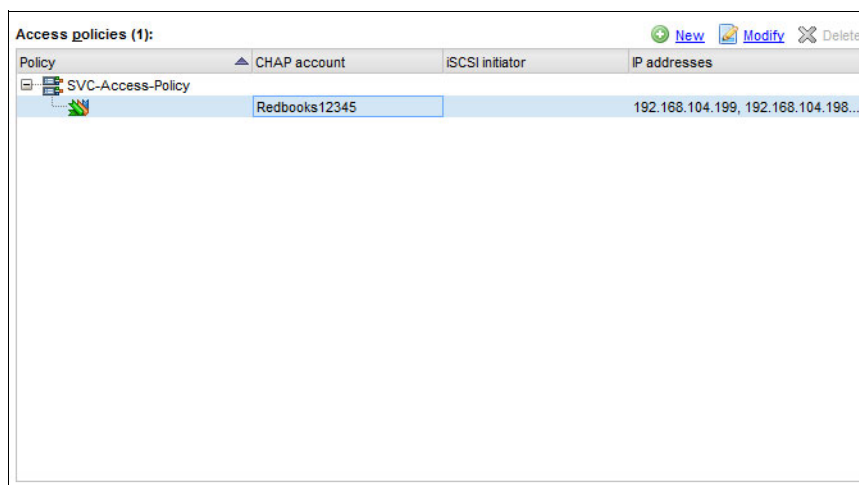


Figure 14-12 Access policy creation completed

14.2.3 Creating volumes and applying access policies

To create the volume of Dell Equallogic PS Series Group, you might need to create a storage pool. You must use Dell product guidelines to plan the storage pools for better performance. The volumes can be created by using a default pool.

To create a volume, complete the following steps:

1. Open the Dell EqualLogic PS Series Group Manager and go to the management IP address in a supported browser.
2. Click **Volumes**. In the Activities window, click **Create volume**, as shown in Figure 14-13.

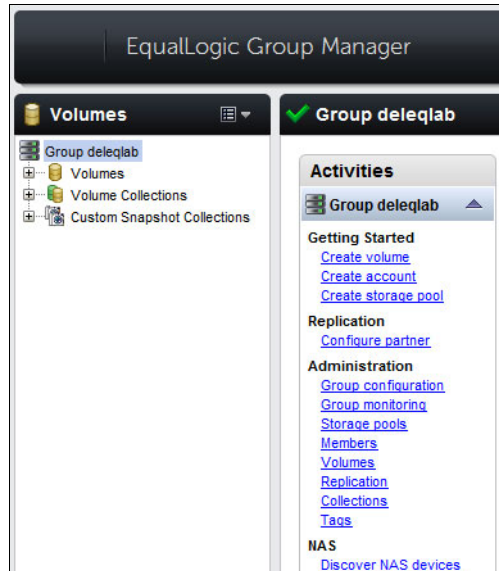


Figure 14-13 Creation of the volume on the Dell system

3. Provide the volume Name and an optional Description. You can optionally select the folder, select the needed storage pool, and click **Next**, as shown in Figure 14-14.

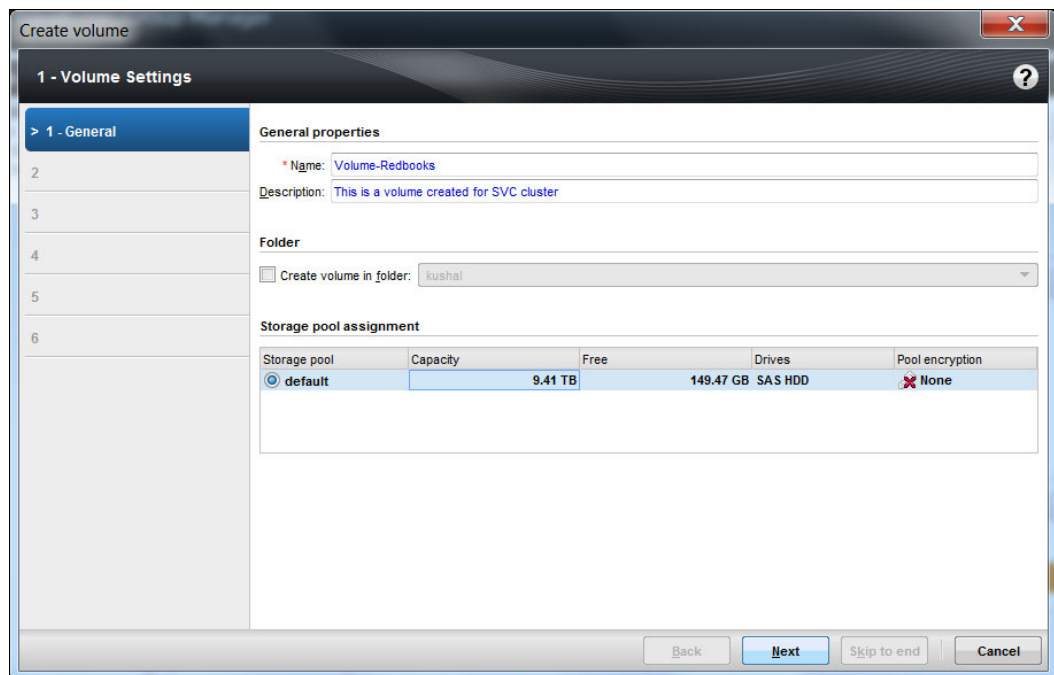
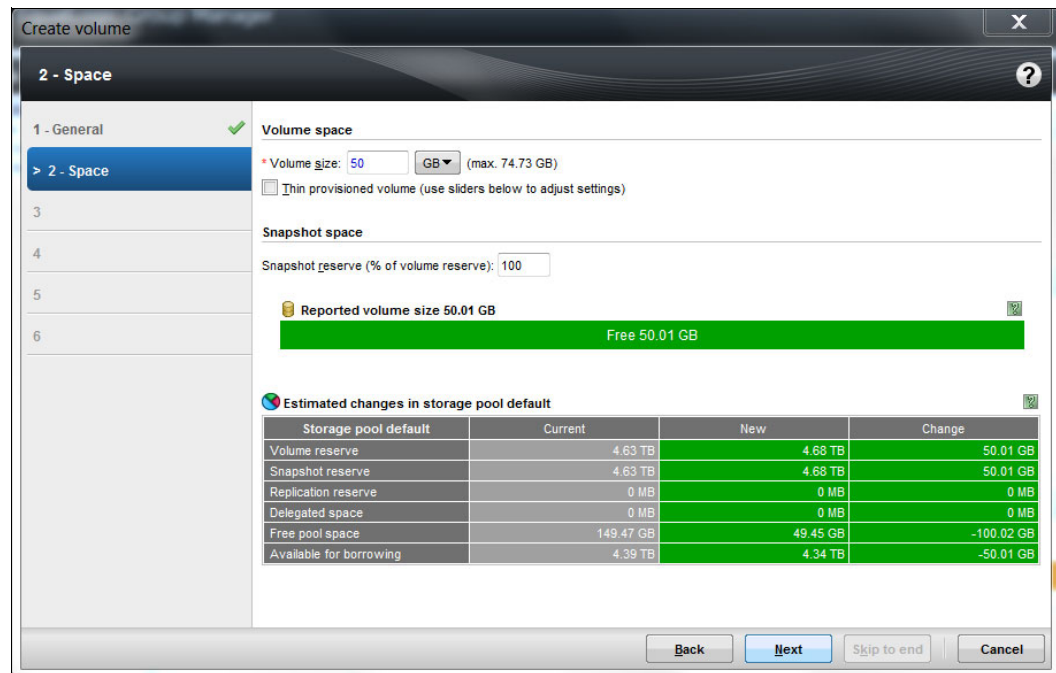


Figure 14-14 Providing Name, optional Description, and storage pool for volume

4. Specify the volume size. Select the check box if you want to create a thin-provisioned volume. Select a suitable snapshot reserve that is based on the requirement and Dell configuration guidelines, and click **Next**. Figure 14-15 shows a sample volume creation with the default pool and no thin provisioning.



Create volume

2 - Space

1 - General ☒ **Volume space**

> 2 - Space

3

4

5

6

* Volume size: GB (max. 74.73 GB)

☐ Thin provisioned volume (use sliders below to adjust settings)

Snapshot space

Snapshot reserve (% of volume reserve):

Reported volume size 50.01 GB

Free 50.01 GB

Estimated changes in storage pool default

Storage pool default	Current	New	Change
Volume reserve	4.63 TB	4.68 TB	50.01 GB
Snapshot reserve	4.63 TB	4.68 TB	50.01 GB
Replication reserve	0 MB	0 MB	0 MB
Delegated space	0 MB	0 MB	0 MB
Free pool space	149.47 GB	49.45 GB	-100.02 GB
Available for borrowing	4.39 TB	4.34 TB	-50.01 GB

Back Next Skip to end Cancel

Figure 14-15 Window to provide volume size and capacity saving details

5. Select the access policy for the volume. If you are planning to connect a Dell Equallogic PS Series system behind a SAN Volume Controller or Storwize cluster, it is advised that you should use the same access policies for all of the LUNs that are virtualized behind the SAN Volume Controller or IBM Storwize cluster. To create access policies, see 14.2.2, “Setting up access policies” on page 271. Select the **Select or define access control policies** radio button. Select the access policy that you want to apply on the volume and click **Add**, as shown in Figure 14-16.

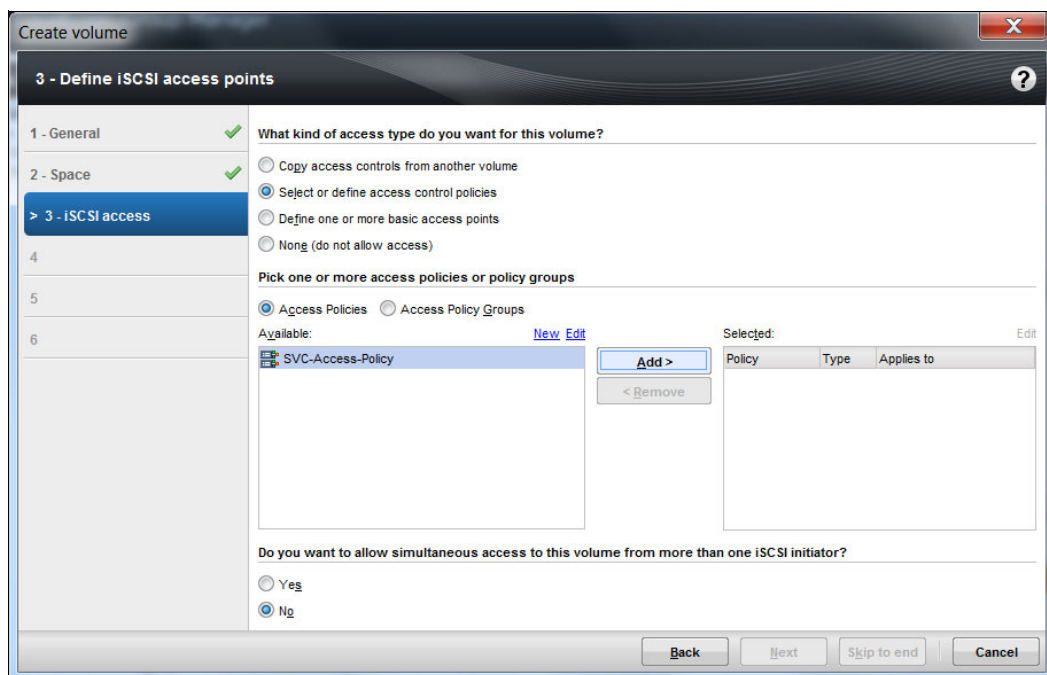


Figure 14-16 Applying the access policy to volume

6. When you add the access policy, it is moved to the Selected box. Because each SAN Volume Controller node has an IQN, for one SAN Volume Controller or IBM Storwize cluster there are multiple iSCSI initiators. Therefore, you must enable simultaneous access to the volume from more than one iSCSI initiator by selecting **Yes**.

7. If you need default Dell controller settings, click **Skip to end**, as shown in Figure 14-17. Otherwise, select **Next** to change the sector size and specify the size that is recommended by Dell configuration guidelines for controller virtualization.

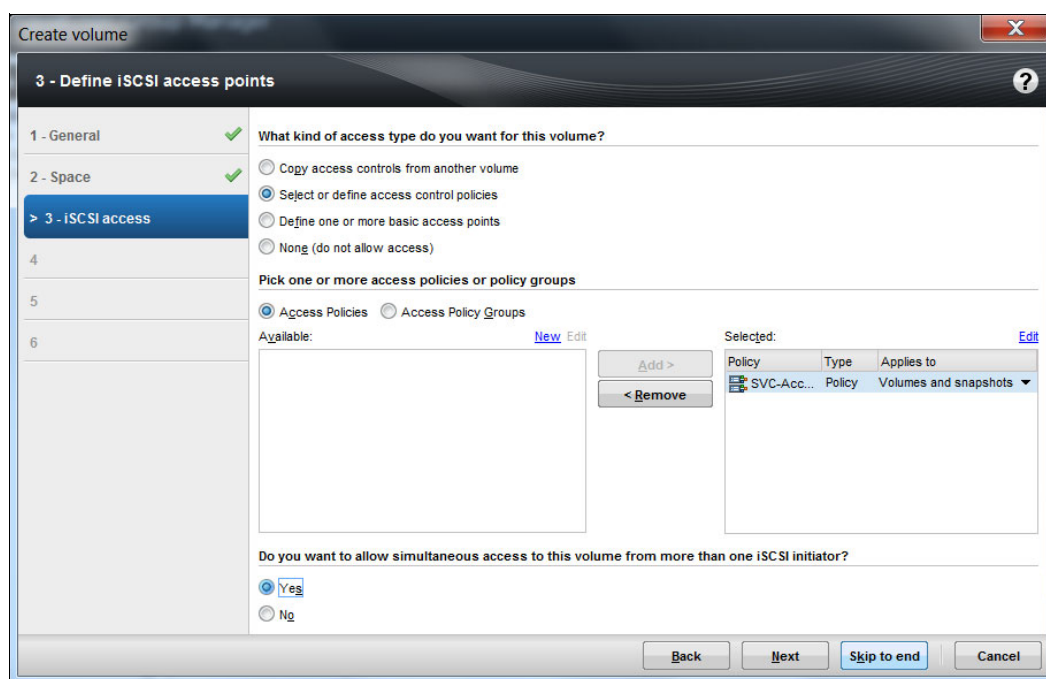


Figure 14-17 Allowing simultaneous access to more than one iSCSI initiator

8. Validate the summary and click **Finish**, as specified in Figure 14-18 to create the volume in the Dell controller.

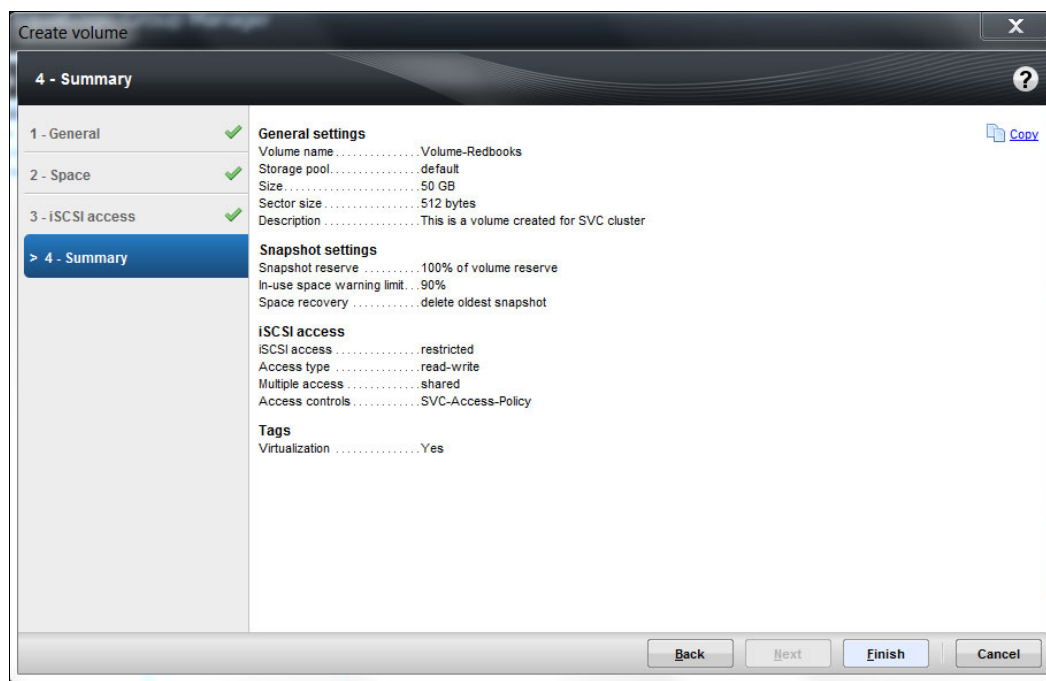


Figure 14-18 Volume summary to validate before creation

Your volume is now created, and it can be viewed in the list of volumes. Figure 14-19 shows the created volume in the volume list. You can validate volume information in this window.

