

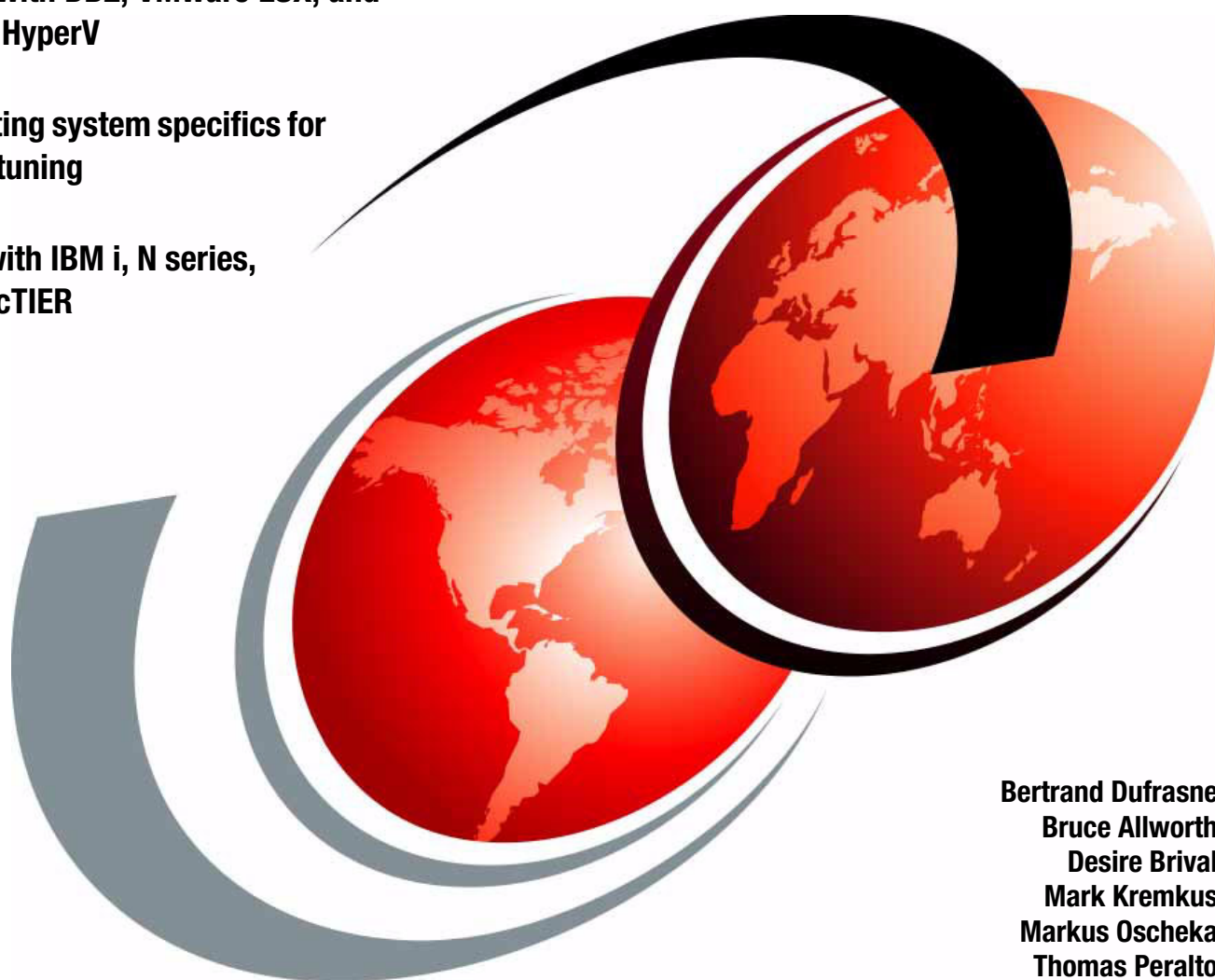
IBM XIV Storage System

Host Attachment and Interoperability

Integrate with DB2, VMware ESX, and Microsoft HyperV

Get operating system specifics for host side tuning

Use XIV with IBM i, N series, and ProtecTIER



Bertrand Dufrasne
Bruce Allworth
Desire Brival
Mark Kremkus
Markus Oscheka
Thomas Peralto

Redbooks



International Technical Support Organization

IBM XIV Storage System Host Attachment and Interoperability

March 2013

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

Third Edition (March 2013)

This edition applies to the IBM XIV Storage System (Machine types 2812-114 and 2810-114) with XIV system software Version 11.1.1.

© Copyright International Business Machines Corporation 2012, 2013. 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
The team who wrote this book	xi
Now you can become a published author, too!	xiii
Comments welcome	xiii
Stay connected to IBM Redbooks	xiv
Summary of changes	xv
March 2013, Third Edition	xv
Chapter 1. Host connectivity	1
1.1 Overview	2
1.1.1 Module, patch panel, and host connectivity	4
1.1.2 Host operating system support	9
1.1.3 Downloading the entire XIV support matrix by using the SSIC	9
1.1.4 Host Attachment Kits	10
1.1.5 Fibre Channel versus iSCSI access	12
1.2 Fibre Channel connectivity	13
1.2.1 Preparation steps	13
1.2.2 Fibre Channel configurations	15
1.2.3 Zoning	18
1.2.4 Identification of FC ports (initiator/target)	20
1.2.5 Boot from SAN on x86/x64 based architecture	22
1.3 iSCSI connectivity	27
1.3.1 Preparation steps	28
1.3.2 iSCSI configurations	28
1.3.3 Network configuration	30
1.3.4 IBM XIV Storage System iSCSI setup	30
1.3.5 Identifying iSCSI ports	33
1.3.6 iSCSI and CHAP authentication	34
1.3.7 iSCSI boot from XIV LUN	35
1.4 Logical configuration for host connectivity	36
1.4.1 Host configuration preparation	36
1.4.2 Assigning LUNs to a host by using the GUI	38
1.4.3 Assigning LUNs to a host by using the XCLI	42
1.5 Troubleshooting	44
Chapter 2. XIV and Windows host connectivity	45
2.1 Attaching a Microsoft Windows 2008 R2 host to XIV	46
2.1.1 Prerequisites	46
2.1.2 Windows host FC configuration	47
2.1.3 Windows host iSCSI configuration	52
2.1.4 Host Attachment Kit utilities	65
2.2 Attaching a Microsoft Windows 2008 R2 cluster to XIV	66
2.2.1 Prerequisites	67
2.2.2 Installing Cluster Services	68
2.2.3 Configuring the IBM Storage Enabler for Windows Failover Clustering	73

2.3	Attaching a Microsoft Hyper-V Server 2008 R2 to XIV	79
2.4	Microsoft System Center Virtual Machine Manager 2012 Storage Automation	80
2.4.1	The XIV Open API overview	80
2.4.2	System Center Virtual Machine Manager overview	82
Chapter 3. XIV and Linux host connectivity		85
3.1	IBM XIV Storage System and Linux support overview	86
3.1.1	Issues that distinguish Linux from other operating systems	86
3.1.2	Reference material	86
3.1.3	Recent storage-related improvements to Linux	89
3.2	Basic host attachment	90
3.2.1	Platform-specific remarks	90
3.2.2	Configuring for Fibre Channel attachment	93
3.2.3	Determining the WWPN of the installed HBAs	97
3.2.4	Attaching XIV volumes to an Intel x86 host using the Host Attachment Kit	98
3.2.5	Checking attached volumes	101
3.2.6	Setting up Device Mapper Multipathing	106
3.2.7	Special considerations for XIV attachment	114
3.3	Non-disruptive SCSI reconfiguration	115
3.3.1	Adding and removing XIV volumes dynamically	115
3.3.2	Adding and removing XIV volumes in Linux on System z	117
3.3.3	Adding new XIV host ports to Linux on System z	118
3.3.4	Resizing XIV volumes dynamically	119
3.3.5	Using snapshots and remote replication targets	120
3.4	Troubleshooting and monitoring	123
3.4.1	Linux Host Attachment Kit utilities	123
3.4.2	Multipath diagnosis	124
3.4.3	Other ways to check SCSI devices	127
3.4.4	Performance monitoring with iostat	128
3.4.5	Generic SCSI tools	128
3.5	Boot Linux from XIV volumes	129
3.5.1	The Linux boot process	129
3.5.2	Configuring the QLogic BIOS to boot from an XIV volume	130
3.5.3	OS loader considerations for other platforms	130
3.5.4	Installing SLES11 SP1 on an XIV volume	131
Chapter 4. XIV and AIX host connectivity		135
4.1	Attaching XIV to AIX hosts	136
4.1.1	Prerequisites	136
4.1.2	AIX host FC configuration	136
4.1.3	AIX host iSCSI configuration	148
4.1.4	Management volume LUN 0	156
4.1.5	Host Attachment Kit utilities	157
4.2	SAN boot in AIX	159
4.2.1	Creating a SAN boot disk by mirroring	160
4.2.2	Installation on external storage from bootable AIX CD-ROM	164
4.2.3	AIX SAN installation with NIM	166
Chapter 5. XIV and HP-UX host connectivity		169
5.1	Attaching XIV to an HP-UX host	170
5.2	HP-UX multi-pathing solutions	172
5.3	Veritas Volume Manager on HP-UX	174
5.3.1	Array Support Library for an IBM XIV Storage System	177
5.4	HP-UX SAN boot	178

5.4.1	Installing HP-UX on external storage	179
5.4.2	Creating a SAN boot disk by mirroring	182
Chapter 6.	XIV and Solaris host connectivity	183
6.1	Attaching a Solaris host to XIV	184
6.2	Solaris host configuration for Fibre Channel	184
6.2.1	Obtaining WWPN for XIV volume mapping.	184
6.2.2	Installing the Host Attachment Kit	184
6.2.3	Configuring the host	185
6.3	Solaris host configuration for iSCSI	188
6.4	Solaris Host Attachment Kit utilities	190
6.5	Creating partitions and file systems with UFS.	191
Chapter 7.	XIV and Symantec Storage Foundation	197
7.1	Introduction	198
7.2	Prerequisites	198
7.2.1	Checking ASL availability and installation.	198
7.2.2	Installing the XIV Host Attachment Kit	199
7.2.3	Configuring the host	200
7.3	Placing XIV LUNs under VxVM control	202
7.4	Configuring multipathing with DMP	206
7.5	Working with snapshots	207
Chapter 8.	IBM i and AIX clients connecting to XIV through VIOS.	209
8.1	Introduction to IBM PowerVM	210
8.1.1	IBM PowerVM overview	210
8.1.2	Virtual I/O Server	211
8.1.3	Node Port ID Virtualization	212
8.2	Planning for VIOS and IBM i	213
8.2.1	Requirements	213
8.2.2	Supported SAN switches	214
8.2.3	Physical Fibre Channel adapters and virtual SCSI adapters	214
8.2.4	Queue depth in the IBM i operating system and Virtual I/O Server	214
8.2.5	Multipath with two Virtual I/O Servers.	215
8.2.6	General guidelines	215
8.3	Connecting an PowerVM IBM i client to XIV	216
8.3.1	Creating the Virtual I/O Server and IBM i partitions	217
8.3.2	Installing the Virtual I/O Server	220
8.3.3	IBM i multipath capability with two Virtual I/O Servers	220
8.3.4	Virtual SCSI adapters in multipath with two Virtual I/O Servers	221
8.4	Mapping XIV volumes in the Virtual I/O Server.	223
8.5	Matching XIV volume to IBM i disk unit.	225
8.6	Performance considerations for IBM i with XIV.	228
8.6.1	Testing environment	229
8.6.2	Testing workload.	232
8.6.3	Test with 154-GB volumes on XIV generation 2	233
8.6.4	Test with 1-TB volumes on XIV generation 2	237
8.6.5	Test with 154-GB volumes on XIV Gen 3	241
8.6.6	Test with 1-TB volumes on XIV Gen 3	245
8.6.7	Test with doubled workload on XIV Gen 3	249
8.6.8	Testing conclusions	254
Chapter 9.	XIV Storage System and VMware connectivity	257
9.1	Integration concepts and implementation guidelines	258

9.1.1 vSphere storage architectural overview	258
9.1.2 XIV and VMware general connectivity guidelines	259
Chapter 10. XIV and N series Gateway connectivity	263
10.1 Overview of N series Gateway	264
10.2 Attaching N series Gateway to XIV	264
10.2.1 Supported versions	265
10.2.2 Other considerations	266
10.3 Cabling	266
10.3.1 Cabling example for single N series Gateway with XIV	266
10.3.2 Cabling example for N series Gateway cluster with XIV	267
10.4 Zoning	267
10.4.1 Zoning example for single N series Gateway attachment to XIV	268
10.4.2 Zoning example for clustered N series Gateway attachment to XIV	268
10.5 Configuring the XIV for N series Gateway	268
10.5.1 Creating a Storage Pool in XIV	269
10.5.2 Creating the root volume in XIV	270
10.5.3 Creating the N series Gateway host in XIV	271
10.5.4 Adding the WWPN to the host in XIV	271
10.5.5 Mapping the root volume to the N series host in XIV GUI	274
10.6 Installing Data ONTAP	275
10.6.1 Assigning the root volume to N series Gateway	275
10.6.2 Installing Data ONTAP	276
10.6.3 Updating Data ONTAP	277
10.6.4 Adding data LUNs to N series Gateway	278
Chapter 11. ProtecTIER Deduplication Gateway connectivity	279
11.1 Overview	280
11.2 Preparing an XIV for ProtecTIER Deduplication Gateway	281
11.2.1 Supported versions and prerequisites	282
11.2.2 Fibre Channel switch cabling	282
11.2.3 Zoning configuration	283
11.2.4 Configuring XIV Storage System for ProtecTIER Deduplication Gateway	284
11.3 IBM SSR installs the ProtecTIER software	290
Chapter 12. XIV in database and SAP application environments	291
12.1 XIV volume layout for database applications	292
12.1.1 Common guidelines	292
12.1.2 Oracle database	293
12.1.3 Oracle ASM	293
12.1.4 IBM DB2	293
12.1.5 DB2 parallelism options for Linux, UNIX, and Windows	294
12.1.6 Microsoft SQL Server	295
12.2 Guidelines for SAP	297
12.2.1 Number of volumes	297
12.2.2 Separation of database logs and data files	297
12.2.3 Storage pools	298
12.3 Database Snapshot backup considerations	298
12.3.1 Snapshot backup processing for Oracle and DB2 databases	298
12.3.2 Snapshot restore	299
Related publications	301
IBM Redbooks	301
Other publications	301

Online resources	302
How to get Redbooks	302
Help from IBM	302

Notices

This information was developed for products and services offered in the U.S.A.

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, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: 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 states 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 supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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.

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 the names and addresses used by an actual business enterprise 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.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

AIX®	Micro-Partitioning®	Redbooks (logo)  ®
BladeCenter®	POWER®	Storwize®
DB2®	Power Architecture®	System i®
developerWorks®	Power Systems™	System Storage®
DS8000®	POWER6®	System z®
Dynamic Infrastructure®	POWER6+™	Tivoli®
FICON®	POWER7®	XIV®
FlashCopy®	POWER7 Systems™	z/VM®
HyperFactor®	PowerVM®	z10™
i5/OS™	ProtecTIER®	
IBM®	Redbooks®	

The following terms are trademarks of other companies:

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.

Ultrium, the LTO Logo and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

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

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

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

Preface

This IBM® Redbooks® publication provides information for attaching the IBM XIV® Storage System to various host operating system platforms, including IBM i.

The book provides information and references for combining the XIV Storage System with other storage platforms, host servers, or gateways, including IBM N Series, and IBM ProtecTIER®. It is intended for administrators and architects of enterprise storage systems.

The book also addresses using the XIV storage with databases and other storage-oriented application software that include:

- ▶ IBM DB2®
- ▶ VMware ESX
- ▶ Microsoft HyperV
- ▶ SAP

The goal is to give an overview of the versatility and compatibility of the XIV Storage System with various platforms and environments.

The information that is presented here is not meant as a replacement or substitute for the Host Attachment kit publications. It is meant as a complement and to provide readers with usage guidance and practical illustrations.

Host Attachment Kits can be downloaded from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

For more information, see *XIV Storage System in VMware Environments*, REDP-4965. For information about XIV deployment in an OpenStack environment, see *Using the IBM XIV Storage System in OpenStack Cloud Environments*, REDP-4971.

The team who wrote this book

This book was produced by a team of specialists from around the world working at the International Technical Support Organization, San Jose Center.

Bertrand Dufrasne is an IBM Certified Consulting I/T Specialist and Project Leader for IBM System Storage® disk products at the International Technical Support Organization, San Jose Center. He has worked at IBM in various I/T areas. He has authored many IBM Redbooks publications, and has also developed and taught technical workshops. Before joining the ITSO, he worked for IBM Global Services as an Application Architect. He holds a Master's degree in Electrical Engineering.

Bruce Allworth is a Senior IT Specialist working as a Client Technical Open Systems Storage Specialist in America's Sales and Distribution. He has over 20 years in IT and extensive experience with midrange storage products in solution design, solution management, advanced problem determination, and disaster recovery. He has worked closely with various IBM divisions to launch new products, create critical documentation including Technical and Delivery Assessment Checklists, and develop and deliver technical training for a wide range of audiences.

Desire Brival is a Certified Architect working with clients and sales teams to architect complex multi-brand solutions. He has deep knowledge of the IBM cross-brand initiatives such as Cloud Computing, Virtualization, IT Optimization, IBM Dynamic Infrastructure®, and Storage Optimization Infrastructure. He also has competitive knowledge of the major storage providers and converged infrastructure solution offerings (Matrix, UCS, Exadata/Exalogic).

Mark Kremkus is a Senior IT Specialist in the Advanced Technical Skills organization. He has 11 years of experience in the design of high performance, high availability solutions. Mark has achieved Consulting-level certification in Actualizing IT solutions. Mark's areas of expertise include enterprise storage performance analysis with emphasis on using empirical data to perform mathematical modeling of disk storage performance, and integrating storage with open systems hypervisors. He writes and presents on these topics. He holds a BS degree in Electrical Engineering from Texas A&M.

Markus Oscheka is an IT Specialist for Proof of Concepts and Benchmarks in the Disk Solution Europe team in Mainz, Germany. His areas of expertise include setup and demonstration of IBM System Storage solutions in open environments. He has written several IBM Redbooks, and has acted as the co-project lead for Redbooks including IBM DS8000® and IBM XIV Storage. He holds a degree in Electrical Engineering from the Technical University in Darmstad.

Thomas Peralto is a principal consultant in the storage solutions engineering group. He has extensive experience in implementing large and complex transport networks and mission-critical data protection throughout the globe. Mr. Peralto also serves as a data replication and data migration expert, and speaks both at national and international levels for IBM on the best practices for corporate data protection.

For their technical advice, support, and other contributions to this project, many thanks to:

Shimon Ben-David,
Brian Carmody
John Cherbini
Tim Dawson
Dave Denny
Rami Elron
Orli Gan
Tedd Gregg
Moriel Lechtman
Allen Marin
Brian Sherman
Aviad Offer
Juan Yanes
IBM

Thanks also to the authors of the previous editions:

Roger Eriksson
Wilhelm Gardt
Andrew Greenfield
Jana Jamsek
Suad Musovich
Nils Nause
Markus Oscheka
Rainer Pansky
In Kyu Park
Francesco Perillo

Paul Rea
Carlo Saba
Hank Sautter
Jim Sedgwick
Eugene Tsypin
Anthony Vandewerdt
Anthony Vattathil
Kip Wagner
Alexander Warmuth
Peter Wendler
Axel Westphal
Ralf Wohlfarth.

Special thanks to ESCC team in IBM Mainz, Germany for hosting the project and making equipment available in their lab.

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 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 SG24-7904-02
for IBM XIV Storage System Host Attachment and Interoperability
as created or updated on June 18, 2014.

March 2013, Third Edition

This revision reflects the addition, deletion, or modification of new and changed information described below.

New information

Information about the following products has been added:

- ▶ Reference information for Microsoft System Center Virtual Machine Manager 2012 Storage Automation
- ▶ Reference for Thin Reclamation Using Veritas Storage Foundation Enterprise HA
- ▶ Portable XIV Host Attach Kit for IBM AIX®

Changed information

The following information has changed:

- ▶ Several updates to reflect latest product information (as of November 2012)
- ▶ SVC attachment is now covered in the *IBM XIV Storage System: Copy Services and Migration*, SG24-7759.
- ▶ Additional VMware ESX and VMware SRM information is now covered in *XIV Storage System in VMware Environments*, REDP-4965
- ▶ Removed SONAS Gateway connectivity chapter



Host connectivity

This chapter addresses host connectivity for the XIV Storage System, in general. It highlights key aspects of host connectivity. It also reviews concepts and requirements for both Fibre Channel (FC) and Internet Small Computer System Interface (iSCSI) protocols.

The term *host* refers to a server that runs a supported operating system such as AIX or Windows. SAN Volume Controller as a host has special considerations because it acts as both a host and a storage device. For more information, see the SAN Volume Controller chapter in *IBM XIV Storage System: Copy Services and Migration*, SG24-7759.

This chapter does not address attachments from a secondary XIV used for Remote Mirroring or data migration from an older storage system. These topics are covered in *IBM XIV Storage System: Copy Services and Migration*, SG24-7759.

This chapter covers common tasks that pertain to most hosts. For operating system-specific information about host attachment, see the subsequent chapters in this book.

For the latest information, see the hosts attachment kit publications at:

[http://www.ibm.com/support/fixcentral/swg/quickorder?parent=Enterprise+Storage+Servers&product=ibm/Storage_Disk/XIV+Storage+System+\(2810,+2812\)&release=All&platform=All&function=all&source=fc](http://www.ibm.com/support/fixcentral/swg/quickorder?parent=Enterprise+Storage+Servers&product=ibm/Storage_Disk/XIV+Storage+System+(2810,+2812)&release=All&platform=All&function=all&source=fc)

This chapter includes the following sections:

- ▶ Overview
- ▶ Fibre Channel connectivity
- ▶ iSCSI connectivity
- ▶ Logical configuration for host connectivity
- ▶ Troubleshooting

1.1 Overview

The XIV Storage System can be attached to various host systems by using the following methods:

- ▶ Fibre Channel adapters using Fibre Channel Protocol (FCP)
- ▶ Fibre Channel over Converged Enhanced Ethernet (FCoCEE) adapters where the adapter connects to a converged network that is bridged to a Fibre Channel network
- ▶ iSCSI software initiator or iSCSI host bus adapter (HBA) using the iSCSI protocol

The XIV is perfectly suited for integration into a new or existing Fibre Channel storage area network (SAN). After the host HBAs, cabling, and SAN zoning are in place, connecting a Fibre Channel host to an XIV is easy. The XIV storage administrator defines the hosts and ports, and then maps volumes to them as LUNs.

You can also implement XIV with iSCSI using an existing Ethernet network infrastructure. However, your workload might require a dedicated network. iSCSI attachment and also iSCSI hardware initiators are not supported by all systems. If you have Ethernet connections between your sites, you can use that setup for a less expensive backup or disaster recovery setup. iSCSI connections are often used for asynchronous replication to a remote site. iSCSI-based mirroring that is combined with XIV snapshots or volume copies can also be used for the following tasks:

- ▶ Migrate servers between sites
- ▶ Facilitate easy off-site backup or software development

The XIV Storage System has up to 15 data modules, of which up to six are also interface modules. The number of interface modules and the activation status of the interfaces on those modules is dependent on the rack configuration. Table 1-1 summarizes the number of active interface modules and the FC and iSCSI ports for different rack configurations. As shown in Table 1-1, a six module XIV physically has three interface modules, but only two of them have active ports. An 11 module XIV physically has six interface modules, five of which have active ports. A 2nd Generation (model A14) XIV and an XIV Gen 3 (model 114) have different numbers of iSCSI ports.

Table 1-1 XIV host ports as capacity grows

Module	6	9	10	11	12	13	14	15
Module 9 host ports	Not present	Inactive ports	Inactive ports	Active	Active	Active	Active	Active
Module 8 host ports	Not present	Active	Active	Active	Active	Active	Active	Active
Module 7 host ports	Not present	Active	Active	Active	Active	Active	Active	Active
Module 6 host ports	Inactive ports	Inactive ports	Inactive ports	Inactive ports	Inactive ports	Active	Active	Active
Module 5 host ports	Active	Active	Active	Active	Active	Active	Active	Active
Module 4 host ports	Active	Active	Active	Active	Active	Active	Active	Active

Module	6	9	10	11	12	13	14	15
Fibre Channel Ports	8	16	16	20	20	24	24	24
iSCSI ports on model A14	0	4	4	6	6	6	6	6
iSCSI ports on model 114	6	14	14	18	18	22	22	22

Regardless of model, each active Interface Module (Modules 4-9, if enabled) has four Fibre Channel ports. The quantity of iSCSI ports varies based on XIV model:

- ▶ For 2nd Generation XIV, up to three Interface Modules (Modules 7-9, if enabled) have two iSCSI ports each, for a maximum of six ports.
- ▶ For XIV Gen 3, each active interface module except module 4 has four iSCSI ports. Module 4 on an XIV Gen 3 has only two iSCSI ports. The maximum is therefore 22 ports.

All of these ports are used to attach hosts, remote XIVs, or older storage systems (for migration) to the XIV. This connection can be through a SAN or iSCSI network that is attached to the internal patch panel.

The patch panel simplifies cabling because the Interface Modules are pre-cabled to it. Therefore, all your SAN and network connections are in one central place at the back of the rack. This arrangement also helps with general cable management.

Hosts attach to the FC ports through an FC switch, and to the iSCSI ports through a Gigabit Ethernet switch.

Restriction: Direct attachment between hosts and the XIV Storage System is not supported.

Figure 1-1 on page 4 shows an example of how to connect to a fully populated 2nd Generation (model A14) XIV Storage System. You can connect through either a storage area network (SAN) or an Ethernet network. For clarity, the patch panel is not shown.

Important: Host traffic can be directed to any of the Interface Modules. The storage administrator must ensure that host connections avoid single points of failure. The server administrator also must ensure that the host workload is adequately balanced across the connections and Interface Modules. This balancing can be done by installing the relevant host attachment kit. Review the balancing periodically and when traffic patterns change.

With XIV, all interface modules and all ports can be used concurrently to access any logical volume in the system. The only affinity is the mapping of logical volumes to host, which simplifies storage management. Balancing traffic and zoning (for adequate performance and redundancy) is more critical, although not more complex, than with traditional storage systems.

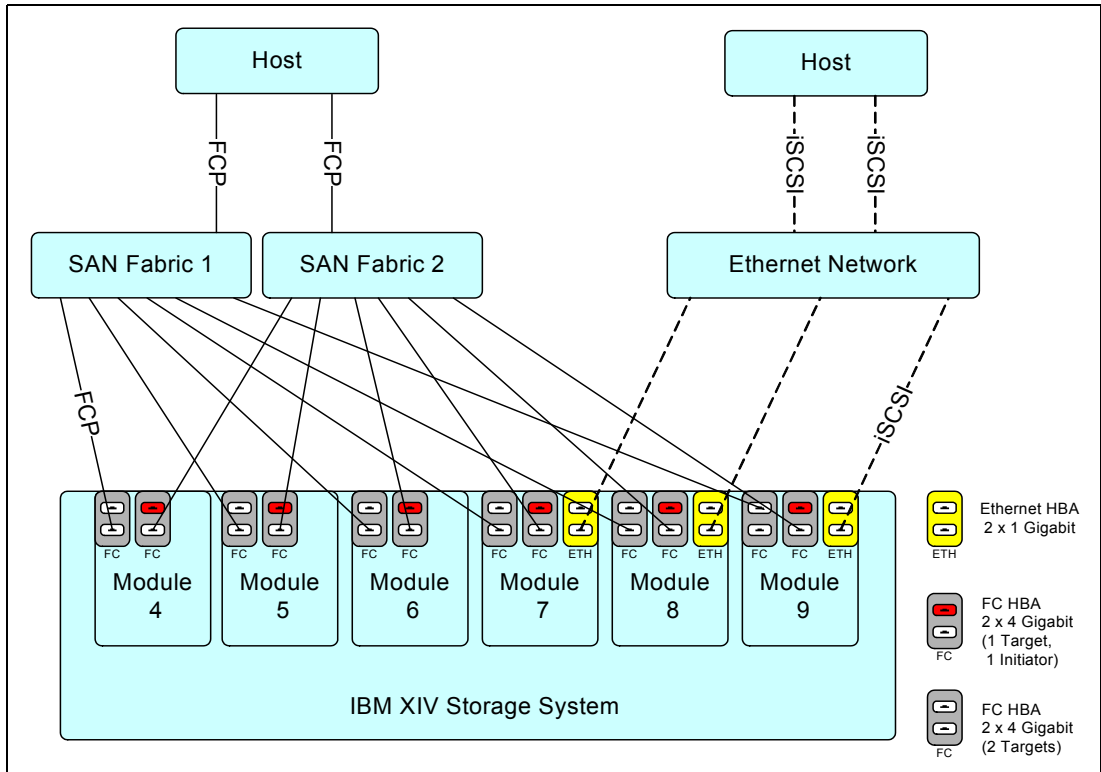


Figure 1-1 Host connectivity overview with 2nd Generation XIV (without patch panel)

1.1.1 Module, patch panel, and host connectivity

This section presents a simplified view of the host connectivity to explain the relationship between individual system components and how they affect host connectivity. For more information and an explanation of the individual components, see Chapter 3 of *IBM XIV Storage System Gen3 Architecture, Implementation, and Usage*, SG24-7659.

When connecting hosts to the XIV, there is no “one size fits all” solution that can be applied because every environment is different. However, follow these guidelines avoid single points of failure and ensure that hosts are connected to the correct ports:

- ▶ FC hosts connect to the XIV patch panel FC ports 1 and 3 (or FC ports 1 and 2 depending on your environment) on Interface Modules.
- ▶ Use XIV patch panel FC ports 2 and 4 (or ports 3 and 4 depending on your environment) for mirroring to another XIV Storage System. They can also be used for data migration from an older storage system.

Tip: Most illustrations in this book show ports 1 and 3 allocated for host connectivity. Likewise, ports 2 and 4 are reserved for more host connectivity, or remote mirror and data migration connectivity. This configuration gives you more resiliency because ports 1 and 3 are on separate adapters. It also gives you more availability. During adapter firmware upgrade, one connection remains available through the other adapter. It also boosts performance because each adapter has its own PCI bus.

For certain environments on 2nd Generation XIV (model A14), you must use ports 1 and 2 for host connectivity and reserve ports 3 and 4 for mirroring. If you do not use mirroring, you can also change port 4 to a target port.

Discuss with your IBM support representative what port allocation would be most desirable in your environment.

- ▶ iSCSI hosts connect to at least one port on each active Interface Module.

Restriction: A six module (27 TB) 2nd Generation XIV (model A14) does not have any iSCSI ports. If iSCSI ports are needed, you must upgrade that XIV to a nine module configuration or any size XIV Gen 3 (model 114).

- ▶ Connect hosts to multiple separate Interface Modules to avoid a single point of failure.

Figure 1-2 shows an overview of FC and iSCSI connectivity for a full rack configuration that uses a 2nd Generation XIV.

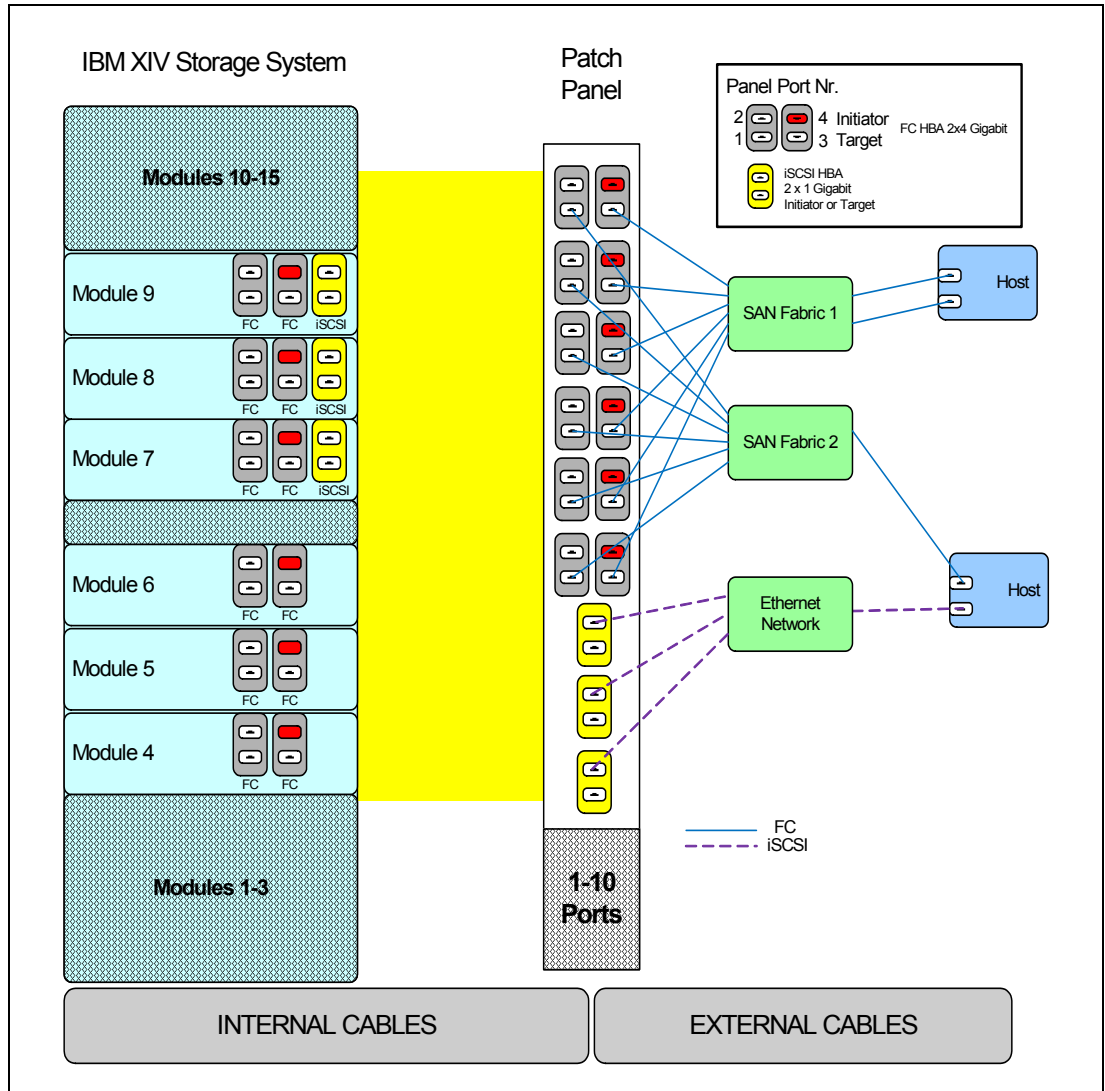


Figure 1-2 Host connectivity end-to-end view

Figure 1-3 shows a 2nd Generation (model A14) XIV patch panel to FC and patch panel to iSCSI adapter mappings. It also shows the worldwide port names (WWPNs) and iSCSI qualified names (IQNs) associated with the ports.

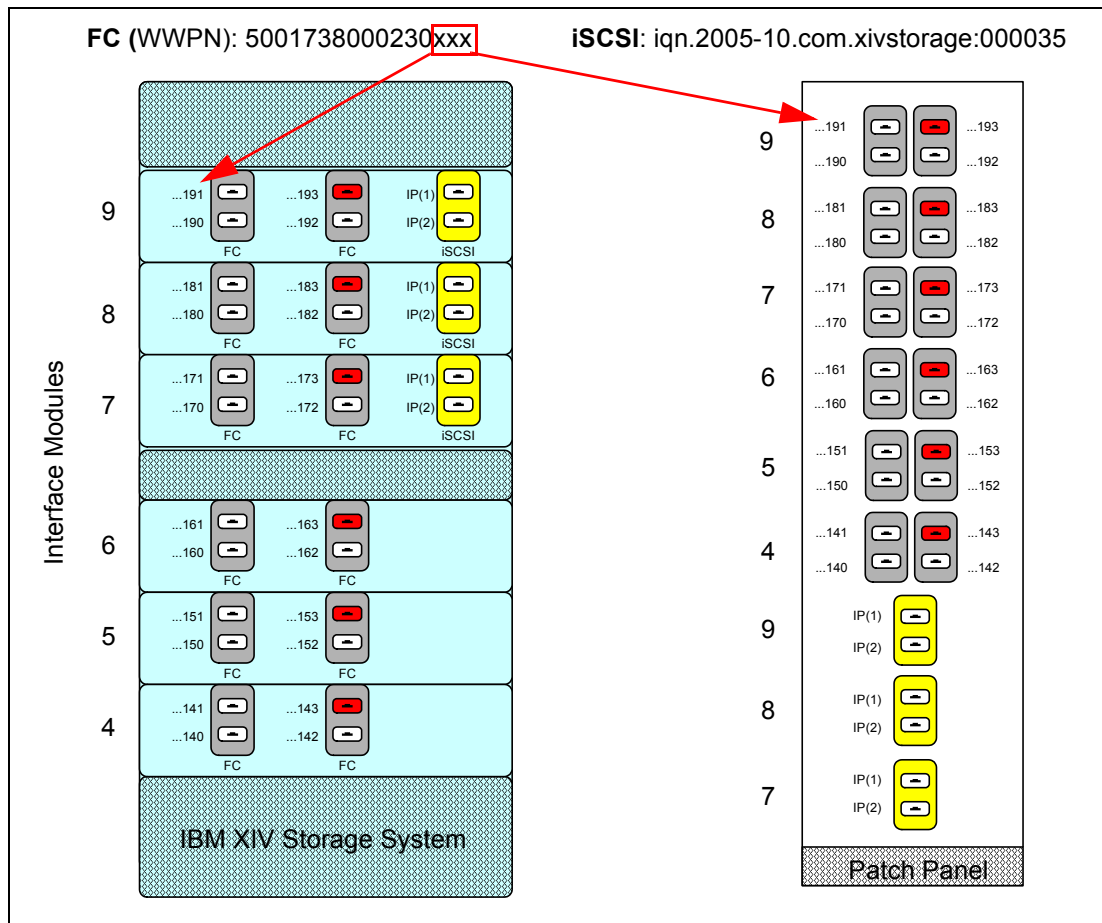


Figure 1-3 2nd Generation (model A14) patch panel to FC and iSCSI port mappings

Figure 1-4 shows an XIV Gen 3 (model 114) patch panel to FC and to iSCSI adapter mappings. It also shows the WWPNs associated with the ports.