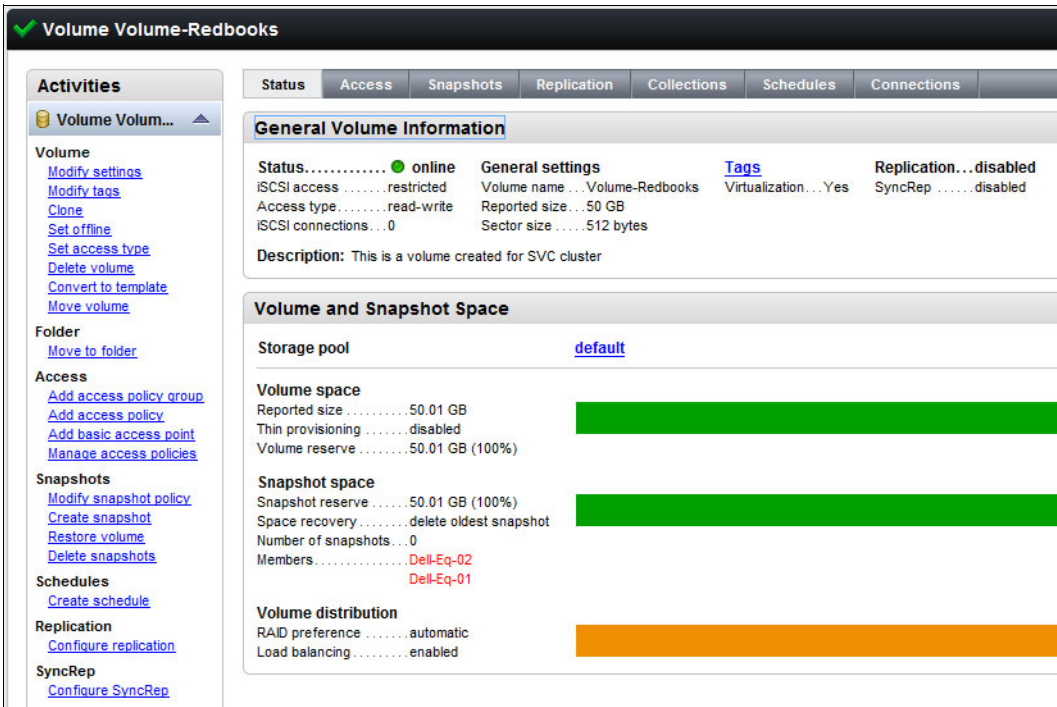


Figure 14-19 Dell volume details after volume creation

14.2.4 Modifying the settings of existing volumes

This section explains how to change the access policy mapping to the existing pre-created volume, and how to allow multiple iSCSI initiator access simultaneously. These steps are needed if you have a volume that is created on Dell Equallogic PS Series and you are planning online data migration to an SAN Volume Controller or IBM Storwize system. These settings also must be applied if you plan to use existing volumes for iSCSI-based virtualization with SAN Volume Controller or IBM Storwize systems.

Applying access policies on an existing volume

If you have a volume that is created on a Dell Equallogic PS Series system, and you are planning to change the access policies due to the reasons mentioned previously, complete the following steps to apply a new access policy on an existing volume. It is advised that you should remove older access policy mappings to prevent unwanted initiators from accessing the volume.

To apply the access policy on the volume, complete the following steps:

1. Open the Dell Equallogic PS Series Group manager by using the management IP address in a supported browser.
2. Select **Volumes** and click the volume name in the list. You see general volume information and volume space information in the window.

3. In the Activities window, click **Add access policy**, as shown in Figure 14-20.

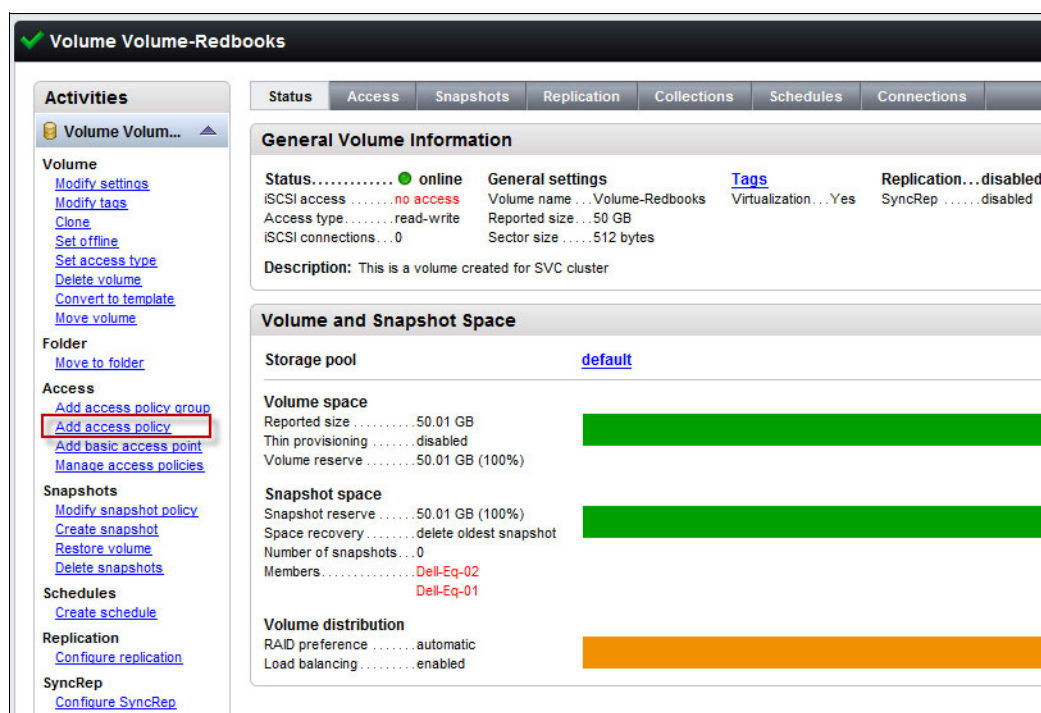


Figure 14-20 Window to add an access policy

4. Select the access policy that you want to apply on the volume. You can optionally create a new access policy by clicking **New**, which takes you to the access policy creation wizard that is described in 14.2.2, “Setting up access policies” on page 271. Click **OK**. The access policy is now applied to the volume. Figure 14-21 shows the access policy that is applied to the volume.

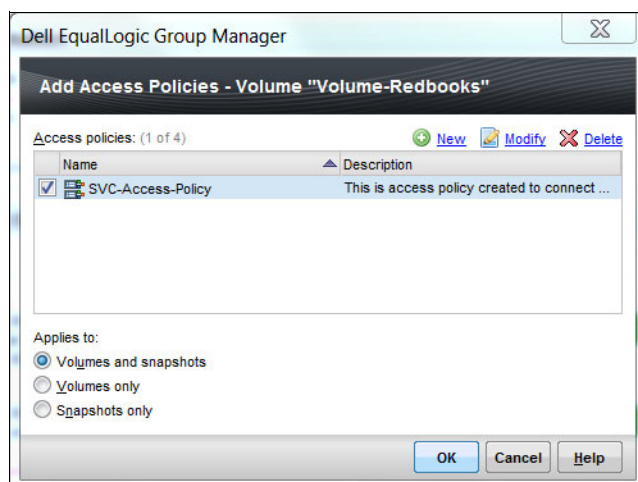


Figure 14-21 Selection of the access policy

- To validate the access policy application, you must view the **Access** tab of the volume, as shown in Figure 14-22. This tab gives you information about the access policies for the selected volume. Other access policies can be added by clicking **Add** for access policies.

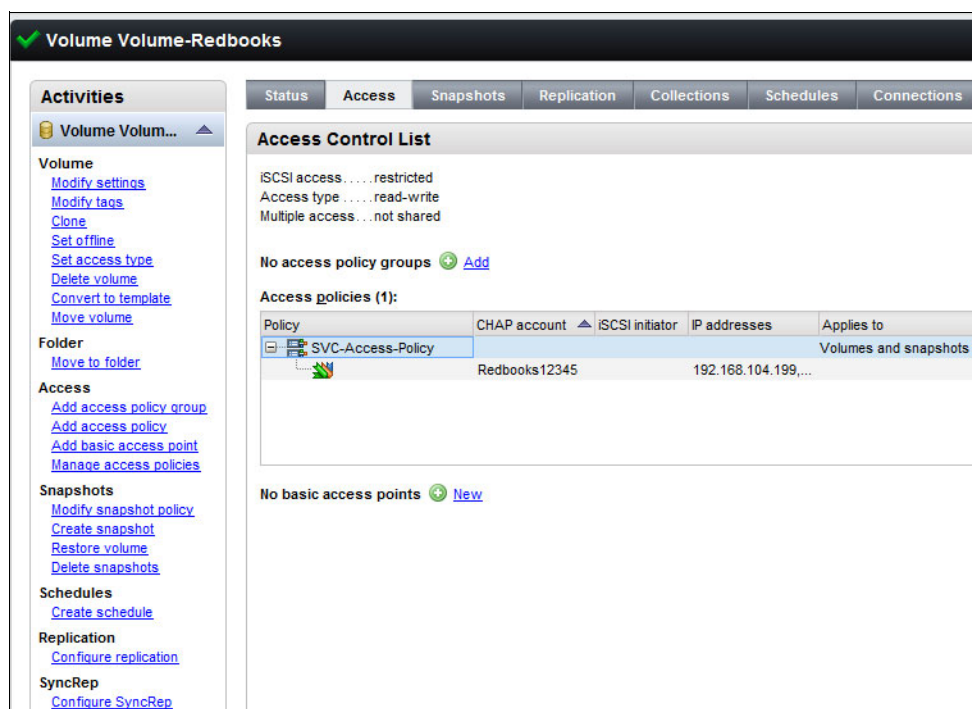


Figure 14-22 Window to validate access policy application

Allowing simultaneous access from multiple initiators

If you are planning to connect the Dell Equallogic PS Series volume to a SAN Volume Controller or IBM Storwize system, a SAN Volume Controller system has multiple iSCSI initiators. Due to this, the volume is accessed by multiple iSCSI initiators at the same time. You must perform the following steps for all the volumes that are supposed to be connected to a SAN Volume Controller or IBM Storwize system for migration or virtualization purposes.

If you must provide simultaneous access through multiple iSCSI initiators, complete the following steps:

- Open the Dell Equallogic PS Series Group manager by using the management IP address in a supported browser.
- Select **Volumes** and click the volume name in the list. You see general volume information and volume space information in the window.

3. In the Activities window, click **Set access type**, as shown in Figure 14-23.

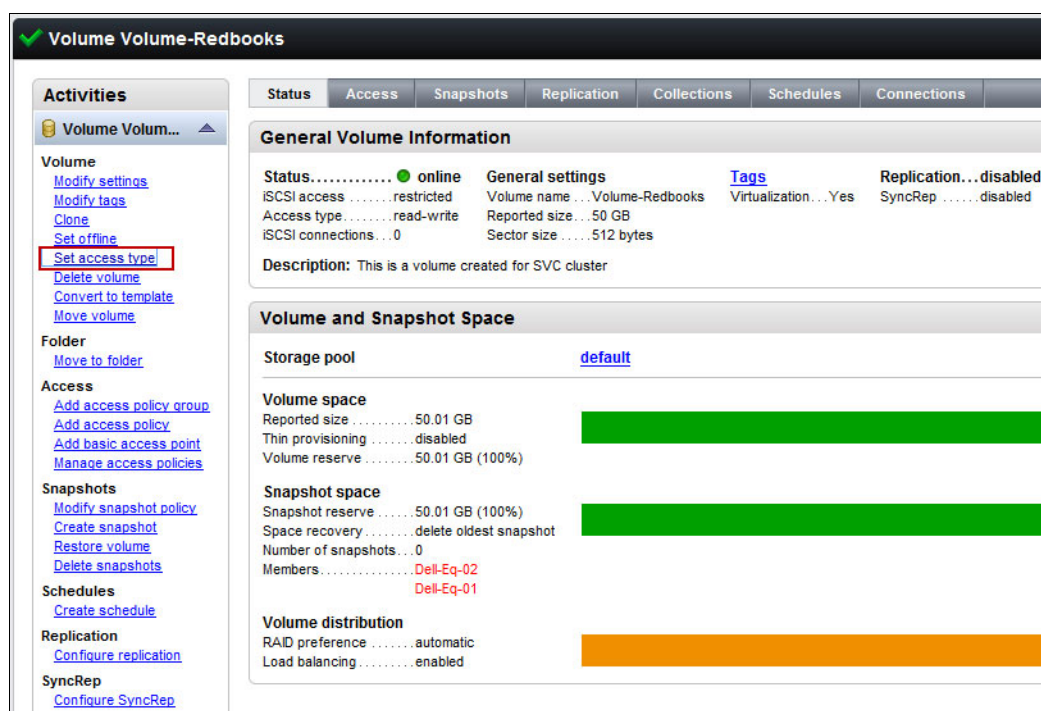


Figure 14-23 Selection of the access type for the volume

4. Select the **Allow simultaneous connections from initiators with different IQNs** check box and click **OK**, as shown in Figure 14-24.

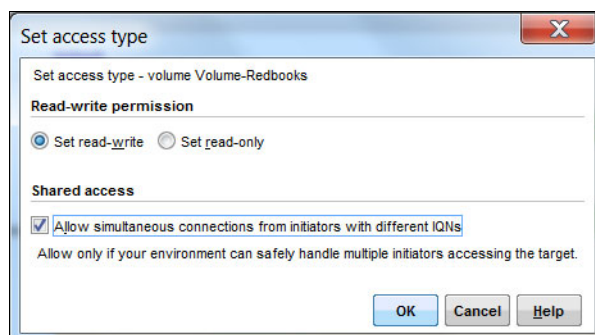


Figure 14-24 Enabling simultaneous access

14.3 Initiator configuration

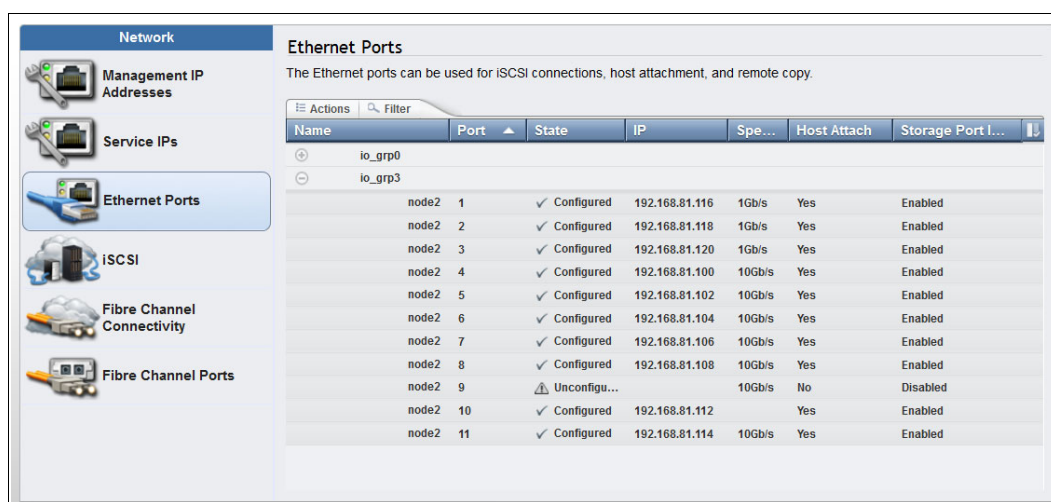
This section deals with SAN Volume Controller system configuration and setting up iSCSI-attached storage with SAN Volume Controller or IBM Storwize Version 7.7.0 and later to virtualize Dell Equallogic PS Series. To virtualize Dell Equallogic PS Series, refer to 14.1, “Planning considerations” on page 266 before connecting Dell to the SAN Volume Controller or IBM Storwize systems. Before this step, users are expected to configure back-end storage by using the guidelines that are described in the previous sections of this chapter. In addition, it is assumed that the IQNs of the SAN Volume Controller or IBM Storwize initiators are added to the back-end system.

This section describes the initiator configuration while connecting a Dell Equallogic PS Series controller for migration or virtualization purpose through the iSCSI protocol. After you ensure that the target configuration is complete, the steps in this section must be followed to complete SAN Volume Controller or IBM Storwize initiators. This section is bifurcated into two subsections elaborating GUI and CLI configurations, respectively.

14.3.1 GUI workflow

When your target is configured and ready to connect, complete the following steps to connect Dell Equallogic PS Series to a SAN Volume Controller or IBM Storwize initiator system by using the GUI:

1. Log in to the SAN Volume Controller cluster user interface in a supported browser.
2. Check that the storage flags are enabled for all of the ports that you are planning for the Dell controller attachments. This information can be viewed in the Ethernet Ports window. Click **Settings** → **Network** → **Ethernet ports** and check for the Storage Port IPv4 or Storage Port IPv6, as shown in Figure 14-25. If ports are not enabled, follow the procedure that is described in 11.2.2, “Source port configuration” on page 220 to enable the storage ports.



The screenshot shows the 'Ethernet Ports' window in the SAN Volume Controller GUI. The left sidebar contains navigation options: Network, Management IP Addresses, Service IPs, Ethernet Ports (selected), iSCSI, Fibre Channel Connectivity, and Fibre Channel Ports. The main area displays a table of Ethernet ports with columns: Name, Port, State, IP, Spe..., Host Attach, and Storage Port I... The table lists two groups of ports: io_grp0 and io_grp3. io_grp0 has one port (node2 1) which is Configured with IP 192.168.81.116, 1Gb/s speed, and Storage Port I... Enabled. io_grp3 has 11 ports (node2 2-11) which are Configured with various IP addresses, 1Gb/s or 10Gb/s speeds, and Storage Port I... Enabled. Port node2 9 is marked as Unconfigured with a warning icon and has Storage Port I... Disabled.

Name	Port	State	IP	Spe...	Host Attach	Storage Port I...
io_grp0						
node2	1	✓ Configured	192.168.81.116	1Gb/s	Yes	Enabled
node2	2	✓ Configured	192.168.81.118	1Gb/s	Yes	Enabled
node2	3	✓ Configured	192.168.81.120	1Gb/s	Yes	Enabled
node2	4	✓ Configured	192.168.81.100	10Gb/s	Yes	Enabled
node2	5	✓ Configured	192.168.81.102	10Gb/s	Yes	Enabled
node2	6	✓ Configured	192.168.81.104	10Gb/s	Yes	Enabled
node2	7	✓ Configured	192.168.81.106	10Gb/s	Yes	Enabled
node2	8	✓ Configured	192.168.81.108	10Gb/s	Yes	Enabled
node2	9	⚠ Unconfigu...		10Gb/s	No	Disabled
node2	10	✓ Configured	192.168.81.112		Yes	Enabled
node2	11	✓ Configured	192.168.81.114	10Gb/s	Yes	Enabled

Figure 14-25 Storage flag settings GUI window

3. Click **Pools** → **External Storage** and click **Add External iSCSI Storage**. A dialog box displays prompts you to input the external storage controller type.

4. Select **Dell** and click **Next**, as shown in Figure 14-26. This action routes you to the Dell storage configuration wizard.

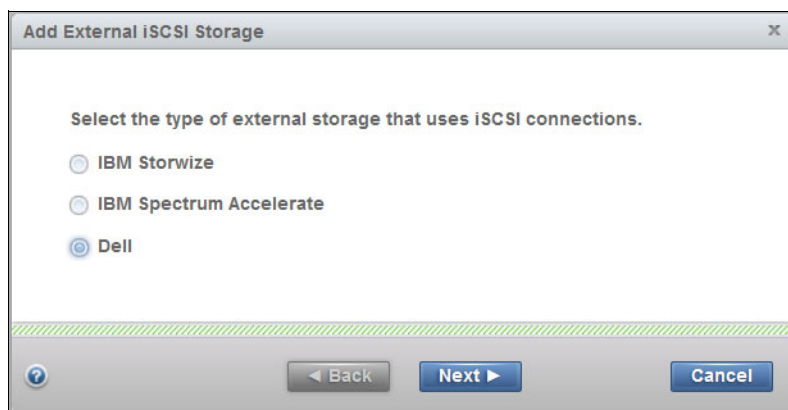


Figure 14-26 GUI dialog box for the selection of the controller type

5. Select the required I/O group from the drop-down menu options and provide the target IP address. Optionally, you can also specify a user name and CHAP secret. Specify the source ports from the drop-down list and click **Next**. Figure 14-27 provides an example of a completed form with all of the mentioned details.

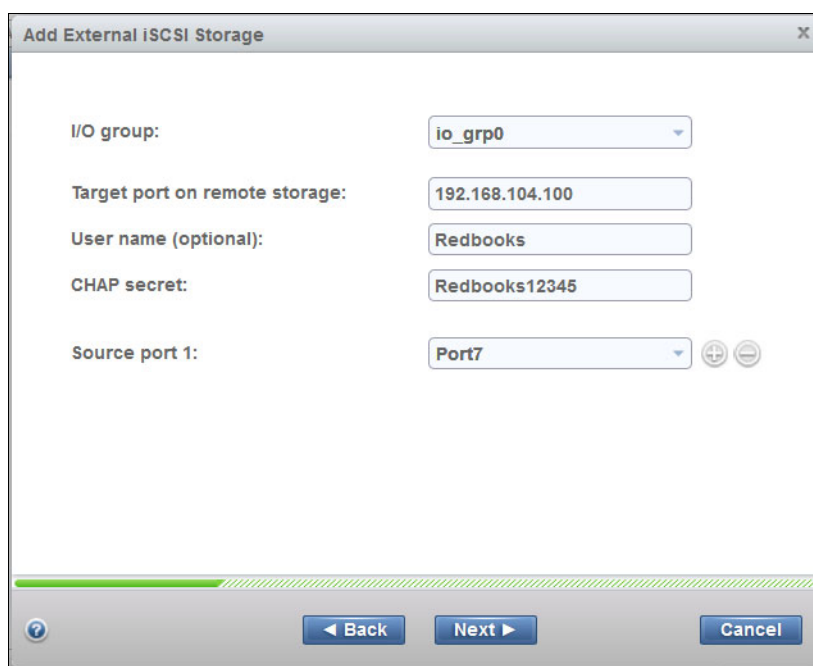


Figure 14-27 Completed form for Dell Equallogic PS Series connection wizard

6. The list of discovered targets is displayed in a window with their respective IQNs and source ports. Figure 14-28 shows the sample window showing all of the discovered Dell LUNs.

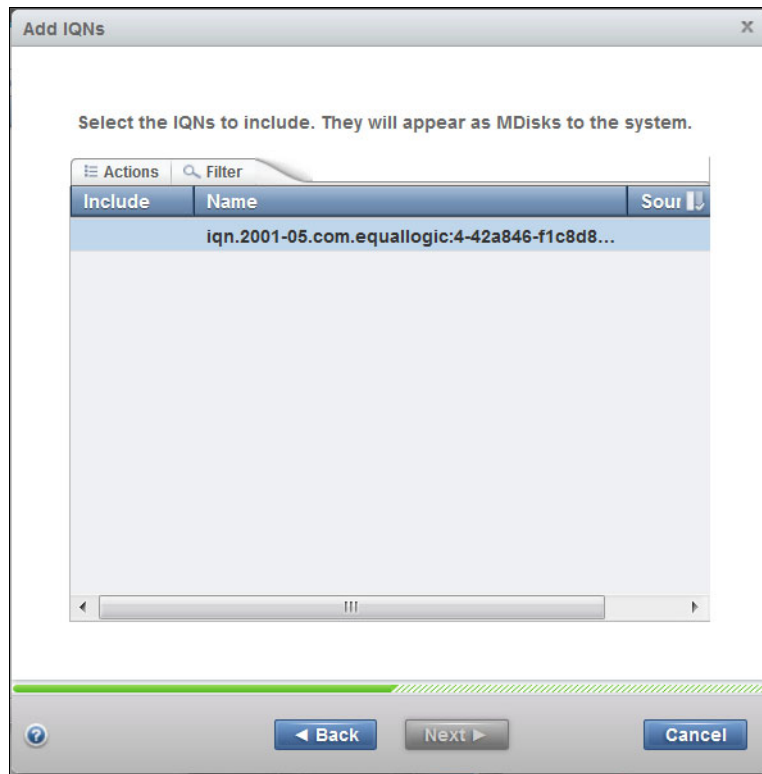


Figure 14-28 Discovered LUNs from the Dell controller

7. Right-click the LUNs that you want to connect to and click **Include**, as shown in Figure 14-29.

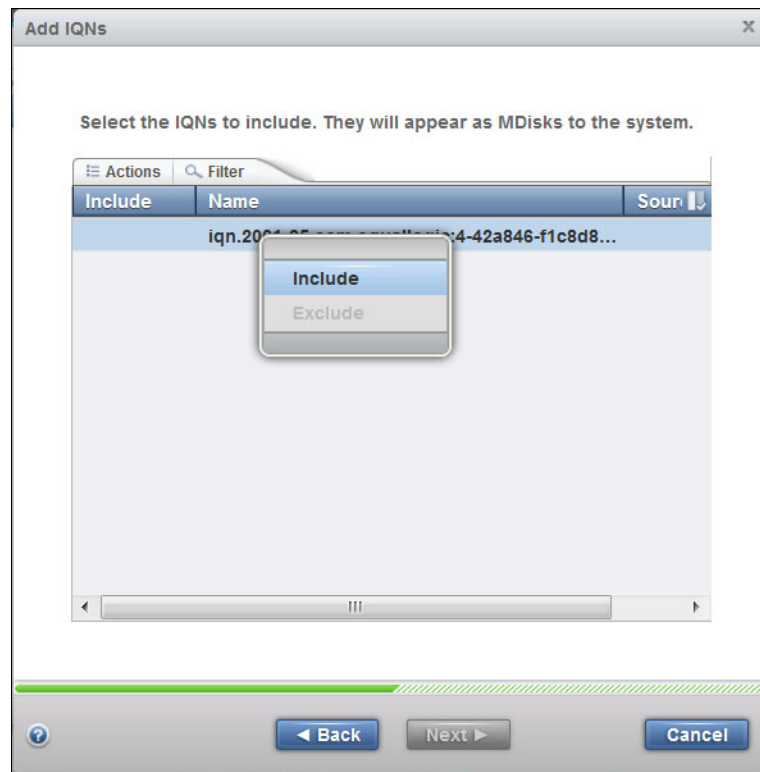


Figure 14-29 Including LUNs for connection

8. After you click **Include**, click the drop-down list to select the source port from which you want to connect the target LUN and click **OK**. Figure 14-30 shows the selected port to connect the back-end LUN from where the back end is discovered.

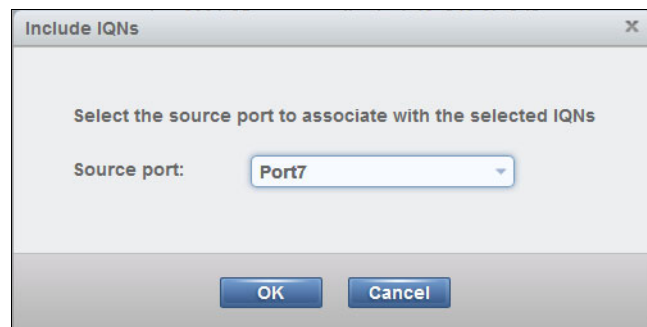


Figure 14-30 Port selection window for back-end LUN connectivity

9. The LUN discovery window shows the included status and the appropriate port from where the LUN is included. In Figure 14-31, port 7 is selected to connect the back-end LUN. Perform this task for all of the LUNs that you are planning to include. Click **Next** to proceed.

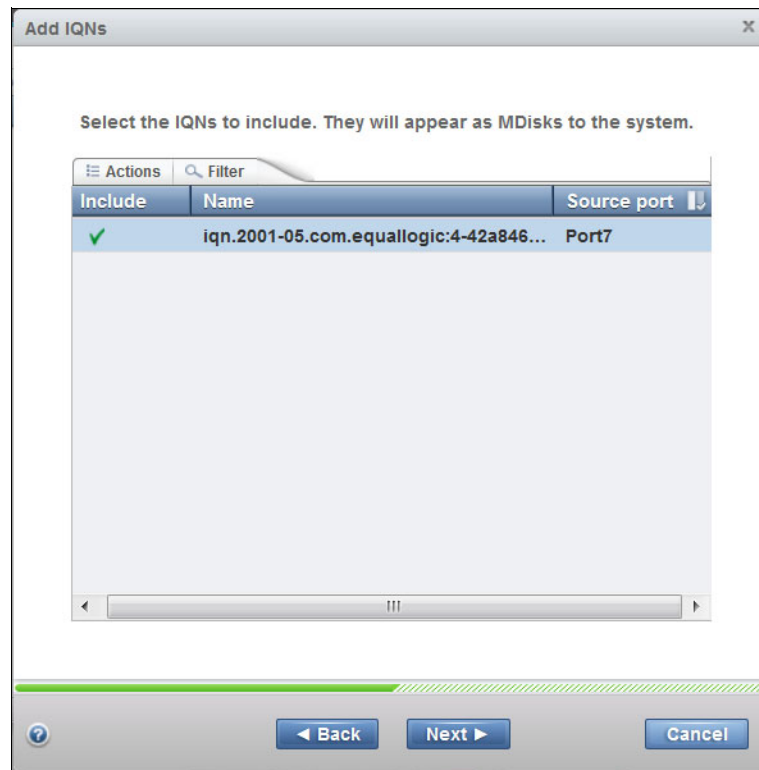


Figure 14-31 Included LUN to connect from the initiator system

10. The summary of the included LUN is displayed in the summary window for validation, as shown in Figure 14-32. Click **Finish** after you review the summary. This action creates sessions to the specified LUNs of Dell controller.