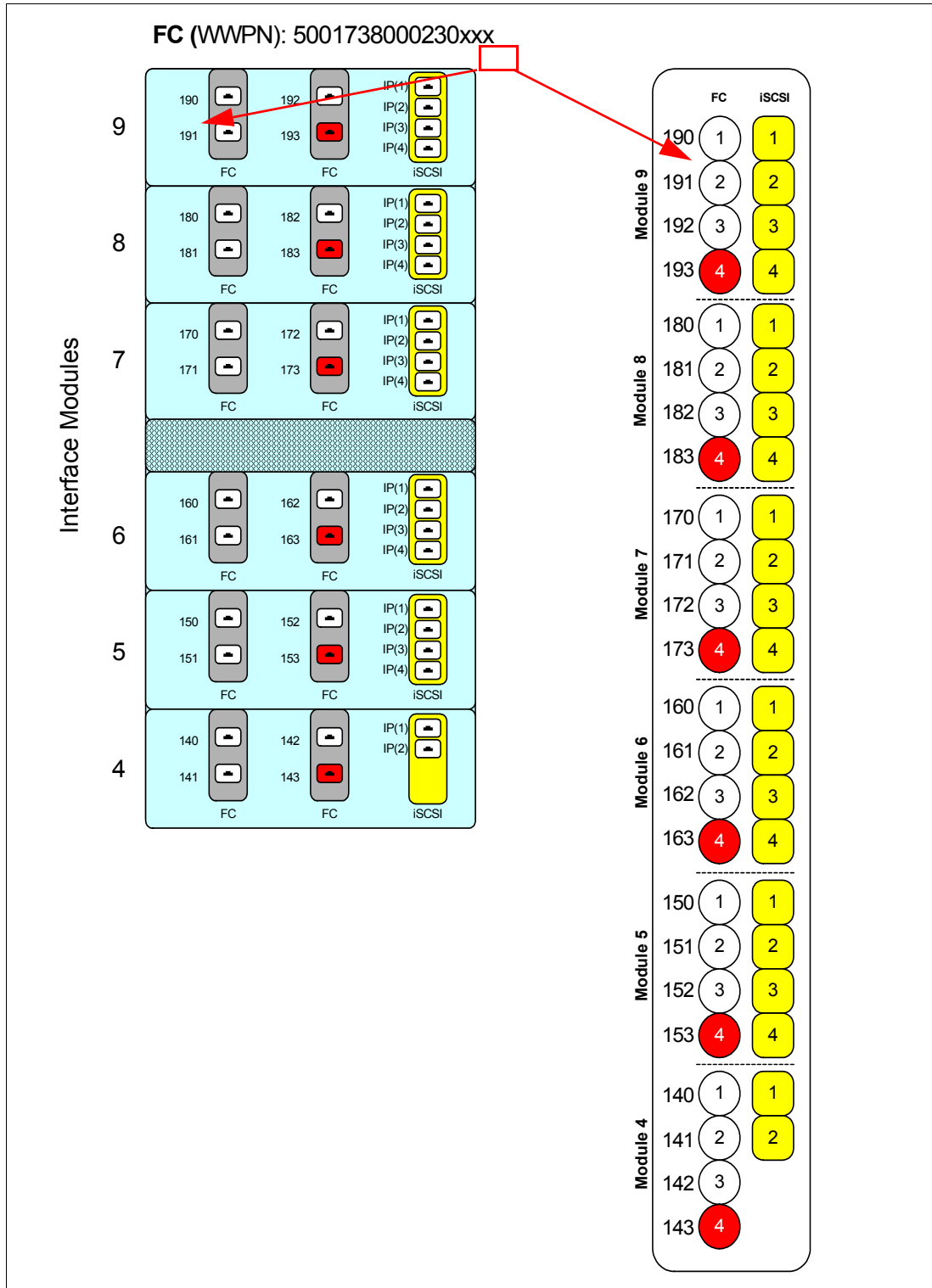


Figure 1-4 XIV Gen 3 Patch panel to FC and iSCSI port mappings

For more information about host connectivity and configuration options, see 1.2, “Fibre Channel connectivity” on page 13 and 1.3, “iSCSI connectivity” on page 28.

1.1.2 Host operating system support

The XIV Storage System supports many operating systems, and the list is constantly growing. The following operating systems are supported, among others:

- ▶ AIX
- ▶ VMware ESX/ESXi
- ▶ Linux (RHEL, SuSE)
- ▶ HP-UX
- ▶ VIOS (a component of Power/VM)
- ▶ IBM i (as a VIOS client)
- ▶ Solaris
- ▶ Windows

To get the current list, see the IBM System Storage Interoperation Center (SSIC) at:

<http://www.ibm.com/systems/support/storage/config/ssic>

From the SSIC, you can select any combination from the available boxes to determine whether your configuration is supported. You do not have to start at the top and work down. The result is a Comma Separated Values (CSV) file to show that you confirmed that your configuration is supported.

If you cannot locate your current (or planned) combination of product versions, talk to your IBM Business Partner, IBM Sales Representative, or IBM Pre-Sales Technical Support Representative. You might need to request a support statement called a Storage Customer Opportunity REquest (SCORE). It is sometimes called a request for price quotation (RPQ).

1.1.3 Downloading the entire XIV support matrix by using the SSIC

If you want to download every interoperability test result for a specific product version, complete the following steps:

1. Open the SSIC
2. Select the relevant version in the **Product Version** box.
3. Select **Export Selected Product Version (xls)**. Figure 1-5 shows all the results for XIV Gen 3, which uses XIV Software version 11.

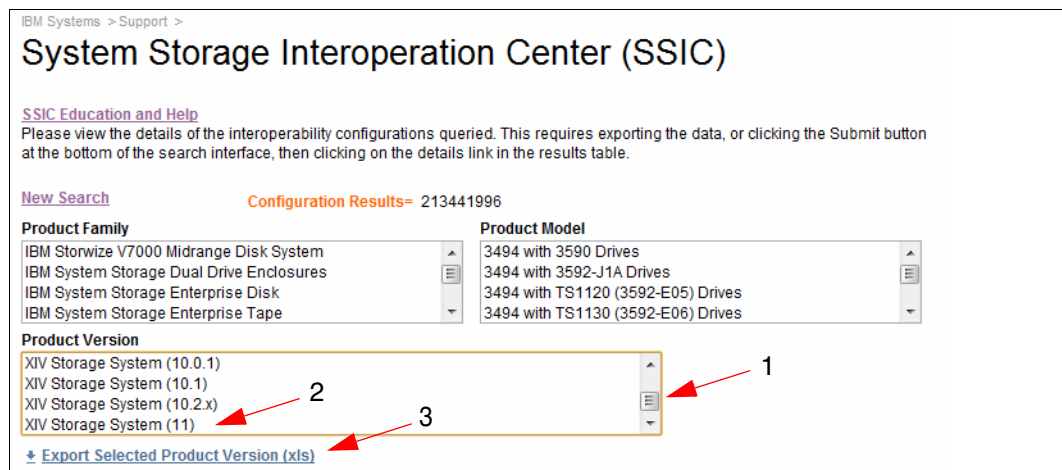


Figure 1-5 Exporting an entire product version in the SSIC

1.1.4 Host Attachment Kits

For high availability, every host that is attached to an XIV must have multiple paths to the XIV. In the past, you had to install vendor-supplied multi-pathing software such as Subsystem Device Driver (SDD) or Redundant Disk Array Controller (RDAC). However, multi-pathing that is native to the host is more efficient. Most operating systems such as AIX, Windows, VMware, and Linux are now capable of providing native multi-pathing. IBM has Host Attachment Kits for most of these supported operating systems. These kits customize the host multipathing. The Host Attachment Kit also supplies powerful tools to assist the storage administrator in day-to-day tasks.

The Host Attachment Kits have the following features:

- ▶ Backwards compatibility to Version 10.1.x of the XIV system software
- ▶ Validates host server patch and driver versions
- ▶ Sets up multipathing on the host using native multipathing
- ▶ Adjusts host system tunable parameters (if required) for performance
- ▶ Provides an installation wizard (which might not be needed if you use the portable version)
- ▶ Provide management utilities such as the `xiv_devlist` command
- ▶ Provide support and troubleshooting utilities such as the `xiv_diag` command
- ▶ A portable version that can be run without installation (starting with release 1.7)

Host Attachment Kits are built on a Python framework, and provide a consistent interface across operating systems. Other XIV tools, such as the Microsoft Systems Center Operations Manager (SCOM) management pack, also install a Python-based framework called xPYV. With release 1.7 of the Host Attachment Kit, the Python framework is now embedded with the Host Attachment Kit code. It is no longer a separate installer.

Before release 1.7 of the Host Attachment Kit, it was *mandatory* to install the Host Attachment Kit to get technical support from IBM. Starting with release 1.7, a portable version allows all Host Attachment Kit commands to be run without installing the Host Attachment Kit.

Host Attachment Kits can be downloaded from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

Commands provided by the XIV Host Attachment Kit

Regardless of which host operating system is in use, the Host Attachment Kit provides a uniform set of commands that create output in a consistent manner. Each chapter in this book includes examples of the appropriate Host Attachment Kit commands. This section lists all of them for completeness. In addition, useful parameters are suggested.

xiv_attach

This command locally configures the operating system and defines the host on the XIV.

Tip: AIX needs extra consideration. For more information, see “Installing the XIV Host Attachment Kit for AIX” on page 140.

Sometimes after you run the `xiv_attach` command, you might be prompted to reboot the host. This reboot might be needed because the command can perform system modifications that force a reboot based on the normal behavior of the operating system. For example, a reboot is required when you install a Windows hot fix. You need to run this command only once, when performing initial host configuration. After the first time, use `xiv_fc_admin -R` to detect newly mapped volumes.

xiv_detach

This command is used on a Windows Server to remove all XIV multipathing settings from the host. For other operating systems, use the uninstallation option. If you are upgrading a server from Windows 2003 to Windows 2008, use **xiv_detach** first to remove the multi-pathing settings.

xiv_devlist

This command displays a list of all volumes that are visible to the system. It also displays the following information:

- ▶ The size of the volume
- ▶ The number of paths (working and detected)
- ▶ The name and ID of each volume on the XIV
- ▶ The ID of the XIV itself
- ▶ The name of the host definition on the XIV

The **xiv_devlist** command is one of the most powerful tools in your toolkit. Make sure that you are familiar with this command and use it whenever performing system administration. The XIV Host Attachment Kit Attachment Guide lists a number of useful parameters that can be run with **xiv_devlist**. The following parameters are especially useful:

xiv_devlist -u GiB	Displays the volume size in binary GB. The -u stands for unit size.
xiv_devlist -V	Displays the Host Attachment Kit version number. The -V stands for version.
xiv_devlist -f filename.csv -t csv	Directs the output of the command to a file.
xiv_devlist -h	Brings up the help page that displays other available parameters. The -h stands for help.

xiv_diag

This command is used to satisfy requests from the IBM support center for log data. The **xiv_diag** command creates a compressed packed file (using tar.gz format) that contains log data. Therefore, you do not need to collect individual log files from your host server.

xiv_fc_admin

This command is similar to **xiv_attach**. Unlike **xiv_attach**, however, the **xiv_fc_admin** command allows you to perform individual steps and tasks. The following **xiv_fc_admin** command parameters are especially useful:

xiv_fc_admin -P	Displays the WWPNs of the host server HBAs. The -P stands for print.
xiv_fc_admin -V	Lists the tasks that xiv_attach would perform if it were run. Knowing the tasks is vital if you are using the portable version of the Host Attachment Kit. You must know what tasks the Host Attachment Kit needs to perform on your system before the change window. The -V stands for verify.
xiv_fc_admin -C	Performs all the tasks that the xiv_fc_admin -V command identified as being required for your operating system. The -C stands for configure.
xiv_fc_admin -R	This command scans for and configures new volumes that are mapped to the server. For a new host that is not yet connected to an XIV, use xiv_attach . However, if more volumes are mapped to such a host later, use xiv_fc_admin -R to detect them. You can use native host methods but the Host Attachment Kit command is an easier way to detect volumes. The -R stands for rescan.

`xiv_fc_admin -h` Brings up the help page that displays other available parameters. The `-h` stands for help.

xiv_iscsi_admin

This command is similar to `xiv_fc_admin`, but is used on hosts with iSCSI interfaces rather than Fibre Channel.

Co-existence with other multipathing software

The Host Attachment Kit is itself not a multi-pathing driver. It enables and configures multipathing rather than providing it. IBM insists that the correct host attachment kit be installed for each OS type.

A mix of different multipathing solution software on the same server is not supported. Each product can have different requirements for important system settings, which can conflict. These conflicts can cause issues that range from poor performance to unpredictable behaviors, and even data corruption.

If you need co-existence and a support statement does not exist, apply for a support statement from IBM. This statement is known as a SCORE, or sometimes an RPQ. There is normally no additional charge for this support request.

1.1.5 Fibre Channel versus iSCSI access

Hosts can attach to XIV over a Fibre Channel or Ethernet network (using iSCSI). The version of XIV system software at the time of writing supports iSCSI using the software initiator only. The only exception is AIX, where an iSCSI HBA is also supported.

Choose the connection protocol (iSCSI or FCP) based on your application requirements. When you are considering IP storage-based connectivity, look at the performance and availability of your existing infrastructure.

Take the following considerations into account:

- ▶ Always connect FC hosts in a production environment to a minimum of two separate SAN switches in independent fabrics to provide redundancy.
- ▶ For test and development, you can choose to have single points of failure to reduce costs. However, you must determine whether this practice is acceptable for your environment. The cost of an outage in a development environment can be high, and an outage can be caused by the failure of a single component.
- ▶ When you are using iSCSI, use a separate section of the IP network to isolate iSCSI traffic using either a VLAN or a physically separated section. Storage access is susceptible to latency or interruptions in traffic flow. Do not mix it with other IP traffic.

Figure 1-6 illustrates the simultaneous access to two different XIV volumes from one host by using both protocols.

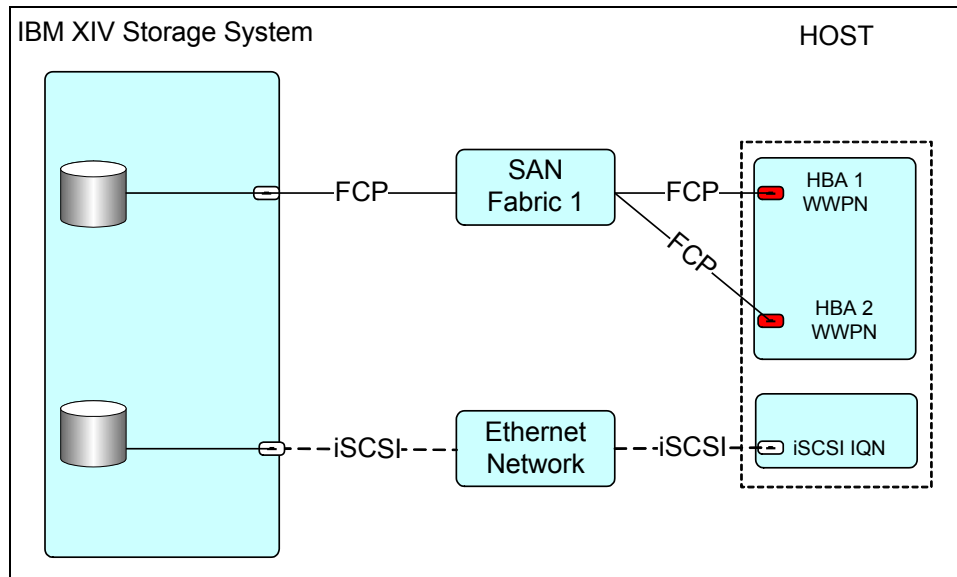


Figure 1-6 Connecting by using FCP and iSCSI simultaneously with separate host objects

A host can connect through FC and iSCSI simultaneously. However, you cannot access the same LUN with both protocols.

1.2 Fibre Channel connectivity

This section highlights information about FC connectivity that applies to the XIV Storage System in general. For operating system-specific information, see the relevant section in the subsequent chapters of this book.

1.2.1 Preparation steps

Before you can attach an FC host to the XIV Storage System, you must complete several procedures. The following general procedures pertain to all hosts. However, you also must review any procedures that pertain to your specific hardware and operating system.

1. Ensure that your HBA is supported. Information about supported HBAs and the firmware and device driver levels is available at the SSIC website at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

For each query, select the XIV Storage System, a host server model, an operating system, and an HBA vendor. Each query shows a list of all supported HBAs. Unless otherwise noted in SSIC, you can use any supported driver and firmware by the HBA vendors. The latest versions are always preferred. For HBAs in Oracle / Sun systems, use Sun-branded HBAs and Sun-ready HBAs only.

Also, review any documentation that comes from the HBA vendor and ensure that any additional conditions are met.

2. Check the LUN limitations for your host operating system and verify that there are enough adapters installed. You need enough adapters on the host server to manage the total number of LUNs that you want to attach.

3. Check the optimum number of paths that must be defined to help determine the zoning requirements.
4. Download and install the latest supported HBA firmware and driver, if needed.

HBA vendor resources

All of the Fibre Channel HBA vendors have websites that provide information about their products, facts, and features, and support information. These sites are useful when you need details that cannot be supplied by IBM resources. IBM is not responsible for the content of these sites.

Brocade

The Brocade website can be found at:

<http://www.brocade.com/services-support/drivers-downloads/adapters/index.page>

QLogic Corporation

The QLogic website can be found at:

<http://www.qlogic.com>

QLogic maintains a page that lists all the HBAs, drivers, and firmware versions that are supported for attachment to IBM storage systems at:

http://support.qlogic.com/support/oem_ibm.asp

Emulex Corporation

The Emulex home page is at:

<http://www.emulex.com>

They also have a page with content specific to IBM storage systems at:

<http://www.emulex.com/products/host-bus-adapters/ibm-branded.html>

Oracle

Oracle ships its own HBAs. They are Emulex and QLogic based. However, these “native” HBAs can be used to attach servers that are running Oracle Solaris to disk systems. In fact, such HBAs can even be used to run StorEdge Traffic Manager software. For more information, see the following websites:

- ▶ For Emulex:
<http://www.oracle.com/technetwork/server-storage/solaris/overview/emulex-corporation-136533.html>
- ▶ For QLogic:
<http://www.oracle.com/technetwork/server-storage/solaris/overview/qlogic-corp--139073.html>

HP

HP ships its own HBAs.

- ▶ Emulex publishes a cross-reference at:
<http://www.emulex-hp.com/interop/matrix/index.jsp?mfgId=26>
- ▶ QLogic publishes a cross-reference at:
http://driverdownloads.qlogic.com/QLogicDriverDownloads_UI/Product_detail.aspx?oemid=21

Platform and operating system vendor pages

The platform and operating system vendors also provide support information for their clients. See this information for general guidance about connecting their systems to SAN-attached storage. However, be aware that you might not be able to find information to help you with third-party vendors. Check with IBM about interoperability and support from IBM in regard to these products. It is beyond the scope of this book to list all of these vendors' websites.

Special consideration for Brocade Fabric

For Brocade Fabric OS versions 6.2.0, 6.2.0a, and 6.2.0b, the default fillword is "ARBF/ARBF" (mode 1).

For Brocade Fabric OS versions 6.2.0c and above, the default fillword changed to "IDLE/IDLE" (mode 0). The "IDLE/IDLE" fillword setting is not supported with 8 Gbps XIV storage system ports and might cause loss of connectivity.

Restriction: The XIV storage system supports any fillword *except* "IDLE/IDLE".

Use the following guidelines:

- ▶ When using Brocade Fabric OS versions 6.2.0, 6.2.0a, or 6.2.0b, keep the default fillword "ARBF/ARBF" (mode 1).
- ▶ When using Brocade Fabric OS versions 6.2.0c to 6.3.0d, change the default fillword to "ARBF/ARBF" (mode 1).
- ▶ When using Brocade Fabric OS versions 6.3.1 or later, use the following fillword "if ARBF/ARBF fails use IDLE/ARBF" (mode 3).

These guidelines apply only when the XIV 8 Gbps FC ports are connected to an 8 Gbps Brocade switch.

Important: Keep in mind that changing the fillword is considered a disruptive action.

Use the following procedure to change the fillword, if necessary:

1. Check the current fillword configuration by using the **portcfgshow** command:

```
IBM_2499_192:FID128:admin> portcfgshow 1/0
Area Number: 0
Speed Level: AUTO(HW)
Fill Word: 3(A-A then SW I-A)
AL_PA Offset 13: OFF
```

2. Change the fillword for a specific port by using the **portcfgfillword** command. For example to set the fillword mode 3 on blade 1 and port 0, issue:

```
portCfgFillWord 1/0 3
```

1.2.2 Fibre Channel configurations

Several configurations using Fibre Channel are technically possible. They vary in terms of their cost, and the degree of flexibility, performance, and reliability that they provide.

Production environments must always have a redundant (high availability) configuration. Avoid single points of failure. Assign as many HBAs to hosts as needed to support the operating system, application, and overall performance requirements.

This section details three typical FC configurations that are supported and offer redundancy. All of these configurations have no single point of failure:

- ▶ If a module fails, each host remains connected to all other interface modules.
- ▶ If an FC switch fails, each host remains connected to at least three interface modules.
- ▶ If a host HBA fails, each host remains connected to at least three interface modules.
- ▶ If a host cable fails, each host remains connected to at least three interface modules.

Redundant configuration with 12 paths to each volume

The fully redundant configuration is illustrated in Figure 1-7.

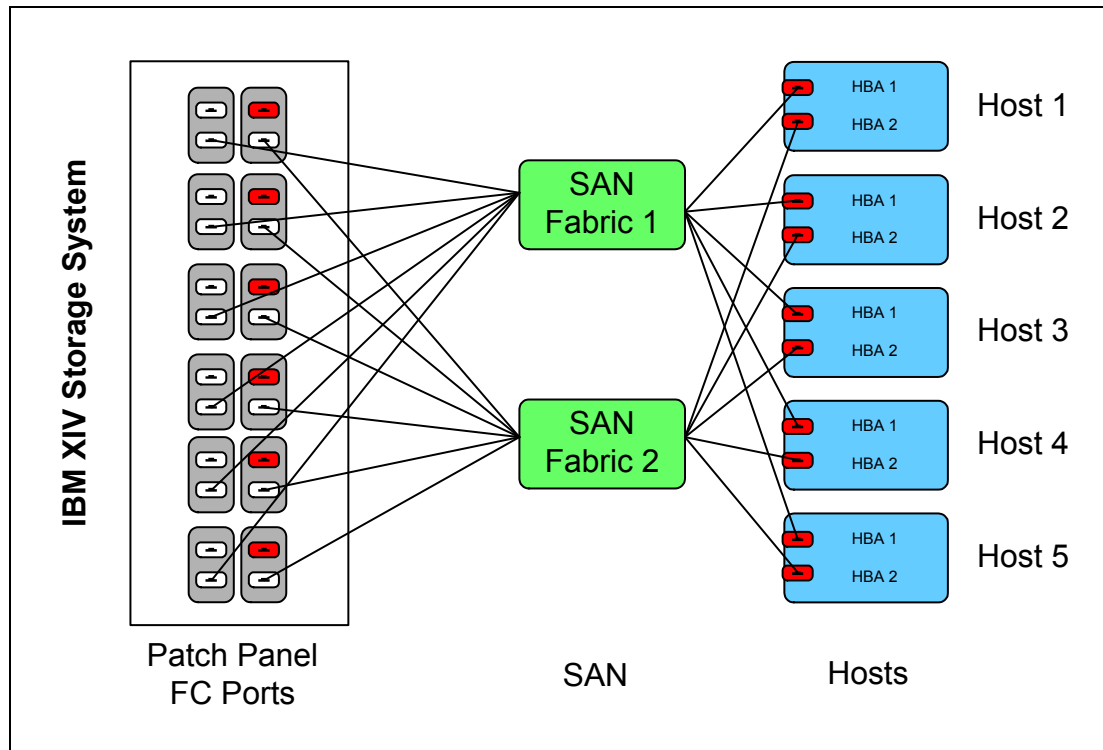


Figure 1-7 Fibre Channel fully redundant configuration

This configuration has the following characteristics:

- ▶ Each host is equipped with dual HBAs. Each HBA (or HBA port) is connected to one of two FC switches.
- ▶ Each of the FC switches has a connection to a separate FC port of each of the six Interface Modules.
- ▶ Each volume can be accessed through 12 paths. There is no benefit in going beyond 12 paths because it can cause issues with host processor utilization and server reliability if a path failure occurs.

Redundant configuration with six paths to each volume

A redundant configuration that accesses all interface modules, but uses the ideal of six paths per LUN on the host, is depicted in Figure 1-8.

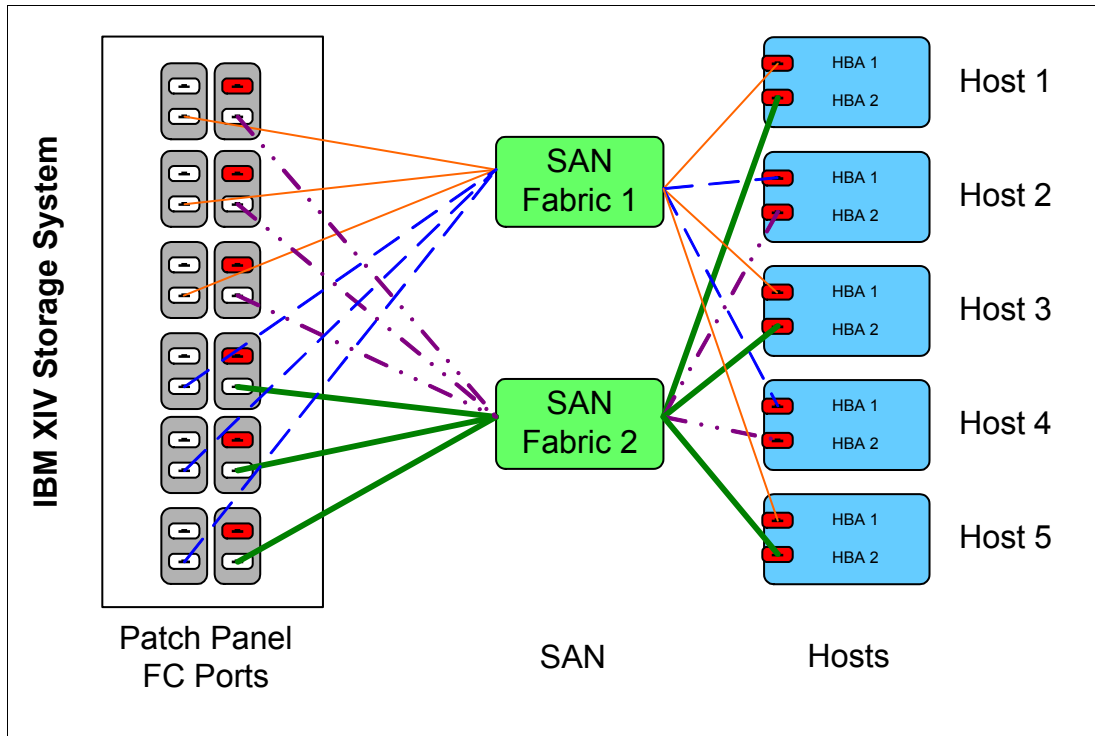


Figure 1-8 Fibre Channel redundant configuration

This configuration has the following characteristics:

- ▶ Each host is equipped with dual HBAs. Each HBA (or HBA port) is connected to one of two FC switches.
- ▶ Each of the FC switches has a connection to a separate FC port of each of the six Interface Modules.
- ▶ One host is using the first three paths per fabric and the other is using the three other paths per fabric.
- ▶ If a fabric fails, all interface modules are still used.
- ▶ Each volume has six paths. Six paths is the ideal configuration.

Important: Six paths per LUN is the best overall multipathing configuration.

Redundant configuration with minimal cabling

An even simpler redundant configuration is illustrated in Figure 1-9.

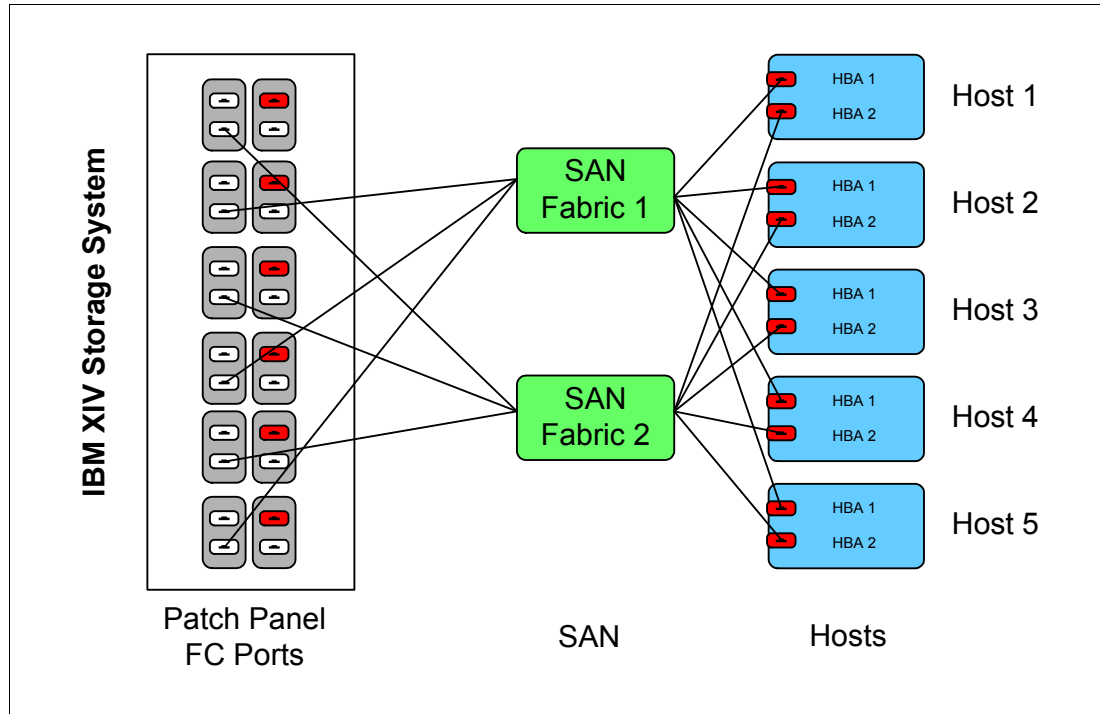


Figure 1-9 Fibre Channel simple redundant configuration

This configuration has the following characteristics:

- ▶ Each host is equipped with dual HBAs. Each HBA (or HBA port) is connected to one of two FC switches.
- ▶ Each of the FC switches has a connection to three separate interface modules.
- ▶ Each volume has six paths.

Determining the ideal path count

In the examples in this chapter, SAN zoning can be used to control the number of paths that are configured per volume. Because the XIV can have up to 24 Fibre Channel ports, you might be tempted to configure many paths. However, using many paths is not a good practice.

Tip: There is no performance or reliability benefit in using too many paths. Going beyond 12 paths per volume has no benefit. More paths add more processor usage and cause longer times for recovery, and going beyond six paths rarely has much benefit. Use four or six paths per volume as a standard.

Consider the configurations that are shown in Table 1-2 on page 19. The columns show the interface modules, and the rows show the number of installed modules. The table does not show how the system is cabled to each redundant SAN fabric, or how many cables are connected to the SAN fabric. You normally connect each module to each fabric and alternate which ports you use on each module.

- ▶ For a six module system, each host has four paths per volume: Two from module 4 and two from module 5. Port 1 on each module is connected to fabric A, whereas port 3 on each module is connected to fabric B. Each host would be zoned to all four ports.
- ▶ For a 9 or 10 module system, each host has four paths per volume (one from each module). Port 1 on each module is connected to fabric A, whereas port 3 on each module is connected to fabric B. Divide the hosts into two groups. Group 1 is zoned to port 1 on modules 4 and 8 in fabric A, and port 3 on modules 5 and 7 in fabric B. Group 2 is zoned to port 3 on modules 4 and 8 in fabric B, and port 1 on modules 5 and 7 in fabric A.
- ▶ For an 11 or 12 module system, each host has five paths per volume. Port 1 on each module is connected to fabric A, whereas port 3 on each module is connected to fabric B. Divide the hosts into two groups. Group 1 is zoned to port 1 on modules 4 and 8 in fabric A, and port 3 on modules 5, 7 and 9 in fabric B. Group 2 is zoned to port 3 on modules 4 and 8 in fabric B, and port 1 on modules 5, 7 and 9 in fabric A. This configuration has a slight disadvantage in that one HBA can get slightly more workload than the other HBA. The extra workload is not usually an issue.
- ▶ For a 13, 14 or 15 module system, each host would have six paths per volume (three paths from each fabric). Port 1 on each module is connected to fabric A, whereas port 3 on each module is connected to fabric B. Divide the hosts into two groups. Group 1 is zoned to port 1 on modules 4, 6 and 8 in fabric A, and port 3 on modules 5, 7 and 9 in fabric B. Group 2 is zoned to port 3 on modules 4, 6 and 8 in fabric B, and port 1 on modules 5, 7 and 9 in fabric A.

Table 1-2 Number of paths per volume per interface module

Modules	4	5	6	7	8	9
6	2 paths	2 paths	Inactive	Not present	Not present	Not present
9 or 10	1 path	1 path	Inactive	1 path	1 path	Inactive
11 or 12	1 path	1 path	Inactive	1 path	1 path	1 path
13, 14 or 15	1 path	1 path	1 path	1 path	1 path	1 path

This path strategy works best on systems that start with nine modules. If you start with six modules, you must reconfigure all hosts when you upgrade to a nine module configuration. Do not go below four paths.

1.2.3 Zoning

Zoning is mandatory when you are connecting FC hosts to an XIV Storage System. Zoning is configured on the SAN switch, and isolates and restricts FC traffic to only those HBAs within a specific zone.

A zone can be either a *hard zone* or a *soft zone*. Hard zones group HBAs depending on the physical ports they are connected to on the SAN switches. Soft zones group HBAs depending on the WWPNs of the HBA. Each method has its merits, and you must determine which is correct for your environment. From a switch perspective, both methods are enforced by the hardware.

Correct zoning helps avoid issues and makes it easier to trace the cause of errors. Here are examples of why correct zoning is important:

- ▶ An error from an HBA that affects the zone or zone traffic is isolated to only the devices that it is zoned to.

- ▶ Any change in the SAN fabric triggers a *registered state change notification* (RSCN). Such changes can be caused by a server restarting or a new product being added to the SAN. An RSCN requires that any device that can “see” the affected or new device to acknowledge the change, interrupting its own traffic flow.

Important: Disk and tape traffic are ideally handled by separate HBA ports because they have different characteristics. If both traffic types use the same HBA port, it can cause performance problems, and other adverse and unpredictable effects

Zoning is affected by the following factors, among others:

- ▶ Host type
- ▶ Number of HBAs
- ▶ HBA driver
- ▶ Operating system
- ▶ Applications

Therefore, it is not possible to provide a solution to cover every situation. The following guidelines can help you to avoid reliability or performance problems. However, also review documentation about your hardware and software configuration for any specific factors that must be considered.

- ▶ Each zone (excluding those for SAN Volume Controller) has one initiator HBA (the host) and multiple target HBA ports from a single XIV
- ▶ Zone each host to ports from at least two Interface Modules.
- ▶ Do not mix disk and tape traffic in a single zone. Also, avoid having disk and tape traffic on the same HBA.

For more information about SAN zoning, see Section 4.7 of *Introduction to Storage Area Networks and System Networking*, SG24-5470. You can download this publication from:

<http://www.redbooks.ibm.com/redbooks/pdfs/sg245470.pdf>

Soft zoning using the “single initiator, multiple targets” method is illustrated in Figure 1-10.

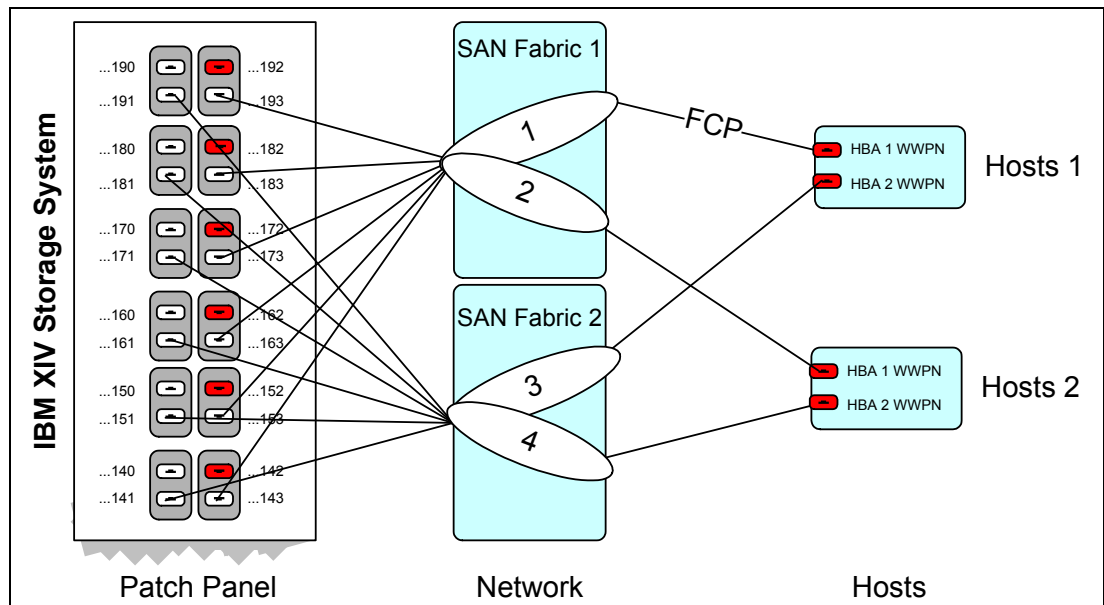


Figure 1-10 FC SAN zoning: single initiator, multiple target

Spread the IO workload evenly between the interfaces. For example, for a host that is equipped with two single port HBA, connect one HBA port to one port on modules 4, 6, and 8. Also, connect the second HBA port to one port on modules 5, 7, and 9. This configuration divides the workload between even and odd-numbered interface modules.

When round-robin is not in use (for example, with VMware ESX 3.5 or AIX 5.3 TL9 and earlier, or AIX 6.1 TL2 and earlier), statically balance the workload between the paths. Monitor the IO workload on the interfaces to make sure that it stays balanced by using the XIV statistics view in the GUI (or XIVTop).