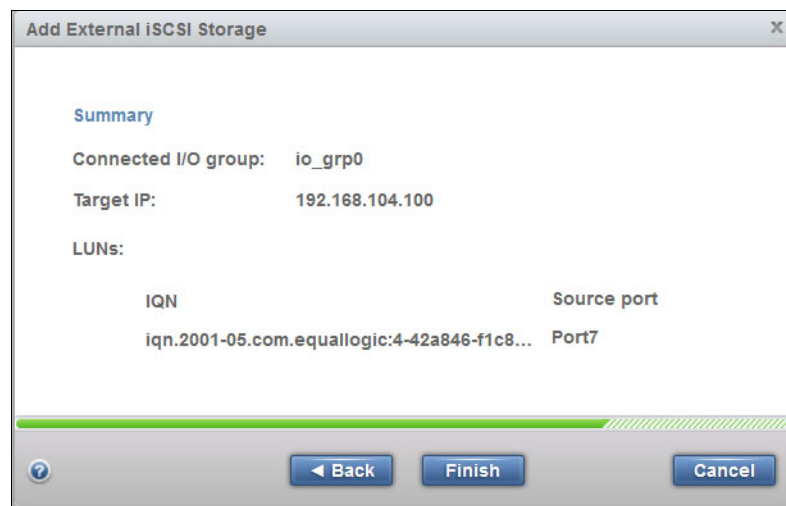


Figure 14-32 Summary for back-end connection establishment

11. A set of commands run, and you see the task completion window indicating that the sessions are established from the initiator system to the target Dell controller.
12. To validate the added external storage controller, click **Pools** → **External Storage**. The newly added Dell Controller window opens, as shown in Figure 14-33. Click the plus sign (+) to see the MDisks that are associated with the controller and validate the correct status of the MDisks.

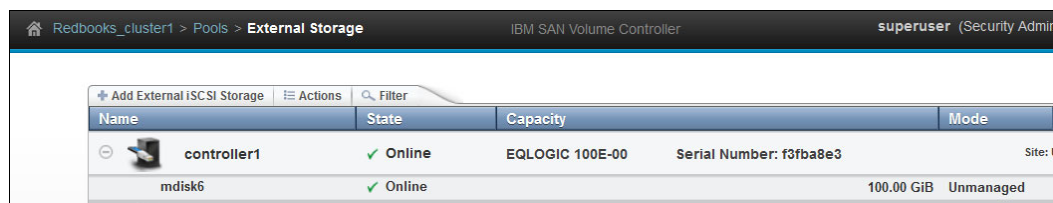


Figure 14-33 Added external Dell controller to the SAN Volume Controller system

After these steps are complete, you can use the added external Dell Equallogic PS Series controller for migration or virtualization purposes. The steps are the same as for external LUNs attached through Fibre Channel (FC): In the **Pools** → **MDisks by Pools** window, create a pool if necessary and add the managed disks to the pool (Figure 14-34).

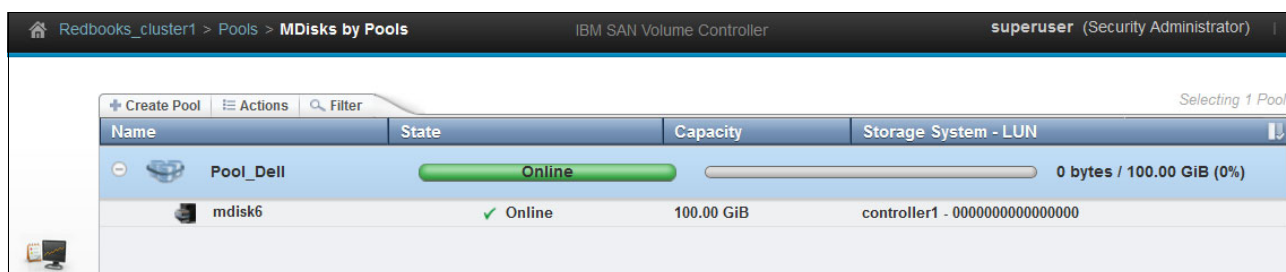


Figure 14-34 MDisk from a Dell controller that is configured in a new pool

14.3.2 CLI workflows

This section provides information about configuring a Dell Equallogic PS Series storage system as an external iSCSI controller by using SAN Volume Controller or IBM Storwize CLI commands:

1. Log in to the SAN Volume Controller cluster command-line interface (CLI) by performing an SSH connection to the cluster IP address.
2. Check that the storage flags are enabled for all of the ports that you are planning for the Dell controller attachments, as shown in the **lsportip** CLI view in Example 14-1.

Example 14-1 The *lsportip* command output

```
IBM_2145:Redbooks_cluster1:superuser> svcinfo lsportip -delim ,
id,node_id,node_name,IP_address,mask,gateway,IP_address_6,prefix_6,gateway_6,MAC,duplex,state,speed,failover,link_state,host,remote_copy,host_6,remote_copy_6,remote_copy_status,remote_copy_status_6,vlan,vlan_6,adapter_location,adapter_port_id,lossless_iscsi,lossless_iscsi6,storage,storage_6
1,1,node1,,,,,,,,,08:94:ef:1b:bc:71,Full,unconfigured,1Gb/s,no,active,,0,,0,,,,,0,1,,,,
1,1,node1,,,,,,,,,08:94:ef:1b:bc:71,Full,unconfigured,1Gb/s,yes,active,,0,,0,,,,,0,1,,,,
2,1,node1,,,,,,,,,08:94:ef:1b:bc:72,Full,unconfigured,1Gb/s,no,active,,0,,0,,,,,0,2,,,,
```



```

2,1,node1,,,,,,,,08:94:ef:1b:bc:72,Full,unconfigured,1Gb/s,yes,active,,0,,0,,,,,
0,2,,,,,
3,1,node1,,,,,,,,08:94:ef:1b:bc:73,,unconfigured,,no,inactive,,0,,0,,,,,0,3,,,,,
3,1,node1,,,,,,,,08:94:ef:1b:bc:73,,unconfigured,,yes,inactive,,0,,0,,,,,0,3,,,,,
4,1,node1,192.168.104.50,255.255.0.0,192.168.104.1,,,,40:f2:e9:e0:de:f4,Full,co
nfigured,10Gb/s,no,active,yes,0,,0,,,,,1,1,off,,yes,
4,1,node1,,,,,,,,40:f2:e9:e0:de:f4,Full,configured,10Gb/s,yes,active,,0,,0,,,,,1
,1,,,,,
5,1,node1,192.168.104.51,255.255.0.0,192.168.104.1,,,,40:f2:e9:e0:de:f5,Full,co
nfigured,10Gb/s,no,active,yes,0,,0,,,,,1,2,off,,yes,
5,1,node1,,,,,,,,40:f2:e9:e0:de:f5,Full,configured,10Gb/s,yes,active,,0,,0,,,,,1
,2,,,,,
6,1,node1,192.168.104.52,255.255.0.0,192.168.104.1,,,,40:f2:e9:e0:de:f6,,config
ured,,no,inactive,yes,0,,0,,,,,1,3,off,,yes,
6,1,node1,,,,,,,,40:f2:e9:e0:de:f6,,configured,,yes,inactive,,0,,0,,,,,1,3,,,,,
7,1,node1,192.168.104.53,255.255.0.0,192.168.104.1,,,,40:f2:e9:e0:de:f7,Full,co
nfigured,10Gb/s,no,active,yes,0,,0,,,,,1,4,off,,yes,
7,1,node1,,,,,,,,40:f2:e9:e0:de:f7,Full,configured,10Gb/s,yes,active,,0,,0,,,,,1
,4,,,,,
1,3,node2,,,,,,,,08:94:ef:1b:b6:29,Full,unconfigured,1Gb/s,no,active,,0,,0,,,,,0
,1,,,,,
1,3,node2,,,,,,,,08:94:ef:1b:b6:29,Full,unconfigured,1Gb/s,yes,active,,0,,0,,,,,
0,1,,,,,
2,3,node2,,,,,,,,08:94:ef:1b:b6:2a,Full,unconfigured,1Gb/s,no,active,,0,,0,,,,,0
,2,,,,,
2,3,node2,,,,,,,,08:94:ef:1b:b6:2a,Full,unconfigured,1Gb/s,yes,active,,0,,0,,,,,
0,2,,,,,
3,3,node2,,,,,,,,08:94:ef:1b:b6:2b,Full,unconfigured,1Gb/s,no,active,,0,,0,,,,,0
,3,,,,,
3,3,node2,,,,,,,,08:94:ef:1b:b6:2b,Full,unconfigured,1Gb/s,yes,active,,0,,0,,,,,
0,3,,,,,
4,3,node2,192.168.104.54,255.255.0.0,192.168.104.1,,,,40:f2:e9:e0:03:c8,Full,co
nfigured,10Gb/s,no,active,yes,0,,0,,,,,1,1,off,,yes,
4,3,node2,,,,,,,,40:f2:e9:e0:03:c8,Full,configured,10Gb/s,yes,active,,0,,0,,,,,1
,1,,,,,
5,3,node2,192.168.104.55,255.255.0.0,192.168.104.1,,,,40:f2:e9:e0:03:c9,Full,co
nfigured,10Gb/s,no,active,yes,0,,0,,,,,1,2,off,,yes,
5,3,node2,,,,,,,,40:f2:e9:e0:03:c9,Full,configured,10Gb/s,yes,active,,0,,0,,,,,1
,2,,,,,
6,3,node2,192.168.104.56,255.255.0.0,192.168.104.1,,,,40:f2:e9:e0:03:ca,Full,co
nfigured,10Gb/s,no,active,yes,0,,0,,,,,1,3,off,,yes,
6,3,node2,,,,,,,,40:f2:e9:e0:03:ca,Full,configured,10Gb/s,yes,active,,0,,0,,,,,1
,3,,,,,
7,3,node2,192.168.104.57,255.255.0.0,192.168.104.1,,,,40:f2:e9:e0:03:cb,Full,co
nfigured,10Gb/s,no,active,yes,0,,0,,,,,1,4,off,,yes,
7,3,node2,,,,,,,,40:f2:e9:e0:03:cb,Full,configured,10Gb/s,yes,active,,0,,0,,,,,1
,4,,,,,
IBM_2145:Redbooks_cluster1:superuser>

```

If ports are not enabled for storage virtualization, you can use the **cfgportip** command to enable the ports for storage, as shown in Example 14-2.

Example 14-2 Using the cfgportip command to change the storage flag

```
IBM_2145:Redbooks_cluster1:superuser>
IBM_2145:Redbooks_cluster1:superuser>svctask cfgportip -node node1 -storage yes
7
IBM_2145:Redbooks_cluster1:superuser>
```

3. Discover the remote iSCSI targets by using the **svctask detectiscsistorageportcandidate** command, as shown in Example 14-3. Using this command, you can discover all of the iSCSI targets to which you are authorized to connect.

Example 14-3 The detectiscsistorageportcandidate CLI example

```
IBM_2145:Redbooks_cluster1:superuser>
IBM_2145:Redbooks_cluster1:superuser>svctask detectiscsistorageportcandidate
-srcportid 7 -iogrp 0 -username Redbooks -chapsecret Redbooks12345 -targetip
192.168.104.100
IBM_2145:Redbooks_cluster1:superuser>
```

4. Confirm the discovered remote targets by using the **svcinfo lsiscsistorageportcandidate** command. This command shows all of the discovered Dell LUNs and their respective IQNs. Example 14-4 shows the Dell LUN that is discovered from I/O group 0 and its configured status as no, indicating that the connection to this LUN is not established.

Example 14-4 The lsiscsistorageportcandidate command example

```
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageportcandidate
id src_port_id target_ipv4 target_ipv6 target_iscsiname
iogroup_list configured status site_id site_name
0 7 192.168.104.100
iqn.2001-05.com.equallogic:4-42a846-f1c8d8531-2170000303858d0f-volumerb1:-:-:-
no full
IBM_2145:Redbooks_cluster1:superuser>
```

5. After you decide to establish sessions with the targets, you can use the **svctask addiscsistorageport** command to add the targets to the SAN Volume Controller or IBM Storwize system and make them appear as managed disks (MDisks). Example 14-5 shows the sample **addiscsistorageport** storage command. Because every Dell LUN has a different IQN, this command needs to be run for all of the LUN IDs listed in the **addiscsistorageportcandidate** output to which you want to connect.

Example 14-5 The addiscsistorageport command example

```
IBM_2145:Redbooks_cluster1:superuser>
IBM_2145:Redbooks_cluster1:superuser>svctask addiscsistorageport -iogrp 0
-username Redbooks -chapsecret Redbooks12345 0
IBM_2145:Redbooks_cluster1:superuser>
```

6. View the iSCSI connection status by using the **svcinfo lsiscsistorageport** command, which provides the details of the iSCSI initiator connections. Example 14-6 shows how to verify the status of the established sessions by using the CLI.

Example 14-6 Status verification by using the lsiscsistorageport command

```
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageport
id src_port_id target_ipv4 target_ipv6 target_iscsiname
controller_id iogroup_list status site_id site_name
0 7 192.168.104.100
iqn.2001-05.com.equallogic:4-42a846-f1c8d8531-2170000303858d0f-volumerb 1
1:-:-:- full
IBM_2145:Redbooks_cluster1:superuser>
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageport 0
id 0
src_port_id 7
target_ipv4 192.168.104.100
target_ipv6
target_iscsiname
iqn.2001-05.com.equallogic:4-42a846-f1c8d8531-2170000303858d0f-volumerb
controller_id 1
iogroup_list 1:-:-:-
status full
site_id
site_name
node_id 1
node_name node1
src_ipv4 192.168.104.53
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node1
connected yes
node_id 3
node_name node2
src_ipv4 192.168.104.57
src_ipv6
src_iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node2
connected yes
IBM_2145:Redbooks_cluster1:superuser>
```

7. Optionally, you can run the **svctask detectmdisk** command on the SAN Volume Controller or IBM Storwize system that you are trying to virtualize from to detect any LUNs that are exported to the system from the external controller, as shown in Example 14-7.

Example 14-7 Discovering the storage at the initiator

```
IBM_2145:Redbooks_cluster1:superuser>svctask detectmdisk
```

8. You can then run the **svcinfo lsmdisk** command to list discovered storage. Example 14-8 shows the iSCSI attached MDisk upon a successful Dell external controller addition.

Example 14-8 Concise and detailed view of MDisk listing the iSCSI-related fields

```
IBM_2145:Redbooks_cluster1:superuser>lsmdisk
id name status mode mdisk_grp_id mdisk_grp_name capacity ctrl_LUN_#
controller_name UID
tier encrypt site_id site_name distributed dedupe
```

```

0 mdisk0 online managed 0 mdiskgrp0 200.0GB
0000000000000000 controller0
600a0b800068b0ca0000128f559c826100000000000000000000000000000000
tier_enterprise no no no
1 mdisk1 online managed 0 mdiskgrp0 200.0GB
00000000000000001 controller0
600a0b800068b0ca00001292559c827f00000000000000000000000000000000
tier_enterprise no no no
3 mdisk3 online unmanaged 100.0GB
0000000000000000 controller1
64842a1453d8c8f10f8d85030300702100000000000000000000000000000000
tier_enterprise no no no
IBM_2145:Redbooks_cluster1:superuser>
IBM_2145:Redbooks_cluster1:superuser>lsmdisk 3
id 3
name mdisk3
status online
mode unmanaged
mdisk_grp_id
mdisk_grp_name
capacity 100.0GB
quorum_index
block_size 512
controller_name controller1
ctrl_type 4
ctrl_WWNN
controller_id 1
path_count 2
max_path_count 2
ctrl_LUN_# 0000000000000000
UID 64842a1453d8c8f10f8d85030300702100000000000000000000000000000000
preferred_WWPN
active_WWPN
fast_write_state empty
raid_status
raid_level
redundancy
strip_size
spare_goal
spare_protection_min
balanced
tier tier_enterprise
slow_write_priority
fabric_type iscsi
site_id
site_name
easy_tier_load medium
encrypt no
distributed no
drive_class_id
drive_count 0
stripe_width 0
rebuild_areas_total
rebuild_areas_available
rebuild_areas_goal

```

```
dedupe no
preferred_iscsi_port_id 0
active_iscsi_port_id 0
replacement_date
IBM_2145:Redbooks_cluster1:superuser>
```

9. View the discovered external iSCSI-attached controllers by using the **svctask lscontroller** command. A detailed view shows the protocol by which the controller is attached to the system, as shown in Example 14-9.