1.2.4 Identification of FC ports (initiator/target)

You must identify ports before you set up the zoning. This identification aids any modifications that might be required, and assists with problem diagnosis. The unique name that identifies an FC port is called the WWPN.

The easiest way to get a record of all the WWPNs on the XIV is to use the XIV Command Line Interface (XCLI). However, this information is also available from the GUI. Example 1-1 shows all WWPNs for one of the XIV Storage Systems that were used in the preparation of this book. It also shows the XCLI command that was used to list them. For clarity, some of the columns have been removed.

Example 1-1 Getting the WWPN of an IBM XIV Storage System (XCLI)

```
>> fc_port_list
```

Component ID	Status	Currently Functioning	WWPN	Port ID	Role
1:FC_Port:4:1	OK	yes	5001738000230140	00030A00	Target
1:FC_Port:4:2	OK	yes	5001738000230141	00614113	Target
1:FC_Port:4:3	OK	yes	5001738000230142	00750029	Target
1:FC_Port:4:4	OK	yes	5001738000230143	00FFFFFF	Initiator
1:FC_Port:5:1	OK	yes	5001738000230150	00711000	Target
.....					
1:FC_Port:6:1	OK	yes	5001738000230160	00070A00	Target
.....					
1:FC_Port:7:1	OK	yes	5001738000230170	00760000	Target
.....					
1:FC_Port:8:1	OK	yes	5001738000230180	00060219	Target
.....					
1:FC_Port:9:1	OK	yes	5001738000230190	00FFFFFF	Target
1:FC_Port:9:2	OK	yes	5001738000230191	00FFFFFF	Target
1:FC_Port:9:3	OK	yes	5001738000230192	00021700	Target
1:FC_Port:9:4	OK	yes	5001738000230193	00021600	Initiator

The `fc_port_list` command might not always print the port list in the same order. Although they might be ordered differently, all the ports are listed.

To get the same information from the XIV GUI, complete the following steps:

1. Select the main view of an XIV Storage System.
2. Use the arrow at the bottom (circled in red) to reveal the patch panel.

3. Move the mouse cursor over a particular port to reveal the port details, which include the WWPN, as shown in Figure 1-11.

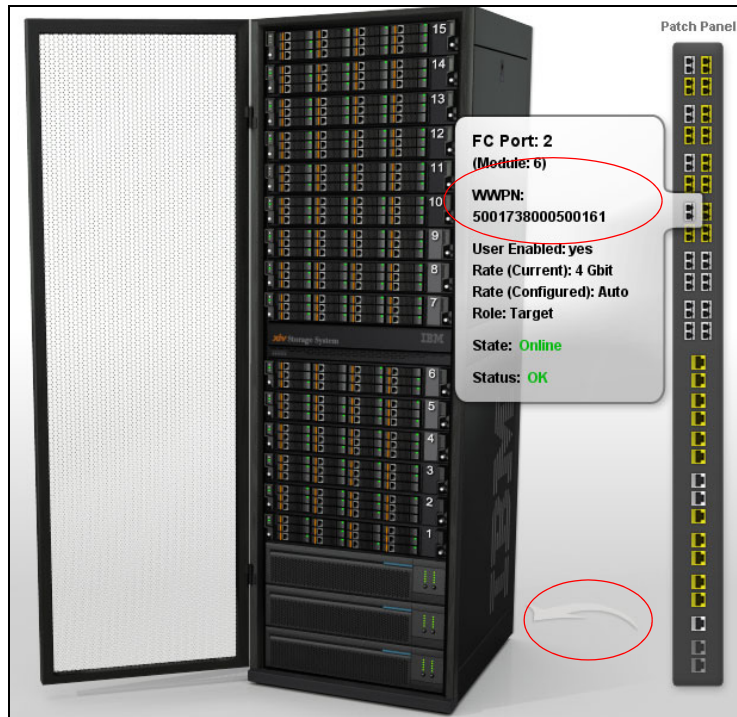


Figure 1-11 Getting the WWPNs of IBM XIV Storage System (GUI)

Tip: The WWPNs of an XIV Storage System are static. The last two digits of the WWPN indicate to which module and port the WWPN corresponds.

As shown in Figure 1-11, the WWPN is 5001738000500161, which means that the WWPN is from module 6, port 2. The WWPNs for the port are numbered from 0 to 3, whereas the physical ports are numbered from 1 to 4.

The values that comprise the WWPN are shown in Example 1-2.

Example 1-2 Composition of the WWPN

If WWPN is 50:01:73:8N:NN:NN:RR:MP

5	NAA (Network Address Authority)
001738	IEEE Company ID from http://standards.ieee.org/regauth/oui/oui.txt
NNNNN	IBM XIV Serial Number in hexadecimal
RR	Rack ID (01-FF, 00 for WNN)
M	Module ID (1-F, 0 for WNN)
P	Port ID (0-7, 0 for WNN)

1.2.5 Boot from SAN on x86/x64 based architecture

Booting from SAN creates a number of possibilities that are not available when booting from local disks. The operating systems and configuration of SAN-based computers can be centrally stored and managed. Central storage is an advantage with regards to deploying servers, backup, and disaster recovery procedures.

To boot from SAN, complete these basic steps:

1. Go into the HBA configuration mode.
2. Set the HBA BIOS to *Enabled*.
3. Detect at least one XIV target port.
4. Select a LUN to boot from.

You typically configure 2-4 XIV ports as targets. You might need to enable the BIOS on two HBAs, depending on the HBA, driver, and operating system. See the documentation that came with your HBA and operating systems.

For information about SAN boot for AIX, see Chapter 4, “XIV and AIX host connectivity” on page 137. For information about SAN boot for HPUX, see Chapter 5, “XIV and HP-UX host connectivity” on page 171.

The procedures for setting up your server and HBA to boot from SAN vary. They are dependent on whether your server has an Emulex or QLogic HBA (or the OEM equivalent). The procedures in this section are for a QLogic HBA. If you have an Emulex card, the configuration panels differ but the logical process is the same.

1. Boot your server. During the boot process, press Ctrl+q when prompted to load the configuration utility and display the **Select Host Adapter** menu (Figure 1-12).

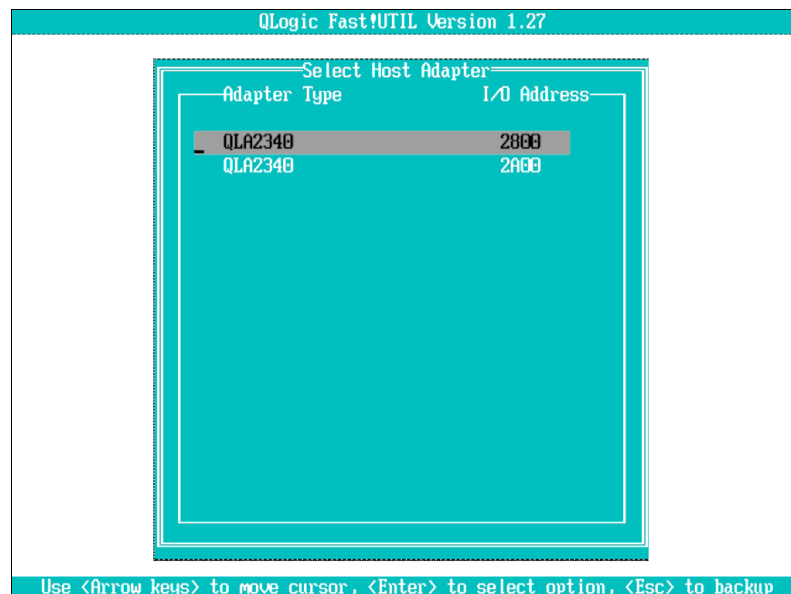


Figure 1-12 Select Host Adapter menu

2. You normally see one or more ports. Select a port and press Enter to display the panel that is shown in Figure 1-13. If you are enabling the BIOS on only one port, make sure to select the correct port.

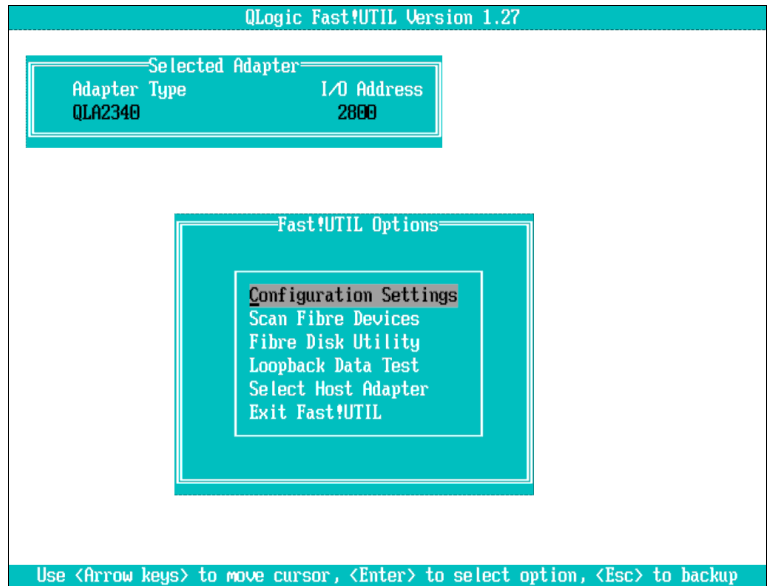


Figure 1-13 Fast!UTIL Options menu

3. Select **Configuration Settings**.
4. In the panel that is shown in Figure 1-14, select **Adapter Settings**.

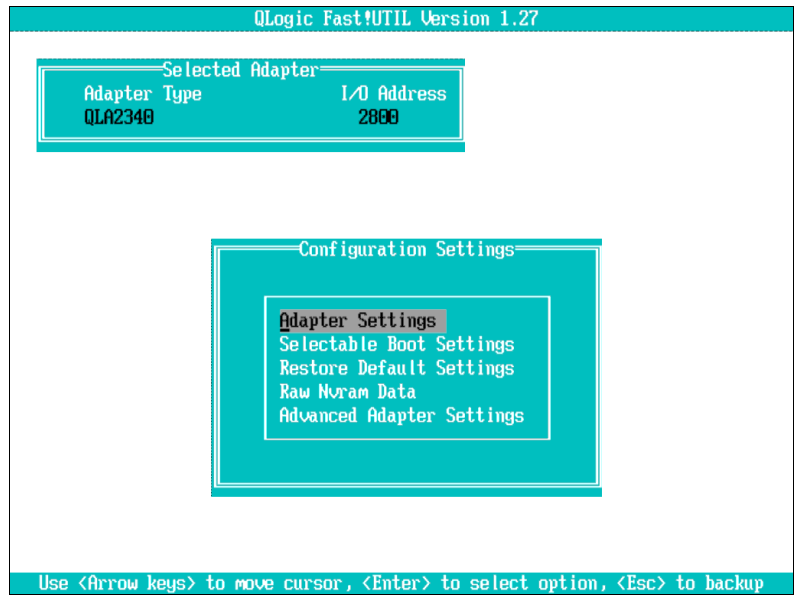


Figure 1-14 Configuration Settings menu

- The **Adapter Settings** menu is displayed as shown in Figure 1-15. Change the **Host Adapter BIOS** setting to **Enabled**, then press Esc to exit and go back to the **Configuration Settings** menu shown in Figure 1-14 on page 24.

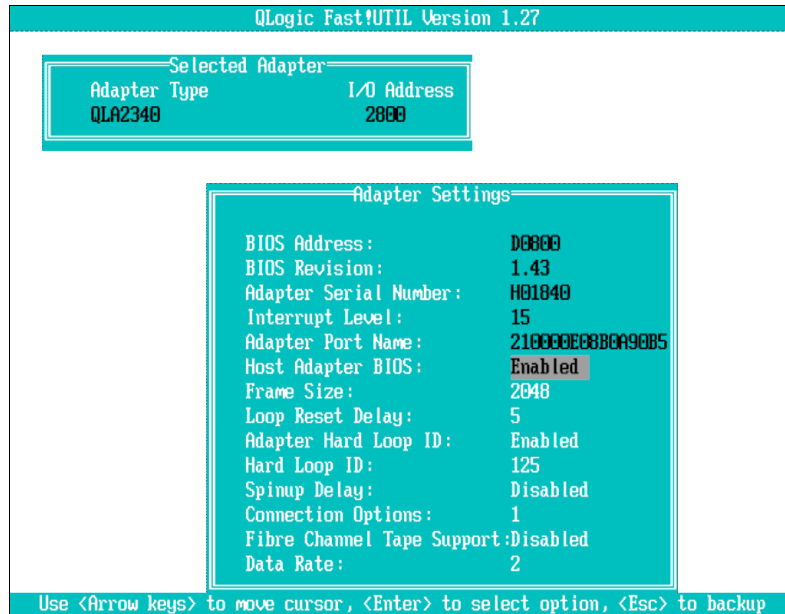


Figure 1-15 Adapter Settings menu

- From the **Configuration Settings** menu, select **Selectable Boot Settings** to get to the panel shown in Figure 1-16.

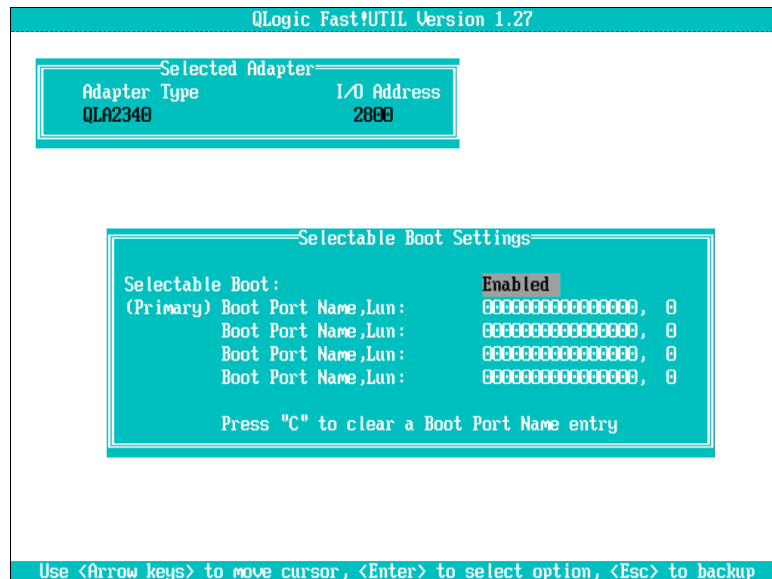


Figure 1-16 Selectable Boot Settings menu

- Change the **Selectable Boot** option to **Enabled**.

- Select **Boot Port Name, Lun** and then press Enter to get the **Select Fibre Channel Device** menu (Figure 1-17).

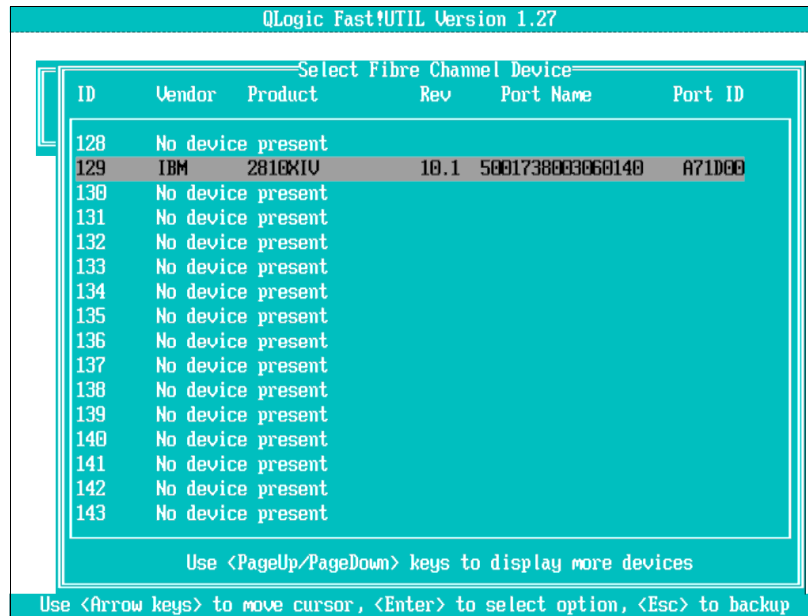


Figure 1-17 Select Fibre Channel Device menu

- Select the **IBM 2810XIV** device, and press Enter to display the **Select LUN** menu shown in Figure 1-18.

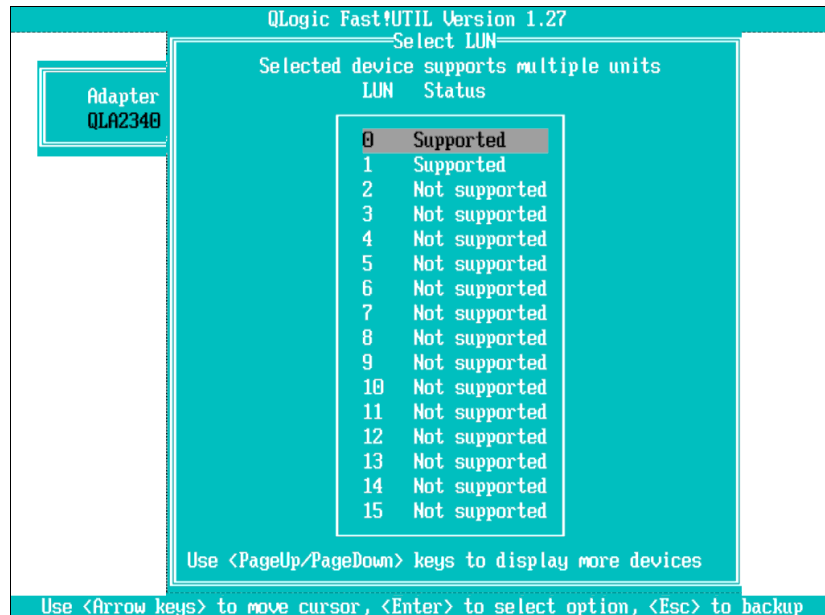


Figure 1-18 Select LUN menu

- Select the boot LUN (in this example, **LUN 0**). You are taken back to the **Selectable Boot Setting** menu, and the boot port with the boot LUN is displayed as shown in Figure 1-19.

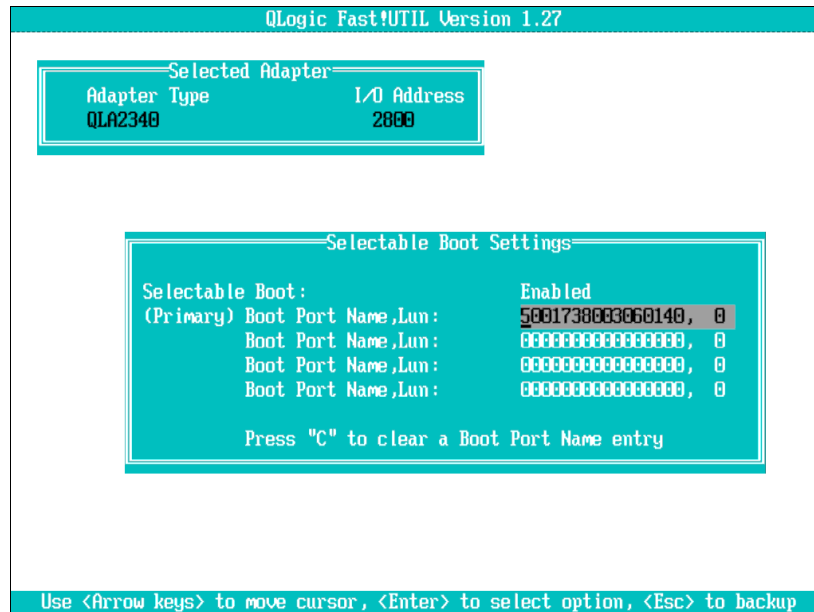


Figure 1-19 Boot port selected

- Repeat the steps 8-10 to add more controllers. Any additional controllers must be zoned so that they point to the same boot LUN.
- After all the controllers are added, press Esc to exit the Configuration Setting panel. Press Esc again to get the **Save changes** option as shown in Figure 1-20.

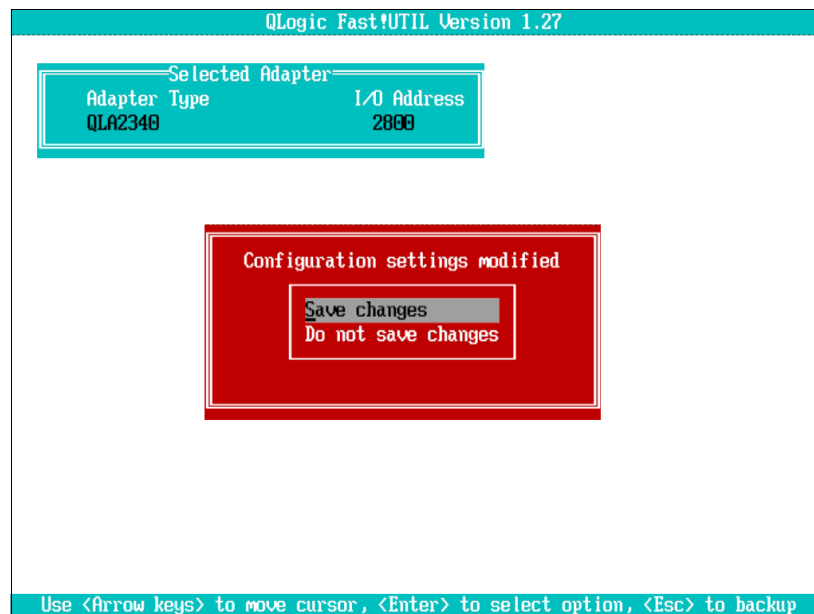


Figure 1-20 Save changes

- Select **Save changes** to go back to the Fast!UTIL option panel. From there, select **Exit Fast!UTIL**.

14. The **Exit Fast!UTIL** menu is displayed as shown in Figure 1-21. Select **Reboot System** to reboot from the newly configured SAN drive.

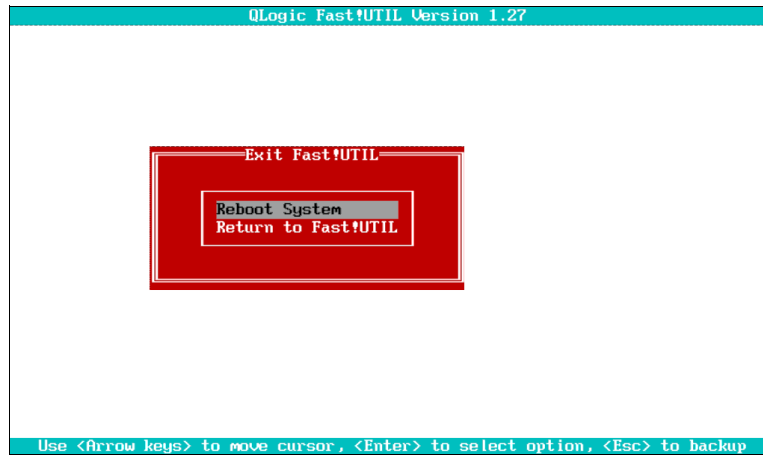


Figure 1-21 Exit Fast!UTIL

Important: Depending on your operating system and multipath drivers, you might need to configure multiple ports as “boot from SAN” ports. For more information, see your operating system documentation.

1.3 iSCSI connectivity

This section focuses on iSCSI connectivity as it applies to the XIV Storage System in general. For operating system-specific information, see the relevant section in the corresponding chapter of this book.

Currently, iSCSI hosts other than AIX are supported by using the software iSCSI initiator. For more information about iSCSI software initiator support, see the SSIC website at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

Table 1-3 shows supported operating systems.

Table 1-3 iSCSI supported operating systems

Operating System	Initiator
AIX	AIX iSCSI software initiator iSCSI HBA FC573B
Linux (CentOS)	Linux iSCSI software initiator Open iSCSI software initiator
Linux (RedHat)	RedHat iSCSI software initiator
Linux SuSE	Novell iSCSI software initiator
Solaris	Oracle / SUN iSCSI software initiator
Windows	Microsoft iSCSI software initiator

1.3.1 Preparation steps

Before you can attach an iSCSI host to the XIV Storage System, you must complete the following procedures. These general procedures pertain to all hosts. However, you must also review any procedures that pertain to your specific hardware and operating system.

1. Connect the host to the XIV over iSCSI using a standard Ethernet port on the host server. Dedicate the port that you choose to iSCSI storage traffic only. This port must also be a minimum of 1 Gbps capable. This port requires an IP address, subnet mask, and gateway. Also, review any documentation that came with your operating system about iSCSI to ensure that any additional conditions are met.
2. Check the LUN limitations for your host operating system. Verify that enough adapters are installed on the host server to manage the total number of LUNs that you want to attach.
3. Check the optimum number of paths that must be defined, which helps determine the number of physical connections that must be made.
4. Install the latest supported adapter firmware and driver. If the latest version was not shipped with your operating system, download it.
5. Maximum transmission unit (MTU) configuration is required if your network supports an MTU that is larger than the default (1500 bytes). Anything larger is known as a *jumbo frame*. Specify the largest possible MTU. Generally, use 4500 bytes (which is the default value on XIV) if supported by your switches and routers.
6. Any device that uses iSCSI requires an IQN and an attached host. The IQN uniquely identifies iSCSI devices. The IQN for the XIV Storage System is configured when the system is delivered and must not be changed. Contact IBM technical support if a change is required.

The XIV Storage System name in this example is `iqn.2005-10.com.xivstorage:000035`.

1.3.2 iSCSI configurations

Several configurations are technically possible. They vary in terms of their cost and the degree of flexibility, performance, and reliability that they provide.

In the XIV Storage System, each iSCSI port is defined with its own IP address.

Restriction: Link aggregation is not supported. Ports cannot be bonded

Redundant configurations

A redundant configuration is illustrated in Figure 1-22 on page 30.

This configuration has the following characteristics:

- ▶ Each host is equipped with dual Ethernet interfaces. Each interface (or interface port) is connected to one of two Ethernet switches.
- ▶ Each of the Ethernet switches has a connection to a separate iSCSI port. The connection is to Interface Modules 7-9 on a 2nd Generation XIV (model A14), and modules 4-9 on an XIV Gen 3 (model 114).

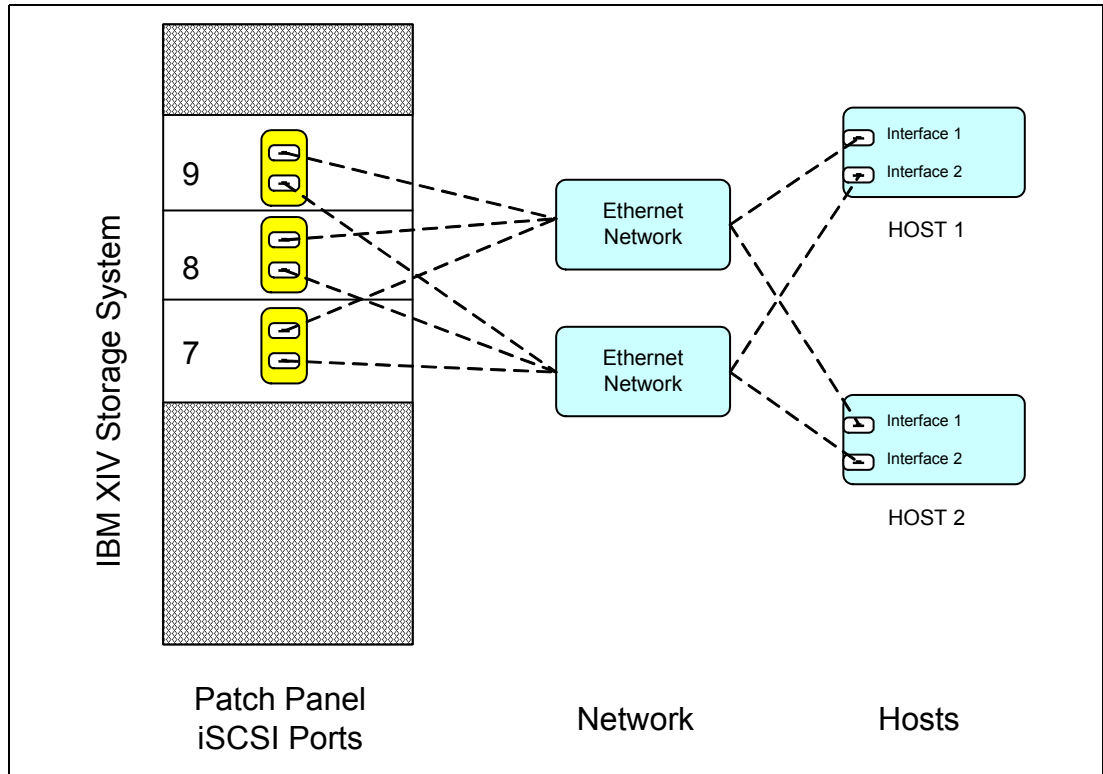


Figure 1-22 iSCSI redundant configuration using 2nd Generation XIV model A14 hardware

This configuration has no single point of failure:

- ▶ If a module fails, each host remains connected to at least one other module. How many depends on the host configuration, but it is typically one or two other modules.
- ▶ If an Ethernet switch fails, each host remains connected to at least one other module. How many depends on the host configuration, but is typically one or two other modules through the second Ethernet switch.
- ▶ If a host Ethernet interface fails, the host remains connected to at least one other module. How many depends on the host configuration, but is typically one or two other modules through the second Ethernet interface.
- ▶ If a host Ethernet cable fails, the host remains connected to at least one other module. How many depends on the host configuration, but is typically one or two other modules through the second Ethernet interface.

Consideration: For the best performance, use a dedicated iSCSI network infrastructure.

Non-redundant configurations

Use non-redundant configurations only where the risks of a single point of failure are acceptable. This configuration is typically acceptable for test and development environments.

Figure 1-23 illustrates a non-redundant configuration.

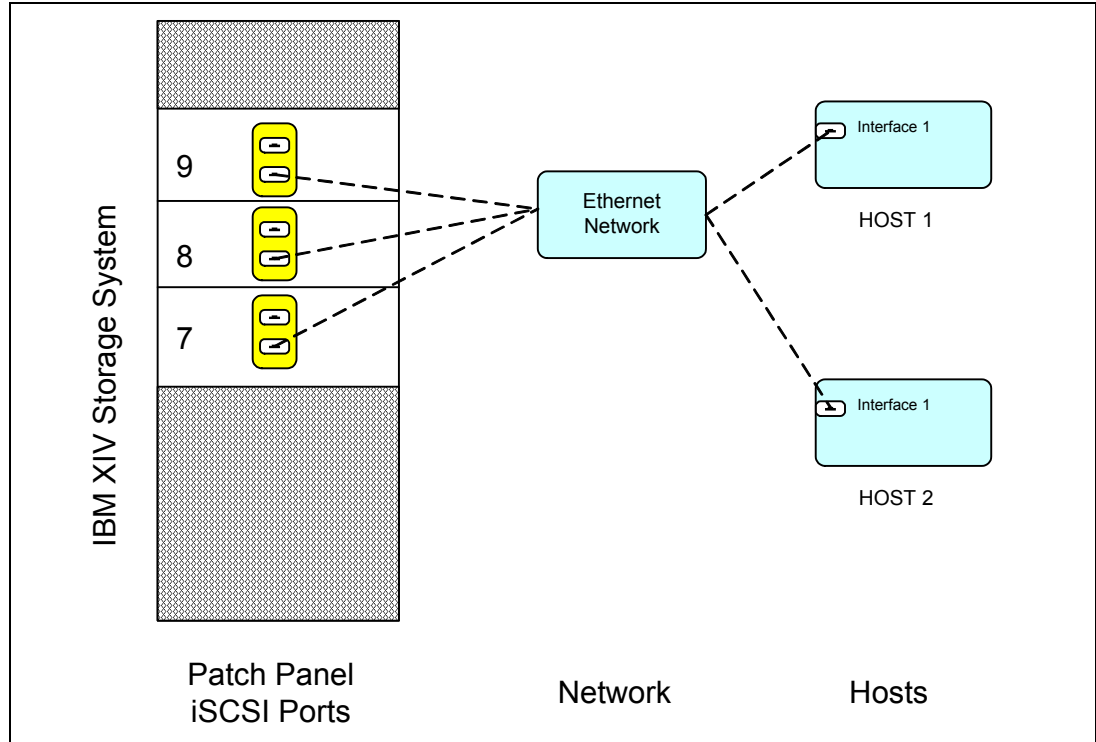


Figure 1-23 iSCSI single network switch configuration with 2nd Generation XIV model A14 hardware

Consideration: Both Figure 1-22 on page 30 and Figure 1-23 show a 2nd Generation XIV (model A14). An XIV Gen 3 has more iSCSI ports on more modules.

1.3.3 Network configuration

Disk access is susceptible to network latency. Latency can cause timeouts, delayed writes, and data loss. To get the best performance from iSCSI, place all iSCSI IP traffic on a dedicated network. Physical switches or VLANs can be used to provide a dedicated network. This network requires be a minimum of 1 Gbps, and the hosts need interfaces that are dedicated to iSCSI only. You might need to purchase more host Ethernet ports.

1.3.4 IBM XIV Storage System iSCSI setup

Initially, no iSCSI connections are configured in the XIV Storage System. The configuration process is simple, but requires more steps than an FC connection setup.

Getting the XIV iSCSI Qualified Name (IQN)

Every XIV Storage 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 XIV Storage System 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 as part of the XIV Storage System, complete the following steps:

1. From the XIV GUI, click **Systems** → **System Settings** → **System**.
2. The Settings dialog box is displayed. Select the **Parameters** tab as shown in Figure 1-24.
If you are displaying multiple XIVs from the All Systems view, you can right-click an XIV, and select **Properties** → **Parameters** to get the same information.

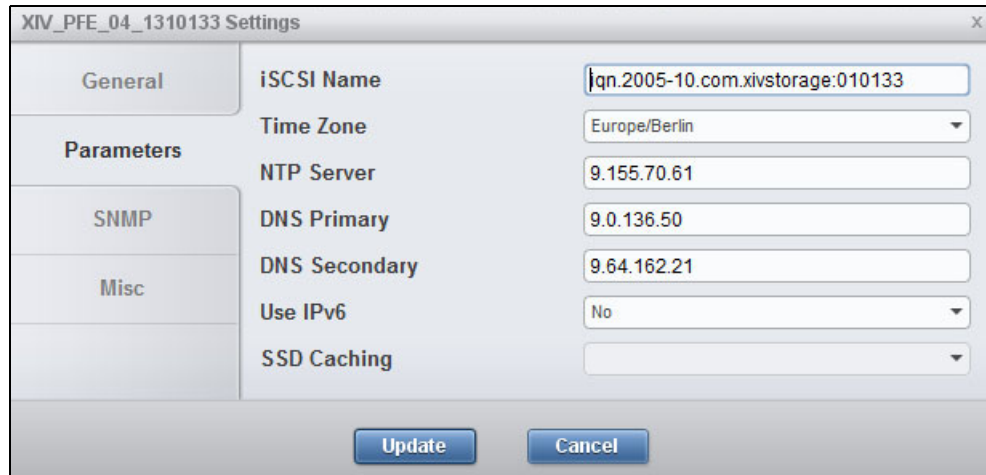


Figure 1-24 Using the XIV GUI to get the iSCSI name (IQN)

To show the same information in the XCLI, run the XCLI **config_get** command as shown in Example 1-3.

Example 1-3 Using the XCLI to get the iSCSI name (IQN)

```
XIV PFE-Gen 3-1310133>>config_get
Name                Value
dns_primary         9.64.162.21
dns_secondary       9.64.163.21
system_name         XIV PFE-Gen 3-1310133
snmp_location       Unknown
snmp_contact        Unknown
snmp_community      XIV
snmp_trap_community XIV
system_id           10133
machine_type        2810
machine_model       114
machine_serial_number 1310133
email_sender_address
email_reply_to_address
email_subject_format {severity}: {description}
iscsi_name          iqn.2005-10.com.xivstorage:010133
ntp_server           9.155.70.61
support_center_port_type Management
isns_server
```

iSCSI XIV port configuration by using the GUI

To set up the iSCSI port by using the GUI, complete these steps:

1. Log on to the XIV GUI.
2. Select the XIV Storage System to be configured.
3. Point to the **Hosts and Clusters** icon and select **iSCSI Connectivity** as shown in Figure 1-25.

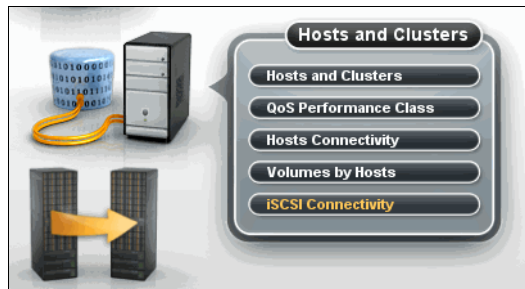


Figure 1-25 iSCSI Connectivity menu option

4. The **iSCSI Connectivity** window opens. Click the **Define** icon at the top of the window (Figure 1-26) to open the Define IP interface window.

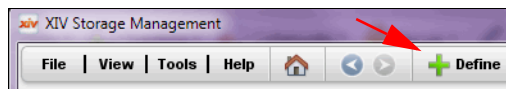


Figure 1-26 iSCSI Define interface icon

5. Enter the following information (Figure 1-27):
 - Name: This is a name that you define for this interface.
 - Address, netmask, and gateway: These are the standard IP address details.
 - MTU: The default is 4500. All devices in a network must use the same MTU. If in doubt, set MTU to 1500, because 1500 is the default value for Gigabit Ethernet. Performance might be affected if the MTU is set incorrectly.
 - Module: Select the module to configure.
 - Port number: Select the port to configure.

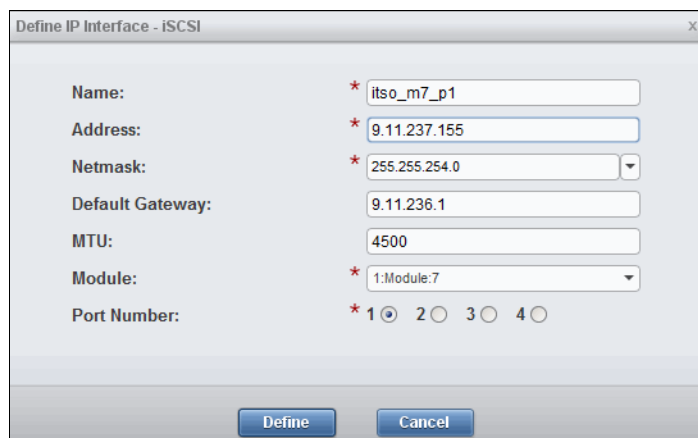


Figure 1-27 Define IP Interface - iSCSI setup window on a XIV Gen 3 (model 114)

Tip: Figure 1-27 was created by using an XIV Gen 3 that has four iSCSI ports per interface module (except module 4, which has only two). A 2nd Generation XIV has only two ports per module.

6. Click **Define** to complete the IP interface and iSCSI setup.

Tip: If the MTU being used by the XIV 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 to specify 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 1472 10.1.1.1` command sends a 1500-byte frame to the 10.1.1.1 IP address (1472 bytes of payload and 28 bytes of headers). If this command succeeds, you can use an MTU of 1500 in the XIV GUI or XCLI.

iSCSI XIV port configuration by using the XCLI

To configure iSCSI ports by using the XCLI session tool, issue the `ipinterface_create` command as shown in Example 1-4.

Example 1-4 iSCSI setup (XCLI)

```
>> ipinterface_create ipinterface="Test" address=10.0.0.10 netmask=255.255.255.0
module=1:Module:5 ports="1" gateway=10.0.0.1 mtu=4500
```

1.3.5 Identifying iSCSI ports

iSCSI ports can be easily identified and configured in the XIV Storage System. Use either the GUI or an XCLI command to display current settings.

Viewing iSCSI configuration by using the GUI

To view the iSCSI configuration, complete the following steps:

1. Log on to the XIV GUI
2. Select the XIV Storage System to be configured
3. Point to the **Hosts and Clusters** icon and select **iSCSI connectivity** (Figure 1-25 on page 33).
4. The iSCSI connectivity panel is displayed as shown in Figure 1-28. Right-click the port and select **Edit** to make the changes.

Name	Address	Netmask	Gateway	MTU	Module	Ports
M6_P4	9.155.56.42	255.255.255.0	9.155.56.1	1500	1:Module:6	4
M5_P4	9.155.56.41	255.255.255.0	9.155.56.1	1500	1:Module:5	4

Figure 1-28 iSCSI connectivity window

In this example, only four iSCSI ports are configured. Non-configured ports are not displayed.

View iSCSI configuration by using the XCLI

The `ipinterface_list` command that is illustrated in Example 1-5 can be used to display configured network ports only. This example shows a Gen 3 XIV (model 114).

Example 1-5 Listing iSCSI ports with the `ipinterface_list` command

```
XIV PFE-Gen 3-1310133>>ipinterface_list
Name      Type      IP Address  Network Mask  Default Gateway  MTU  Module  Ports
M5_P4    iSCSI     9.155.56.41 255.255.255.0 9.155.56.1      1500 1:Module:5 4
M6_P4    iSCSI     9.155.56.42 255.255.255.0 9.155.56.1      1500 1:Module:6 4
management Management 9.155.56.38 255.255.255.0 9.155.56.1      1500 1:Module:1
VPN      VPN       255.0.0.0    1500 1:Module:1
VPN      VPN       255.0.0.0    1500 1:Module:3
management Management 9.155.56.39 255.255.255.0 9.155.56.1      1500 1:Module:2
management Management 9.155.56.40 255.255.255.0 9.155.56.1      1500 1:Module:3
```

The rows might be in a different order each time you run this command. To see a complete list of IP interfaces, use the command `ipinterface_list_ports`.

1.3.6 iSCSI and CHAP authentication

Starting with microcode level 10.2, the IBM XIV Storage System supports industry-standard unidirectional iSCSI Challenge Handshake Authentication Protocol (CHAP). The iSCSI target of the IBM XIV Storage System can validate the identity of the iSCSI Initiator that attempts to log on to the system.

The CHAP configuration in the IBM XIV Storage System is defined on a per-host basis. There are no global configurations for CHAP that affect all the hosts that are connected to the system.

Tip: By default, hosts are defined without CHAP authentication.

For the iSCSI initiator to log in with CHAP, both the `iscsi_chap_name` and `iscsi_chap_secret` parameters must be set. After both of these parameters are set, the host 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 apply the next time that the host is connected.

Configuring CHAP

CHAP can be configured either through the XIV GUI when you add the host.

Figure 1-29 shows an example of setting CHAP name and secret.

The screenshot shows a window titled "Add Host" with a close button (X) in the top right corner. The window contains the following fields and values:

- System / Cluster:** Chapcluster (XIV_PFE_04_1310133)
- Name:** Chap (with a red asterisk indicating a required field)
- Type:** default
- CHAP Name:** Monkey
- CHAP Secret:** Chimpanzeebanana

At the bottom of the window, there are two buttons: "Add" and "Cancel".

Figure 1-29 Adding CHAP Name and CHAP Secret

Alternatively, the following XCLI commands can be used to configure CHAP:

- ▶ If you are defining a new host, use the following XCLI command to add CHAP parameters:

```
host_define host=[hostName] iscsi_chap_name=[chapName]
iscsi_chap_secret=[chapSecret]
```

- ▶ If the host already exists, use the following XCLI command to add CHAP parameters:

```
host_update host=[hostName] iscsi_chap_name=[chapName]
iscsi_chap_secret=[chapSecret]
```

- ▶ If you no longer want to use CHAP authentication, use the following XCLI command to clear the CHAP parameters:

```
host_update host=[hostName] iscsi_cha_name= iscsi_chap_secret=
```

1.3.7 iSCSI boot from XIV LUN

At the time of writing, you cannot boot through iSCSI, even if an iSCSI HBA is used.

1.4 Logical configuration for host connectivity

This section shows the tasks that are required to define a volume (LUN) and assign it to a host. The following sequence of steps is generic and intended to be operating system independent. The exact procedures for your server and operating system might differ somewhat.

1. Gather information about hosts and storage systems (WWPN or IQN).
2. Create SAN zoning for the FC connections.
3. Create a storage pool.
4. Create a volume within the storage pool.
5. Define a host.
6. Add ports to the host (FC or iSCSI).
7. Map the volume to the host.
8. Check host connectivity at the XIV Storage System.
9. Complete any operating-system-specific tasks.
10. If the server is going to SAN boot, install the operating system.
11. Install multipath drivers if required. For information about installing multi-path drivers, see the appropriate section from the host-specific chapters of this book.
12. Reboot the host server or scan new disks.

Important: For the host system to effectively see and use the LUN, more and operating system-specific configuration tasks are required. The tasks are described in operating-system-specific chapters of this book.

1.4.1 Host configuration preparation

The environment shown in Figure 1-30 on page 38 is used to illustrate the configuration tasks. The example uses a 2nd Generation XIV. There are two hosts: One host uses FC connectivity and the other host uses iSCSI. The diagram also shows the unique names of components that are used in the configuration steps.

The following assumptions are made for the scenario that is shown in Figure 1-30 on page 38:

- ▶ One host is set up with an FC connection. It has two HBAs and a multi-path driver installed.
- ▶ One host is set up with an iSCSI connection. It has one connection, and has the software initiator loaded and configured.

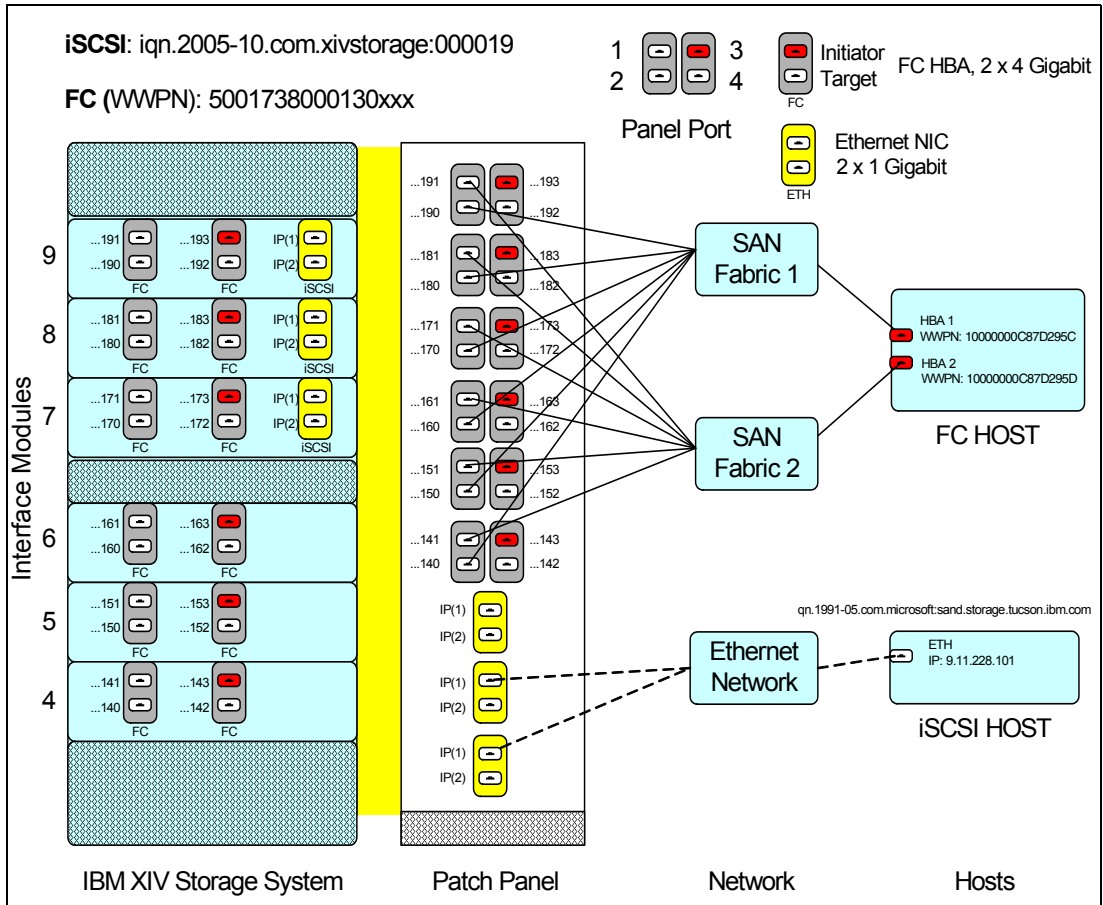


Figure 1-30 Overview of base host connectivity setup

Hardware information

Write down the component names and IDs because doing so saves time during the implementation. An example is illustrated in Table 1-4 for the example scenario.

Table 1-4 Required component information

Component	FC environment	iSCSI environment
IBM XIV FC HBAs	WWPN: 5001738000130nnn nnn for Fabric1: 140, 150, 160, 170, 180, and 190 nnn for Fabric2: 142, 152, 162, 172, 182, and 192	N/A
Host HBAs	HBA1 WWPN: 21000024FF24A426 HBA2 WWPN: 21000024FF24A427	N/A
IBM XIV iSCSI IPs	N/A	Module7 Port1: 9.11.237.155 Module8 Port1: 9.11.237.156
IBM XIV iSCSI IQN (do not change)	N/A	iqn.2005-10.com.xivstorage:000019

Component	FC environment	iSCSI environment
Host IPs	N/A	9.11.228.101
Host iSCSI IQN	N/A	iqn.1991-05.com.microsoft:sand.storage.tucson.ibm.com
OS Type	Default	Default

Remember: The OS Type is *default* for all hosts except HP-UX and IBM z/VM®.

FC host-specific tasks

Configure the SAN (Fabrics 1 and 2) and power on the host server first. These actions populate the XIV Storage System with a list of WWPNs from the host. This method is preferable because it is less prone to error when adding the ports in subsequent procedures.

For more information about configuring zoning, see your FC switch manual. The following is an example of what the zoning details might look like for a typical server HBA zone. If you are using SAN Volume Controller as a host, there are more requirements that are not addressed here.

Fabric 1 HBA 1 zone

Log on to the Fabric 1 SAN switch and create a host zone:

```
zone: prime_sand_1
    prime_4_1; prime_6_1; prime_8_1; sand_1
```

Fabric 2 HBA 2 zone

Log on to the Fabric 2 SAN switch and create a host zone:

```
zone: prime_sand_2
    prime_5_3; prime_7_3; prime_9_3; sand_2
```

In the previous examples, the following aliases are used:

- ▶ sand is the name of the server, sand_1 is the name of HBA1, and sand_2 is the name of HBA2.
- ▶ prime_sand_1 is the zone name of fabric 1, and prime_sand_2 is the zone name of fabric 2.
- ▶ The other names are the aliases for the XIV patch panel ports.

iSCSI host-specific tasks

For iSCSI connectivity, ensure that any configurations such as VLAN membership or port configuration are completed so the hosts and the XIV can communicate over IP.

1.4.2 Assigning LUNs to a host by using the GUI

There are a number of steps that are required to define a new host and assign LUNs to it. The volumes must have been created in a Storage System.

Defining a host

To define a host, complete these steps:

1. In the XIV Storage System main GUI window, mouse over the **Hosts and Clusters** icon and select **Hosts and Clusters** (Figure 1-31).



Figure 1-31 Hosts and Clusters menu

2. The Hosts window is displayed showing a list of hosts (if any) that are already defined. To add a host or cluster, click either the **Add Host** or **Add Cluster** in the menu bar (Figure 1-32). The difference between the two is that **Add Host** is for a single host that is assigned a LUN or multiple LUNs. **Add Cluster** is for a group of hosts that shares a LUN or multiple LUNs.



Figure 1-32 Add new host

3. The **Add Host** dialog is displayed as shown in Figure 1-33. Enter a name for the host.

Figure 1-33 Add Host details

4. To add a server to a cluster, select a cluster name. If a cluster definition was created in the step 2, it is available in the cluster menu. In this example, no cluster was created, so **None** is selected.
5. Select the **Type**. In this example, the type is **default**. If you have an HP-UX or z/VM host, you must change the Type to match your host type. For all other hosts (such as AIX, Linux, Solaris, VMWare, and Windows), the **default** option is correct.
6. Repeat steps 2 through 5 to create more hosts if needed.

- Host access to LUNs is granted depending on the host adapter ID. For an FC connection, the host adapter ID is the FC HBA WWPN. For an iSCSI connection, the host adapter ID is the host IQN. To add a WWPN or IQN to a host definition, right-click the host and select **Add Port** from the menu (Figure 1-34).

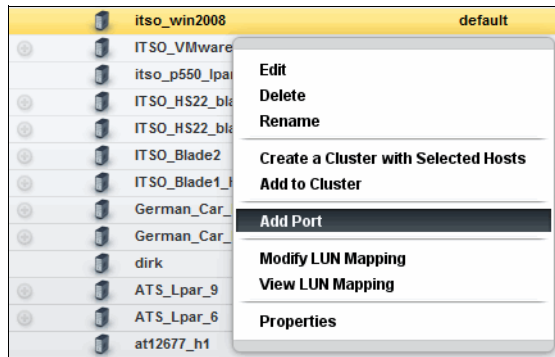


Figure 1-34 Add port to host definition (GUI)

- The Add Port window is displayed as shown in Figure 1-35. Select port type FC or iSCSI. In this example, an FC host is defined. Add the WWPN for HBA1 as listed in Table 1-4 on page 38. If the host is correctly connected and has done a port login to the SAN switch at least once, the WWPN is shown in the menu. Otherwise, you must manually enter the WWPN. Adding ports from the menu is preferable because it is less prone to error. However, if hosts are not yet connected to the SAN or zoned, manually adding the WWPNs is the only option.

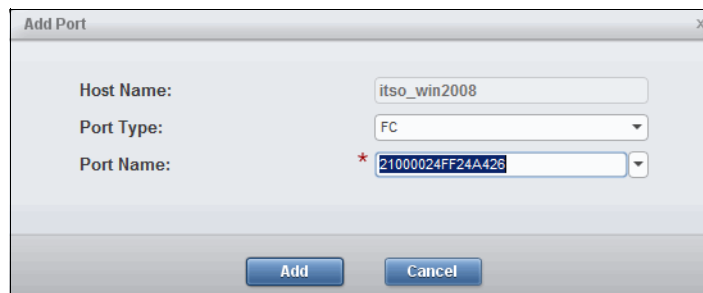


Figure 1-35 Add FC port WWPN (GUI)

Repeat steps 7 and 8 to add the second HBA WWPN. Ports can be added in any order.

- To add an iSCSI host, specify the port type as iSCSI and enter the IQN of the HBA as the iSCSI Name (Figure 1-36).



Figure 1-36 Add iSCSI port (GUI)

10. The host is displayed with its ports in the Hosts window as shown in Figure 1-37.

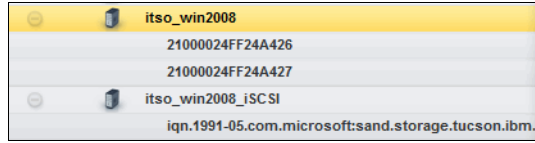


Figure 1-37 List of hosts and ports

In this example, the hosts *its_o_win2008* and *its_o_win2008_iscsi* are in fact the same physical host. However, they are entered as separate entities so that when mapping LUNs, the FC, and iSCSI protocols do not access the same LUNs.

Mapping LUNs to a host

The final configuration step is to map LUNs to the host by completing the following steps:

1. In the **Hosts and Clusters** configuration window, right-click the host to which the volume is to be mapped and select **Modify LUN Mappings** (Figure 1-38).

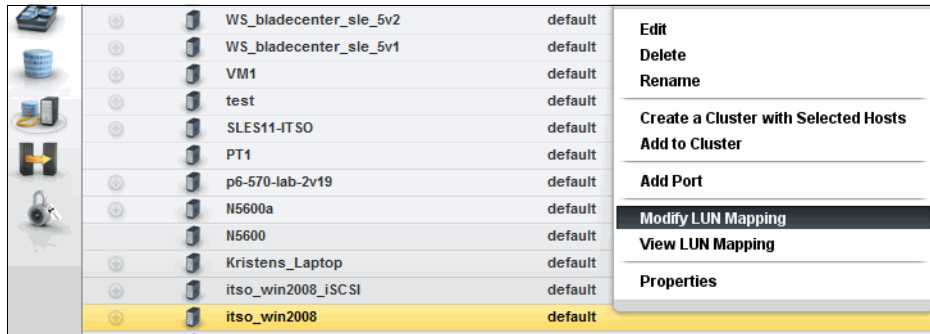


Figure 1-38 Mapping LUN to host

2. The Volume to LUN Mapping window opens as shown in Figure 1-39. Select an available volume from the left window. The GUI suggests a LUN ID to which to map the volume. However, this ID can be changed to meet your requirements. Click **Map** and the volume is assigned immediately.

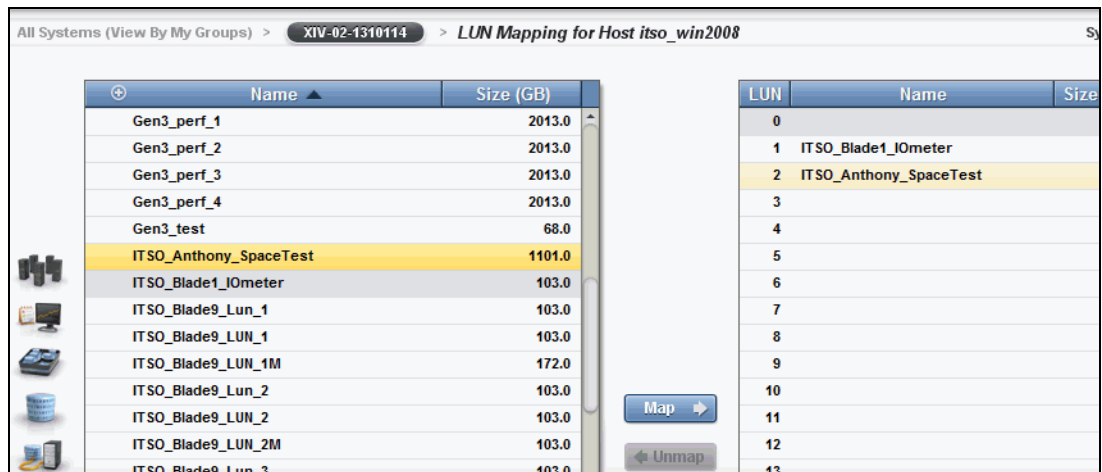


Figure 1-39 Mapping FC volume to FC host

There is no difference between mapping a volume to an FC or iSCSI host in the XIV GUI Volume to LUN Mapping view.

3. Power up the host server and check connectivity. The XIV Storage System has a real-time connectivity status overview. Select **Hosts Connectivity** from the **Hosts and Clusters** menu to access the connectivity status (Figure 1-40).

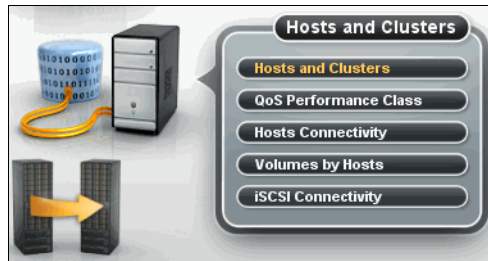


Figure 1-40 Selecting Hosts and Clusters

4. Make sure that the LUN is displayed as connected in Host Connectivity window (Figure 1-41).

Name	1:4	1:5	1:6	1:7	1:8	1:9
PT1						
p6-570-lab-2v19						
N5600a						
N5600						
Kristens_Laptop						
itso_win2008_iSCSI						
itso_win2008						
21000024FF24A427	✓1	✓2		✓2		✓2
21000024FF24A426	✓1		✓1		✓1	

Figure 1-41 Host connectivity matrix (GUI)

At this stage, there might be operating-system-dependent steps that must be performed. These steps are described in the operating system chapters.

1.4.3 Assigning LUNs to a host by using the XCLI

There are a number of steps that are required to define a new host and assign LUNs to it. Volumes must have already been created in a Storage Pool.

Defining a new host

To use the XCLI to prepare for a new host, complete these steps:

1. Create a host definition for your FC and iSCSI hosts by using the **host_define** command as shown in Example 1-6.

Example 1-6 Creating host definition (XCLI)

```
>>host_define host=itso_win2008
Command executed successfully.
```

```
>>host_define host=itso_win2008_iscsi
Command executed successfully.
```

2. Host access to LUNs is granted depending on the host adapter ID. For an FC connection, the host adapter ID is the FC HBA WWPN. For an iSCSI connection, the host adapter ID is the IQN of the host.

In Example 1-7, the WWPN of the FC host for HBA1 and HBA2 is added with the `host_add_port` command by specifying a `fcaddress`.

Example 1-7 Creating FC port and adding it to host definition

```
>> host_add_port host=itso_win2008 fcaddress=21000024FF24A426
Command executed successfully.
```

```
>> host_add_port host=itso_win2008 fcaddress=21000024FF24A427
Command executed successfully.
```

In Example 1-8, the IQN of the iSCSI host is added. This is the same `host_add_port` command, but with the `iscsi_name` parameter.

Example 1-8 Creating iSCSI port and adding it to the host definition

```
>> host_add_port host=itso_win2008 iscsi
iscsi_name=iqn.1991-05.com.microsoft:sand.storage.tucson.ibm.com
Command executed successfully
```

Mapping LUNs to a host

To map the LUNs, complete these steps:

1. Map the LUNs to the host definition. For a cluster, the volumes are mapped to the cluster host definition. There is no difference between FC and iSCSI mapping to a host. Both commands are shown in Example 1-9.

Example 1-9 Mapping volumes to hosts (XCLI)

```
>> map_vol host=itso_win2008 vol=itso_win2008_vol1 lun=1
Command executed successfully.
```

```
>> map_vol host=itso_win2008 vol=itso_win2008_vol2 lun=2
Command executed successfully.
```

```
>> map_vol host=itso_win2008 iscsi vol=itso_win2008_vol3 lun=1
Command executed successfully.
```

2. Power up the server and check the host connectivity status from the XIV Storage System point of view. Example 1-10 shows the output for both hosts.

Example 1-10 Checking host connectivity (XCLI)

```
XIV-02-1310114>>host_connectivity_list host=itso_win2008
Host          Host Port          Module          Local FC port  Local iSCSI port
itso_win2008  21000024FF24A427  1:Module:5     1:FC_Port:5:2
itso_win2008  21000024FF24A427  1:Module:7     1:FC_Port:7:2
itso_win2008  21000024FF24A427  1:Module:9     1:FC_Port:9:2
itso_win2008  21000024FF24A426  1:Module:4     1:FC_Port:4:1
itso_win2008  21000024FF24A426  1:Module:6     1:FC_Port:6:1
itso_win2008  21000024FF24A426  1:Module:8     1:FC_Port:8:1

>> host_connectivity_list host=itso_win2008_iscsi
Host          Host Port          Module          Local FC port  Type
itso_win2008_iscsi  iqn.1991-05.com.microsoft:sand.storage.tucson.ibm.com  1:Module:8     iSCSI
itso_win2008_iscsi  iqn.1991-05.com.microsoft:sand.storage.tucson.ibm.com  1:Module:7     iSCSI
```

In Example 1-10 on page 44, there are two paths per host FC HBA and two paths for the single Ethernet port that was configured.

At this stage, there might be operating system-dependent steps that must be performed, these steps are described in the operating system chapters.

1.5 Troubleshooting

Troubleshooting connectivity problems can be difficult. However, the XIV Storage System does have built-in troubleshooting tools. Table 1-5 lists some of the built-in tools. For more information, see the XCLI manual, which can be downloaded from the IBM XIV Storage System Information Center at:

<http://publib.boulder.ibm.com/infocenter/ibmxiv/r2/index.jsp>

Table 1-5 XIV in-built tools

Tool	Description
fc_connectivity_list	Discovers FC hosts and targets on the FC network
fc_port_list	Lists all FC ports, their configuration, and their status
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
host_connectivity_list	Lists FC and iSCSI connectivity to hosts



XIV and Windows host connectivity

This chapter addresses the specific considerations for attaching XIV to various Microsoft host servers.

Important: The procedures and instructions that are provided here are based on code that was available at the time of writing. For the latest support information and instructions, see the System Storage Interoperation Center (SSIC) at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

In addition, Host Attachment Kit and related publications can be downloaded from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

This chapter includes the following sections:

- ▶ Attaching a Microsoft Windows 2008 R2 host to XIV
- ▶ Attaching a Microsoft Windows 2008 R2 cluster to XIV
- ▶ Attaching a Microsoft Hyper-V Server 2008 R2 to XIV
- ▶ Microsoft System Center Virtual Machine Manager 2012 Storage Automation

2.1 Attaching a Microsoft Windows 2008 R2 host to XIV

This section highlights specific instructions for Fibre Channel (FC) and Internet Small Computer System Interface (iSCSI) connections. All the information here relates only to Windows Server 2008 R2 unless otherwise specified.

Important: The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, ALWAYS see the SSIC at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

The Host Attachment Kit and related publications can be downloaded from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

2.1.1 Prerequisites

To successfully attach a Windows host to XIV and access storage, a number of prerequisites must be met. The following is a generic list. Your environment might have extra requirements.

- ▶ Complete the cabling.
- ▶ Complete the zoning.
- ▶ Install Service Pack 1 to Windows 2008 R2.
- ▶ Install hot fix KB2468345. Otherwise, your server will not be able to reboot.
- ▶ Create volumes to be assigned to the host.

Supported versions of Windows

At the time of writing, the following versions of Windows (including cluster configurations) are supported:

- ▶ Windows Server 2008 R2 and later (x64)
- ▶ Windows Server 2008 SP1 and later (x86, x64)
- ▶ Windows Server 2003 R2 SP2 and later (x86, x64)
- ▶ Windows Server 2003 SP2 and later (x86, x64)

Supported FC HBAs

Supported FC HBAs are available from Brocade, Emulex, IBM, and QLogic. Further details about driver versions are available from the SSIC website at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

Unless otherwise noted in SSIC, use any supported driver and firmware by the HBA vendors. For best performance, install the latest firmware and drivers for the HBAs you are using.

Multipath support

Microsoft provides a multi-path framework that is called Microsoft Multipath I/O (MPIO). The driver development kit allows storage vendors to create Device Specific Modules (DSMs) for MPIO. You can use DSMs to build interoperable multi-path solutions that integrate tightly with the Microsoft Windows family of products. MPIO allows the host HBAs to establish multiple sessions with the same target LUN, but present them to Windows as a single LUN. The Windows MPIO driver enables a true active/active path policy, allowing I/O over multiple paths simultaneously. Starting with Microsoft Windows 2008, the MPIO device driver is part of the operating system. Using the former XIVDSM with Windows 2008 R2 is not supported.

For more information about Microsoft MPIO, see:

<http://technet.microsoft.com/en-us/library/ee619778%28WS.10%29.aspx>

Boot from SAN support

SAN boot is supported (over FC only) in the following configurations:

- ▶ Windows 2008 R2 with MSDSM
- ▶ Windows 2008 with MSDSM
- ▶ Windows 2003 with XIVDSM

2.1.2 Windows host FC configuration

This section describes attaching to XIV over Fibre Channel. It provides detailed descriptions and installation instructions for the various software components required.

Installing HBA drivers

Windows 2008 R2 includes drivers for many HBAs. However, they probably are not the latest versions. Install the latest available driver that is supported. HBA drivers are available from the IBM, Emulex, and QLogic websites, and come with instructions.

With Windows operating systems, the queue depth settings are specified as part of the host adapter configuration. These settings can be specified through the BIOS settings or by using software that is provided by the HBA vendor.

The XIV Storage System can handle a queue depth of 1400 per FC host port, and up to 256 per volume.

Optimize your environment by evenly spreading the I/O load across all available ports. Take into account the load on a particular server, its queue depth, and the number of volumes.

Installing Multi-Path I/O (MPIO) feature

MPIO is provided as a built-in feature of Windows 2008 R2. To install it, complete the following steps:

1. Open Server Manager.
2. Select **Features Summary**, then right-click and select **Add Features**.

3. In the **Select Feature** window, select **Multi-Path I/O** as shown in Figure 2-1.

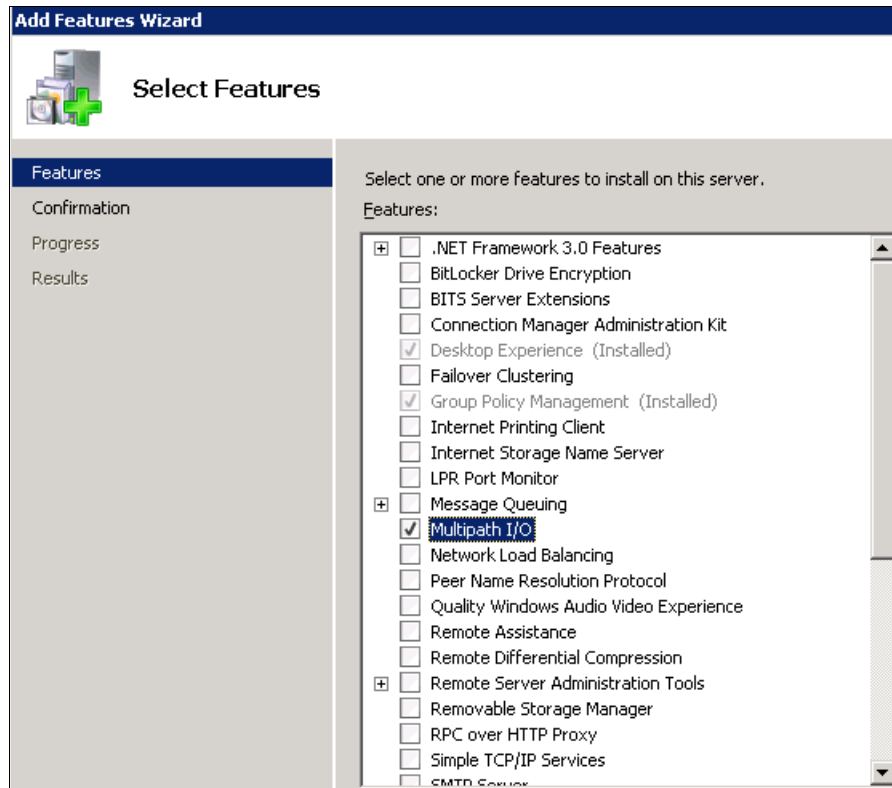


Figure 2-1 Selecting the Multipath I/O feature

4. Follow the instructions on the panel to complete the installation. This process might require a reboot.
5. Check that the driver is installed correctly by loading **Device Manager**. Verify that it now includes **Microsoft Multi-Path Bus Driver** as illustrated in Figure 2-2.

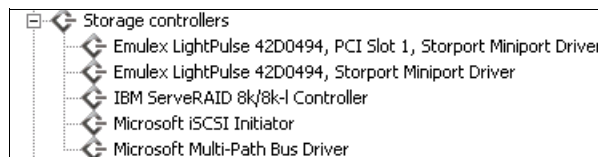


Figure 2-2 Microsoft Multi-Path Bus Driver

Windows Host Attachment Kit installation

The Windows 2008 Host Attachment Kit must be installed to gain access to XIV storage. Specific versions of the Host Attachment Kit are available for specific versions of Windows. These versions come in 32-bit and 64-bit versions. Host Attachment Kits can be downloaded from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

Portable XIV Host Attachment Kit Install and usage

The IBM XIV Host Attachment Kit is now offered in a portable format. The portable package allows you to use the Host Attachment Kit without having to install the utilities locally on the host. You can run all Host Attachment Kit utilities from a shared network drive or from a portable USB flash drive. This is the preferred method for deployment and management.

Performing a local installation

The following instructions are based on the installation performed at the time of writing. For more information, see the instructions in the *Windows Host Attachment Guide*. These instructions show the GUI installation, and can change over time. For information about command-line instructions, see the *Windows Host Attachment Guide*.

Before you install the Host Attachment Kit, remove any multipathing software that was previously installed. Failure to do so can lead to unpredictable behavior or even loss of data.

Install the XIV Host Attachment Kit (it is a mandatory prerequisite for support) by completing these steps:

1. Run the installation setup executable file (IBM_XIV_Host_Attachment_Kit_1.10.0-b1221_for_Windows-x64.exe at the time of writing). When the setup file is run, it starts the Python engine (*xpyv*). Select your language when prompted, then proceed with the installation by completing the installation wizard instructions (Figure 2-3).



Figure 2-3 Welcome to XIV Host Attachment Kit installation wizard

2. When the installation completes, click **Finish** as shown in Figure 2-4). The IBM XIV Host Attachment Kit is added to the list of installed Windows programs.

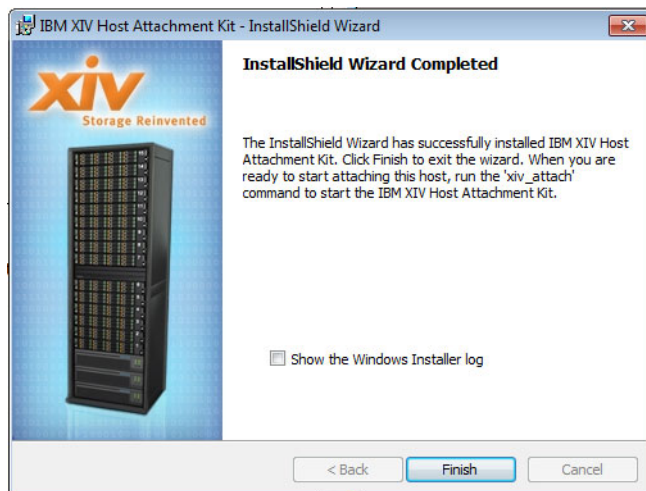


Figure 2-4 XIV Host Attachment Kit installation wizard completed

The installation directory is C:\Program Files\XIV\host_attach.

Running the xiv_attach program

Complete the procedure that is illustrated in Example 2-1 for a Fibre Channel connection.

Example 2-1 Running xiv_attach

This wizard will help you attach this host to the XIV storage system.

The wizard will now validate the host configuration for the XIV storage system. Press [ENTER] to proceed.

Please specify the connectivity type: [f]c / [i]scsi : f

Please wait while the wizard validates your existing configuration...

Verifying Previous HAK versions	OK
Verifying Disk timeout setting	OK
Verifying Built-In MPIIO feature	OK
Verifying Multipath I/O feature compatibility with XIV storage devices	OK
Verifying XIV MPIIO Load Balancing (service)	OK
Verifying XIV MPIIO Load Balancing (agent)	OK
Verifying Windows Hotfix 2460971	OK
Verifying Windows Hotfix 2522766	OK
Verifying LUNO device driver	OK

This host is already configured for the XIV storage system.

Please define zoning for this host and add its World Wide Port Names (WWPNs) to the XIV storage system:

21:00:00:24:ff:28:c1:50: [QLogic QMI2582 Fibre Channel Adapter]: QMI2582

21:00:00:24:ff:28:c1:51: [QLogic QMI2582 Fibre Channel Adapter]: QMI2582

Press [ENTER] to proceed.

Would you like to rescan for new storage devices? [default: yes]:

Please wait while rescanning for XIV storage devices...

This host is connected to the following XIV storage arrays:

Serial	Version	Host Defined	Ports Defined	Protocol	Host Name(s)
1310114	11.1.1.0	Yes	All	FC	ITSO_Blade7_Win
6000105	10.2.4.6	Yes	All	FC	ITSO_Blade7_Win
1310133	11.1.1.0	Yes	All	FC	ITSO_Blade7_Win

This host is defined on all FC-attached XIV storage arrays.

Press [ENTER] to proceed.

The IBM XIV host attachment wizard has successfully configured this host.

Press [ENTER] to exit.

Scanning for new LUNs

Before you can scan for new LUNs, your host must be created, configured, and have LUNs assigned. For more information, see Chapter 1, “Host connectivity” on page 1. The following instructions assume that these operations are complete.

To scan for LUNs, complete these steps:

1. Click **Server Manager** → **Device Manager** → **Action** → **Scan for hardware changes**.
Your XIV LUNs are displayed in the Device Manager tree under **Disk Drives** (Figure 2-5).