Example 14-9 Listing the discovered external Dell controller

```
IBM_2145:Redbooks_cluster1:superuser>lscontroller 1
id 1
controller_name controller1
WWNN
mdisk_link_count 1
max_mdisk_link_count 1
degraded no
vendor_id EQLOGIC
product_id_low 100E-00
product_id_high
product_revision 9.0
ctrl_s/n f3fba8e3
allow_quorum no
fabric_type iscsi
site_id
site_name
WWPN
path_count 2
max_path_count 2
iscsi_port_id 0
ip 192.168.104.100
IBM_2145:Redbooks_cluster1:superuser>
```

Now, you can configure the Dell LUNs that are mapped to the cluster of initiator nodes as managed disks in the SAN Volume Controller or IBM Storwize system. The steps are the same as for external LUNs attached through FC. Using the **mkmdiskgrp** CLI command, you can create a pool, if necessary, and add the managed disks to the pool (Example 14-10).

Example 14-10 The mkmdiskgrp command

```
IBM_2145:Redbooks_cluster1:superuser> svctask mkmdiskgrp -ext 1024 -mdisk mdisk3
-name Pool_Dell
MDisk Group, id [0], successfully created
IBM_2145:Redbooks_cluster1:superuser>
```



Configuration and administration of iSCSI

This chapter contains special considerations for when you are administering or changing the configuration of an Internet Small Computer System Interface (iSCSI) SAN.

This chapter describes the following topics:

- ▶ 15.1, “Changing the iSCSI port configuration” on page 298
- ▶ 15.2, “Adding or removing nodes or I/O groups” on page 300
- ▶ 15.3, “Changing the system name or node name” on page 304
- ▶ 15.4, “Changing the CHAP configuration” on page 312
- ▶ 15.5, “Changing the number of LUNs, ports, and IQNs in an IBM Storwize system” on page 315

15.1 Changing the iSCSI port configuration

This section describes how to make configuration changes to a SAN Volume Controller or IBM Storwize system's iSCSI ports. In particular, it describes changing the iSCSI ports' IP addresses and enabling or disabling iSCSI capabilities from IP ports. The impact of changing an iSCSI port's IP address depends on whether that port is an initiator port or a target port. Section 15.1.1, "Changing the iSCSI initiator ports' IP addresses" on page 298 describes changing the initiator ports' IP addresses. Section 15.1.2, "Changing the iSCSI target ports' IP addresses" on page 298 describes changing the target ports' IP addresses. If a port is both an initiator port and a target port, the considerations in both sections apply.

Attention: A SAN Volume Controller or IBM Storwize system's IP ports can be used both for iSCSI and IP replication. This section considers only IP ports that are being used for iSCSI. If the system is also using these ports for IP replication, you must also consider the effects of reconfiguring the port on the system's IP replication partnerships.

15.1.1 Changing the iSCSI initiator ports' IP addresses

You can change the IP address of an iSCSI initiator port by using the **cfgportip** command. If that port is the source port for any iSCSI storage port objects, this command fails unless the **-force** flag is used. For more information about this command, see [IBM Knowledge Center](#).

Changing the IP address of an iSCSI initiator port is nondisruptive if the new IP is accessible to all of the target ports with which the initiator port in question has a session. You can check which IP addresses these are by using the **lsiscsistorageport** view; look in the target IP fields of the lines of the view for which the source port ID field contains the index of the iSCSI port object of interest.

You typically do not need to make configuration changes to any storage controllers that are connected to the initiator port because the storage controller identifies an initiator by its IQN.

15.1.2 Changing the iSCSI target ports' IP addresses

You can change the IP address of an iSCSI target port by using the **cfgportip** command. For more information about this command, see [IBM Knowledge Center](#).

If you change the IP address of a target port, you also must change the configurations of any hosts that are connected to that port. They must be reconfigured to establish iSCSI sessions that use that port's new IP address.

Changing the IP address of an iSCSI target port is disruptive because it brings down any iSCSI sessions with hosts that are using that port. However, if the host has dual-redundant connections to the SAN Volume Controller or IBM Storwize system that use different target ports, it might still be possible to change the IP addresses without loss of access to data.

Because changing the IP address of a target port involves reconfiguring the host, the procedure to change the IP address without loss of access to data depends on the type of hosts that are connected to the target ports.

The following example describes how to do this task when the hosts that are connected to those target ports are other SAN Volume Controller or IBM Storwize systems. To change target port IP addresses to which other types of hosts are connected, you must adapt these instructions for the types of hosts in question. These instructions describe how to use the CLI both on the system whose target ports you want to reconfigure and on the SAN Volume Controller and IBM Storwize systems that are acting as hosts that are attached to these ports. You can use either the CLI or the GUI to change the target system configuration, but you *must* use the CLI to make the changes on the SAN Volume Controller or IBM Storwize host systems without loss of access to data.

Attention:

- Failure to follow these instructions correctly can result in loss of access to data. Do not attempt to proceed past step 1 without resolving any connectivity problems that are uncovered in that step. Do not attempt to repeat step 2 for a new target port until the previous target port is successfully reconfigured and all hosts have good connectivity to it.
- When you are doing this task, the system has a single point of failure. While one of the target ports is being reconfigured in step 2, each host has a connection to only one target port of the system.

Complete the following steps:

1. Ensure that all the hosts that are connected to the system that is using the target ports in question have a dual-redundant connection to the system by using at least two target ports. Also, ensure that all of those connections have good connectivity. To do this task when the hosts are other SAN Volume Controller or IBM Storwize systems, complete the following steps on each of the hosts in question:
 - a. Check that the controller object that represents the target system is not degraded. The degraded field of the detailed **lscontroller** view for the controller object that represents it should contain no.
 - b. Check that each iSCSI storage port object that represents a set of connections to the target system has full connectivity. The status field of the **lsiscsistorageport** view should contain full.

If any host does not have full and dual-redundant connectivity to the target system, diagnose and fix the problem before you continue.
2. Change the IP address of each of the target ports in question and reconfigure the hosts to use the new IP address by completing the following steps. To prevent loss of access to data, do not start reconfiguring a target port until you are finished reconfiguring the previous target port and the hosts all have good connectivity to it.
 - a. Change the IP address of the target port in question with the **cfgportip** command on the target SAN Volume Controller or IBM Storwize system.
 - b. Reconfigure any hosts' connections with this port to connect to it by using the new IP address. When the hosts in question are other SAN Volume Controller or IBM Storwize systems, do this task by carrying out the following steps on the hosts in question:
 - i. Identify the connections that must be reconfigured by using **lsiscsistorageport**; look for the iSCSI storage port object for which the target IP field contains the old IP address of the target port in question. Note the details of that iSCSI storage port object.
 - ii. Remove the iSCSI storage port object that represents those connections by using the **rmiscsistorageport** command.

- iii. Re-create the iSCSI storage port object with the same details as before, but with the new IP target port IP address by using the **detectiscsistorageportcandidate** and **addiscsistorageport** commands.
- iv. Check that the new connections have good connectivity to the target port by using the **lsiscsistorageport** command. The status field for the corresponding iSCSI storage port object should contain full.

If you discover during step iv that any host does not have good connectivity to the target port, diagnose and fix the problem before you continue.

For full instructions about using the CLI commands in this step, see the IBM Knowledge Center [IBM Knowledge Center](#).

3. When you have reconfigured the IP addresses of all the target ports in question and reconfigured all the hosts that are connected to those ports, repeat the checks in step 1 on page 299 to ensure that the procedure has restored full and dual-redundant connectivity.

15.1.3 Enabling or disabling iSCSI on IP ports

In certain circumstances, you might want to enable or disable iSCSI capability on IP ports, for example, to dedicate a port to IP replication or to use a port that was previously dedicated to IP replication for I/O. To enable or disable iSCSI host attachment or iSCSI external virtualization on a port, use the **cfgportip** command. For the full documentation of this command, see [IBM Knowledge Center](#).

To use a port for iSCSI host attachment, in addition to enabling iSCSI host attachment on that port, you must also reconfigure your hosts to establish iSCSI sessions that use that port as the target port. Before you remove host attachment capabilities from a port, you should first ensure that any hosts that are connected to that port still have a dual-redundant connection after it is reconfigured, and then reconfigure these hosts to stop using that port before reconfiguring it. This process depends on the type of hosts in use.

To use a port for iSCSI external virtualization, in addition to enabling iSCSI external virtualization on it, you must also create iSCSI storage port objects to use that port to connect to storage controllers. Typically, you do not need to reconfigure the storage controllers to do this task. If you want to remove external virtualization capabilities from a port, you should first ensure that the system still has a dual-redundant connection to any storage controllers that are connected to that port after it is reconfigured. Typically, you do not need to reconfigure those storage controllers.

15.2 Adding or removing nodes or I/O groups

MDisks that are visible through iSCSI connections to storage controllers can be connected in three different ways, as described in 15.2.1, “Adding nodes” on page 300. Adding nodes to a SAN Volume Controller initiator cluster can have different impacts on MDisks that are connected differently.

15.2.1 Adding nodes

This section describes how to add nodes.

MDisks that are connected through all nodes in the SAN Volume Controller cluster

For the MDisks that are connected through cluster-wide connectivity, adding a node causes the new iSCSI sessions to be automatically established from the newly added nodes to the configured LUNs or target iSCSI ports on the storage controller. You must ensure that the new nodes are correctly configured with the correct IP addresses and flags such that the initiator iSCSI ports on those nodes can access the configured LUNs and iSCSI targets on the storage controllers. For key considerations while you are configuring a new node, see “Key considerations for adding a node” on page 301.

MDisks that are connected through an I/O group of the SAN Volume Controller cluster

For the MDisks that are connected through I/O group-wide connectivity, adding a node does not have any impact. No action is necessary from users in such cases because connectivity to each iSCSI target port (or LUN) on the storage controller must be available from two nodes of a single SAN Volume Controller initiator IO group for the corresponding MDisk to be in the *online* state.

MDisks that are connected through a site of the SAN Volume Controller cluster

For the MDisks that are connected through site-wide connectivity, adding a node has an impact only if some (or all) of the new nodes are added to that site. The new nodes that are part of that site automatically initiate an iSCSI session with the storage controllers, if connectivity is cluster-wide. You must ensure that the new nodes are correctly configured with the correct IP addresses and flags such that the initiator iSCSI ports on those nodes can access the existing LUNs and iSCSI targets on the existing storage controllers. For more information, see “Key considerations for adding a node” on page 301.

Key considerations for adding a node

Consider the following things when you add a node:

1. Modify the LUN mapping policy on the storage controller.

The LUN mapping policy must be enhanced such that all the new SAN Volume Controller nodes can access the existing LUNs and storage controllers. If the storage controller is IBM Storwize or XIV, the IQNs of the new nodes must be added to the host object. If the storage controller is Dell, the IP addresses, IQNs, or user name and CHAP secret (depending upon the type of access policy that is used) of the new nodes must be added to the access list of the existing LUNs.

2. Configure ports on the SAN Volume Controller initiator.

All the new nodes must be configured such that they can access all existing iSCSI storage controllers that expose MDisks. This requirement means that all iSCSI initiator ports on the new nodes must be configured with IP addresses that are accessible to the storage controller ports. In addition, the storage flag must be set to yes for those ports that are used to access the storage controllers. For more information, see 15.2.3, “Adding ports to the SAN Volume Controller initiator” on page 302.

3. Check the connectivity status on the SAN Volume Controller initiator.

If steps 1 and 2 are done correctly, adding a node causes an increase in the number of iSCSI sessions to the storage controller because node addition automatically triggers an iSCSI session from the new nodes. However, it can take up to 1 minute to establish a session with storage controller after the IP addresses are configured on the new nodes.

Check the connectivity status by using the `lsiscsistorageport` CLI command. If the status for a target IQN shows full, it means that the new nodes can access the storage controller. If the status is partial, wait for 1 minute for the session to become established. If the status continues to be partial, then new nodes cannot establish an iSCSI session with the storage controller, which means that step to debug the issue. For more information, see 11.2.7, “Viewing the discovery results” on page 222.

4. Troubleshoot degraded MDisks.

If any MDisk goes to a degraded state while you are adding a node, it means that either an iSCSI session is not established to the storage controller from the new nodes or LUN mapping is not done on the storage controller. You can check the connectivity status by doing step 3 on page 301. If the field status shows FULL but the MDisk state is degraded, LUN mapping is not configured for the new nodes correctly. For more information, see the corresponding storage controller’s documentation.

15.2.2 Removing nodes from a SAN Volume Controller initiator cluster

Removing I/O groups or nodes from the SAN Volume Controller initiator cluster does not have any impact on MDisks while there is at least one path to the storage controller. The MDisks continue to be in the same state. No action is required from the user. However, if removing nodes removes all the iSCSI sessions to a storage controller, then the MDisk from that storage controller goes offline. Ensure that the last path to an MDisk is available before you remove a node.

15.2.3 Adding ports to the SAN Volume Controller initiator

If a user wants to add iSCSI ports to the SAN Volume Controller, new ports must be configured with the storage flag set to true.

When an IP is already configured on the ports and you want to enable storage functions on that port by using the GUI, connect to the SAN Volume Controller initiator GUI and click **Settings** → **Network**. Go to the Ethernet Ports tab, right-click the required IP, and select **Modify Storage Ports**, as shown in Figure 15-1.

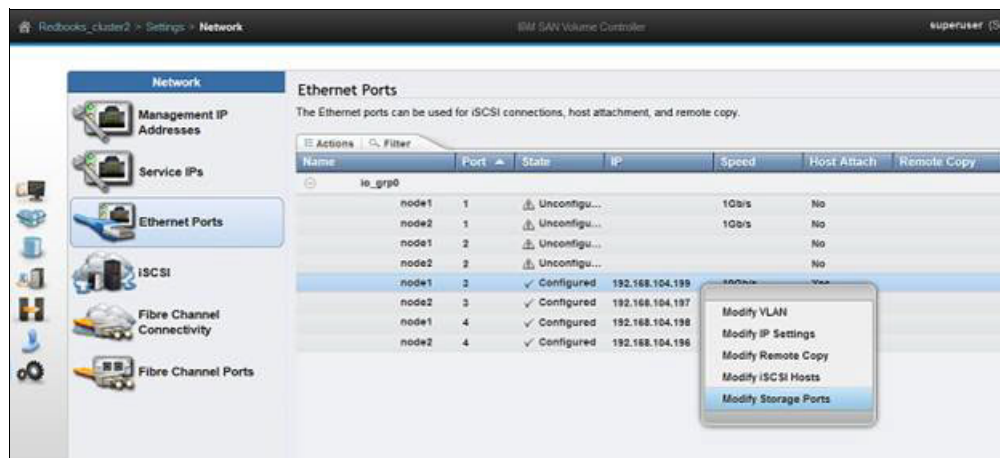


Figure 15-1 Modifying the storage attribute of an IP

Select **Enabled** from the drop-down menu for the corresponding IP type and click **Modify**, as shown in Figure 15-2.

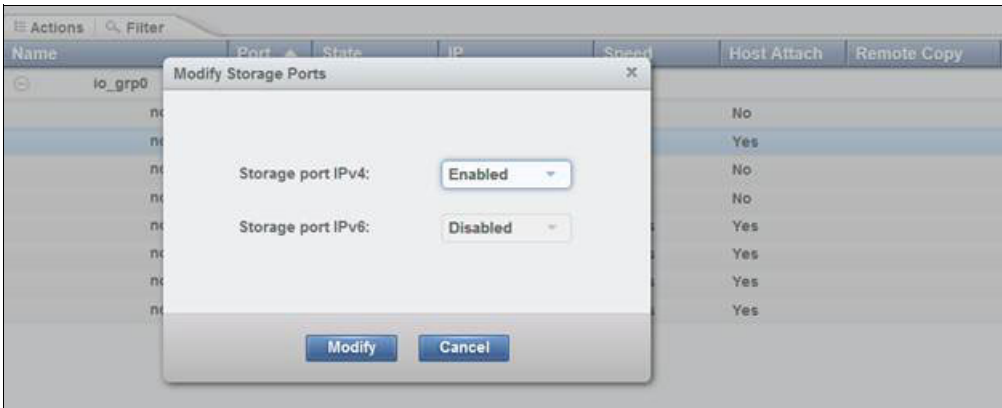


Figure 15-2 Enabling the storage attribute of an IP

Alternatively, you can use the **cfgportip** CLI command:

```
svctask cfgportip -node <node id> -storage yes <port id>
svctask cfgportip -node <node id> -storage_6 yes <port id>
```

When iSCSI ports are enabled for storage use, you can use them for connecting to the iSCSI storage controller.