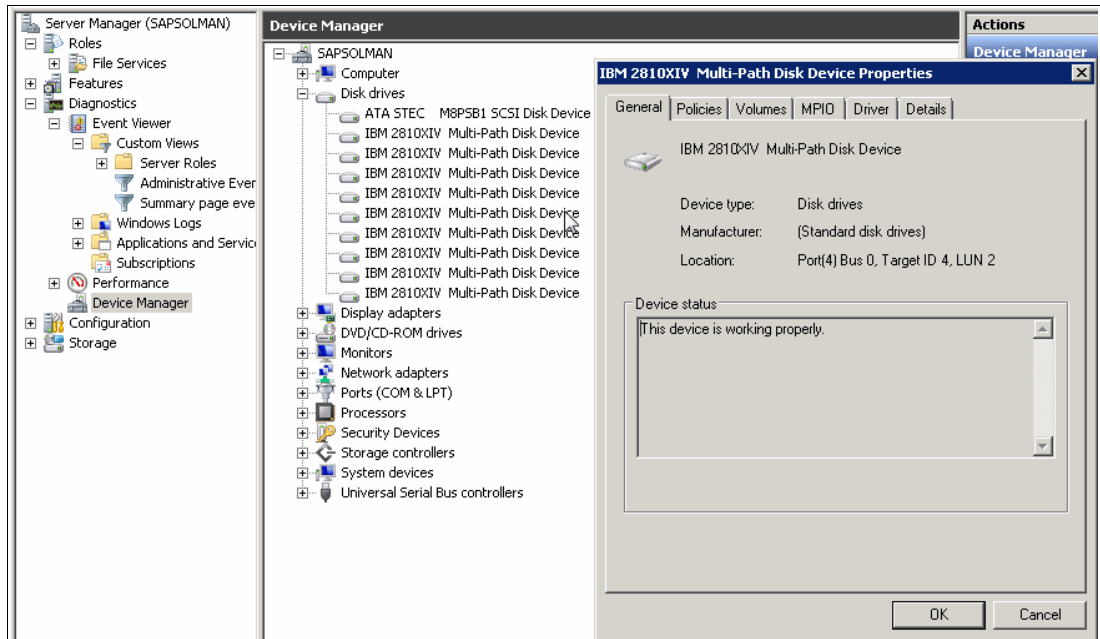


Figure 2-5 Multi-Path disk devices in Device Manager

The number of objects that are named **IBM 2810XIV SCSI Disk Device** depends on the number of LUNs mapped to the host.

2. Right-click one of the **IBM 2810XIV SCSI Device** objects and select **Properties**.
3. Click the **MPIO** tab to set the load balancing as shown in Figure 2-6.

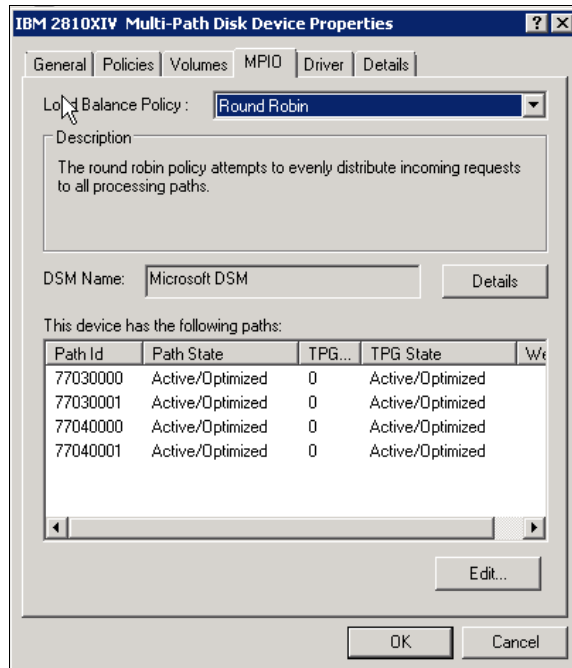


Figure 2-6 MPIO load balancing

The default setting here is **Round Robin**. Change this setting only if you are confident that another option is better suited to your environment.

Load balancing has these possible options:

- **Fail Over Only**
- **Round Robin** (default)
- **Round Robin With Subset**
- **Least Queue Depth**
- **Weighted Paths**

4. The mapped LUNs on the host can be seen under **Disk Management** as illustrated in Figure 2-7.

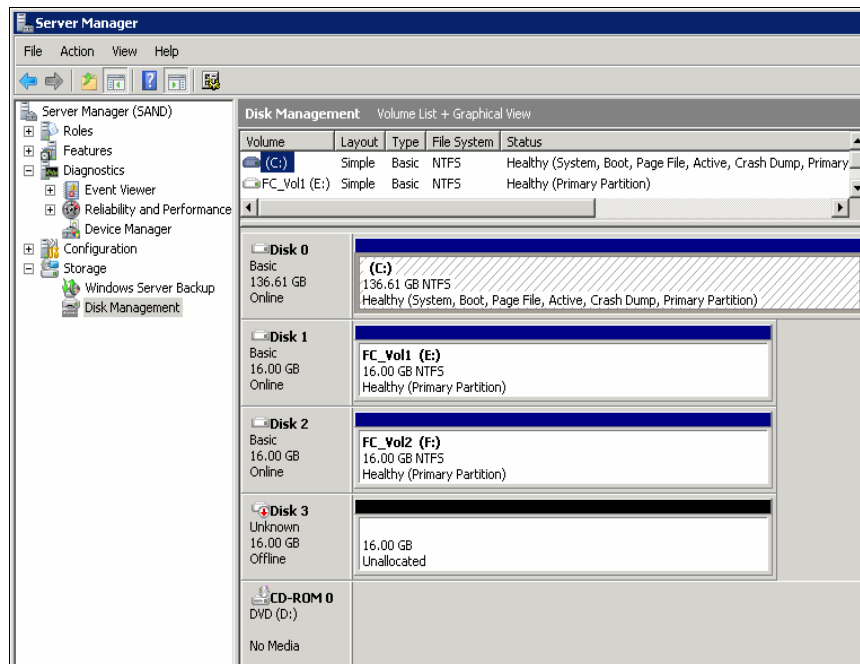


Figure 2-7 Mapped LUNs as displayed in Disk Management

2.1.3 Windows host iSCSI configuration

In Windows 2008, the iSCSI Software Initiator is part of the operating system. For Windows 2003 Servers, you must install the MS iSCSI Initiator version 2.08 or later. You must also establish the physical iSCSI connection to the XIV Storage System. For more information, see 1.3, “iSCSI connectivity” on page 28.

IBM XIV Storage System supports the iSCSI Challenge Handshake Authentication Protocol (CHAP). These examples assume that CHAP is not required. If it is, specify the settings for the required CHAP parameters on both the host and XIV sides.

Supported iSCSI HBAs

For Windows, XIV does not support hardware iSCSI HBAs. The only adapters that are supported are standard Ethernet interface adapters that use an iSCSI software initiator.

Windows multipathing feature and host attachment kit installation

To install the Windows multipathing feature, follow the procedure given in “Installing Multi-Path I/O (MPIO) feature” on page 49.

To install the Windows Host Attachment Kit, use the procedure explained under “Windows Host Attachment Kit installation” on page 50.

Running the xiv_attach program

Run the `xiv_attach` program as shown in Example 2-2.

Example 2-2 Using the XIV Host Attachment Wizard to attach to XIV over iSCSI

```
-----
Welcome to the IBM XIV host attachment wizard, version 1.10.0.
This wizard will help you attach this host to the XIV storage system.

The wizard will now validate the host configuration for the XIV storage system.
Press [ENTER] to proceed.
-----
Please choose a connectivity type, [f]c / [i]scsi : i
-----
Please wait while the wizard validates your existing configuration...
Verifying Previous HAK versions                                OK
Verifying Disk timeout setting                                OK
Verifying iSCSI service                                        NOT OK
Verifying Built-In MPIIO feature                              OK
Verifying Multipath I/O feature compatibility with XIV storage devices OK
Verifying XIV MPIIO Load Balancing (service)                 OK
Verifying XIV MPIIO Load Balancing (agent)                   OK
Verifying Windows Hotfix 2460971                             OK
Verifying Windows Hotfix 2522766                             OK
Verifying LUNO device driver                                  OK
-----
The wizard needs to configure this host for the XIV storage system.
Do you want to proceed? [default: yes ]:
Please wait while the host is being configured...
-----
Configuring Previous HAK versions                                OK
Configuring Disk timeout setting                                OK
Configuring iSCSI service                                        OK
Configuring Built-In MPIIO feature                              OK
Configuring Multipath I/O feature compatibility with XIV storage devices OK
Configuring XIV MPIIO Load Balancing (service)                 OK
Configuring XIV MPIIO Load Balancing (agent)                   OK
Configuring Windows Hotfix 2460971                             OK
Configuring Windows Hotfix 2522766                             OK
Configuring LUNO device driver                                  OK
The host is now configured for the XIV storage system
-----
Would you like to discover a new iSCSI target? [default: yes ]:
Please enter an XIV iSCSI discovery address (iSCSI interface): 9.155.51.72
Is this host defined in the XIV system to use CHAP? [default: no ]:
Would you like to discover a new iSCSI target? [default: yes ]:
Enter an XIV iSCSI discovery address (iSCSI interface): 9.155.51.73
Is this host defined in the XIV system to use CHAP? [default: no ]:
Would you like to discover a new iSCSI target? [default: yes ]: 9.155.51.74
Is this host defined in the XIV system to use CHAP? [default: no ]:
Would you like to discover a new iSCSI target? [default: yes ]: n
Would you like to rescan for new storage devices now? [default: yes ]:
-----
```

The host is connected to the following XIV storage arrays:

Serial	Version	Host Defined	Ports Defined	Protocol	Host Name(s)
1310114	0000	Yes	All	iSCSI	VM1

This host is defined on all iSCSI-attached XIV storage arrays
 Press [ENTER] to proceed.

You can now map the XIV volumes to the defined Windows host.

Configuring Microsoft iSCSI software initiator

The iSCSI connection must be configured on both the Windows host and the XIV Storage System. Follow these instructions to complete the iSCSI configuration:

1. Click **Control Panel** and select **iSCSI Initiator** to display the **iSCSI Initiator Properties** window that is shown in Figure 2-8.

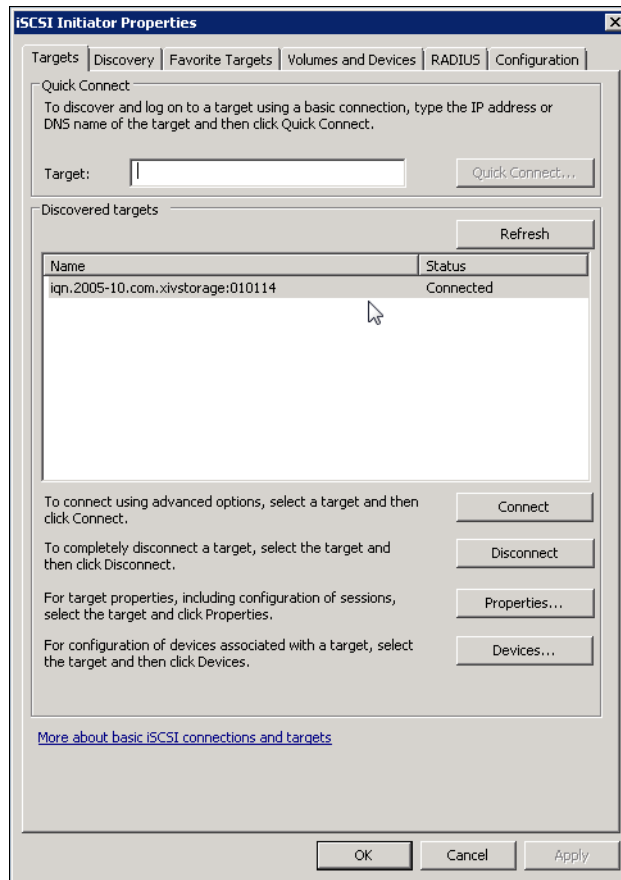


Figure 2-8 iSCSI Initiator Properties window

2. Get the iSCSI Qualified Name (IQN) of the server from the Configuration tab (Figure 2-9). In this example, it is `iqn.1991-05.com.microsoft:win-8h202jnaffa`. Copy this IQN to your clipboard and use it to define this host on the XIV Storage System.

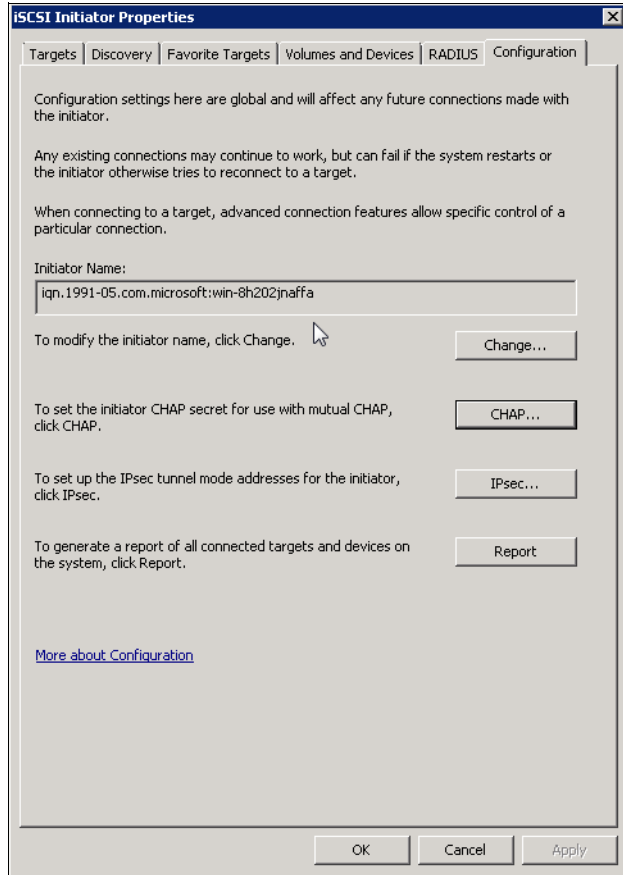


Figure 2-9 iSCSI Configuration tab

3. Define the host in the XIV as shown in Figure 2-10.

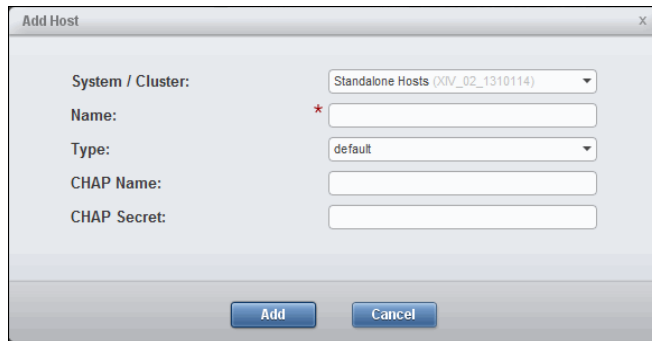


Figure 2-10 Defining the host

4. Add port to host as illustrated in Figure 2-11.

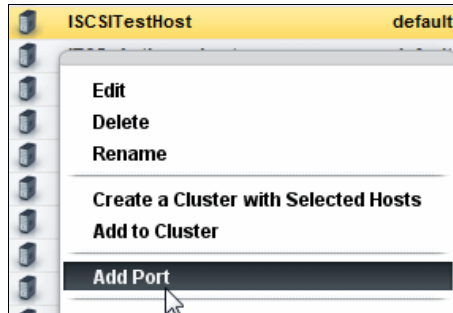


Figure 2-11 Adding the port to the host

5. Configure as an iSCSI connection as shown in Figure 2-12.

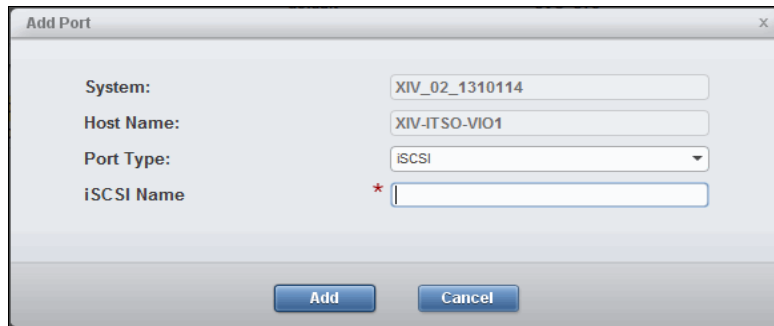


Figure 2-12 Configuring the iSCSI connection

6. Click the Discovery tab.

- Click **Discover Portal** in the Target Portals window. Use one of the iSCSI IP addresses from your XIV Storage System. Repeat this step for more target portals. Figure 2-13 shows the results.

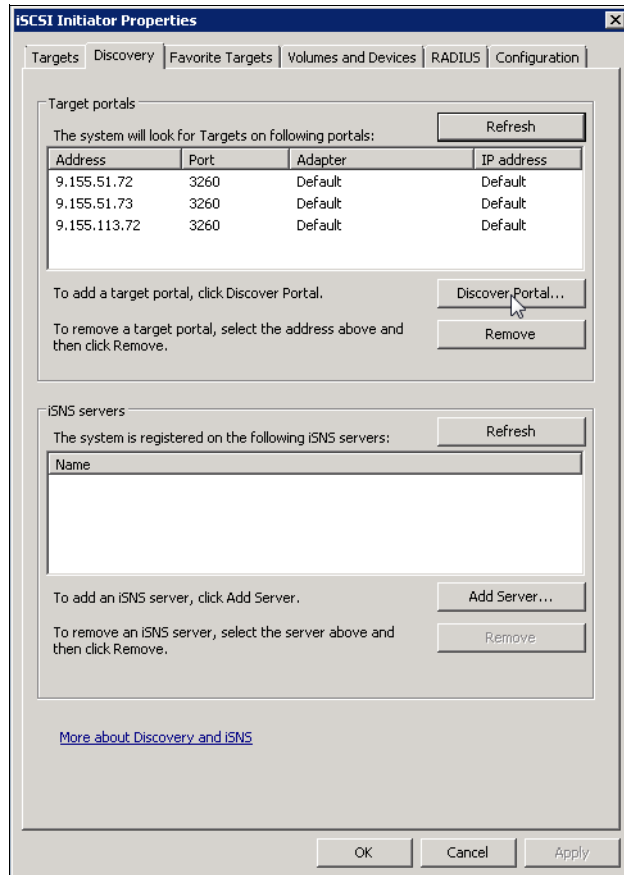


Figure 2-13 iSCSI targets portals defined

- In the XIV GUI, mouse over the **Hosts and Clusters** icon and select **iSCSI Connectivity** as shown in Figure 2-14.



Figure 2-14 Selecting iSCSI Connectivity

The iSCSI Connectivity window shows which LAN ports are connected using iSCSI, and which IP address is used (Figure 2-15).

Name	Address	Netmask	Gateway
m4p1	9.155.51.72	255.255.255.0	9.155.51.1
m7p1	9.155.51.73	255.255.255.0	9.155.51.1
m9p1	9.155.51.74	255.255.255.0	9.155.51.1

Figure 2-15 iSCSI Connectivity window

To improve performance, you can change the MTU size to 4500 if your network supports it as shown in Figure 2-16.

Properties

Name: m4p1

Type: iSCSI

Address: 9.155.51.72

Netmask: 255.255.255.0

Gateway: 9.155.51.1

MTU: 4500

Module: 1:Module:4

Ports: 1

OK

Figure 2-16 Updating iSCSI properties

Alternatively, you can use the Command Line Interface (XCLI) command as shown in Example 2-3.

Example 2-3 Listing iSCSI interfaces

```
>>ipinterface_list
Name      Type      IP Address  Network Mask  Default Gateway  MTU  Module  Ports
m4p1     iSCSI    9.155.51.72 255.255.255.0 9.155.51.1      4500 1:Module:4 1
m7p1     iSCSI    9.155.51.73 255.255.255.0 9.155.51.1      4500 1:Module:7 1
m9p1     iSCSI    9.155.51.74 255.255.255.0 9.155.51.1      4500 1:Module:9 1
management  Management 9.155.51.68 255.255.255.0 9.155.51.1      1500 1:Module:1
VPN      VPN      10.0.20.104 255.255.255.0 10.0.20.1       1500 1:Module:1
management  Management 9.155.51.69 255.255.255.0 9.155.51.1      1500 1:Module:2
management  Management 9.155.51.70 255.255.255.0 9.155.51.1      1500 1:Module:3
VPN      VPN      10.0.20.105 255.255.255.0 10.0.20.1       1500 1:Module:3
XIV-02-1310114>>
```

The iSCSI IP addresses that were used in the test environment are 9.155.51.72, 9.155.51.73, and 9.155.51.74.

9. The XIV Storage System is discovered by the initiator and displayed in the **Favorite Targets** tab as shown in Figure 2-17. At this stage, the Target shows as **Inactive**.

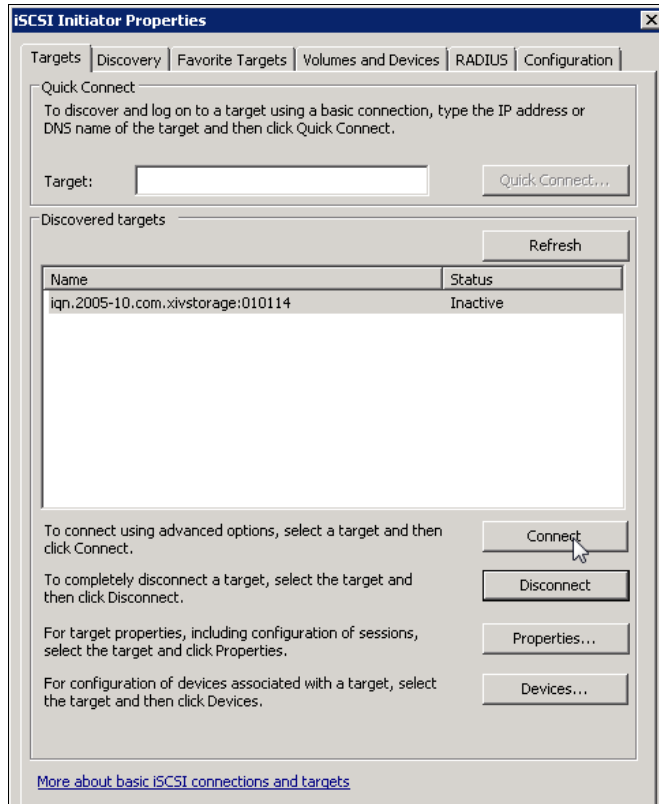


Figure 2-17 A discovered XIV Storage with Inactive status

10. To activate the connection, click **Connect**.
11. In the Connect to Target window, select **Enable multi-path**, and **Add this connection to the list of Favorite Targets** as shown in Figure 2-18. These settings automatically restore this connection when the system boots.

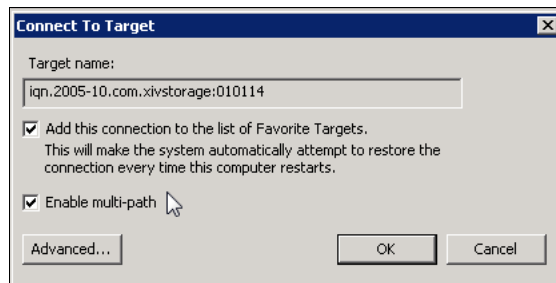


Figure 2-18 Connect to Target window

The iSCSI Target connection status now shows as **Connected** as illustrated in Figure 2-19.

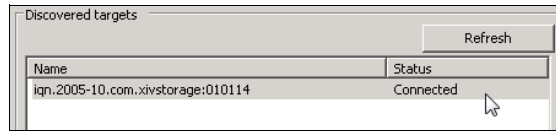


Figure 2-19 Connect to Target is active

12. Click the **Discovery** tab.

13. Select the Discover Portal IP address of the XIV Storage System as shown in Figure 2-20.

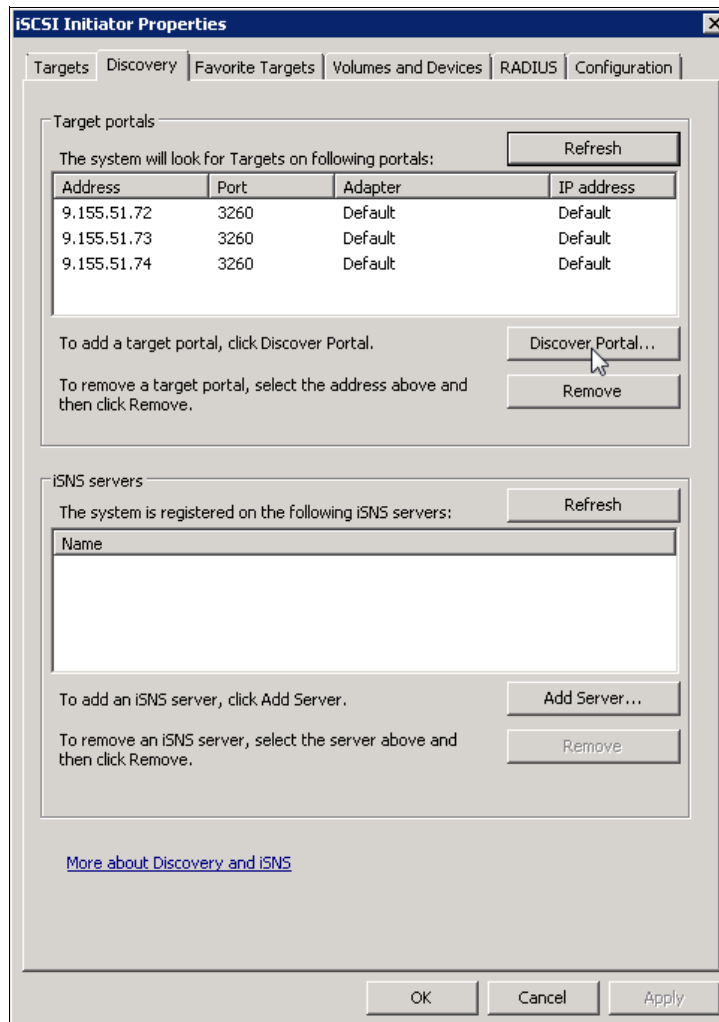


Figure 2-20 Discovering the iSCSI connections

14. Enter the XIV iSCSI IP address and repeat this step for all connection paths as shown in Figure 2-21.

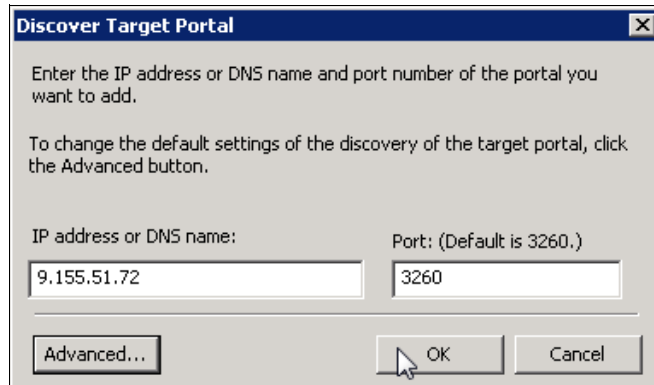


Figure 2-21 Discovering the XIV iSCSI IP addresses

The **FavoriteTargets** tab shows the connected IP addresses as shown in Figure 2-22.

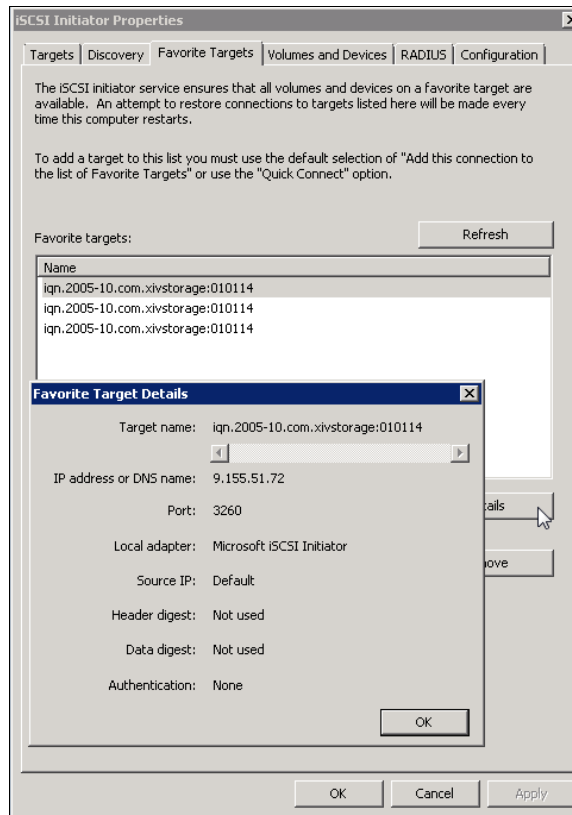


Figure 2-22 A discovered XIV Storage with Connected status

- View the iSCSI sessions by clicking the **Targets** tab, highlighting the target, and clicking **Properties**. Verify the sessions of the connection as seen in Figure 2-23.

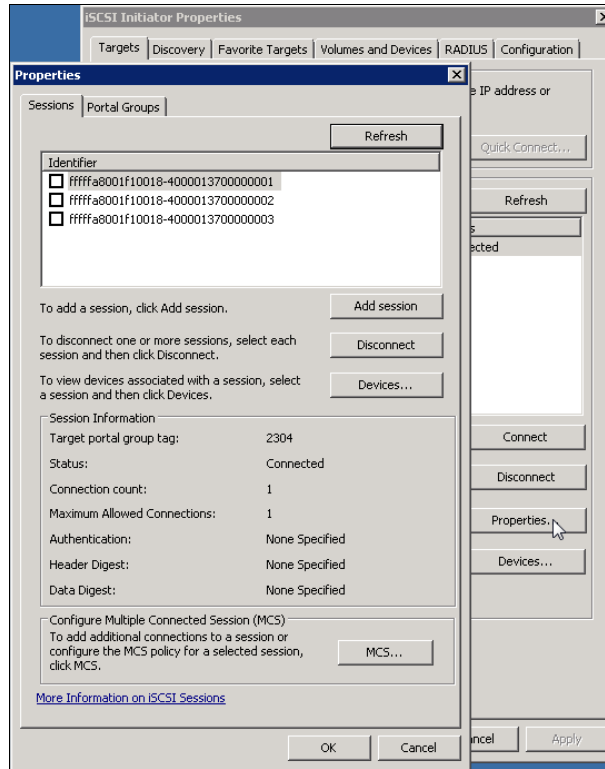


Figure 2-23 Target connection details

- To see further details or change the load balancing policy, click **Connections** as shown in Figure 2-24.

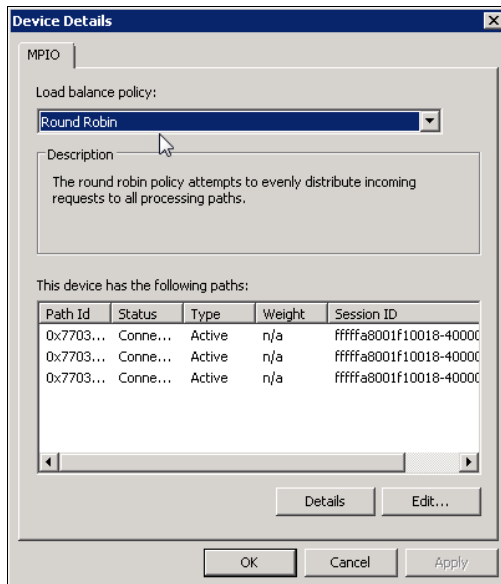


Figure 2-24 Connected sessions

Use the default load balancing policy, **Round Robin**. Change this setting only if you are confident that another option is better suited to your environment.

The following are the available options:

- **Fail Over Only**
- **Round Robin** (default)
- **Round Robin With Subset**
- **Least Queue Depth**
- **Weighted Paths**

17. If you have already mapped volumes to the host system, you see them under the **Devices** tab. If no volumes are mapped to this host yet, assign them now.

Another way to verify your assigned disk is to open the Windows Device Manager as shown in Figure 2-25.

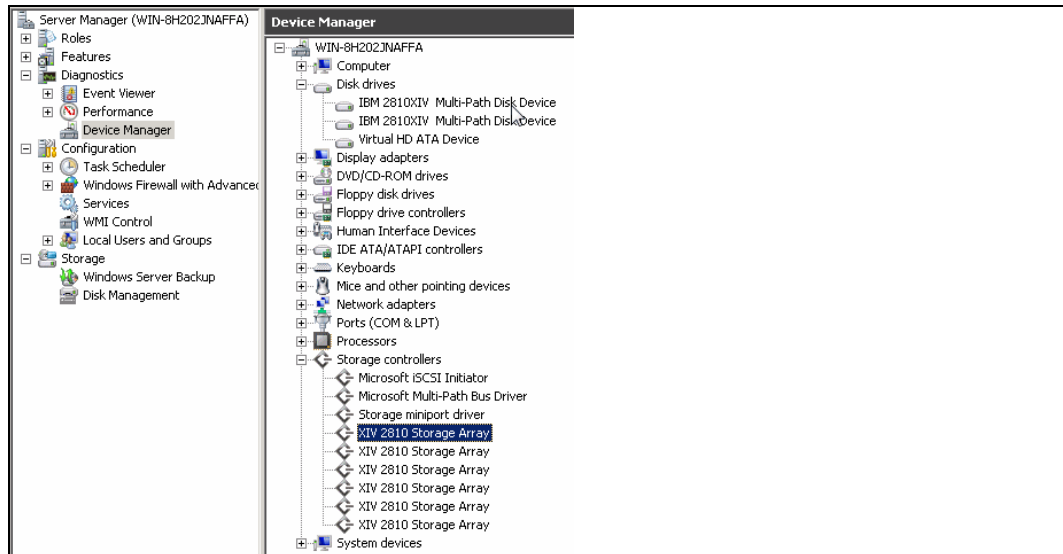


Figure 2-25 Windows Device Manager with XIV disks connected through iSCSI

The mapped LUNs on the host can be seen in Disk Management window as illustrated in Figure 2-26.

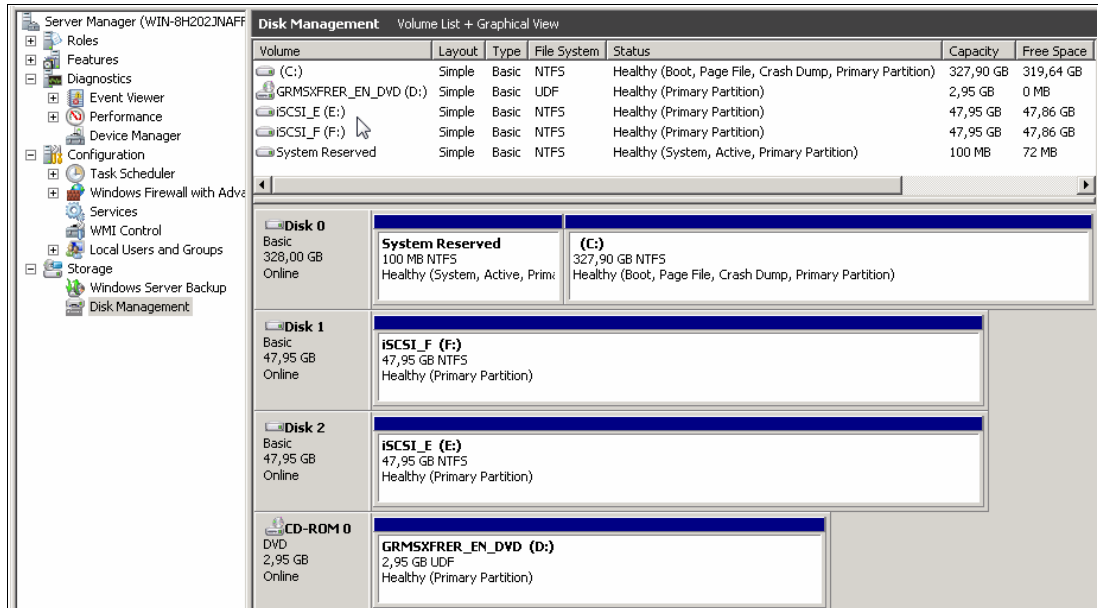


Figure 2-26 Mapped LUNs are displayed in Disk Management

18. Click **Control Panel** and select **iSCSI Initiator** to display the **iSCSI Initiator Properties** window that is shown in Figure 2-27.

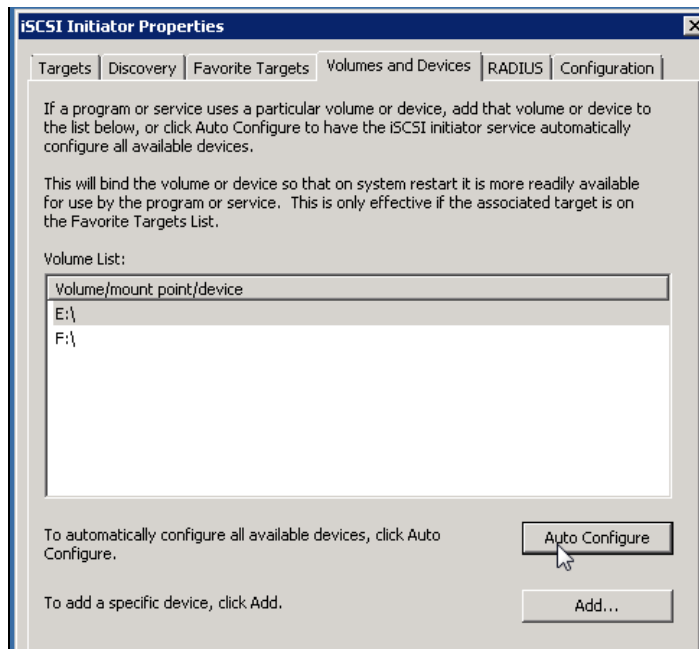


Figure 2-27 Connected Volumes list

19. Click the **Volumes and Devices** tab to verify the freshly created volumes.

2.1.4 Host Attachment Kit utilities

The Host Attachment Kit includes the following utilities:

- ▶ `xiv_devlist`
- ▶ `xiv_diag`

`xiv_devlist`

This utility requires Administrator privileges. The utility lists the XIV volumes available to the host. Non-XIV volumes are also listed separately. To run it, go to a command prompt and enter `xiv_devlist` as shown in Example 2-4.

Example 2-4 `xiv_devlist` command results

```
C:\Users\Administrator.SAND>xiv_devlist
XIV Devices
-----
Device                Size    Paths  Vol Name          Vol Id  XIV Id  XIV Host
-----
\\.\PHYSICALDRIVE1    17.2GB  4/4    itso_win2008_vol1 2746    1300203 sand
-----
\\.\PHYSICALDRIVE2    17.2GB  4/4    itso_win2008_vol2 194     1300203 sand
-----
\\.\PHYSICALDRIVE3    17.2GB  4/4    itso_win2008_vol3 195     1300203 sand
-----
Non-XIV Devices
-----
Device                Size    Paths
-----
\\.\PHYSICALDRIVE0    146.7GB N/A
-----
```

`xiv_diag`

This utility gathers diagnostic information from the operating system. It requires Administrator privileges. The resulting compressed file can then be sent to IBM-XIV support teams for review and analysis. To run, go to a command prompt and enter `xiv_diag` as shown in Example 2-5.