If a new hardware adapter is being added, follow the instructions that are given in [IBM Knowledge Center](#).

15.2.4 Removing ports

iSCSI ports can be removed only if the ports are not actively used for iSCSI virtualization sessions. You can check whether a port is being using in iSCSI virtualization or not by using the **lsiscsistorageport** command, as shown in Figure 15-3.

```
[11:21:45] node1008:~ # lsiscsistorageport
```

id	src_port_id	target_ipv4	target_ip6	target_iscsiname	controller_id	io_group_list	status
0	3	192.168.104.3		iqn.2001-05.com.equallogic:4-42a846-91b677f31-2788f59261458340-redbook9	2	01-i-i-	none
1	3	192.168.100.121		iqn.2005-10.com.xivstorage:041529	3	01-i-i-	none
2	3	192.168.104.190		iqn.1986-03.com.ibm:2145.redbooksbackendscluster.node1	4	11-i-i-	full
3	4	192.168.104.191		iqn.1986-03.com.ibm:2145.redbooksbackendscluster.node1	4	11-i-i-	full
4	3	192.168.104.192		iqn.1986-03.com.ibm:2145.redbooksbackendscluster.node2	4	11-i-i-	full
5	4	192.168.104.193		iqn.1986-03.com.ibm:2145.redbooksbackendscluster.node2	4	11-i-i-	full

Figure 15-3 Checking whether a port is being used in iSCSI virtualization

The second column, **src_port_id**, shows the port ID. In this example, ports 3 and 4 are being used for iSCSI virtualization, so they cannot be removed. All other ports can be removed.

If any port is already in use for connecting to iSCSI-connected MDisks, the storage flag for that port cannot be turned off, and the IP address for that port cannot be removed. Before removing and disabling a hardware adapter, you must ensure that there are no active sessions from those ports to the iSCSI storage controller.

15.3 Changing the system name or node name

This section explains how to change the system name or the node name.

15.3.1 Changing the system name or node name of the initiator (SAN Volume Controller system)

Changing the initiator's system name or node name causes the initiator's IQN or IQNs to change. For example, consider a scenario where you want the old system's name is "redbookscluster1" and you want to change it to "redbookscluster2".

Each node's present IQNs are `iqn.1986-03.com.ibm:2145.redbookscluster1.nodeX`, where X is the node index. After you change the system name, new IQNs for each node will be `iqn.1986-03.com.ibm:2145.redbookscluster2.nodeX`, where X is the node index. When the initiator's IQN changes, SAN Volume Controller does not log out all the existing sessions that are established with the old IQN, so all the logged in session remain logged in and I/O continues until the iSCSI session is reestablished. After the session is reestablished, all the LUNs that were visible through sessions that were established with old IQNs are no longer accessible.

Therefore, before changing the system name/node name, you must perform some administrative steps to ensure that the LUNs remain accessible after sessions are reestablished with the new IQNs and everything continues to work correctly. These administrative steps depend on which iSCSI controllers to which the SAN Volume Controller is connected.

Steps for changing the initiator's system or node name when it is connected to SAN Volume Controller or IBM Storwize controllers

If you are using SAN Volume Controller or IBM Storwize controllers, complete the following steps to change the initiator's name:

1. On the SAN Volume Controller or IBM Storwize controller, add the IQN or IQNs to the host object so that all existing MDisk can be accessed through these IQNs. Complete the following steps:
 - a. Connect to the SAN Volume Controller GUI.
 - b. Select **Hosts** → **Hosts** and select the host object that represents the SAN Volume Controller initiators.

- c. Right-click the host object and click **Properties**, as shown in Figure 15-4.

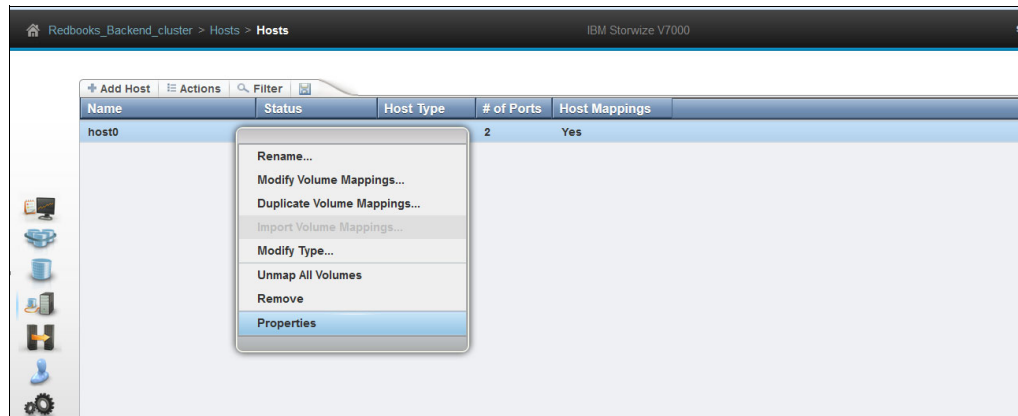


Figure 15-4 Changing the properties of the host object

A window opens.

- d. Go to the Port Definitions tab, as shown in Figure 15-5.

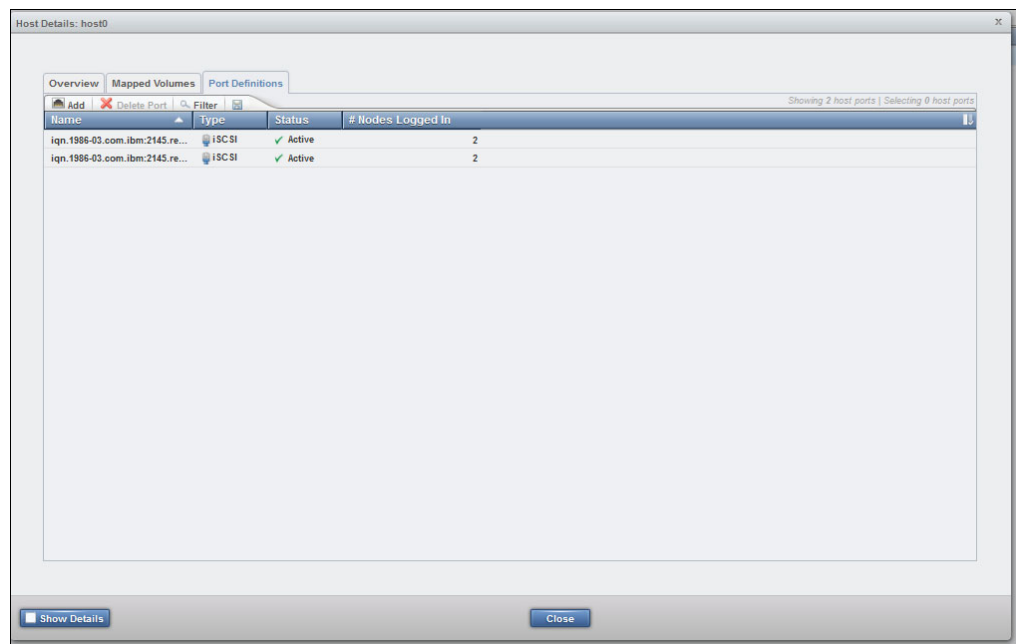


Figure 15-5 IQN list of a host object

- e. Click **Add** → **iSCSI Port**. Provide the new IQN of the iSCSI initiator, and then click **Add Port to List**. Similarly, add all the IQNs to the port list. After adding all the IQNs, click **Add ports to Host**, as shown in Figure 15-6.

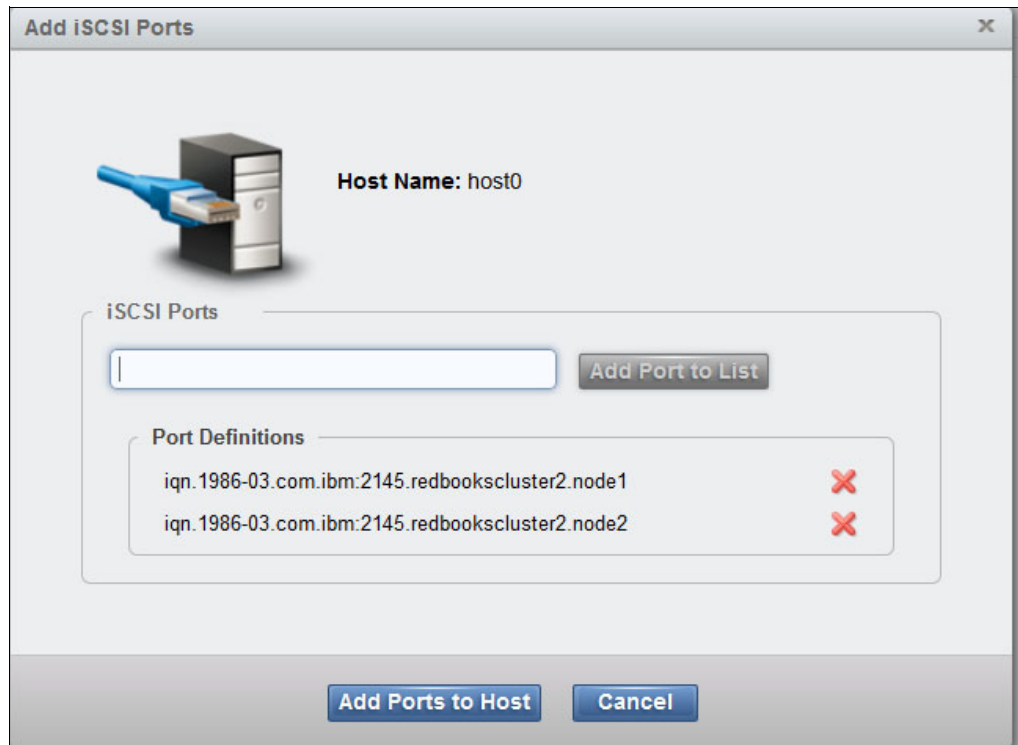


Figure 15-6 Adding the IQNs to the host object

Now, the host object has both the old IQNs and the new IQNs, as shown in Figure 15-7.

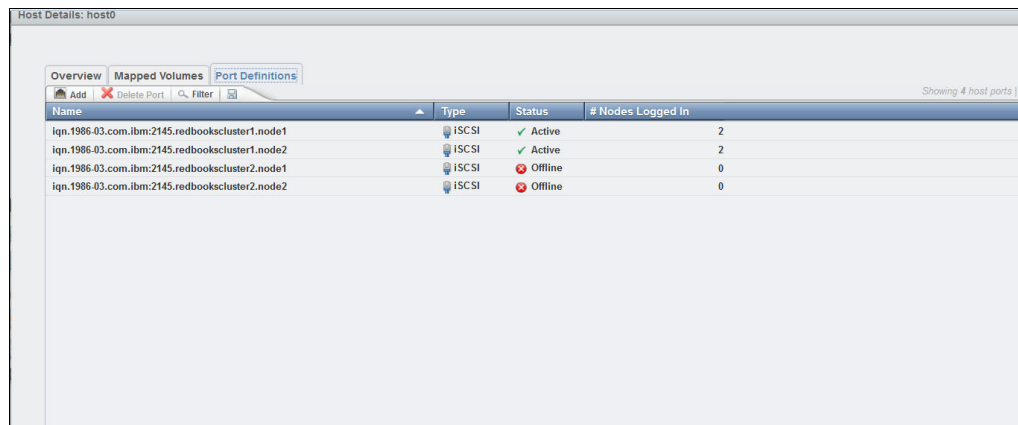


Figure 15-7 IQN list of the host object

Attention: Do not remove the old IQNs yet because there are active iSCSI sessions from those IQNs. These IQNs can be removed after you change the system or node name on the initiator.

Alternatively, you can add these new IQNs by using the **addhostport** command:

```
addhostport -iscsiname iqn.1986-03.com.ibm:2145.redbookscluster2.node1 0
addhostport -iscsiname iqn.1986-03.com.ibm:2145.redbookscluster2.node2 0
```

After this step, the new node IQNs can access all the MDisks that the old IQNs were accessing.

Note: The new IQNs that must be added are not set on the initiator yet. They must be deduced from the IQN naming format. If the node name is changed, the last part of the IQN of that initiator node is replaced by the new node name. If the system name is changed, the portion of the IQN between the last two dots is replaced by the new system name.

2. On the SAN Volume Controller or IBM Storwize initiator, change the system name. To do this task, complete the following steps:
 - a. Connect to the SAN Volume Controller initiator system and click **Monitoring** → **System**.
 - b. Click **Actions**.
 - c. Click **Rename system**, as shown in Figure 15-8.

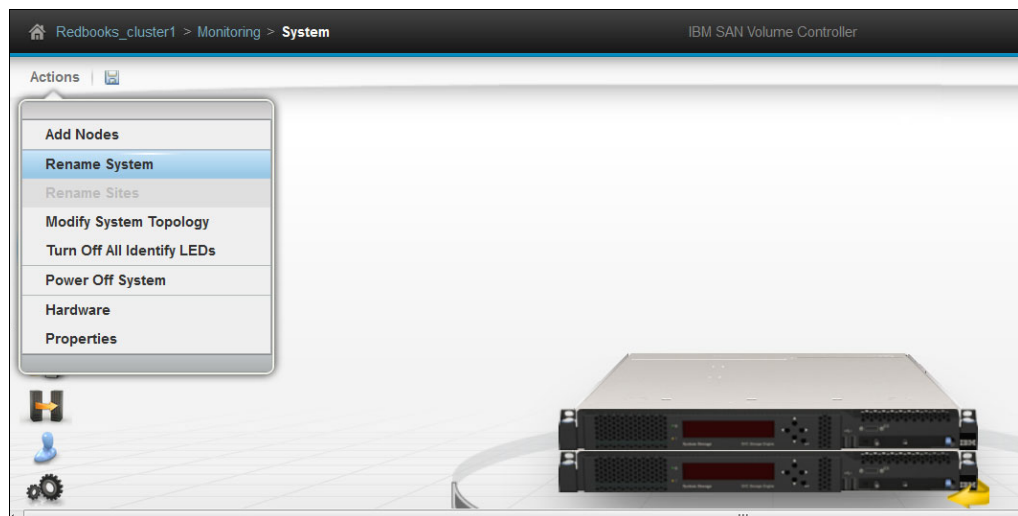


Figure 15-8 Changing the system name

- d. Enter a name for the system in the window that opens and click **Rename**, as shown in Figure 15-9.