Example 2-5 `xiv_diag` command results

```
C:\Users\Administrator.SAND>xiv_diag
xiv_diag
Welcome to the XIV diagnostics tool, version 1.10.0.
This tool will gather essential support information from this host.
Please type in a path to place the xiv_diag file in [default: /tmp]:
Creating archive xiv_diag-results_2012-10-22_14-12-1
INFO: Gathering System Information (2/2)...           DONE
INFO: Gathering System Event Log...                 DONE
INFO: Gathering Application Event Log...            DONE
INFO: Gathering Cluster Log Generator...            SKIPPED
INFO: Gathering Cluster Reports...                  SKIPPED
INFO: Gathering Cluster Logs (1/3)...                SKIPPED
INFO: Gathering Cluster Logs (2/3)...                SKIPPED
INFO: Gathering DISKPART: List Disk...               DONE
INFO: Gathering DISKPART: List Volume...             DONE
INFO: Gathering Installed HotFixes...               DONE
INFO: Gathering DSMXIV Configuration...              DONE
INFO: Gathering Services Information...              DONE
INFO: Gathering Windows Setup API (1/3)...          SKIPPED
```

```

INFO: Gathering Windows Setup API (2/3)... DONE
INFO: Gathering Windows Setup API (3/3)... DONE
INFO: Gathering Hardware Registry Subtree... DONE
INFO: Gathering xiv_devlist... DONE
INFO: Gathering Host Attachment Kit version...
DONE
INFO: Gathering xiv_fc_admin -L... DONE
INFO: Gathering xiv_fc_admin -V... DONE
INFO: Gathering xiv_fc_admin -P... DONE
INFO: Gathering xiv_iscsi_admin -L... DONE
INFO: Gathering xiv_iscsi_admin -V... DONE
INFO: Gathering xiv_iscsi_admin -P... DONE
INFO: Gathering inquiry.py... DONE
INFO: Gathering drivers.py... DONE
INFO: Gathering mpio_dump.py... DONE
INFO: Gathering wmi_dump.py... DONE
INFO: Gathering xiv_mscs_admin --report... SKIPPED
INFO: Gathering xiv_mscs_admin --report --debug ... SKIPPED
INFO: Gathering xiv_mscs_admin --verify... SKIPPED
INFO: Gathering xiv_mscs_admin --verify --debug ... SKIPPED
INFO: Gathering xiv_mscs_admin --version... SKIPPED
INFO: Gathering build-revision file... DONE
INFO: Gathering host_attach logs... DONE
INFO: Gathering xiv logs... DONE
INFO: Gathering ibm products logs... DONE
INFO: Gathering vss provider logs... DONE
INFO: Closing xiv_diag archive file DONE
Deleting temporary directory... DONE
INFO: Gathering is now complete.
INFO: You can now send
c:\users\admini~1\appdata\local\temp\2\xiv_diag-results_2012-10-22_14-12-1.tar.gz to
IBM-XIV for review.
INFO: Exiting.

```

2.2 Attaching a Microsoft Windows 2008 R2 cluster to XIV

This section addresses the attachment of Microsoft Windows 2008 R2 cluster nodes to the XIV Storage System.

Important: The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, see the System Storage Interoperability Center (SSIC) at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

For more information, see the XIV Storage System *Host System Attachment Guide for Windows - Installation Guide*, which is available at:

http://publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/xiv_pubsrelatedinfoic.html

This section addresses the implementation of a two node Windows 2008 R2 Cluster by using FC connectivity.

2.2.1 Prerequisites

To successfully attach a Windows cluster node to XIV and access storage, a number of prerequisites must be met. This is a generic list. Your environment might have extra requirements.

- ▶ Complete the FC cabling.
- ▶ Configure the SAN zoning.
- ▶ Two network adapters and a minimum of five IP addresses.
- ▶ Install Windows 2008 R2 SP1 or later.
- ▶ Install any other updates, if required.
- ▶ Install hot fix KB2468345 if Service Pack 1 is used.
- ▶ Install the Host Attachment Kit to enable the Microsoft Multipath I/O Framework.
- ▶ Ensure that all nodes are part of the same domain.
- ▶ Create volumes to be assigned to the XIV Host/Cluster group, not to the nodes.

Supported versions of Windows Cluster Server

At the time of writing, the following versions of Windows Cluster Server are supported:

- ▶ Windows Server 2008 R2 (x64)
- ▶ Windows Server 2008 (x32 and x64)

Supported configurations of Windows Cluster Server

Windows Cluster Server was tested in the following configurations:

- ▶ Up to eight nodes: Windows 2008 (x32 and x64)
- ▶ Up to 10 nodes: Windows 2008 R2 (x64)

If other configurations are required, you need a Storage Customer Opportunity REquest (SCORE), which replaces the older request for price quotation (RPQ) process. IBM then tests your configuration to determine whether it can be certified and supported. Contact your IBM representative for more information.

Supported FC HBAs

Supported FC HBAs are available from Brocade, Emulex, IBM, and QLogic. More information about driver versions is available from the SSIC at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

Unless otherwise noted in SSIC, use any supported driver and firmware by the HBA vendors. The latest versions are always preferred.

Multi-path support

Microsoft provides a multi-path framework and development kit that is called the MPIO. The driver development kit allows storage vendors to create DSMs for MPIO. You can use DSMs to build interoperable multi-path solutions that integrate tightly with Microsoft Windows.

MPIO allows the host HBAs to establish multiple sessions with the same target LUN, but present them to Windows as a single LUN. The Windows MPIO drivers enable a true active/active path policy that allows I/O over multiple paths simultaneously.

Further information about Microsoft MPIO support is available at:

<http://download.microsoft.com/download/3/0/4/304083f1-11e7-44d9-92b9-2f3cdbf01048/mpio.doc>

2.2.2 Installing Cluster Services

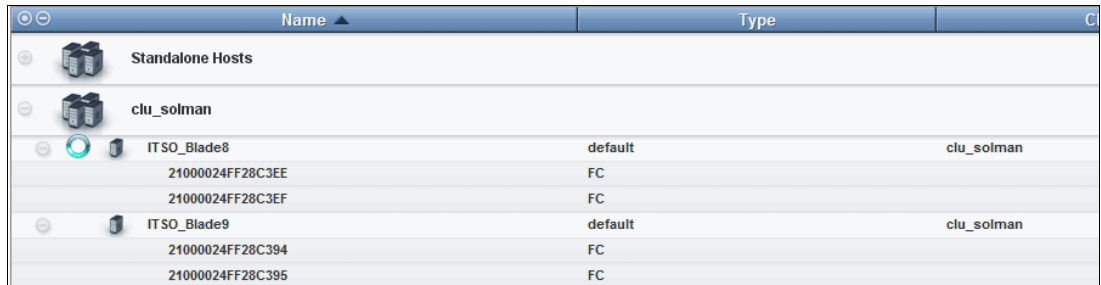
This scenario covers a two node Windows 2008 R2 Cluster. The procedures assume that you are familiar with Windows 2008 Cluster. Therefore, they focus on specific requirements for attaching to XIV.

For more information about installing a Windows 2008 Cluster, see:

<http://technet.microsoft.com/en-us/windowsserver/bb310558.aspx>

To install the cluster, complete these steps:

1. On the XIV system main GUI, select **Hosts and Clusters**. Create a cluster and put both nodes into the cluster as depicted in Figure 2-28.




Name	Type	
Standalone Hosts		
clu_solman		
ITSO_Blade8	default	clu_solman
21000024FF28C3EE	FC	
21000024FF28C3EF	FC	
ITSO_Blade9	default	clu_solman
21000024FF28C394	FC	
21000024FF28C395	FC	

Figure 2-28 XIV cluster with both nodes

In this example, an XIV cluster named *clu_solman* was created and both nodes were placed in it.

2. Map all the LUNs to the cluster as shown in Figure 2-29.



LUN	Volume
8	ITSO_Blade9_LUN_4M
7	ITSO_Blade9_LUN_3M
6	ITSO_Blade9_LUN_2M
5	ITSO_Blade9_LUN_1M
4	ITSO_Blade9_Lun_4
3	ITSO_Blade9_Lun_3
2	ITSO_Blade9_Lun_2
1	ITSO_Blade9_Lun_1

Figure 2-29 Mapped LUNs list

All LUNs are mapped to the XIV cluster, but not to the individual nodes.

3. Set up a cluster-specific configuration that includes the following characteristics:
 - All nodes are in the same domain
 - Have network connectivity
 - Private (heartbeat) network connectivity

- Node2 must not do any Disk I/O
 - Run the cluster configuration check
4. On Node1, scan for new disks. Then, initialize, partition, and format them with NTFS. The following requirements are for shared cluster disks:
- These disks must be basic disks.
 - For Windows 2008, you must decide whether they are Master Boot Record (MBR) disks or GUID Partition Table (GPT) disks.

Figure 2-30 shows what this configuration looks like on node 1.

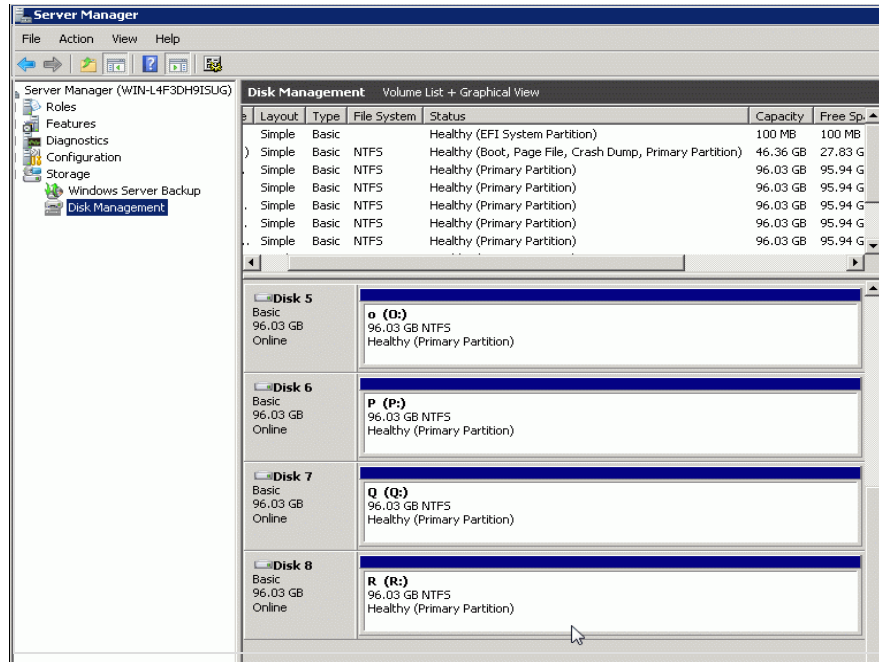


Figure 2-30 Initialized, partitioned, and formatted disks

5. Ensure that only one node accesses the shared disks until the cluster service is installed on all nodes. This restriction must be done before you continue to the Cluster wizard (Figure 2-31). You no longer must turn off all nodes as you did with Windows 2003. You can bring all nodes into the cluster in a single step. However, no one is allowed to work on the other nodes.

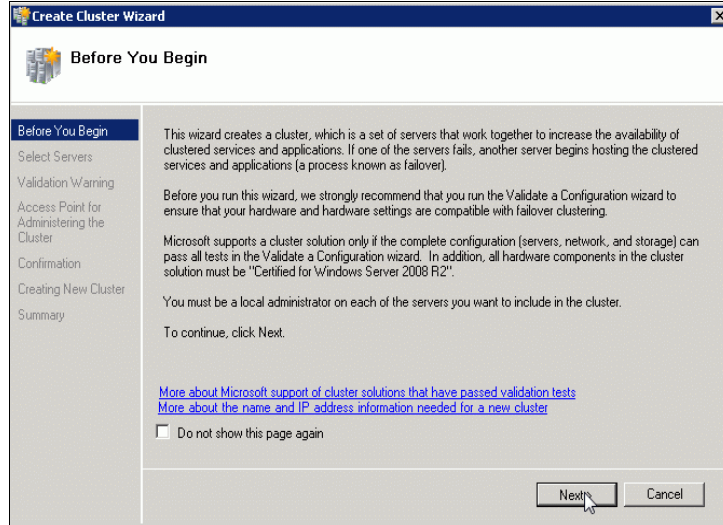


Figure 2-31 Create Cluster Wizard welcome window

6. Select all nodes that belong to the cluster as shown in Figure 2-32.

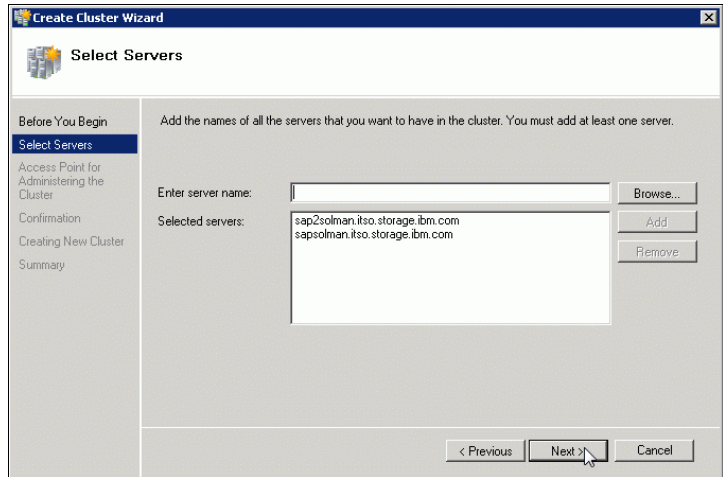


Figure 2-32 Selecting your nodes

7. After the Create Cluster wizard completes, a summary panel shows you that the cluster was created successfully (Figure 2-33). Keep this report for documentation purposes.

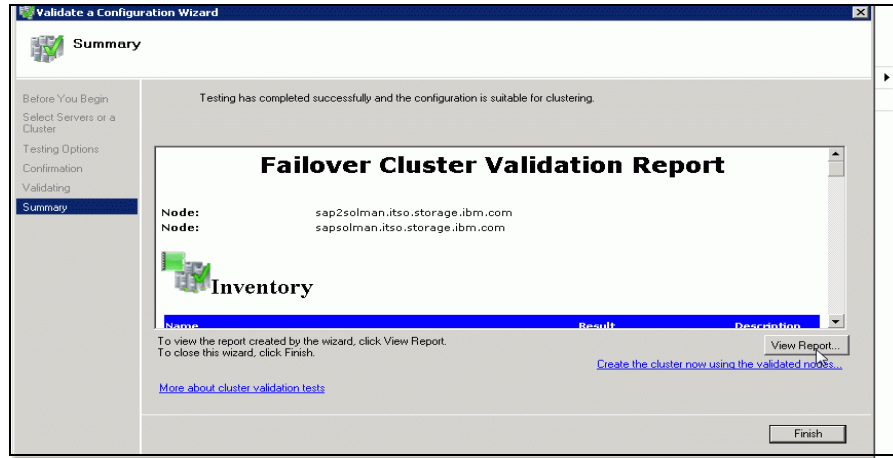


Figure 2-33 Failover Cluster Validation Report window

8. Check access to at least one of the shared drives by creating a document. For example, create a text file on one of them, and then turn off node 1.
9. Check the access from Node2 to the shared disks and power node 1 on again.

10. Make sure that you have the correct cluster witness model as illustrated in Figure 2-34. The old cluster model had a quorum as a single point of failure.

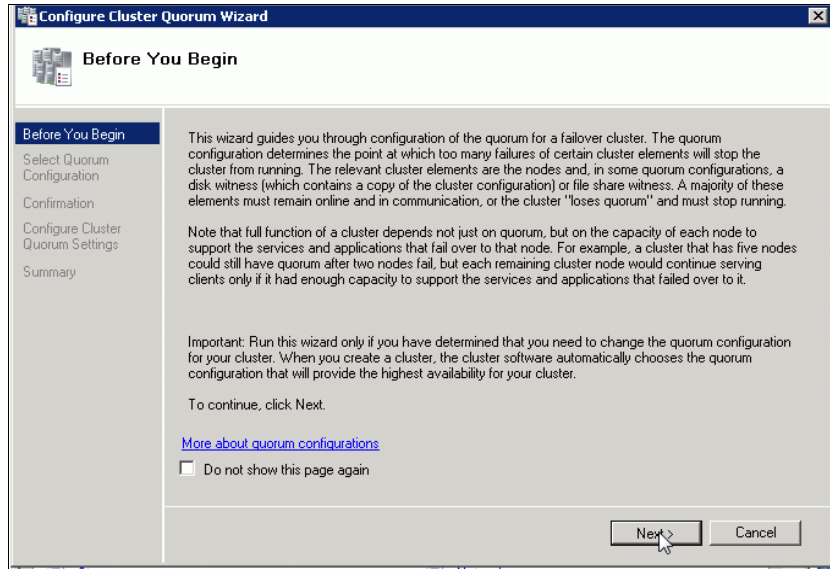


Figure 2-34 Configure Cluster Quorum Wizard window

In this example, the cluster witness model is changed, which assumes that the witness share is in a third data center (see Figure 2-35).

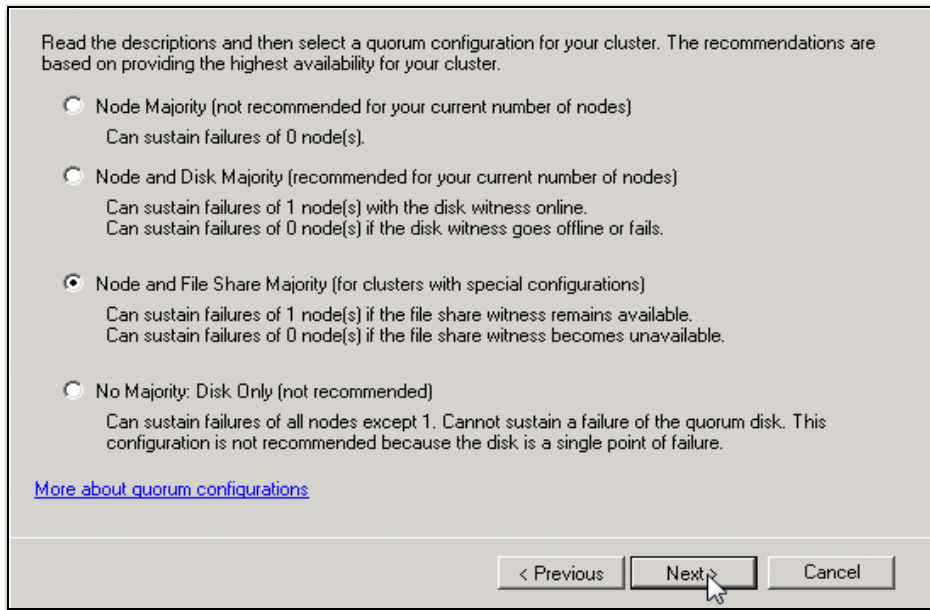


Figure 2-35 Selecting the witness model

2.2.3 Configuring the IBM Storage Enabler for Windows Failover Clustering

The IBM Storage Enabler for Windows Failover Clustering is a software agent that runs as a Microsoft Windows Server service. It runs on two geographically dispersed cluster nodes, and provides failover automation for XIV storage provisioning on them. This agent enables deployment of these nodes in a geo cluster configuration.

The software, Release Notes, and the User Guide can be found at:

http://www.ibm.com/support/fixcentral/swg/selectFixes?parent=ibm~Storage_Disk&product=ibm/Storage_Disk/XIV+Storage+System+%282810,+2812%29&release=11.0.0&platform=A11&function=all

Supported Windows Server versions

The IBM Storage Enabler for Windows Failover Clustering supports the Windows Server versions or editions that are listed in Table 2-1.

Table 2-1 Storage Enabler for Windows Failover Clustering supported servers

Operating system	Architecture	Service Pack
Microsoft Windows Server 2003	x86, x64	Service Pack 1, Service Pack 2
Microsoft Windows Server 2003 R2	x86, x64	Service Pack 1, Service Pack 2
Microsoft Windows Server 2008	x86, x64	Service Pack 1, Service Pack 2
Microsoft Windows Server 2008 R2	x64	Tested with Service Pack 1
Microsoft Windows Server 2012	x64	None

Installing and configuring the Storage Enabler

The following instructions are based on the installation that was performed at the time of writing. For more information, see the instructions in the *Release Notes and the User Guide*. These instructions are subject to change over time.

1. Start the installer as administrator (Figure 2-36).



Figure 2-36 Starting the installation

2. Follow the wizard instructions.

After the installation is complete, observe a new service that is called XIVmcsAgent as shown in Figure 2-37.

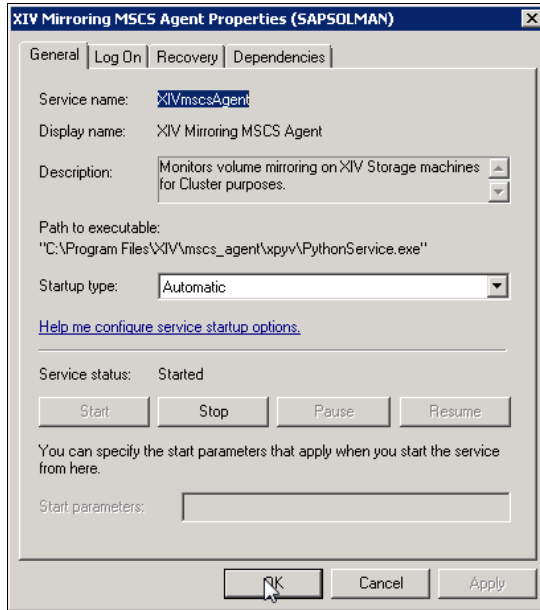


Figure 2-37 XIVmcsAgent as a service

No configuration took place until now. Therefore, the dependencies of the Storage LUNs did not change as shown in Figure 2-38.

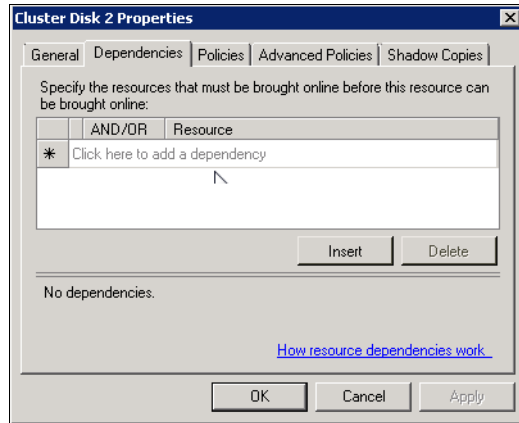


Figure 2-38 Dependencies of drive properties

3. Define the mirror connections for your LUNs between the two XIVs, as shown in Figure 2-39. For more information about how to define the mirror pairs, see *IBM XIV Storage System: Copy Services and Migration, SG24-7759*.

ITSO_Blade9_Test	M	↔	Synchronized	ITSO_Blade9_Test	XIV PFE-GEN3-13
ITSO_Blade9_LUN_4M	M	↔	Synchronized	ITSO_Blade9_LUN_4M	XIV PFE-GEN3-13
ITSO_Blade9_Lun_4	M	↔	Synchronized	ITSO_Blade9_Lun_4	XIV PFE-GEN3-13
ITSO_Blade9_LUN_3M	M	↔	Synchronized	ITSO_Blade9_LUN_3M	XIV PFE-GEN3-13
ITSO_Blade9_Lun_3	M	↔	Synchronized	ITSO_Blade9_Lun_3	XIV PFE-GEN3-13
ITSO_Blade9_LUN_2M	M	↔	Synchronized	ITSO_Blade9_LUN_2M	XIV PFE-GEN3-13
ITSO_Blade9_Lun_2	M	↔	Synchronized	ITSO_Blade9_Lun_2	XIV PFE-GEN3-13
ITSO_Blade9_LUN_1M	M	↔	Synchronized	ITSO_Blade9_LUN_1M	XIV PFE-GEN3-13
ITSO_Blade9_Lun_1	M	↔	Synchronized	ITSO_Blade9_Lun_1	XIV PFE-GEN3-13

Figure 2-39 Mirror definitions at the master side and side of node 1

Also, define the connections on the subordinate side as shown in Figure 2-40.

ITSO_Blade9_Test	sj	Consistent	ITSO_Blade9_Test
ITSO_Blade9_LUN_4M	sj	Consistent	ITSO_Blade9_LUN_4M
ITSO_Blade9_Lun_4	sj	Consistent	ITSO_Blade9_Lun_4
ITSO_Blade9_LUN_3M	sj	Consistent	ITSO_Blade9_LUN_3M
ITSO_Blade9_Lun_3	sj	Consistent	ITSO_Blade9_Lun_3
ITSO_Blade9_LUN_2M	sj	Consistent	ITSO_Blade9_LUN_2M
ITSO_Blade9_Lun_2	sj	Consistent	ITSO_Blade9_Lun_2
ITSO_Blade9_LUN_1M	sj	Consistent	ITSO_Blade9_LUN_1M
ITSO_Blade9_Lun_1	sj	Consistent	ITSO_Blade9_Lun_1

Figure 2-40 Mirror definitions on the subordinate side

4. Redefine the host mapping of the LUNs on both XIVs. For a working cluster, both nodes and their HBAs must be defined in a cluster group. All of the LUNs provided to the cluster must be mapped to the cluster group itself, not to the nodes (Figure 2-29 on page 70). When using the XIVmcsAgent, you must remap those LUNs to their specific XIV/node combination. Figure 2-41 shows the mapping for node 1 on the master side.

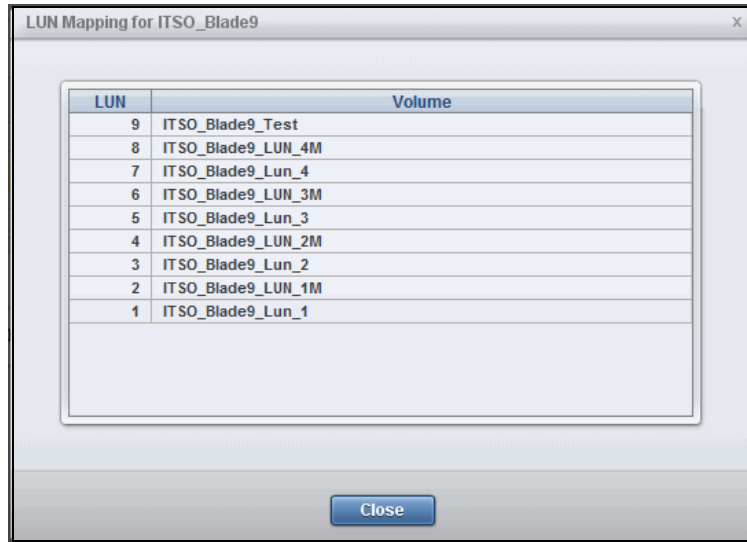


Figure 2-41 Selecting the private mapping for node 1 on master side

Figure 2-42 shows the mapping for node 2 on the master side.



Figure 2-42 Changing the default mapping: node 2 has no access to the master side

Figure 2-43 shows the mapping for node 2 on the subordinate side.

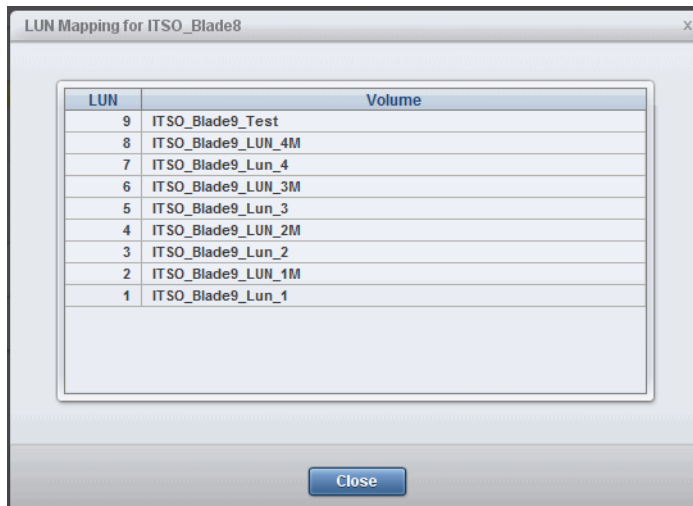


Figure 2-43 Selecting the private mapping on the subordinate side for node 2

Figure 2-44 shows the mapping for node 1 on the subordinate side.



Figure 2-44 Node 1 has no access to XIV2

5. Check that all resources are on node 1, where the Mirror Master side is defined, as illustrated in Figure 2-45.

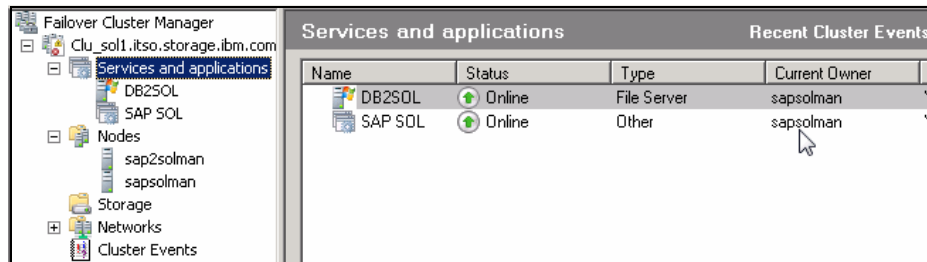


Figure 2-45 All resources are on node 1

6. To configure the XIVmcsAgent, run the **admin** tool with the **-install** option as shown in Example 2-6.

Example 2-6 How to use mscs_agent

```

C:\Users\Administrator.ITS0>cd C:\Program Files\XIV\mscs_agent\bin
C:\Program Files\XIV\mscs_agent\bin>dir
Volume in drive C has no label.
Volume Serial Number is CA6C-8122
Directory of C:\Program Files\XIV\mscs_agent\bin
09/29/2011  11:13 AM    <DIR>          .
09/29/2011  11:13 AM    <DIR>          ..
09/14/2011  03:37 PM                1,795 project_specific_pyrunner.py
09/13/2011  07:20 PM                2,709 pyrunner.py
09/14/2011  11:48 AM            134,072 xiv_mscs_admin.exe
09/14/2011  11:48 AM            134,072 xiv_mscs_service.exe
                4 File(s)      272,648 bytes
                2 Dir(s)  14,025,502,720 bytes free
C:\Program Files\XIV\mscs_agent\bin>xiv_mscs_admin.exe
Usage: xiv_mscs_admin [options]
Options:
  --version          show program's version number and exit
  -h, --help        show this help message and exit
  --install          installs XIV MSCS Agent components on this node and
cluster Resource Type
  --upgrade         upgrades XIV MSCS Agent components on this node
  --report          generates a report on the cluster
  --verify          verifies XIV MSCS Agent deployment
  --fix-dependencies fixes dependencies between Physical Disks and XIV
Mirror resources
  --deploy-resources deploys XIV Mirror resources in groups that contain
Physical Disk Resources
  --delete-resources deletes all existing XIV Mirror resources from the
cluster
  --delete-resourcetype deletes the XIV mirror resource type
  --uninstall       uninstalls all XIV MSCS Agent components from this
node
  --change-credentials change XIV credentials
  --debug           enables debug logging
  --verbose         enables verbose logging
  --yes            confirms distructive operations
XCLI Credentials Options:
  --xcli-username=USERNAME
  --xcli-password=PASSWORD
C:\Program Files\XIV\mscs_agent\bin>xiv_mscs_admin.exe --install --verbose
--xcli-username=itso --xcli-password=<PASSWORD>
2011-09-29 11:19:12 INFO classes.py:76 checking if the resource DLL exists
2011-09-29 11:19:12 INFO classes.py:78 resource DLL doesn't exist, installing
it
2011-09-29 11:19:12 INFO classes.py:501 The credentials MSCS Agent uses to
connect to the XIV Storage System have been change
d. Check the guide for more information about credentials.
Installing service XIVmcsAgent
Service installed
Changing service configuration
Service updated

```

2011-09-29 11:19:14 INFO classes.py:85 resource DLL exists
 C:\Program Files\XIV\mscs_agent\bin>

- To deploy the resources into the geo cluster, run the xiv_mcs_admin.exe utility:
 C:\Program Files\XIV\mscs_agent\bin>xiv_mcs_admin.exe --deploy-resources
 --verbose --xcli-username=itso --xcli-password=<PASSWORD> --yes

Figure 2-46 illustrates the cluster dependencies that result.

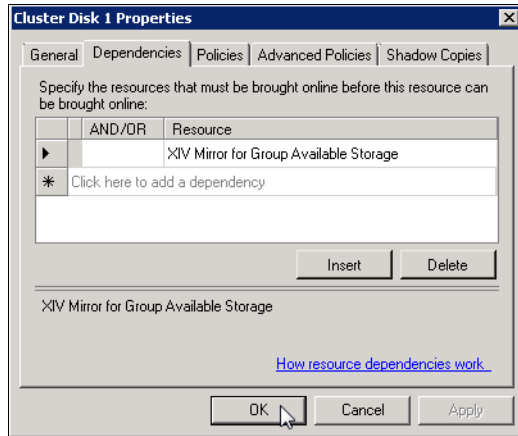


Figure 2-46 Dependencies after deploying the resources

- Power node 1 down and repeat the previous steps on node 2.
- A switch of the cluster resource group from node 1 to node 2 leads to a change of the replication direction. XIV2 becomes the master and XIV1 becomes the subordinate, as shown in Figure 2-47 and Figure 2-48.

Name	RPO	Status	Remote Volume	Remote System
sync_test_a	VT	Inactive	sync_test_a	XIV-02-1310114
ITSO_VM_Datastore2	ST	Consistent	ITSO_VM_Datastore2	XIV-02-1310114
ITSO_Blade9_Test	ST	Consistent	ITSO_Blade9_Test	XIV-02-1310114
ITSO_Blade9_LUN_4M	ST	Consistent	ITSO_Blade9_LUN_4M	XIV-02-1310114
ITSO_Blade9_Lun_4	ST	Consistent	ITSO_Blade9_Lun_4	XIV-02-1310114
ITSO_Blade9_LUN_3M	ST	Consistent	ITSO_Blade9_LUN_3M	XIV-02-1310114
ITSO_Blade9_Lun_3	VT	Synchronized	ITSO_Blade9_Lun_3	XIV-02-1310114
ITSO_Blade9_LUN_2M	ST	Synchronized	ITSO_Blade9_LUN_2M	XIV-02-1310114
ITSO_Blade9_Lun_2	VT	Synchronized	ITSO_Blade9_Lun_2	XIV-02-1310114
ITSO_Blade9_LUN_1M	ST	Consistent	ITSO_Blade9_LUN_1M	XIV-02-1310114
ITSO_Blade9_Lun_1	ST	Consistent	ITSO_Blade9_Lun_1	XIV-02-1310114

Figure 2-47 XIVmscsAgent changing the replication direction

The results are shown in Figure 2-48.

Name	RPO	Status	Remote Volume	Remote System
sync_test_a	ST	Inactive	sync_test_a	XIV PFE-GEN3-1310133
ITSO_VM_Datastore2	VT	Synchronized	ITSO_VM_Datastore2	XIV PFE-GEN3-1310133
ITSO_Blade9_Test	VT	Synchronized	ITSO_Blade9_Test	XIV PFE-GEN3-1310133
ITSO_Blade9_LUN_4M	VT	Synchronized	ITSO_Blade9_LUN_4M	XIV PFE-GEN3-1310133
ITSO_Blade9_Lun_4	VT	Synchronized	ITSO_Blade9_Lun_4	XIV PFE-GEN3-1310133
ITSO_Blade9_LUN_3M	VT	Synchronized	ITSO_Blade9_LUN_3M	XIV PFE-GEN3-1310133
ITSO_Blade9_Lun_3	ST	Consistent	ITSO_Blade9_Lun_3	XIV PFE-GEN3-1310133
ITSO_Blade9_LUN_2M	ST	Consistent	ITSO_Blade9_LUN_2M	XIV PFE-GEN3-1310133
ITSO_Blade9_Lun_2	ST	Consistent	ITSO_Blade9_Lun_2	XIV PFE-GEN3-1310133
ITSO_Blade9_LUN_1M	VT	Synchronized	ITSO_Blade9_LUN_1M	XIV PFE-GEN3-1310133
ITSO_Blade9_Lun_1	VT	Synchronized	ITSO_Blade9_Lun_1	XIV PFE-GEN3-1310133

Figure 2-48 A switch of the cluster resources leads to a “change role” on XIV

2.3 Attaching a Microsoft Hyper-V Server 2008 R2 to XIV

This section addresses a Microsoft Hyper-V2008 R2 environment with XIV. Hyper-V Server 2008 R2 is the hypervisor-based server virtualization product from Microsoft that consolidates workloads on a single physical server.

System hardware requirements

To run Hyper-V, you must fulfill the following hardware requirements:

- ▶ Processors can include virtualization hardware assists from Intel (Intel VT) and AMD (AMD-V). To enable Intel VT, enter System Setup and click **Advanced Options** → **CPU Options**, then select **Enable Intel Virtualization Technology**. AMD-V is always enabled. The processors must have the following characteristics:
 - Processor cores: Minimum of four processor cores.
 - Memory: A minimum of 16 GB of RAM.
 - Ethernet: At least one physical network adapter.
 - Disk space: One volume with at least 50 GB of disk space and one volume with at least 20 GB of space.
 - BIOS: Enable the Data Execution Prevention option in System Setup. Click **Advanced Options** → **CPU Options** and select **Enable Processor Execute Disable Bit**. Ensure that you are running the latest version of BIOS.
- ▶ Server hardware that is certified by Microsoft to run Hyper-V. For more information, see the Windows Server Catalog at:
<http://go.microsoft.com/fwlink/?LinkID=111228>
Select **Hyper-V** and **IBM** from the categories on the left side.

Installing Hyper-V in Windows Server 2008 R2 with Server Core

The Server Core option on Windows Server 2008 R2 provides a subset of the features of Windows Server 2008 R2. This option runs these supported server roles without a full Windows installation:

- ▶ Dynamic Host Configuration Protocol (DHCP)
- ▶ Domain Name System (DNS)
- ▶ Active Directory
- ▶ Hyper-V

With the Server Core option, the setup program installs only the files that are needed for the supported server roles.

Using Hyper-V on a Server Core installation reduces the *attack surface*. The attack surface is the scope of interfaces, services, APIs, and protocols that a hacker can use to attempt to gain entry into the software. As a result, a Server Core installation reduces management requirements and maintenance. Microsoft provides management tools to remotely manage the Hyper-V role and virtual machines (VMs). Hyper-V servers can be managed from Windows Vista or other Windows Server 2008 or 2008 R2 systems (both x86 and x64). You can download the management tools at:

<http://support.microsoft.com/kb/952627>

For more information, see *Implementing Microsoft Hyper-V on IBM System x and IBM BladeCenter*, REDP-4481, at:

<http://www.redbooks.ibm.com/abstracts/redp4481.html?Open>

For more information about using the XIV Storage System with Microsoft Hyper-V, see *Implementing a High Performance, End-to-End Virtualized Solution Using Windows Server 2008 R2 Hyper-V, IBM XIV Storage, and Brocade Fabric Solutions - Configuration and Best Practices Guide*, available at:

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101741>

2.4 Microsoft System Center Virtual Machine Manager 2012 Storage Automation

With *Microsoft System Center Virtual Machine Manager (SCVMM) 2012* new announcements alongside XIV Storage System Gen3, administrators have more features to extend Microsoft Hyper-V virtualization for Cloud. They can now perform storage management tasks through an integrated administrative interface. For example, within the SCVMM's graphical user interface (GUI), administrators can discover, classify, allocate, provision, map, assign, and decommission storage. This storage can be associated with clustered and stand-alone virtualization hosts. Usually, these tasks would require multiple steps that use different applications and skills.

IBM has complied with the Storage Networking Industry Association (SNIA) standards requirements for a long time. In addition, the Common Information Model (CIM) framework is embedded in XIV Storage System Gen3. Therefore, XIV Storage System Gen3 is an SCVMM 2012 supported device, and is fully compliant with the SNIA Storage Management Initiative Specification (SMI-S) 1.4.

The benefits are immediate when you attach the XIV Storage System Gen3 to the Hyper-V hypervisor and manage the Cloud with SCVMM 2012.

This section addresses storage automation concepts and provides some implementation guidelines for SCVMM 2012 with XIV Gen3.

2.4.1 The XIV Open API overview

The XIV Open API complies to SMI-S, as specified by the SNIA to manage the XIV Storage System Gen. 3. This API can be used by any storage resource management application to configure and manage the XIV.

Microsoft SCVMM 2012 uses this API to communicate with the XIV System CIM Agent as shown in Figure 2-49. The main components are the CIM object manager (CIMOM), the Service Location Protocol (SLP), and the device provider.

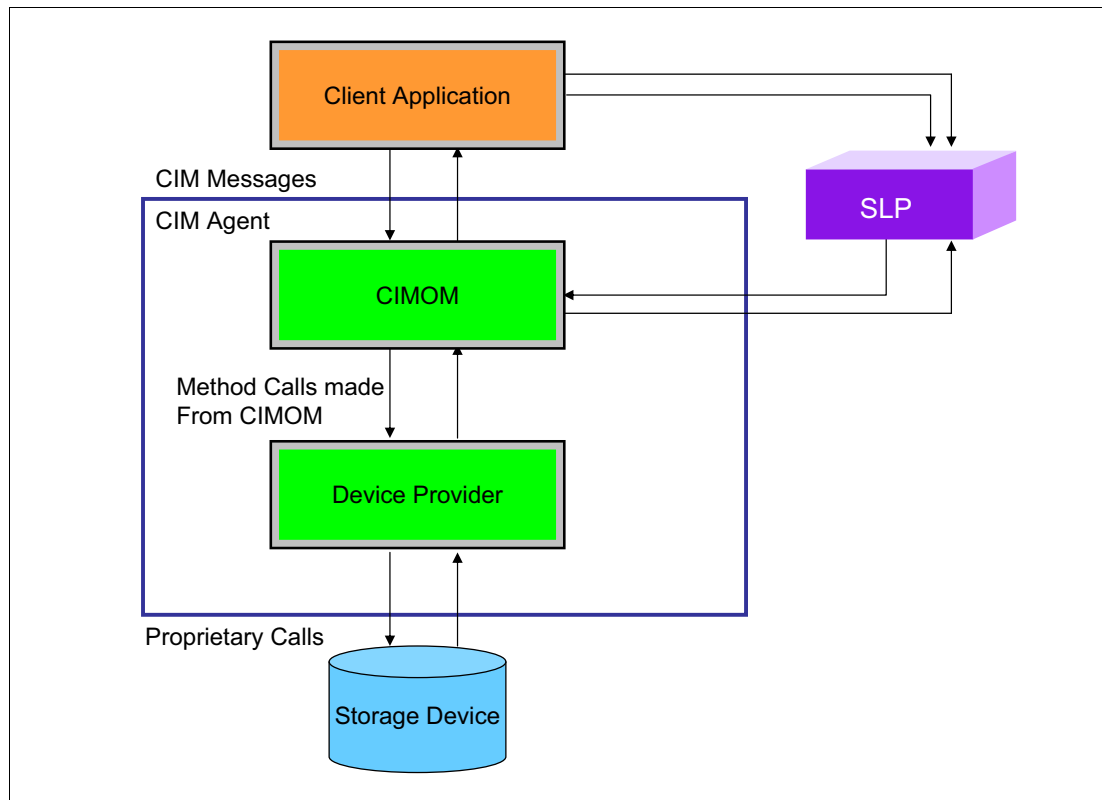


Figure 2-49 XIV System CIM Agent

Within a Cloud environment, security is key. The CIM agent can operate in two security modes:

- ▶ Secure Mode: Requests are sent over HTTP or HTTP over SSL. The Secure Mode is the preferred mode with Microsoft SCVMM.
- ▶ Non-secure Mode: A basic configuration that authorizes communication with a user name and password.

As previously said, the CIM Agent is embedded in the administrative module and does not require configuration. It is also enabled by default.

The CIM agent has the following limitations:

- ▶ The CIM agent is able to manage only the system on which the administrative module is installed.
- ▶ The secure mode must be used over port 5989.
- ▶ The CIM Agent uses the same account that is used to manage the XIV System with the GUI or the XCLI.
- ▶ The SLP requires Internet Group Management Protocol (IGMP) enabled if the network traffic goes through an Ethernet router.

For further XIV Open API and CIM framework details, see the following web page:

<https://www.ibm.com/support/docview.wss?uid=ssg1S7003246>

2.4.2 System Center Virtual Machine Manager overview

SCVMM is a Microsoft virtualization management solution with eight components. It enables centralized administration of both physical and virtual servers, and provides rapid storage provisioning. The availability of VMM 2008 R2 was announced in August 2009. The following enhancements, among others, were introduced in SCVMM 2012:

- ▶ Quick Storage Migration for running virtual machines with minimal downtime
- ▶ Template-based rapid provisioning for new virtual machines
- ▶ Support for Live Migration of virtual machines
- ▶ Support for SAN migration in and out of failover clusters
- ▶ Multi-vendor Storage provisioning

Within a Cloud, private or public, fabric is a key concept. Fabric is composed by hosts, host groups, library servers, networking, and storage configuration. In SCVMM 2012, this concept simplifies the operations that are required to use and define resource pools. From a user's perspective, it becomes possible to deploy VMs and provision storage capacity without performing storage administrative tasks. Figure 2-50 shows the SCVMM fabric management window.

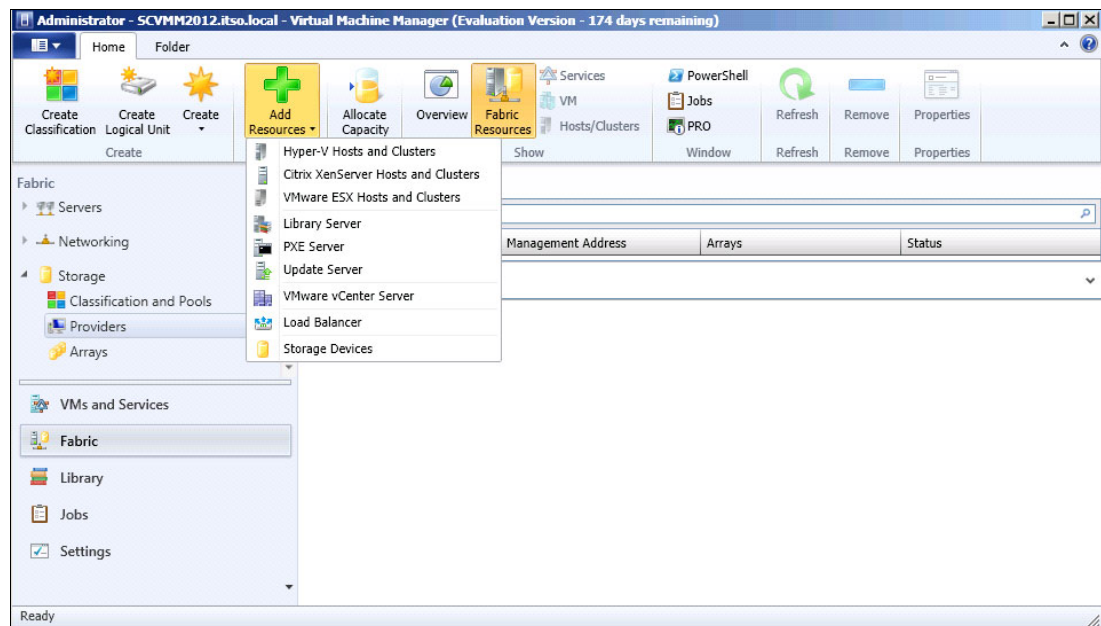


Figure 2-50 SCVMM Fabric management

For more information about configuring and deploying fabric resources with XIV Storage System Gen3 using SCVMM 2012, see the IBM white paper at:

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102071>

Important: As specified in the white paper on page 12, you might face a Transport Layer Security (TLS) issue. This problem might require you to change some parameters on the Windows registry. Additionally, depending on the Windows Service Pack Level used, you must install corrective “fixes” to connect properly to the XIV CIM using SSL and TLS Protocols. The corrective fixes can be found at:

<http://support.microsoft.com/kb/2643584>

For more information about the procedures to define storage resources, see “SCVMM 2012 Storage Automation Step-by-Step Processes”. These resources include storage

classifications, logical units, and storage pools that are made available to Hyper-V hosts and host clusters. The following SCVMM 2012 key features that use IBM XIV Storage System Gen3 are also documented:

- ▶ Storage Device Discovery
- ▶ Storage Pool Classification
- ▶ Allocation
- ▶ Provisioning

For more information, “Configuring Storage in VMM Overview” at:

<http://technet.microsoft.com/en-us/library/gg610600.aspx>



XIV and Linux host connectivity

This chapter addresses the specifics for attaching IBM XIV Storage System to host systems that are running Linux. Although it does not cover every aspect of connectivity, it addresses all of the basics. The examples usually use the Linux console commands because they are more generic than the GUIs provided by vendors.

This guide covers all hardware architectures that are supported for XIV attachment:

- ▶ Intel x86 and x86_64, both Fibre Channel and iSCSI
- ▶ IBM Power Systems™
- ▶ IBM System z®

Older Linux versions are supported to work with the IBM XIV Storage System. However, the scope of this chapter is limited to the most recent enterprise level distributions:

- ▶ Novell SUSE Linux Enterprise Server 11, Service Pack 1 (SLES11 SP1)
- ▶ Red Hat Enterprise Linux 5, Update 6 (RH-EL 5U6).
- ▶ Red Hat Enterprise Linux 6, Update 1 (RH-EL 6U1).

Important: The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, ALWAYS see the System Storage Interoperation Center (SSIC) at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

You can retrieve the Host Attachment Kit publications from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

This chapter includes the following sections:

- ▶ IBM XIV Storage System and Linux support overview
- ▶ Basic host attachment
- ▶ Non-disruptive SCSI reconfiguration
- ▶ Troubleshooting and monitoring
- ▶ Boot Linux from XIV volumes

3.1 IBM XIV Storage System and Linux support overview

Linux is an open source, UNIX-like kernel. The term *Linux* is used in this chapter to mean the whole operating system of GNU/Linux.

3.1.1 Issues that distinguish Linux from other operating systems

Linux is different from the other proprietary operating systems in many ways:

- ▶ There is no one person or organization that can be held responsible or called for support.
- ▶ The distributions differ widely in the amount of support that is available.
- ▶ Linux is available for almost all computer architectures.
- ▶ Linux is rapidly evolving.

All these factors make it difficult to provide generic support for Linux. As a consequence, IBM decided on a support strategy that limits the uncertainty and the amount of testing.

IBM supports only these Linux distributions that are targeted at enterprise clients:

- ▶ Red Hat Enterprise Linux (RH-EL)
- ▶ SUSE Linux Enterprise Server (SLES)

These distributions have major release cycles of about 18 months. They are maintained for five years, and require you to sign a support contract with the distributor. They also have a schedule for regular updates. These factors mitigate the issues that are listed previously. The limited number of supported distributions also allows IBM to work closely with the vendors to ensure interoperability and support. Details about the supported Linux distributions can be found in the SSIC at:

<http://www.ibm.com/systems/support/storage/config/ssic>

3.1.2 Reference material

A wealth of information is available to help you set up your Linux server and attach it to an XIV Storage System. The Linux Host Attachment Kit release notes and user guide contain up-to-date materials. You can retrieve them from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

Primary references

The following are other useful references about Linux distributions:

- ▶ Red Hat Online Storage Reconfiguration Guide for RH-EL5

This guide is part of the documentation that is provided by Red Hat for Red Hat Enterprise Linux 5. Although written specifically for Red Hat Enterprise Linux 5, most of the information is valid for Linux in general. It covers the following topics for Fibre Channel and iSCSI attached devices:

- Persistent device naming
- Dynamically adding and removing storage devices
- Dynamically resizing storage devices
- Low-level configuration and troubleshooting

This publication is available at:

http://docs.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux/5/html/Online_Storage_Reconfiguration_Guide/index.html

► Red Hat Online Storage Administration Guide for RH-EL6

This guide is part of the documentation that is provided by Red Hat for Red Hat Enterprise Linux 6. There were some important changes that were made in this version of Linux:

- **iscsiadm** using `iface.transport` and `iface` configurations
- XIV Host Attachment Kit version 1.7 or later is required

This publication is available at:

http://docs.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux/6/html/Storage_Administration_Guide/index.html

► RH-EL 5 DM Multipath Configuration and Administration

Another part of the Red Hat Enterprise Linux 5 documentation. It contains useful information for anyone who works with Device Mapper Multipathing (DM-MP). Most of the information is valid for Linux in general.

- Understanding how Device Mapper Multipathing works
- Setting up and configuring DM-MP within Red Hat Enterprise Linux 5
- Troubleshooting DM-MP

This publication is available at:

http://docs.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux/5/html/DM_Multipath/index.html

► RH-EL 6 DM Multipath Configuration and Administration

The DM Multipath has the following changes in RH-EL 6:

- The **mpathconf** utility can be used to change the configuration file, **multipathd** daemon, and **chkconfig**.
- The new path selection algorithms **queue-length** and **service-time** provide benefits for certain workloads.
- The location of the bindings file `etc/multipath/bindings` is different.
- Output of **user_friendly_names=yes**, results in `mpathn` being an alphabetic character and not numeric as in RH-EL5.

This publication is available at:

http://docs.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux/6/html/DM_Multipath/index.html

► SLES 11 SP1: Storage Administration Guide

This publication is part of the documentation for Novell SUSE Linux Enterprise Server 11, Service Pack 1. Although written specifically for SUSE Linux Enterprise Server, it contains useful information for any Linux user who is interested in storage-related subjects. The following are the most useful topics in the book:

- Setting up and configuring multipath I/O
- Setting up a system to boot from multipath devices
- Combining multipathing with Logical Volume Manager and Linux Software RAID

This publication is available at:

http://www.novell.com/documentation/sles11/stor_admin/?page=/documentation/sles11/stor_admin/data/bookinfo.html

► IBM Linux for Power Architecture® wiki

This wiki site hosted by IBM contains information about Linux on Power Architecture. It includes the following sections:

- A discussion forum
- An announcement section
- Technical articles

It can be found at:

<https://www.ibm.com/developerworks/wikis/display/LinuxP/Home>

► *Fibre Channel Protocol for Linux and z/VM on IBM System z, SG24-7266*

This is a comprehensive guide to storage attachment using Fibre Channel to z/VM and Linux on z/VM. It describes the following concepts:

- General Fibre Channel Protocol (FCP) concepts
- Setting up and using FCP with z/VM and Linux
- FCP naming and addressing schemes
- FCP devices in the 2.6 Linux kernel
- N-Port ID Virtualization
- FCP Security topics

It is available at:

<http://www.redbooks.ibm.com/abstracts/sg247266.html>

Other sources of information

The Linux distributor documentation pages are good starting points when it comes to installation, configuration, and administration of Linux servers. These documentation pages are especially useful for server-platform-specific issues.

► Documentation for Novell SUSE Linux Enterprise Server is available at:

<http://www.novell.com/documentation/suse.html>

► Documentation for Red Hat Enterprise Linux is available at:

<http://www.redhat.com/docs/manuals/enterprise/>

IBM System z has its own web page to storage attachment using FCP at:

<http://www.ibm.com/systems/z/connectivity/products/>

The *IBM System z Connectivity Handbook*, SG24-5444, describes the connectivity options available for use within and beyond the data center for IBM System z servers. It has a section for FC attachment, although it is outdated with regards to multipathing. You can download this book at:

<http://www.redbooks.ibm.com/redbooks.nsf/RedbookAbstracts/sg245444.html>

3.1.3 Recent storage-related improvements to Linux

This section provides a summary of storage-related improvements that have been recently introduced to Linux. Details about usage and configuration are available in the subsequent sections.

Past issues

The following is a partial list of storage-related issues in older Linux versions that are addressed in recent versions:

- ▶ Limited number of devices that could be attached
- ▶ Gaps in LUN sequence that led to incomplete device discovery
- ▶ Limited dynamic attachment of devices
- ▶ Non-persistent device naming that might lead to reordering
- ▶ No native multipathing

Dynamic generation of device nodes

Linux uses special files, also called device nodes or special device files, for access to devices. In earlier versions, these files were created statically during installation. The creators of a Linux distribution had to anticipate all devices that would ever be used for a system and create nodes for them. This process often led to a confusing number of existing nodes and missing ones.

In recent versions of Linux, two new subsystems were introduced, *hotplug* and *udev*. Hotplug detects and registers newly attached devices without user intervention. Udev dynamically creates the required device nodes for the newly attached devices according to predefined rules. In addition, the range of major and minor numbers, the representatives of devices in the kernel space, was increased. These numbers are now dynamically assigned.

With these improvements, the required device nodes exist immediately after a device is detected. In addition, only device nodes that are needed are defined.

Persistent device naming

As mentioned, udev follows predefined rules when it creates the device nodes for new devices. These rules are used to define device node names that relate to certain device characteristics. For a disk drive or SAN-attached volume, this name contains a string that uniquely identifies the volume. This string ensures that every time this volume is attached to the system, it gets the same name.

Multipathing

Linux has its own built-in multipathing solution. It is based on *Device Mapper*, a block device virtualization layer in the Linux kernel. Therefore, it is called *Device Mapper Multipathing* (DM-MP). The Device Mapper is also used for other virtualization tasks, such as the logical volume manager, data encryption, snapshots, and software RAID.

DM-MP overcomes these issues that are caused by proprietary multipathing solutions:

- ▶ Proprietary multipathing solutions were only supported for certain kernel versions. Therefore, systems followed the update schedule of the distribution.
- ▶ They were often binary only. Linux vendors did not support them because they were not able to debug them.
- ▶ A mix of different storage systems on the same server usually was not possible because the multipathing solutions could not coexist.

Today, DM-MP is the only multipathing solution that is fully supported by both Red Hat and Novell for their enterprise Linux distributions. It is available on all hardware systems, and supports all block devices that can have more than one path. IBM supports DM-MP wherever possible.

Add and remove volumes online

With the new hotplug and udev subsystems, it is now possible to easily add and remove disks from Linux. SAN-attached volumes are usually not detected automatically. Adding a volume to an XIV host object does not create a hotplug trigger event like inserting a USB storage device does. SAN-attached volumes are discovered during user-initiated device scans. They are then automatically integrated into the system, including multipathing.

To remove a disk device, make sure that it is not used anymore, then remove it logically from the system before you physically detach it.

Dynamic LUN resizing

Improvements were recently introduced to the SCSI layer and DM-MP that allow resizing of SAN-attached volumes while they are in use. However, these capabilities are limited to certain cases.

Write Barrier availability for ext4 file system

RHEL6 by default uses the ext4 file system. This file system uses the new *Write Barriers* feature to improve performance. A write barrier is a kernel mechanism that is used to ensure that file system metadata is correctly written and ordered on persistent storage. The write barrier continues to do so even when storage devices with volatile write caches lose power.

Write barriers are implemented in the Linux kernel by using storage write cache flushes before and after the I/O, which is order-critical. After the transaction is written, the storage cache is flushed, the commit block is written, and the cache is flushed again. The constant flush of caches can significantly reduce performance. You can disable write barriers at mount time by using the `-o nobarrier` option for `mount`.

Important: IBM has confirmed that Write Barriers have a negative impact on XIV performance. Ensure that all of your mounted disks use the following switch:

```
mount -o nobarrier /fs
```

For more information, see the Red Hat Linux Enterprise 6 Storage documentation at:

http://docs.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux/6/html/Storage_Administration_Guide/

3.2 Basic host attachment

This section addresses the steps to make XIV volumes available to your Linux host. It addresses attaching storage for the different hardware architectures. It also describes configuration of the Fibre Channel HBA driver, setting up multipathing, and any required special settings.

3.2.1 Platform-specific remarks

The most popular hardware system for Linux is the Intel x86 (32 or 64 bit) architecture. However, this architecture allows only direct mapping of XIV volumes to hosts through Fibre

Channel fabrics and HBAs, or IP networks. IBM System z and IBM Power Systems provide extra mapping methods so you can use their much more advanced virtualization capabilities.

IBM Power Systems

Linux, running in a logical partition (LPAR) on an IBM Power system, can get storage from an XIV through one of these methods:

- ▶ Directly through an exclusively assigned Fibre Channel HBA
- ▶ Through a Virtual I/O Server (VIOS) running on the system

Direct attachment is not described because it works the same way as with the other systems. VIOS attachment requires specific considerations. For more information about how VIOS works and how it is configured, see Chapter 8, “IBM i and AIX clients connecting to XIV through VIOS” on page 211. More information is available in the following IBM Redbooks:

- ▶ *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
<http://www.redbooks.ibm.com/abstracts/sg247940.html>
- ▶ *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
<http://www.redbooks.ibm.com/abstracts/sg247590.html>

Virtual vscsi disks through VIOS

Linux on Power distributions contain a kernel module (driver) for a virtual SCSI HBA. This driver is called `ibmvscsi`, and attaches the virtual disks that are provided by the VIOS to the Linux system. The devices as seen by the Linux system are shown in Example 3-1.

Example 3-1 Virtual SCSI disks

```
p6-570-lpar13:~ # lsscsi
[0:0:1:0]   disk    AIX      VDASD      0001   /dev/sda
[0:0:2:0]   disk    AIX      VDASD      0001   /dev/sdb
```

In this example, the SCSI vendor ID is AIX, and the device model is VDASD. Apart from that, they are treated like any other SCSI disk. If you run a redundant VIOS setup on the system, the virtual disks can be attached through both servers. They then show up twice, and must be managed by DM-MP to ensure data integrity and path handling.

Virtual Fibre Channel adapters through NPIV

IBM PowerVM® is the hypervisor of the IBM Power system. It uses the N-Port ID Virtualization (NPIV) capabilities of modern SANs and Fibre Channel HBAs. These capacities allow PowerVM to provide virtual HBAs for the LPARs. The mapping of these HBAs is done by the VIOS.

Virtual HBAs register to the SAN with their own *worldwide port names* (WWPNs). To the XIV they look exactly like physical HBAs. You can create Host Connections for them and map volumes. This process allows easier, more streamlined storage management, and better isolation of the LPAR in an IBM Power system.

Linux on Power distributions come with a kernel module for the virtual HBA called `ibmvfc`. This module presents the virtual HBA to the Linux operating system as though it were a real FC HBA. XIV volumes that are attached to the virtual HBA are displayed as though they are connected through a physical adapter (Example 3-2).

Example 3-2 Volumes that are mapped through NPIV virtual HBAs

```
p6-570-lpar13:~ # lsscsi
[1:0:0:0]   disk    IBM      2810XIV    10.2   /dev/sdc
```

[1:0:0:1]	disk	IBM	2810XIV	10.2	/dev/sdd
[1:0:0:2]	disk	IBM	2810XIV	10.2	/dev/sde
[2:0:0:0]	disk	IBM	2810XIV	10.2	/dev/sdm
[2:0:0:1]	disk	IBM	2810XIV	10.2	/dev/sdn
[2:0:0:2]	disk	IBM	2810XIV	10.2	/dev/sdo

To maintain redundancy, you usually use more than one virtual HBA, each one running on a separate real HBA. Therefore, XIV volumes show up more than once (once per path) and must be managed by a DM-MP.

System z

Linux running on an IBM System z server has the following storage attachment choices:

- ▶ Linux on System z running natively in a System z LPAR
- ▶ Linux on System z running in a virtual machine under z/VM

Linux on System z running natively in a System z LPAR

When you run Linux on System z directly on a System z LPAR, there are two ways to attach disk storage.

Tip: In IBM System z, the term “adapters” is better suited than the more common “channels” term that is often used in the System z environment.

The Fibre Channel connection (IBM FICON®) channel in a System z server can operate individually in *Fibre Channel Protocol* (FCP) mode. FCP transports SCSI commands over the Fibre Channel interface. It is used in all open systems implementations for SAN-attached storage. Certain operating systems that run on a System z mainframe can use this FCP capability to connect directly to fixed block (FB) storage devices. Linux on System z provides the kernel module `zfc` to operate the FICON adapter in FCP mode. A channel can run either in FCP or FICON mode. Channels can be shared between LPARs, and multiple ports on an adapter can run in different modes.

To maintain redundancy, you usually use more than one FCP channel to connect to the XIV volumes. Linux sees a separate disk device for each path, and needs DM-MP to manage them.

Linux on System z running in a virtual machine under z/VM

Running a number of virtual Linux instances in a z/VM environment is a common solution. z/VM provides granular and flexible assignment of resources to the virtual machines (VMs). You can also use it to share resources between VMs. z/VM offers even more ways to connect storage to its VMs:

- ▶ Fibre Channel (FCP) attached SCSI devices

z/VM can assign a Fibre Channel card that runs in FCP mode to a VM. A Linux instance that runs in this VM can operate the card by using the `zfc` driver and access the attached XIV FB volumes.

To maximize use of the FCP channels, share them between more than one VM. However, z/VM cannot assign FCP attached volumes individually to virtual machines. Each VM can theoretically access all volumes that are attached to the shared FCP adapter. The Linux instances that run in the VMs must ensure that each VM uses only the volumes that it is supposed to.

- ▶ FCP attachment of SCSI devices through NPIV

To overcome the issue described previously, *N_Port ID Virtualization (NPIV)* was introduced for System z, z/VM, and Linux on System z. It allows creation of multiple virtual Fibre Channel HBAs running on a single physical HBA. These virtual HBAs are assigned individually to virtual machines. They log on to the SAN with their own WWPNs. To the XIV, they look exactly like physical HBAs. You can create Host Connections for them and map volumes. This process allows you to assign XIV volumes directly to the Linux virtual machine. No other instance can access these HBAs, even if it uses the same physical adapter.

Tip: Linux on System z can also use *count-key-data devices (CKDs)*. CKDs are the traditional mainframe method to access disks. However, the XIV Storage System does not support the CKD protocol, so it is not described in this book.

3.2.2 Configuring for Fibre Channel attachment

This section describes how Linux is configured to access XIV volumes. A *Host Attachment Kit* is available for the Intel x86 system to ease the configuration. Therefore, many of the manual steps that are described are only necessary for the other supported systems. However, the description might be helpful because it provides insight in the Linux storage stack. It is also useful if you must resolve a problem.

Loading the Linux Fibre Channel drivers

There are four main brands of *Fibre Channel host bus adapters (FC HBAs)*:

- ▶ QLogic: The most used HBAs for Linux on the Intel X86 system. The kernel module `qla2xxx` is a unified driver for all types of QLogic FC HBAs. It is included in the enterprise Linux distributions. The shipped version is supported for XIV attachment.
- ▶ Emulex: Sometimes used in Intel x86 servers and, rebranded by IBM, the standard HBA for the Power platform. The kernel module `lpfc` is a unified driver that works with all Emulex FC HBAs. A supported version is also included in the enterprise Linux distributions for both Intel x86 and Power Systems.
- ▶ Brocade: Provides *Converged Network Adapters (CNAs)* that operate as FC and Ethernet adapters, which are relatively new to the market. They are supported on Intel x86 for FC attachment to the XIV. The kernel module version that is provided with the current enterprise Linux distributions is not supported. You must download the supported version from the Brocade website. The driver package comes with an installation script that compiles and installs the module. The script might cause support issues with your Linux distributor because it modifies the kernel. The FC kernel module for the CNAs is called `bfa`. The driver can be downloaded at:
<http://www.brocade.com/services-support/drivers-downloads/index.page>
- ▶ IBM FICON Express: These are the HBAs for the System z system. They can either operate in FICON mode for traditional CKD devices, or FCP mode for FB devices. Linux deals with them directly only in FCP mode. The driver is part of the enterprise Linux distributions for System z, and is called `zfc`.

Kernel modules (drivers) are loaded with the `modprobe` command. They can be removed if they are not in use as shown in Example 3-3.

Example 3-3 Loading and unloading a Linux Fibre Channel HBA Kernel module

```
x36501ab9:~ # modprobe qla2xxx
x36501ab9:~ # modprobe -r qla2xxx
```

After the driver is loaded, the FC HBA driver examines the FC fabric, detects attached volumes, and registers them in the operating system. To discover whether a driver is loaded or not, and what dependencies exist for it, use the command **lsmod** (Example 3-4).

Example 3-4 Filter list of running modules for a specific name

```
x36501ab9:~ #lsmod | tee >(head -n 1) >(grep qla) > /dev/null
Module                Size  Used by
qla2xxx                293455  0
scsi_transport_fc     54752  1 qla2xxx
scsi_mod               183796  10 qla2xxx,scsi_transport_fc,scsi_tgt,st,ses, ....
```

To get detailed information about the Kernel module itself, such as the version number and what options it supports, use the **modinfo** command. You can see a partial output in Example 3-5.

Example 3-5 Detailed information about a specific kernel module

```
x36501ab9:~ # modinfo qla2xxx
filename:
/lib/modules/2.6.32.12-0.7-default/kernel/drivers/scsi/qla2xxx/qla2xxx.ko
...
version:      8.03.01.06.11.1-k8
license:      GPL
description:  QLogic Fibre Channel HBA Driver
author:       QLogic Corporation
...
depends:       scsi_mod,scsi_transport_fc
supported:    yes
vermagic:     2.6.32.12-0.7-default SMP mod_unload modversions
parm:         ql2xlogintimeout:Login timeout value in seconds. (int)
parm:         qlport_down_retry:Maximum number of command retries to a port ...
parm:         ql2xplogiabsentdevice:Option to enable PLOGI to devices that ...
...
```

Restriction: The zfcpl driver for Linux on System z automatically scans and registers the attached volumes, but only in the most recent Linux distributions and only if NPIV is used. Otherwise, you must tell it explicitly which volumes to access. The reason is that the Linux virtual machine might not be intended to use all volumes that are attached to the HBA. For more information, see “Linux on System z running in a virtual machine under z/VM” on page 94, and “Adding XIV volumes to a Linux on System z system” on page 106.

Using the FC HBA driver at installation time

You can use XIV volumes that are already attached to a Linux system at installation time. Using already attached volumes allows you to install all or part of the system to the SAN-attached volumes. The Linux installers detect the FC HBAs, load the necessary kernel modules, scan for volumes, and offer them in the installation options.

When you have an unsupported driver version included with your Linux distribution, either replace it immediately after installation, or use a driver disk during the installation. This issue is current for Brocade HBAs. A driver disk image is available for download from the Brocade website. For more information, see “Loading the Linux Fibre Channel drivers” on page 95.

Considerations:

- ▶ Installing a Linux system on a SAN-attached disk does not mean that it is able to start from it. Usually you must complete more steps to configure the boot loader or boot program.
- ▶ You must take special precautions about multipathing if you want to run Linux on SAN-attached disks.

For more information, see 3.5, “Boot Linux from XIV volumes” on page 131.

Making the FC driver available early in the boot process

If the SAN-attached XIV volumes are needed early in the Linux boot process, include the HBA driver into the *Initial RAM file system* (initramfs) image. You must include this driver, for example, if all or part of the system is on these volumes. The initramfs allows the Linux boot process to provide certain system resources before the real system disk is set up.

Linux distributions contain a script that is called `mkinitrd` that creates the initramfs image automatically. They automatically include the HBA driver if you already used a SAN-attached disk during installation. If not, you must include it manually. The ways to tell `mkinitrd` to include the HBA driver differ depending on the Linux distribution used.

Tip: The *initramfs* was introduced years ago and replaced the *Initial RAM Disk* (initrd). People sometimes say `initrd` when they actually mean `initramfs`.

SUSE Linux Enterprise Server

Kernel modules that must be included in the initramfs are listed in the file `/etc/sysconfig/kernel` on the line that starts with `INITRD_MODULES`. The order that they show up on this line is the order that they are loaded at system startup (Example 3-6).

Example 3-6 Telling SLES to include a kernel module in the initramfs

```
x36501ab9:~ # cat /etc/sysconfig/kernel
...
# This variable contains the list of modules to be added to the initial
# ramdisk by calling the script "mkinitrd"
# (like drivers for scsi-controllers, for lvm or reiserfs)
#
INITRD_MODULES="thermal aacraid ata_piix ... processor fan jbd ext3 edd qla2xxx"
...
```

After you add the HBA driver module name to the configuration file, rebuild the initramfs with the `mkinitrd` command. This command creates and installs the image file with standard settings and to standard locations as illustrated in Example 3-7.

Example 3-7 Creating the initramfs

```
x36501ab9:~ # mkinitrd

Kernel image:  /boot/vmlinuz-2.6.32.12-0.7-default
Initrd image:  /boot/initrd-2.6.32.12-0.7-default
Root device:   /dev/disk/by-id/scsi-SServeRA_Drive_1_2D0DE908-part1 (/dev/sda1)..
Resume device: /dev/disk/by-id/scsi-SServeRA_Drive_1_2D0DE908-part3 (/dev/sda3)
Kernel Modules: hwmon thermal_sys ... scsi_transport_fc qla2xxx ...
```

```
(module qla2xxx.ko firmware /lib/firmware/ql2500_fw.bin) (module qla2xxx.ko ...
Features:      block usb resume.userspace resume.kernel
Bootsplash:   SLES (800x600)
30015 blocks
```

If you need nonstandard settings, for example a different image name, use parameters for **mkinitrd**. For more information, see the man page for **mkinitrd** on your Linux system.

Red Hat Enterprise Linux 5 (RH-EL5)

Kernel modules that must be included in the **initramfs** are listed in the file `/etc/modprobe.conf`. The order that they show up in the file is the order that they are loaded at system startup as seen in Example 3-8.

*Example 3-8 Telling RH-EL to include a kernel module in the **initramfs***

```
[root@x36501ab9 ~]# cat /etc/modprobe.conf
```

```
alias eth0 bnx2
alias eth1 bnx2
alias eth2 e1000e
alias eth3 e1000e
alias scsi_hostadapter aacraid
alias scsi_hostadapter1 ata_piix
alias scsi_hostadapter2 qla2xxx
alias scsi_hostadapter3 usb-storage
```

After you add the HBA driver module to the configuration file, rebuild the **initramfs** with the **mkinitrd** command. The Red Hat version of **mkinitrd** requires as the following information as parameters (Example 3-9):

- ▶ The name of the image file to create
- ▶ The location of the image file
- ▶ The kernel version that the image file is built for

*Example 3-9 Creating the **initramfs***

```
[root@x36501ab9 ~]# mkinitrd /boot/initrd-2.6.18-194.el5.img 2.6.18-194.el5
```

If the image file with the specified name exists, use the **-f** option to force **mkinitrd** to overwrite the existing one. The command shows more detailed output with the **-v** option.

You can discover the kernel version that is running on the system with the **uname** command as illustrated in Example 3-10.

Example 3-10 Determining the kernel version

```
[root@x36501ab9 ~]# uname -r
2.6.18-194.el5
```

Red Hat Enterprise Linux 6 (RH-EL6)

Dracut is a new utility for RH-EL6 that is important to the boot process. In previous versions of RH-EL, the initial RAM disk image preinstalled the block device modules, such as for SCSI or RAID. The root file system, on which those modules are normally located, can then be accessed and mounted.

With Red Hat Enterprise Linux 6 (RH-EL6) systems, the **dracut** utility is always called by the installation scripts to create an `initramfs`. This process occurs whenever a new kernel is installed by using the Yum, PackageKit, or Red Hat Package Manager (RPM).

On all architectures other than IBM i, you can create an `initramfs` by running the **dracut** command. However, you usually do not need to create an `initramfs` manually. This step is automatically completed if the kernel and its associated packages are installed or upgraded from the RPM packages that are distributed by Red Hat.

Verify that an `initramfs` corresponding to your current kernel version exists and is specified correctly in the `grub.conf` configuration file by using the following procedure:

1. As root, list the contents in the `/boot/` directory.
2. Find the kernel (`vmlinuz-<kernel_version>`) and `initramfs-<kernel_version>` with the most recent version number, as shown in Figure 3-1.

```
[root@bc-h-15-b7 ~]# ls /boot/
config-2.6.32-131.0.15.el6.x86_64
efi
grub
initramfs-2.6.32-131.0.15.el6.x86_64.img
initrd-2.6.32-131.0.15.el6.x86_64kdump.img
lost+found
symvers-2.6.32-131.0.15.el6.x86_64.gz
System.map-2.6.32-131.0.15.el6.x86_64
vmlinuz-2.6.32-131.0.15.el6.x86_64
```

Figure 3-1 Red Hat 6 (RH-EL6) display of matching `initramfs` and kernel

Optionally, if your `initramfs-<kernel_version>` file does not match the version of the latest kernel in `/boot/`, generate an `initramfs` file with the **dracut** utility. Starting **dracut** as root, without options generates an `initramfs` file in the `/boot/` directory for the latest kernel present in that directory. For more information about options and usage, see **man dracut** and **man dracut.conf**.