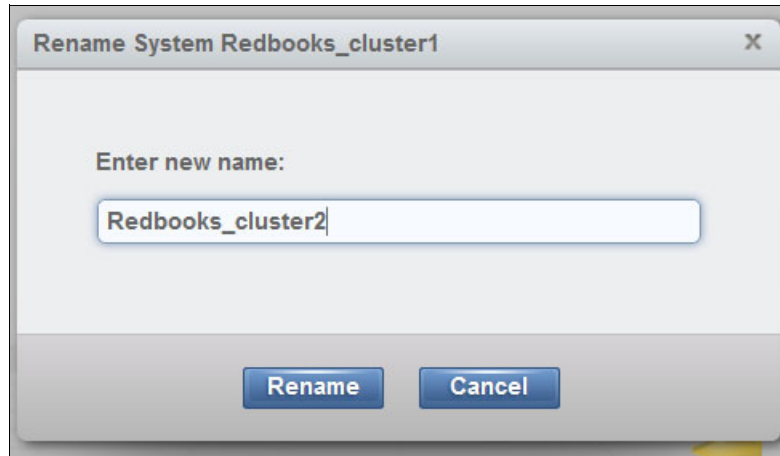


Figure 15-9 Renaming the system

3. On the SAN Volume Controller or IBM Storwize initiator, after you change the system name, remove the iSCSI sessions from the SAN Volume Controller or IBM Storwize target controllers and add them back again one by one. That is, remove sessions from one port, add them back, remove sessions from the second port, add them back, and so on. It is important to do this task one port at a time so that access to a LUN is not lost.

Note: For a cluster-wide or I/O Group-wide discovery, a warning appears in the SAN Volume Controller or IBM Storwize initiator's GUI window. This warning is automatically corrected when the session is added back.

Figure 15-10 shows the controller connectivity list. Ports 3 and 4 of both nodes are connected to both the controllers.

```
[11:21:45] node1CG8:~ # lsiscsistorageport
```

id	src_port_id	target_ipv4	target_ipv6	target_iscsiname	controller_id	iogroup_list	status
0	3	192.168.104.3		iqn.2001-05.com.equallogic:4-42a846-91b677f31-2788f59261458340-redbook9	2	0:-:-:-	none
1	3	192.168.100.121		iqn.2005-10.com.xivstorage:041529	3	0:-:-:-	none
2	3	192.168.104.190		iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1	4	1:-:-:-	full
3	4	192.168.104.191		iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1	4	1:-:-:-	full
4	3	192.168.104.192		iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node2	4	1:-:-:-	full
5	4	192.168.104.193		iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node2	4	1:-:-:-	full

Figure 15-10 Controller connectivity list

Complete the following steps:

- a. Remove sessions from port 3 to SAN Volume Controller or IBM Storwize controller iqn.1986-03.com.ibm:2145.redbooksbackendcluster2.node1 and add it back:

```
rmiscsistorageport 2
svctask detectiscsistorageportcandidate -srcportid 3 -targetip
192.168.104.190
addiscsistorageport 0
```

- b. Repeat the same task for sessions from port4:

```
rmiscsistorageport 3
svctask detectiscsistorageportcandidate -srcportid 4 -targetip
192.168.104.191
addiscsistorageport 0
```

- c. Repeat it for sessions to controller
iqn.1986-03.com.ibm:2145.redbooksbackendcluster2.node2:

rmiscsistorageport 4
svctask detectiscsistorageportcandidate -srcportid 3 -targetip
192.168.104.192
addiscsistorageport 0

rmiscsistorageport 5
svctask detectiscsistorageportcandidate -srcportid 4 -targetip
192.168.104.193
addiscsistorageport 0

4. On the SAN Volume Controller or IBM Storwize Controller, remove the old iSCSI IQNs from the host object:
 - a. Connect to the controller GUI and click **Hosts** → **Hosts**.
 - b. Select the host object and click **Properties**.
 - c. Go to the Port Definitions tab. It displays all the IQNs. The new IQNs are active and old IQNs are offline, as shown in Figure 15-11.

Host Details: host0

Overview Mapped Volumes Port Definitions			
Add Delete Port Filter			
Name	Type	Status	# Nodes Logged In
iqn.1986-03.com.ibm:2145.redbookscluster1.node1	iSCSI	Offline	0
iqn.1986-03.com.ibm:2145.redbookscluster1.node2	iSCSI	Offline	0
iqn.1986-03.com.ibm:2145.redbookscluster2.node1	iSCSI	Active	1
iqn.1986-03.com.ibm:2145.redbookscluster2.node2	iSCSI	Active	1

Figure 15-11 IQN list of the host object

- d. Select the old IQNs and click **Delete Port**. A window opens, as shown in Figure 15-12. In the Verify the number of ports to delete field, enter the number of old IQNs that you want to delete.

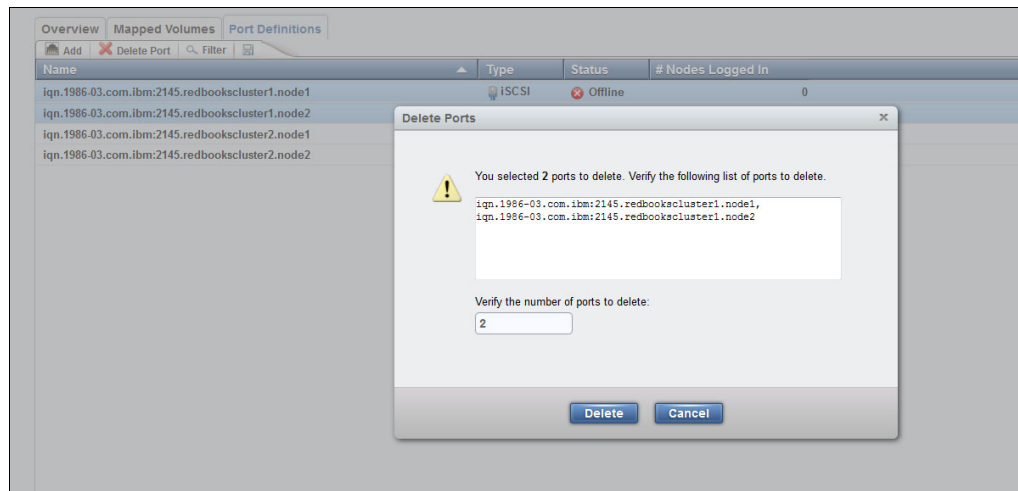


Figure 15-12 Deleting old IQNs from the host object list

When this task is complete, the host object contains only the new IQNs with an active status, as shown in Figure 15-13.

Overview Mapped Volumes Port Definitions			
Add Delete Port Filter			
Name	Type	Status	# Nodes Logged In
iqn.1986-03.com.ibm:2145.redbookscluster2.node1	ISCSI	Active	1
iqn.1986-03.com.ibm:2145.redbookscluster2.node2	ISCSI	Active	1

Figure 15-13 IQN list of the host object

You can also use the **rmhostport** command to do this task:

```
rmhostport -iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node1 0
rmhostport -iscsiname iqn.1986-03.com.ibm:2145.redbookscluster1.node2 0
```

Changing the initiator's system or node name when it is connected to Dell controllers

Each LUN on the Dell EqualLogic controller has an independent IQN and each LUN can have different access methods. Here are the access methods:

- IP list access** Provide a list of initiator IPs that can access a LUN.
- IQN list access** Provide a list of initiator IQNs that can access a LUN.
- CHAP access** All initiators that authenticate themselves with a specified user name and CHAP secret combination can access the LUN.
- Mixed access** Mix of IP and IQN access, CHAP access, or both.

If accessibility of LUNs is configured by initiator IP or CHAP access, nothing must be done. Any change in the initiator's IQN has no impact when the IP or CHAP on the initiator is unchanged because when the iSCSI session is reestablished, the initiator retries the login with the new IQN but with the same IP and CHAP secret, and its access to the LUNs is enabled.

If accessibility of LUNs is configured by using the initiator IQN, you must add the IQNs of the initiators to the access list before you change the initiator system name.

When this step is complete, follow the same steps to change the system name as specified for IBM Storwize, and then reestablish iSCSI sessions with the Dell controller one at a time. Now, you can remove the old IQNs from the LUNs access list on the Dell controller.

Changing the initiator's system or node name when connected to an XIV controller

The process to change the SAN Volume Controller or IBM Storwize initiator's system or node name when connected to an XIV controller is similar to the process for when it is connected to the IBM Storwize controller. To add access to new IQNs for the LUNs, see [IBM Knowledge Center](#), especially Chapter 2, "Adding a port to a host."

Changing the target controller IQN

You can change the IQN of a target SAN Volume Controller or IBM Storwize controller without any downtime. However, changing the IQNs on a Dell EqualLogic controller without losing access to LUNs is not feasible. XIV does not allow its target iSCSI name (IQN) to be changed. [IBM Knowledge Center](#) explains how to define a target object in XIV 10.2.4 by using target_define. In the CLI, you can change only the local target name, not its iSCSI name (IQN).

If you must change the IQN of a target SAN Volume Controller or IBM Storwize controller, you must remove iSCSI sessions to the affected target ports and add them back one at a time. This task must be done for all target ports with new IQNs.

Figure 15-14 displays source ports 3 and 4 of all the initiator nodes that are connected to two target IBM Storwize controller nodes.

```
[11:17:55] node1CG8:~ # lsiscsistorageport
```

id	src_port_id	target_ipv4	target_ipv6	target_iscsiname	controller_id	iogroup_list	status	sit
0	3	192.168.104.3		iqn.2001-05.com.equallogic:4-42a846-91b677f31-2788f59261458340-redbook9	2	0:-:-:-	none	
1	3	192.168.100.121		iqn.2005-10.com.xivstorage:041529	3	0:-:-:-	none	
2	3	192.168.104.190		iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1	4	1:-:-:-	full	
3	4	192.168.104.191		iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node1	4	1:-:-:-	full	
4	3	192.168.104.192		iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node2	4	1:-:-:-	full	
5	4	192.168.104.193		iqn.1986-03.com.ibm:2145.redbooksbackendcluster.node2	4	1:-:-:-	full	

Figure 15-14 Source ports 3 and 4 connected to IBM Storwize server redbookbackendcluster

Complete the following steps:

1. Remove the sessions from port 3 to controller node1 and add it back:

```
rmiscsistorageport 2
svctask detectiscsistorageportcandidate -srcportid 3 -targetip 192.168.104.190
-chapsecret secret1
addiscsistorageport -chapsecret secret1 0
```

2. Repeat the process for port 4, and then for controller node 2:

```
rmiscsistorageport 3
svctask detectiscsistorageportcandidate -srcportid 4 -targetip 192.168.104.191
-chapsecret secret1
```

```

addiscsistorageport -chapsecret secret1 0
rmiscsistorageport 4
svctask detectiscsistorageportcandidate -srcportid 3 -targetip 192.168.104.192
-chapsecret secret1
addiscsistorageport -chapsecret secret1 0
rmiscsistorageport 5
svctask detectiscsistorageportcandidate -srcportid 3 -targetip 192.168.104.193
-chapsecret secret1
addiscsistorageport -chapsecret secret1 0

```

Figure 15-15 shows the results of these steps.

```
[11:17:55] node1CG8:~ # lsiscsistorageport
```

id	src_port_id	target_ipv4	target_ipv6	target_iscsiname	controller_id	iogroup_list	status	si
0	3	192.168.104.3		iqn.2001-05.com.equallogic:4-42a846-91b677f31-2788f59261458340-redbook9	2	0:-:-:-	none	
1	3	192.168.100.121		iqn.2005-10.com.xivstorage:041529	3	0:-:-:-	none	
2	3	192.168.104.190		iqn.1986-03.com.ibm:2145.redbooksbackendcluster2.node1	4	1:-:-:-	full	
3	4	192.168.104.191		iqn.1986-03.com.ibm:2145.redbooksbackendcluster2.node1	4	1:-:-:-	full	
4	3	192.168.104.192		iqn.1986-03.com.ibm:2145.redbooksbackendcluster2.node2	4	1:-:-:-	full	
5	4	192.168.104.193		iqn.1986-03.com.ibm:2145.redbooksbackendcluster2.node2	4	1:-:-:-	full	

Figure 15-15 Source ports 3 and 4 connected to Storwize server redbookbackendcluster2

15.4 Changing the CHAP configuration

In certain circumstances, you can change the CHAP configuration for a SAN Volume Controller or IBM Storwize system's iSCSI sessions without a loss of access to data. This configuration change can be introducing CHAP authentication, removing CHAP authentication, or changing the CHAP secret for a connection that already uses CHAP authentication.

A CHAP configuration applies to a connection between an initiator system and a target system, so to change the CHAP configuration you must make configuration changes to both the target system and initiator system. Therefore, whether the CHAP configuration of a SAN Volume Controller or IBM Storwize system's iSCSI connections can be changed without disruption depends on the properties of the host or storage controller to which it is connected.

Section 15.4.1, "General considerations" on page 312 describes the general principles of changing the CHAP configuration without reference to any particular host or storage controller. Section 15.4.2, "Instructions for a SAN Volume Controller or IBM Storwize initiator system with an IBM Storwize target system" on page 313 gives specific instructions for changing the CHAP configuration for the connection between a SAN Volume Controller or IBM Storwize initiator system and an IBM Storwize target system; this is a case in which the configuration can be changed without loss of access to data.

15.4.1 General considerations

The process for changing the CHAP configuration of a SAN Volume Controller or IBM Storwize system's iSCSI sessions is different for external virtualization and host-attached sessions. The considerations that are involved in making the change nondisruptively are also different.

External virtualization iSCSI sessions

SAN Volume Controller and IBM Storwize systems support only one-way (target authenticates initiator) CHAP authentication for iSCSI connections for external virtualization.

The CHAP configuration that is used for the connection with a back-end storage controller is set when the corresponding iSCSI storage port object is created.

After an iSCSI storage port object is created, its CHAP configuration cannot be changed. So, to reconfigure CHAP, you must delete the object and re-create it. This is a disruptive process because it involves dropping the iSCSI sessions.

However, if the system has dual-redundant connections to the storage controller, it might be possible to change the CHAP configuration without loss of access to data. By reconfiguring the iSCSI storage port objects one at a time, the system can maintain access to data by using one of the objects while the other is being re-created. This situation works only if the storage controller either allows its CHAP configuration to be changed without dropping its iSCSI sessions or allows the CHAP configuration of those of its iSCSI sessions corresponding to a single iSCSI storage port object on the initiator system without disrupting the others. With an IBM Storwize system as the target storage controller, the CHAP configuration can be changed nondisruptively because the Storwize target system does not drop its host-attached iSCSI session when those connections' CHAP configurations are changed.

Host-attached iSCSI sessions

Changing the CHAP configuration on a SAN Volume Controller or IBM Storwize target system does not disrupt its host-attached iSCSI sessions, even if the host's CHAP configuration is not changed. This is because CHAP authentication occurs only when an iSCSI session is established and a SAN Volume Controller or IBM Storwize system does not drop its host-attached iSCSI sessions when their CHAP configurations are changed. If the host's CHAP configuration does not match the SAN Volume Controller or IBM Storwize system configuration, the host cannot re-establish iSCSI sessions if they drop, for example, because of network disruption.

Whether it is possible to change the host's CHAP configuration without disruption depends on the properties of the host. When the host is another SAN Volume Controller or IBM Storwize system with dual-redundant connections to the target IBM Storwize system, its CHAP configuration can be changed without loss of access to data by reconfiguring one iSCSI storage port at a time.

15.4.2 Instructions for a SAN Volume Controller or IBM Storwize initiator system with an IBM Storwize target system

It is possible to change the CHAP configuration between a SAN Volume Controller or IBM Storwize initiator system and an IBM Storwize target system without loss of access to data if each node of the initiator system has two redundant connections to the target system. This is possible because SAN Volume Controller or IBM Storwize systems do not drop their iSCSI sessions with hosts when the CHAP configuration for the connection with those hosts is changed. So, where there are dual-redundant connections between the target system and the initiator system, the initiator system's CHAP settings can be changed one connection at a time while always retaining access to data by using the other connections.

This task can be completed only with the CLI. The following instructions describe how to change the CHAP settings in such a setup without losing access to data.

Attention:

- ▶ Failure to follow these instructions correctly can result in loss of access to data. Do not attempt to proceed past step 1 without resolving any connectivity problems that are uncovered in that step. Do not attempt to repeat step 3 for a new iSCSI storage port object until the previous iSCSI storage object is successfully reconfigured and has good connectivity.
- ▶ While you are following these instructions, the system has a single point of failure. While one of the iSCSI storage port objects is being reconfigured in step 3, each node of the initiator system has only one iSCSI session with the target system.
- ▶ During this procedure, the initiator cannot automatically reestablish iSCSI sessions if they drop. This situation increases the impact of network stability problems.

Complete the following steps:

1. Ensure that the initiator system has full and dual-redundant connectivity to the target system. Complete the following steps:
 - a. Check that the controller object that represents the target system is not degraded (the degraded field of the detailed **lscontroller** view for that controller should contain no).
 - b. Check that each iSCSI storage port object representing a set of connections to the target system has full connectivity (the status field of the **lsiscsistorageport** view should contain full).

If the initiator system does not have full and redundant connectivity to the target system, diagnose and fix the problem before continuing.
2. On the target system, change the one-way (the target authenticates the initiator) CHAP configuration as needed by using the **chhost** command. Section 5.2, “Configuring CHAP for an IBM Storwize storage system” on page 57 contains full instructions for configuring and authenticating CHAP for host attachment.
3. On the initiator system, re-create each iSCSI storage port object with the new CHAP configuration. To do this task, complete the following steps a - d for each iSCSI storage port object that represents a set of iSCSI sessions with the target system. Do not start re-creating an iSCSI storage port until the previous one is successfully re-created. For full instructions about using the CLI commands in this step, see [IBM Knowledge Center](#).
 - a. Note the settings for the iSCSI storage port object in question, which are shown in the **lsiscsistorageport** view.
 - b. Remove the iSCSI storage port object by using the **rmiscsistorageport** command.
 - c. Create a iSCSI storage port object by using the **detectiscsistorageportcandidate** and **addiscsistorageport** commands. Use the same settings as the old object that noted in step a, but use the new CHAP configuration set in step 2.
 - d. Check that the new iSCSI storage port object has the correct settings and full connectivity by using the **lsiscsistorageport** view.
4. Check that the controller object representing the target system is not in a degraded state by using the detailed **lscontroller** view.

15.5 Changing the number of LUNs, ports, and IQNs in an IBM Storwize system

This section describes how to add and remove LUNs.

15.5.1 Adding and removing LUNs exposed from IBM Storwize or XIV controllers

To add/remove LUNs that are exposed from IBM Storwize or XIV controllers, complete the following steps:

1. On the storage controller, to add/remove LUNs from pre-configured IBM Storwize or XIV controllers to SAN Volume Controller storage pools, map (for adding new LUNs) or unmap (for removing LUNs) those LUNs to or from the host object. In this case, the host object is one or more SAN Volume Controller nodes. SAN Volume Controller nodes are identified by their IQNs. To modify the LUN mappings, see IBM Knowledge Center for IBM Storwize [IBM Storwize](#) and [XIV](#).
2. On the SAN Volume Controller initiator, after the mapping is updated, refresh the LUN list on the SAN Volume Controller initiator side. To refresh the LUN list by using GUI, connect to the initiator GUI and click **Pools** → **External Storage**. Right-click the iSCSI controller (IBM Storwize or XIV controller) and click **Discover Storage**, as shown in Figure 15-16.



Figure 15-16 Refreshing the LUN list from a pre-configured controller

After new LUNs are available for use, they are displayed under that controller. For example, Figure 15-17 shows that a new LUN (mdisk8) is added for controller2. This LUN is configured on an IBM Storwize controller.

Name	State	Capacity	Mode
controller0	Online	IBM 1726-4xx FASTT	
mdisk0	Online		136.23 GiB Unmanaged
controller2	Online	IBM 2145 Serial Number: 2076	
mdisk4	Online		500.00 GiB Unmanaged
mdisk6	Online		500.00 GiB Unmanaged
mdisk7	Online		500.00 GiB Unmanaged
mdisk5	Online		500.00 GiB Unmanaged
mdisk8	Online		500.00 GiB Unmanaged
controller1	Degraded	IBM 1726-4xx FASTT	

Figure 15-17 The mdisk8 LUN is added to the LUNs list

Alternatively, you can use the **detectmdisk** command to refresh the LUN list. Figure 15-18 shows that mdisk8 was added for controller2.

```
[15:09:04] node1CG8:~ # lsmdisk
id name status mode mdisk_grp_id mdisk_grp_name capacity ctrl_LUN # controller_name
0 mdisk0 online unmanaged 136.2GB 0000000000000000 controller0
1 mdisk4 online unmanaged 500.0GB 0000000000000001 controller2
2 mdisk5 online unmanaged 500.0GB 0000000000000000 controller2
3 mdisk7 online unmanaged 500.0GB 0000000000000002 controller2
4 mdisk5 online unmanaged 500.0GB 0000000000000003 controller2
[15:09:54] node1CG8:~ # detectmdisk
[15:09:56] node1CG8:~ # lsmdisk
id name status mode mdisk_grp_id mdisk_grp_name capacity ctrl_LUN # controller_name
0 mdisk0 online unmanaged 136.2GB 0000000000000000 controller0
1 mdisk4 online unmanaged 500.0GB 0000000000000001 controller2
2 mdisk6 online unmanaged 500.0GB 0000000000000000 controller2
3 mdisk7 online unmanaged 500.0GB 0000000000000002 controller2
4 mdisk5 online unmanaged 500.0GB 0000000000000003 controller2
5 mdisk8 online unmanaged 500.0GB 0000000000000003 controller2
[15:10:00] node1CG8:~ #
```

Figure 15-18 Example detectmdisk CLI command output

15.5.2 Adding LUNs from a Dell EqualLogic controller

Each LUN on Dell EqualLogic controller has an independent IQN. To add a LUN, complete the following steps:

1. On the Dell controller, modify the access policy of the newly created LUNs so that the required nodes of SAN Volume Controller initiator can access that LUN.
2. On the SAN Volume Controller initiator, run the **detectiscsistorageportcandidate** command to discover the newly created LUNs. The IQNs of the newly created LUNs are displayed by the **1siscsistorageportcandidate** command.

3. On the SAN Volume Controller initiator, run the **addiscsistorageportcandidate** CLI command to establish an iSCSI session with those new IQNs. The **isiscsistorageport** command shows the status of the iSCSI session.
4. On the SAN Volume Controller initiator, newly added LUNs are displayed by the **lsmdisk** command.

15.5.3 Removing LUNs from a Dell EqualLogic controller

To remove LUNs from the Dell controller, complete the following steps:

1. On the SAN Volume Controller initiator, each LUN in the Dell EqualLogic controller has an independent IQN. To remove LUNs, remove the iSCSI session to those IQNs. The **lsiscsistorageport** command displays the list of iSCSI sessions; from that list, find the entries of the Dell EqualLogic controller LUNs by searching for the IQNs of those LUNs. Use the row indexes of those entries to remove the sessions by using the **rmiscsistorageport** command.
2. On the Dell controller, modify the access policy of the LUNs so that the SAN Volume Controller nodes cannot access (discover and log in to) those LUNs.



Troubleshooting iSCSi virtualization

This chapter is focused on troubleshooting problems with Internet Small Computer System (iSCSi) targets. For an introduction to troubleshooting and the applicable data collection procedures for the SAN Volume Controller and IBM Storwize products, see Chapter 10, “Troubleshooting” on page 175.

This chapter describes the following topics:

- ▶ 16.1, “Troubleshooting iSCSi target discovery” on page 320
- ▶ 16.2, “Troubleshooting a degraded or offline status” on page 321
- ▶ 16.3, “Performance issues” on page 322

16.1 Troubleshooting iSCSI target discovery

This section describes the basic troubleshooting methodology for iSCSI target discovery issues.

16.1.1 Problems with initial discovery

This section details the troubleshooting process for iSCSI target discovery.

A failed discovery for iSCSI storage indicates that using the **detectiscsistorageport** resulted in a failure to communicate with the specified IP address on the interfaces that are specified in the command. Therefore, the first item to investigate is why the Storwize initiator cannot access the target.

The **lsiscsistorageportcandidate** command displays the result of the last run **detectiscsistorageportcandidate** command. This information can be useful in troubleshooting discovery issues.

Assume that you have a two-I/O group cluster in a standard topology. Initiate a cluster-level discovery operation on port 1 for each node in this cluster to target IP address 192.168.70.121. After running this command, the **lsiscsistorageportcandidate** command output for `iogroup_list` is 1:0:-:-, as shown in Example 16-1.

Example 16-1 The `lsiscsistorageportcandidate` command output

```
IBM_2145:Redbooks_cluster1:superuser>lsiscsistorageportcandidate
```

id	src_port_id	target_ipv4	target_ipv6	target_iscsiname	iogroup_list	configured	status
0	1	192.168.70.121		iqn.2005-10.com.xivstorage:041529	1:0:-:-	no	partial

By viewing this output, you know that I/O group 0 successfully discovered the storage port on both nodes in that group. However, I/O group 1 failed to discover either one or both of the nodes in this I/O group. To ensure a successful discovery, you must discover why port 1 on the node or nodes in I/O group 1 cannot contact the IP address 192.168.70.121. To troubleshoot this type of issue, complete the following steps:

1. Ensure that the required interfaces are all online and are connected.
2. Validate the connectivity between the SAN Volume Controller or IBM Storwize system and the target storage controller. Connectivity can be validated by using the **ping** command.
3. Validate that the CHAP authentication parameters are correct.
4. Ensure that firewall port 3260 is enabled between the devices.
5. Validate that the target storage controller is supported for iSCSI virtualization.

16.1.2 Problems adding a storage port

If you run **detectiscsistorageportcandidate** and wait for an extended period, there might have been changes that were made to the back-end storage controller. If changes were made since the last time that you ran **detectiscsistorageportcandidate**, you have stale data that is no longer valid in the output of **lsiscsistorageportcandidate**. If you run **addiscsistorageport** against this stale data, then the specified storage port candidate is added to the persistent storage port table, and the SAN Volume Controller or IBM Storwize system attempts to connect to that storage port, and might fail to create an iSCSI session.

To rediscover targets immediately before attempting to add a storage port and to validate that the connections are as expected, run `lsiscsistorageport`.

16.2 Troubleshooting a degraded or offline status

This section focuses on troubleshooting post-configuration issues with target storage controllers, including degraded and offline conditions.

16.2.1 Restoring an offline MDisk or storage controller

A storage port is marked offline by the IBM Storwize cluster if the target controller fails to respond to a heartbeat that is sent by the SAN Volume Controller or IBM Storwize system within 5 seconds. This is typically caused by a connectivity issue between the back-end storage controller and the SAN Volume Controller or IBM Storwize system. To troubleshoot this situation, complete the following steps:

1. Validate that all the required ports are connected.
2. Validate that the SAN Volume Controller or IBM Storwize system can communicate with the target storage controller.
3. Validate that port 3260 is open between the SAN Volume Controller or IBM Storwize device and the target storage controller.
4. Validate that no changes to the target controller were made that affect its availability to be virtualized.

The SAN Volume Controller or IBM Storwize software automatically retries the paths every few seconds to recover the storage port after it is available again. However, you can also force a rediscovery with the `detectmdisk` command.

16.2.2 Restoring degraded MDisk or storage controllers

This section describes how to restore degraded MDisk or storage controllers.

Remote port that is excluded (error code 1230)

If the system detects an excessive number of I/O errors when accessing an MDisk from a particular storage port, the system might exclude this path and generate error code 1230, which indicates that there is an issue in the communication between the SAN Volume Controller or IBM Storwize system and the back-end storage controller. To troubleshoot this type of issue, complete the following steps:

1. Ensure that no changes were made to the back-end storage controller that affect the SAN Volume Controller or IBM Storwize system's ability to use the managed resources.
2. Check the health of the SAN Volume Controller or IBM Storwize interfaces. Because a port is excluded and not offline, all the ports are probably connected, so you must rely on interface statistics to assist you with your troubleshooting. To learn how to review the interface statistics on the SAN Volume Controller or IBM Storwize device, see 10.2.4, "Ethernet logs and statistics on IBM Storwize nodes" on page 193.
3. If the SAN Volume Controller or IBM Storwize device's interfaces do not appear to be the problem, then you should open an investigation into your networking devices and back-end storage controller.

After the communication issue is resolved, force a retry of the path or paths by either running the directed maintenance procedure (DMP) or by using the following command:

```
includemdisk <mdisk name/id>
```

For more information about this command, see [IBM Knowledge Center](#).

Single path failure of a redundant controller

If a path or storage port to a target controller goes offline but other paths or storage ports remain online, then the controller and associated MDisks are likely degraded. Resolve this condition by completing the following steps:

1. Validate that all the required ports are connected.
2. Validate that the SAN Volume Controller or IBM Storwize system can communicate with the target storage controller.
3. Validate that port 3260 is open between the SAN Volume Controller or IBM Storwize device and the target storage controller.
4. Validate that no changes to the target controller were made that affects its ability to be virtualized.

16.3 Performance issues

If the SAN Volume Controller or Storwize initiator system is reporting performance problems from the back-end storage system, follow a similar troubleshooting method for any other performance problem, except that in this case the SAN Volume Controller or IBM Storwize device is the host system. Section 10.4.7, “Problem determination: Checking for performance problems” on page 207 might be useful in determining the source of a performance issue. When the performance problem appears to be coming from a controller that is virtualized over iSCSI, complete the following steps:

1. Ensure that the MTU size is set correctly on all devices.
2. Ensure that the port speeds match on every device in the path and on the end points, including the storage controllers, switches, and routers.
3. Review the SAN Volume Controller or IBM Storwize system performance data to try and identify the bottleneck:
 - a. Use IBM Spectrum Control if possible to review data over long periods.
 - b. If IBM Spectrum Control is not available, review the data, as shown in 8.2.2, “Real-time performance monitoring with the GUI” on page 163.
4. Review the interface counters in the Ethernet trace files that are found in the snap file of the SAN Volume Controller or Storwize system, as shown in 10.2.5, “iSCSI logs on IBM Storwize nodes” on page 195.
5. Review the switch statistics to see whether the network is causing the delay.
6. Review the back-end storage controller to see whether this device is having performance problems. If the back-end storage controller is having performance problems, then this situation is noticed by the SAN Volume Controller or IBM Storwize device.

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 additional information about the topic in this document. Some publications that are referenced in this list might be available in softcopy only.

- ▶ *IBM Private, Public, and Hybrid Cloud Storage Solutions*, REDP-4873
- ▶ *IBM SAN Volume Controller 2145-DH8 Introduction and Implementation*, SG24-8229
- ▶ *Implementing the IBM Storwize V3500*, SG24-8125
- ▶ *Implementing the IBM Storwize V3700*, SG24-8107
- ▶ *Implementing the IBM Storwize V5000 Gen2 (including the Storwize V5010, V5020, and V5030)*, SG24-8162
- ▶ *Implementing the IBM Storwize V7000 Gen2*, SG24-8244
- ▶ *Implementing the IBM Storwize V7000 and IBM Spectrum Virtualize V7.8*, SG24-7938
- ▶ *Implementing the IBM Storwize V7000 Unified Disk System*, SG24-8010

You can search for, view, download, or order these documents and other Redbooks, Redpapers, web docs, draft and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- ▶ IBM Knowledge Center
<https://www.ibm.com/support/knowledgecenter>
- ▶ IBM Support
<https://www.ibm.com/support>
- ▶ IBM System Storage Interoperability Center
<https://www.ibm.com/systems/support/storage/ssic/interoperability.wss>
- ▶ IETF RFC 3720, MPLS Support of Differentiated Services, found at:
<http://www.ietf.org/rfc/rfc3720.txt>
- ▶ Securing Block Storage Protocols over IP
<http://www.ietf.org/rfc/rfc3723.txt>
- ▶ OpenStack
<https://www.openstack.org/>

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

iSCSI Implementation and Best Practices on IBM Storwize Storage Systems

SG24-8327-01

ISBN 0738442755



(0.5" spine)

0.475" <-> 0.873"

250 <-> 459 pages



SG24-8327-01

ISBN 0738442755

Printed in U.S.A.

Get connected