On IBM i servers, the initial RAM disk and kernel files are combined into a single file that is created with the **addRamDisk** command. This step is completed automatically if the kernel and its associated packages are installed or upgraded from the RPM packages that are distributed by Red Hat. Therefore, it does not need to be run manually.

To verify that it was created, use the command `ls -l /boot/` and make sure that the `/boot/vmlinitrd-<kernel_version>` file exists. The `<kernel_version>` must match the version of the installed kernel.

3.2.3 Determining the WWPN of the installed HBAs

To create a host port on the XIV that can map volumes to an HBA, you need the WWPN of the HBA. The WWPN is shown in `sysfs`, a Linux pseudo file system that reflects the installed hardware and its configuration. Example 3-11 shows how to discover which SCSI host instances are assigned to the installed FC HBAs. You can then determine their WWPNs.

Example 3-11 Finding the WWPNs of the FC HBAs

```
[root@x3650lab9 ~]# ls /sys/class/scsi_host/
host1 host2
```

```
# cat /sys/class/fc_host/host1/port_name
0x10000000c93f2d32
# cat /sys/class/fc_host/host2/port_name
0x10000000c93d64f5
```

Map volumes to a Linux host as described in 1.4, “Logical configuration for host connectivity” on page 37.

Tip: For Intel host systems, the XIV Host Attachment Kit can create the XIV host and host port objects for you automatically from the Linux operating system. For more information, see 3.2.4, “Attaching XIV volumes to an Intel x86 host using the Host Attachment Kit” on page 100.

3.2.4 Attaching XIV volumes to an Intel x86 host using the Host Attachment Kit

You can attach the XIV volumes to an Intel x86 host by using a Host Attachment Kit.

Installing the Host Attachment Kit

For multipathing with Linux, IBM XIV provides a *Host Attachment Kit*. This section explains how to install the Host Attachment Kit on a Linux server.

Consideration: Although it is possible to configure Linux on Intel x86 servers manually for XIV attachment, use the Host Attachment Kit. The Host Attachment Kit and its binary files are required in case you need support from IBM. The kit provides data collection and troubleshooting tools.

At the time of writing, Host Attachment Kit version 1.7 is the minimum required version for RH-EL6.

Some additional troubleshooting checklists and tips are available in 3.4, “Troubleshooting and monitoring” on page 125.

Download the latest Host Attachment Kit for Linux from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

To install the Host Attachment Kit, extra Linux packages are required. These software packages are supplied on the installation media of the supported Linux distributions. If required software packages are missing on your host, the installation terminates. You are notified of the missing package.

The required packages are listed in Figure 3-2.

RHEL	SLES 10	SLES 11
device-mapper-multipath	multipath-tools	multipath-tools
sg3_utils	sg3_utils	scsi
optional for iSCSI		
iscsi-initiator-utils	open-iscsi	open-iscsi

Figure 3-2 Required Linux packages

Ensure that all of the listed packages are installed on your Linux system before you install the Host Attachment Kit.

To install the Host Attachment Kit, complete the following steps:

1. Copy the downloaded package to your Linux server
2. Open a terminal session
3. Change to the directory where the package is located.
4. Unpack and install Host Attachment Kit by using the commands that are shown in Example 3-12.

Consideration: Illustrations and example are based on version 1.7, a former version of the Linux Host Attachment kit.

Example 3-12 Installing the Host Attachment Kit package

```
# tar -zxvf XIV_host_attach-1.7-sles11-x86.tar.gz
# cd XIV_host_attach-1.7-sles11-x86
# /bin/sh ./install.sh
```

The name of the archive, and thus the name of the directory that is created when you unpack it, differs depending on the following items:

- ▶ Your Host Attachment Kit version
- ▶ Linux distribution
- ▶ Hardware platform

The installation script prompts you for this information. After you run the script, review the installation log file `install.log` in the same directory.

The Host Attachment Kit provides the utilities that you need to configure the Linux host for XIV attachment. They are in the `/opt/xiv/host_attach` directory.

Remember: You must be logged in as root or have root privileges to use the Host Attachment Kit. The Host Attachment Kit uses Python for both the installation and uninstallation actions. Python is part of most installation distributions.

The main executable files and scripts are in the directory `/opt/xiv/host_attach/bin`. The installation script includes this directory in the command search path of the user root. Therefore, the commands can be run from every working directory.

Configuring the host for Fibre Channel using the Host Attachment Kit

Use the `xiv_attach` command to configure the Linux host. You can also create the XIV host object and host ports on the XIV itself. To do so, you must have a user ID and password for an XIV storage administrator account. Example 3-13 illustrates how `xiv_attach` works for Fibre Channel attachment. Your output can differ depending on your configuration.

Example 3-13 Fibre Channel host attachment configuration using the `xiv_attach` command

```
[/]# xiv_attach
-----
Welcome to the XIV Host Attachment wizard, version 1.7.0
This wizard will assist you to attach this host to the XIV system.

The wizard will now validate host configuration for the XIV system.
Press [ENTER] to proceed.
-----
iSCSI software was not detected. see the guide for more info.
Only fibre-channel is supported on this host.
Would you like to set up an FC attachment? [default: yes ]: yes
-----
Please wait while the wizard validates your existing configuration...
The wizard needs to configure the host for the XIV system.
Do you want to proceed? [default: yes ]: yes
Please wait while the host is being configured...
The host is now being configured for the XIV system
-----
Please zone this host and add its WWPNs with the XIV storage system:
10:00:00:00:c9:3f:2d:32: [EMULEX]: N/A
10:00:00:00:c9:3d:64:f5: [EMULEX]: N/A
Press [ENTER] to proceed.

Would you like to rescan for new storage devices now? [default: yes ]: yes
Please wait while rescanning for storage devices...
-----
The host is connected to the following XIV storage arrays:
Serial   Ver   Host Defined  Ports Defined  Protocol  Host Name(s)
1300203  10.2  No           None           FC       --
This host is not defined on some of the FC-attached XIV storage systems
Do you wish to define this host these systems now? [default: yes ]: yes

Please enter a name for this host [default: tic-17.mainz.de.ibm.com ]:
Please enter a username for system 1300203 : [default: admin ]: itso
Please enter the password of user itso for system 1300203:*****
Press [ENTER] to proceed.
-----
The XIV host attachment wizard successfully configured this host
Press [ENTER] to exit.
#.
```

Configuring the host for iSCSI using the Host Attachment Kit

Use the `xiv_attach` command to configure the host for iSCSI attachment of XIV volumes. First, make sure that the iSCSI service is running as illustrated in Figure 3-3.

```
#service iscsi start
```

Figure 3-3 Ensuring that the iSCSI service is running

Example 3-14 shows example output when you run `xiv_attach`. Again, your output can differ depending on your configuration.

Example 3-14 iSCSI host attachment configuration using the `xiv_attach` command

```
[/]# xiv_attach
-----
Welcome to the XIV Host Attachment wizard, version 1.7.0.
This wizard will assist you to attach this host to the XIV system.

The wizard will now validate host configuration for the XIV system.
Press [ENTER] to proceed.

-----
Please choose a connectivity type, [f]c / [i]scsi : i
-----
Please wait while the wizard validates your existing configuration...
This host is already configured for the XIV system
-----
Would you like to discover a new iSCSI target? [default: yes ]:
Enter an XIV iSCSI discovery address (iSCSI interface): 9.155.90.183
Is this host defined in the XIV system to use CHAP? [default: no ]: no
Would you like to discover a new iSCSI target? [default: yes ]: no
Would you like to rescan for new storage devices now? [default: yes ]: yes

-----
The host is connected to the following XIV storage arrays:
Serial   Ver   Host Defined  Ports Defined  Protocol  Host Name(s)
1300203  10.2  No            None           FC        --

This host is not defined on some of the iSCSI-attached XIV storage systems.
Do you wish to define this host these systems now? [default: yes ]: yes
Please enter a name for this host [default: tic-17.mainz.de.ibm.com]: tic-17_iscsi
Please enter a username for system 1300203 : [default: admin ]: itso
Please enter the password of user itso for system 1300203:*****

Press [ENTER] to proceed.

-----
The XIV host attachment wizard successfully configured this host

Press [ENTER] to exit.
```

3.2.5 Checking attached volumes

The Host Attachment Kit provides tools to verify mapped XIV volumes. You can also use native Linux commands to do so.

Example 3-15 shows using the Host Attachment Kit to verify the volumes for an iSCSI attached volume. The `xiv_devlist` command lists all XIV devices that are attached to a host.

Example 3-15 Verifying mapped XIV LUNs using the Host Attachment Kit tool with iSCSI

```
[/]# xiv_devlist
XIV Devices
-----
Device                Size  Paths  Vol Name  Vol Id  XIV Id  XIV Host
```

```
-----  
/dev/mapper/mpath0 17.2GB 4/4   residency 1428   1300203 tic-17_iscsi  
-----
```

Non-XIV Devices

...

Tip: The `xiv_attach` command already enables and configures multipathing. Therefore, the `xiv_devlist` command shows only multipath devices.

If you want to see the individual devices that represent each of the paths to an XIV volume, use the `lsscsi` command. This command shows any XIV volumes that are attached to the Linux system.

Example 3-16 shows that Linux recognized 16 XIV devices. By looking at the SCSI addresses in the first column, you can determine that there actually are four XIV volumes. Each volume is connected through four paths. Linux creates a SCSI disk device for each of the paths.

Example 3-16 Listing attached SCSI devices

```
[root@x3650lab9 ~]# lsscsi  
[0:0:0:1]   disk    IBM      2810XIV    10.2 /dev/sda  
[0:0:0:2]   disk    IBM      2810XIV    10.2 /dev/sdb  
[0:0:0:3]   disk    IBM      2810XIV    10.2 /dev/sdg  
[1:0:0:1]   disk    IBM      2810XIV    10.2 /dev/sdc  
[1:0:0:2]   disk    IBM      2810XIV    10.2 /dev/sdd  
[1:0:0:3]   disk    IBM      2810XIV    10.2 /dev/sde  
[1:0:0:4]   disk    IBM      2810XIV    10.2 /dev/sdf
```

Requirement: The RH-EL installer does not install `lsscsi` by default. It is shipped with the distribution, but must be selected explicitly for installation.

Linux SCSI addressing explained

The quadruple in the first column of the `lsscsi` output is the internal Linux SCSI address. It is, for historical reasons, like a traditional parallel SCSI address. It consists of these fields:

- ▶ **HBA ID:** Each HBA in the system gets a host adapter instance when it is initiated. The instance is assigned regardless of whether it is parallel SCSI, Fibre Channel, or even a SCSI emulator.
- ▶ **Channel ID:** This field is always zero. It was formerly used as an identifier for the channel in multiplexed parallel SCSI HBAs.
- ▶ **Target ID:** For parallel SCSI, this is the real target ID that you set by using a jumper on the disk drive. For Fibre Channel, it represents a remote port that is connected to the HBA. This ID distinguishes between multiple paths, and between multiple storage systems.
- ▶ **LUN:** LUNs (logical unit numbers) are rarely used in parallel SCSI. In Fibre Channel, they are used to represent a single volume that a storage system offers to the host. The LUN is assigned by the storage system.

Figure 3-4 illustrates how the SCSI addresses are generated.

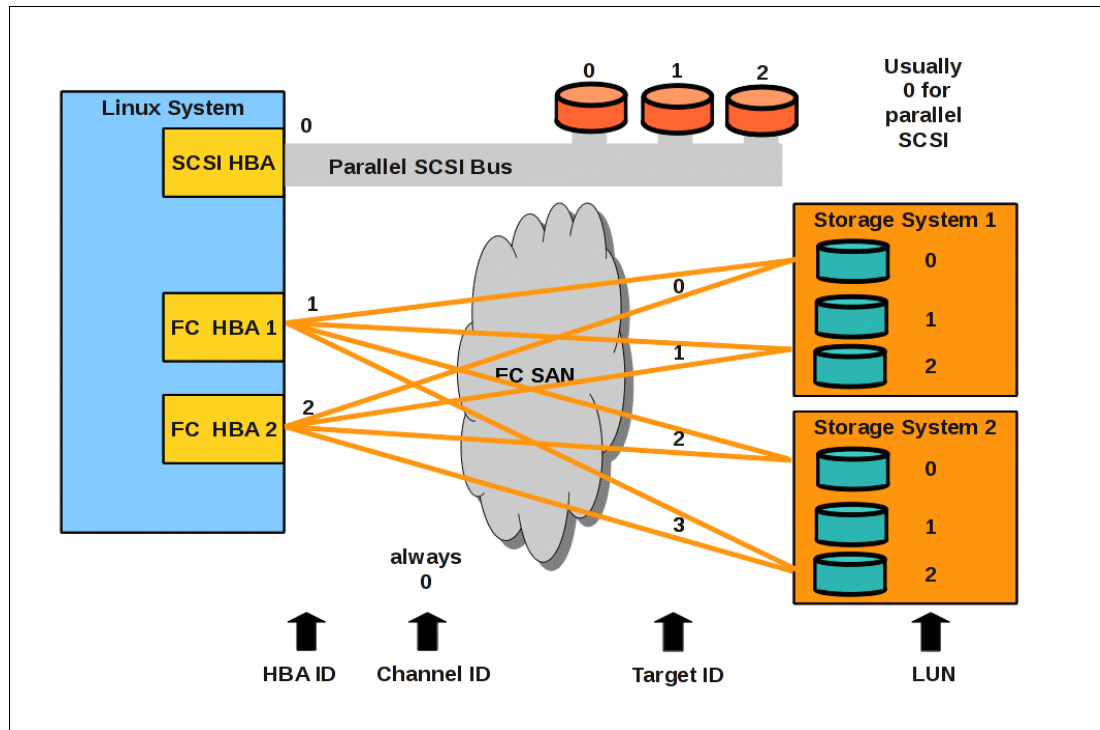


Figure 3-4 Composition of Linux internal SCSI addresses

Identifying a particular XIV Device

The udev subsystem creates device nodes for all attached devices. For disk drives, it not only sets up the traditional `/dev/sdx` nodes, but also some other representatives. The most useful ones can be found in `/dev/disk/by-id` and `/dev/disk/by-path`.

The nodes for XIV volumes in `/dev/disk/by-id` show a unique identifier. This identifier is composed of parts of the following numbers (Example 3-17):

- ▶ The worldwide node name (WWNN) of the XIV system
- ▶ The XIV volume serial number in hexadecimal notation

Example 3-17 The `/dev/disk/by-id` device nodes

```
x36501ab9:~ # ls -l /dev/disk/by-id/ | cut -c 44-
...
scsi-2001738000cb051f -> ../../sde
scsi-2001738000cb0520 -> ../../sdf
scsi-2001738000cb2d57 -> ../../sdb
scsi-2001738000cb3af9 -> ../../sda
scsi-2001738000cb3af9-part1 -> ../../sda1
scsi-2001738000cb3af9-part2 -> ../../sda2
...
```

Remember: The WWNN of the XIV system that is used in the examples is `0x5001738000cb0000`. It has three zeros between the vendor ID and the system ID, whereas the representation in `/dev/disk/by-id` has four zeros

The XIV volume with the serial number **0x3af9** has two partitions. It is the system disk. Partitions show up in Linux as individual block devices.

The udev subsystem already recognizes that there is more than one path to each XIV volume. It creates only one node for each volume instead of four.

Important: The device nodes in `/dev/disk/by-id` are persistent, whereas the `/dev/sdx` nodes are not. They can change when the hardware configuration changes. Do not use `/dev/sdx` device nodes to mount file systems or specify system disks.

The `/dev/disk/by-path` file contains nodes for all paths to all XIV volumes. Here you can see the physical connection to the volumes. This connection starts with the PCI identifier of the HBAs through the remote port, represented by the XIV WWPN, to the LUN of the volumes (Example 3-18).

Example 3-18 The `/dev/disk/by-path` device nodes

```
x36501ab9:~ # ls -l /dev/disk/by-path/ | cut -c 44-
...
pci-0000:1c:00.0-fc-0x5001738000cb0191:0x0001000000000000 -> ../../sda
pci-0000:1c:00.0-fc-0x5001738000cb0191:0x0001000000000000-part1 -> ../../sda1
pci-0000:1c:00.0-fc-0x5001738000cb0191:0x0001000000000000-part2 -> ../../sda2
pci-0000:1c:00.0-fc-0x5001738000cb0191:0x0002000000000000 -> ../../sdb
pci-0000:1c:00.0-fc-0x5001738000cb0191:0x0003000000000000 -> ../../sdg
pci-0000:1c:00.0-fc-0x5001738000cb0191:0x0004000000000000 -> ../../sdh
pci-0000:24:00.0-fc-0x5001738000cb0160:0x0001000000000000 -> ../../sdc
pci-0000:24:00.0-fc-0x5001738000cb0160:0x0001000000000000-part1 -> ../../sdc1
pci-0000:24:00.0-fc-0x5001738000cb0160:0x0001000000000000-part2 -> ../../sdc2
pci-0000:24:00.0-fc-0x5001738000cb0160:0x0002000000000000 -> ../../sdd
pci-0000:24:00.0-fc-0x5001738000cb0160:0x0003000000000000 -> ../../sde
pci-0000:24:00.0-fc-0x5001738000cb0160:0x0004000000000000 -> ../../sdf
```

Adding XIV volumes to a Linux on System z system

Only in recent Linux distributions for System z does the `zfc` driver automatically scan for connected volumes. This section shows how to configure the system so that the driver automatically makes specified volumes available when it starts. Volumes and their path information (the local HBA and XIV ports) are defined in configuration files.

Remember: Because of hardware restraints, SLES10 SP3 is used for the examples. The procedures, commands, and configuration files of other distributions can differ.

In this example, Linux on System z has two FC HBAs assigned through `z/VM`. Determine the device numbers of these adapters as shown in Example 3-19.

Example 3-19 FCP HBA device numbers in `z/VM`

```
#CP QUERY VIRTUAL FCP
FCP 0501 ON FCP 5A00 CHPID 8A SUBCHANNEL = 0000
...
FCP 0601 ON FCP 5B00 CHPID 91 SUBCHANNEL = 0001
...

```

The Linux on System z tool to list the FC HBAs is `lszfc`. It shows the enabled adapters only. Adapters that are not listed correctly can be enabled by using the `chccwdev` command as illustrated in Example 3-20.

Example 3-20 Listing and enabling Linux on System z FCP adapters

```
lnxvm01:~ # lszfc
0.0.0501 host0

lnxvm01:~ # chccwdev -e 601
Setting device 0.0.0601 online
Done

lnxvm01:~ # lszfc
0.0.0501 host0
0.0.0601 host1
```

For SLES 10, the volume configuration files are in the `/etc/sysconfig/hardware` directory. There must be one for each HBA. Example 3-21 shows their naming scheme.

Example 3-21 HBA configuration files naming scheme example

```
lnxvm01:~ # ls /etc/sysconfig/hardware/ | grep zfc
hwcfg-zfc-bus-ccw-0.0.0501
hwcfg-zfc-bus-ccw-0.0.0601
```

Important: The configuration file that is described here is used with SLES9 and SLES10. SLES11 uses udev rules. These rules are automatically created by YAST when you use it to discover and configure SAN-attached volumes. They are complicated and not well documented yet, so use YAST.

The configuration files contain a remote (XIV) port and LUN pair for each path to each volume. Example 3-22 defines two XIV volumes to the HBA 0.0.0501, going through two XIV host ports.

Example 3-22 HBA configuration file example

```
lnxvm01:~ # cat /etc/sysconfig/hardware/hwcfg-zfc-bus-ccw-0.0.0501
#!/bin/sh
#
# hwcfg-zfc-bus-ccw-0.0.0501
#
# Configuration for the zfc adapter at CCW ID 0.0.0501
#
...
# Configured zfc disks
ZFCP_LUNS="
0x5001738000cb0191:0x0001000000000000
0x5001738000cb0191:0x0002000000000000
0x5001738000cb0191:0x0003000000000000
0x5001738000cb0191:0x0004000000000000"
```

The `ZFCP_LUNS=""` statement in the file defines all the remote port to volume relations (paths) that the `zfc` driver sets up when it starts. The first term in each pair is the WWPN of the XIV host port. The second term (after the colon) is the LUN of the XIV volume. The LUN

provided here is the LUN that found in the XIV LUN map that is shown in Figure 3-5. It is padded with zeros until it reaches a length of 8 bytes.

The screenshot shows a window titled "LUN Mapping for ITSO_zLinux". Inside, there is a tab labeled "Mapped Volumes" and a table with two columns: "LUN" and "Volume".

LUN	Volume
1	ITSO_zLinux_1
2	ITSO_zLinux_2
3	ITSO_zLinux_3
4	ITSO_zLinux_4

Figure 3-5 XIV LUN map

RH-EL uses the file `/etc/zfcp.conf` to configure SAN-attached volumes. It contains the same information in a different format, as shown in Example 3-23. The three final lines in the example are comments that explain the format. They do not have to be present in the file.

Example 3-23 Format of the `/etc/zfcp.conf` file for RH-EL

```
lnxvm01:~ # cat /etc/zfcp.conf
0x0501 0x5001738000cb0191 0x0001000000000000
0x0501 0x5001738000cb0191 0x0002000000000000
0x0501 0x5001738000cb0191 0x0003000000000000
0x0501 0x5001738000cb0191 0x0004000000000000
0x0601 0x5001738000cb0160 0x0001000000000000
0x0601 0x5001738000cb0160 0x0002000000000000
0x0601 0x5001738000cb0160 0x0003000000000000
0x0601 0x5001738000cb0160 0x0004000000000000
# | | |
#FCP HBA | LUN
# Remote (XIV) Port
```

3.2.6 Setting up Device Mapper Multipathing

To gain redundancy and optimize performance, connect a server to a storage system through more than one HBA, fabric, and storage port. This results in multiple paths from the server to each attached volume. Linux detects such volumes more than once, and creates a device node for every instance. You need an extra layer in the Linux storage stack to recombine the multiple disk instances into one device.

Linux now has its own native multipathing solution. It is based on the *Device Mapper*, a block device virtualization layer in the Linux kernel, and is called DM-MP. The Device Mapper is also used for other virtualization tasks such as the logical volume manager, data encryption, snapshots, and software RAID.

DM-MP is able to manage path failover and failback, and load balancing for various storage architectures. Figure 3-6 illustrates how DM-MP is integrated into the Linux storage stack.

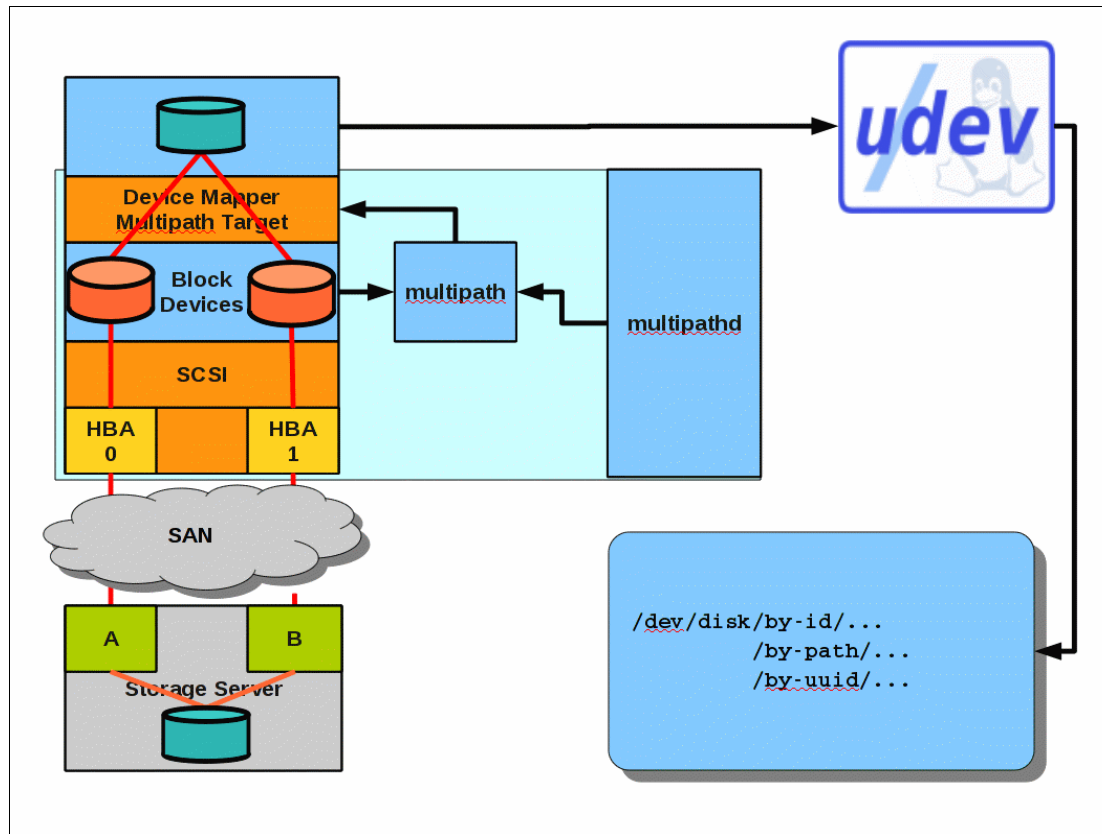


Figure 3-6 Device Mapper Multipathing in the Linux storage stack

In simplified terms, DM-MP consists of four main components:

- ▶ The `dm-multipath` kernel module takes the IO requests that go to the multipath device and passes them to the individual devices that represent the paths.
- ▶ The `multipath` tool scans the device (path) configuration and builds the instructions for the Device Mapper. These instructions include the composition of the multipath devices, failover and failback patterns, and load balancing behavior. This tool is being moved to the multipath background daemon, and will disappear in the future.
- ▶ The multipath background daemon `multipathd` constantly monitors the state of the multipath devices and the paths. If an event occurs, it triggers failover and failback activities in the `dm-multipath` module. It also provides a user interface for online reconfiguration of the multipathing. In the future, it will take over all configuration and setup tasks.
- ▶ A set of rules that tells `udev` what device nodes to create so that multipath devices can be accessed and are persistent.

Configuring DM-MP

You can use the file `/etc/multipath.conf` to configure DM-MP according to your requirements:

- ▶ Define new storage device types
- ▶ Exclude certain devices or device types
- ▶ Set names for multipath devices
- ▶ Change error recovery behavior

The `/etc/multipath.conf` file is not described in detail here. For more information, see the publications in 3.1.2, “Reference material” on page 88. For more information about the settings for XIV attachment, see 3.2.7, “Special considerations for XIV attachment” on page 116.

One option, however, that shows up several times in the next sections needs some explanation. You can tell DM-MP to generate “user-friendly” device names by specifying this option in `/etc/multipath.conf` as illustrated in Example 3-24.

Example 3-24 Specifying user-friendly names in `/etc/multipath.conf`

```
defaults {  
    ...  
    user_friendly_names yes  
    ...  
}
```

The names created this way are persistent. They do not change even if the device configuration changes. If a volume is removed, its former DM-MP name is not used again for a new one. If it is reattached, it gets its old name. The mappings between unique device identifiers and DM-MP user-friendly names are stored in the file `/var/lib/multipath/bindings`.

Tip: The user-friendly names are different for SLES 11 and RH-EL 5. They are explained in their respective sections.

Enabling multipathing for SLES 11

Important: If you install and use the Host Attachment Kit on an Intel x86 based Linux server, you do not have to set up and configure DM-MP. The Host Attachment Kit tools configure DM-MP for you.

You can start Device Mapper Multipathing by running two start scripts as shown in Example 3-25.

Example 3-25 Starting DM-MP in SLES 11

```
x3650lab9:~ # /etc/init.d/boot.multipath start  
Creating multipath target                               done  
x3650lab9:~ # /etc/init.d/multipathd start  
Starting multipathd                                   done
```

To have DM-MP start automatically at each system start, add these start scripts to the SLES 11 system start process (Example 3-26).

Example 3-26 Configuring automatic start of DM-MP in SLES 11

```
x36501ab9:~ # inserv boot.multipath
x36501ab9:~ # inserv multipathd
```

Enabling multipathing for RH-EL 5

RH-EL comes with a default `/etc/multipath.conf` file. It contains a section that blacklists all device types. You must remove or comment out these lines to make DM-MP work. A `#` sign in front of them will mark them as comments so they are ignored the next time DM-MP scans for devices (Example 3-27).

Example 3-27 Disabling blacklisting all devices in `/etc/multipath.conf`

```
...
# Blacklist all devices by default. Remove this to enable multipathing
# on the default devices.
#blacklist {
#devnode "*"
#}
...
```

Start DM-MP as shown in Example 3-28.

Example 3-28 Starting DM-MP in RH-EL 5

```
[root@x36501ab9 ~]# /etc/init.d/multipathd start
Starting multipathd daemon: [ OK ]
```

To have DM-MP start automatically at each system start, add the following start script to the RH-EL 5 system start process (Example 3-29).

Example 3-29 Configuring automatic start of DM-MP in RH-EL 5

```
[root@x36501ab9 ~]# chkconfig --add multipathd
[root@x36501ab9 ~]# chkconfig --levels 35 multipathd on
[root@x36501ab9 ~]# chkconfig --list multipathd
multipathd    0:off  1:off  2:off  3:on   4:off  5:on   6:off
```

Checking and changing the DM-MP configuration

The multipath background daemon provides a user interface to print and modify the DM-MP configuration. It can be started as an interactive session with the `multipathd -k` command. Within this session, various options are available. Use the `help` command to get a list. Some of the more important ones are shown in the following examples. For more information, see 3.3, “Non-disruptive SCSI reconfiguration” on page 117.

The `show topology` command that is illustrated in Example 3-30 prints a detailed view of the current DM-MP configuration, including the state of all available paths.

Example 3-30 Showing multipath topology

```
x36501ab9:~ # multipathd -k"show top"
20017380000cb0520 dm-4 IBM,2810XIV
[size=16G] [features=0] [hwandler=0]
```

```

\_ round-robin 0 [prio=1][active]
  \_ 0:0:0:4 sdh 8:112 [active][ready]
\_ round-robin 0 [prio=1][enabled]
  \_ 1:0:0:4 sdf 8:80 [active][ready]
20017380000cb051f dm-5 IBM,2810XIV
[size=16G][features=0][hwandler=0]
\_ round-robin 0 [prio=1][active]
  \_ 0:0:0:3 sdg 8:96 [active][ready]
\_ round-robin 0 [prio=1][enabled]
  \_ 1:0:0:3 sde 8:64 [active][ready]
20017380000cb2d57 dm-0 IBM,2810XIV
[size=16G][features=0][hwandler=0]
\_ round-robin 0 [prio=1][active]
  \_ 1:0:0:2 sdd 8:48 [active][ready]
\_ round-robin 0 [prio=1][enabled]
  \_ 0:0:0:2 sdb 8:16 [active][ready]
20017380000cb3af9 dm-1 IBM,2810XIV
[size=32G][features=0][hwandler=0]
\_ round-robin 0 [prio=1][active]
  \_ 1:0:0:1 sdc 8:32 [active][ready]
\_ round-robin 0 [prio=1][enabled]
  \_ 0:0:0:1 sda 8:0 [active][ready]

```

The multipath topology in Example 3-30 shows that the paths of the multipath devices are in separate path groups. Thus, there is no load balancing between the paths. DM-MP must be configured with a XIV `multipath.conf` file to enable load balancing. For more information, see 3.2.7, “Special considerations for XIV attachment” on page 116 and “Multipathing” on page 91. The Host Attachment Kit does this configuration automatically if you use it for host configuration.

You can use `reconfigure` as shown in Example 3-31 to tell DM-MP to update the topology after it scans the paths and configuration files. Use it to add new multipath devices after you add new XIV volumes. For more information, see 3.3.1, “Adding and removing XIV volumes dynamically” on page 117.

Example 3-31 Reconfiguring DM-MP

```

multipathd> reconfigure
ok

```

Important: The `multipathd -k` command prompt of SLES11 SP1 supports the `quit` and `exit` commands to terminate. The command prompt of RH-EL 5U5 is a little older and must still be terminated by using the Ctrl + d key combination.

Although the `multipath -l` and `multipath -ll` commands can be used to print the current DM-MP configuration, use the `multipathd -k` interface. The `multipath` tool is being removed from DM-MP, and all further development and improvements will go into `multipathd`.

Tip: You can also issue commands in a “one-shot-mode” by enclosing them in quotation marks and typing them directly, without space, behind the `multipath -k`.

An example would be `multipathd -k“show paths”`

Enabling multipathing for RH-EL 6

Unlike RH-EL 5, RH-EL 6 comes with a new utility, **mpathconf**, that creates and modifies the `/etc/multipath.conf` file.

This command, illustrated in Figure 3-7, enables the multipath configuration file.

```
#mpathconf --enable --with_multipathd y
```

Figure 3-7 The `mpathconf` command

Be sure to start and enable the **multipathd** daemon as shown in Figure 3-8.

```
#service multipathd start
#chkconfig multipathd on
```

Figure 3-8 Commands to ensure that `multipathd` is started and enabled at boot

Because the value of `user_friendly_name` in RH-EL6 is set to `yes` in the default configuration file, the multipath devices are created as:

```
/dev/mapper/mpathn
```

where *n* is an alphabetic letter that designates the path.

Red Hat has released numerous enhancements to the device-mapper-multipath drivers that were shipped with RH 6. Make sure to install and update to the latest version, and download any bug fixes.

Accessing DM-MP devices in SLES 11

The device nodes that you use to access DM-MP devices are created by `udev` in the directory `/dev/mapper`. If you do not change any settings, SLES 11 uses the unique identifier of a volume as device name as seen in Example 3-32.

Example 3-32 Multipath devices in SLES 11 in `/dev/mapper`

```
x36501ab9:~ # ls -l /dev/mapper | cut -c 48-
```

```
20017380000cb051f
20017380000cb0520
20017380000cb2d57
20017380000cb3af9
```

```
...
```

Important: The Device Mapper itself creates its default device nodes in the `/dev` directory. They are called `/dev/dm-0`, `/dev/dm-1`, and so on. These nodes are not persistent. They can change with configuration changes and should not be used for device access.

SLES 11 creates an extra set of device nodes for multipath devices. It overlays the former single path device nodes in `/dev/disk/by-id`. Any device mappings you did for one of these nodes before starting DM-MP are not affected. It uses the DM-MP device instead of the SCSI disk device as illustrated in Example 3-33.

Example 3-33 SLES 11 DM-MP device nodes in /dev/disk/by-id

```
x36501ab9:~ # ls -l /dev/disk/by-id/ | cut -c 44-  
  
scsi-2001738000cb051f -> ../../dm-5  
scsi-2001738000cb0520 -> ../../dm-4  
scsi-2001738000cb2d57 -> ../../dm-0  
scsi-2001738000cb3af9 -> ../../dm-1  
...
```

If you set the `user_friendly_names` option in `/etc/multipath.conf`, SLES 11 creates DM-MP devices with the names `mpatha`, `mpathb`, and so on, in `/dev/mapper`. The DM-MP device nodes in `/dev/disk/by-id` are not changed. They also have the unique IDs of the volumes in their names.

Accessing DM-MP devices in RH-EL

RH-EL sets the `user_friendly_names` option in its default `/etc/multipath.conf` file. The devices that it creates in `/dev/mapper` look as shown in Example 3-34.

Example 3-34 Multipath devices in RH-EL 5 in /dev/mapper

```
[root@x36501ab9 ~]# ls -l /dev/mapper/ | cut -c 45-  
  
mpath1  
mpath2  
mpath3  
mpath4
```

Example 3-35 show the output from an RH-EL 6 system.

Example 3-35 RH-EL 6 device nodes in /dev/mpath

```
[root@x36501ab9 ~]# ls -l /dev/mpath/ | cut -c 39-  
  
2001738000cb051f -> ../../dm-5  
2001738000cb0520 -> ../../dm-4  
2001738000cb2d57 -> ../../dm-0  
2001738000cb3af9 -> ../../dm-1
```

A second set of device nodes contains the unique IDs of the volumes in their name, regardless of whether user-friendly names are specified or not.

In RH-EL5, you find them in the directory `/dev/mpath` as shown in Example 3-36.

Example 3-36 RH-EL 6 Multipath devices

```
mpatha -> ../dm-2  
mpathap1 -> ../dm-3  
mpathap2 -> ../dm-4  
mpathap3 -> ../dm-5
```

```
mpathc -> ../dm-6
mpathd -> ../dm-7
```

In RH-EL6, you find them in `/dev/mapper` as shown in Example 3-37.

Example 3-37 RH-EL 6 device nodes in /dev/mapper

```
# ls -l /dev/mapper/ | cut -c 43-
```

```
mpatha -> ../dm-2
mpathap1 -> ../dm-3
mpathap2 -> ../dm-4
mpathap3 -> ../dm-5
mpathc -> ../dm-6
```

```
mpathd ->
../dm-7
```

Using multipath devices

You can use the device nodes that are created for multipath devices just like any other block device:

- ▶ Create a file system and mount it
- ▶ Use them with the *Logical Volume Manager (LVM)*
- ▶ Build software RAID devices

You can also partition a DM-MP device by using the `fdisk` command or any other partitioning tool. To make new partitions on DM-MP devices available, use the `partprobe` command. It triggers `udev` to set up new block device nodes for the partitions as illustrated in Example 3-38.

Example 3-38 Using the partprobe command to register newly created partitions

```
x3650lab9:~ # fdisk /dev/mapper/20017380000cb051f
...
<all steps to create a partition and write the new partition table>
...
x3650lab9:~ # ls -l /dev/mapper/ | cut -c 48-
```

```
20017380000cb051f
20017380000cb0520
20017380000cb2d57
20017380000cb3af9
```

```
...
x3650lab9:~ # partprobe
x3650lab9:~ # ls -l /dev/mapper/ | cut -c 48-
```

```
20017380000cb051f
20017380000cb051f-part1
20017380000cb0520
20017380000cb2d57
20017380000cb3af9
```

Example 3-38 on page 115 was created with SLES 11. The method works as well for RH-EL 5, but the partition names might be different.

Remember: This limitation, that LVM by default would not work with DM-MP devices, does not exist in recent Linux versions.

3.2.7 Special considerations for XIV attachment

This section addresses special considerations that apply to XIV.

Configuring multipathing

The XIV Host Attachment Kit normally updates the `/etc/multipath.conf` file during installation to optimize use for XIV. If you must manually update the file, the following are the contents of this file as it is created by the Host Attachment Kit. The settings that are relevant for XIV are shown in Example 3-39.

Example 3-39 DM-MP settings for XIV

```
x3650lab9:~ # cat /etc/multipath.conf
devices {
    device {
        vendor "IBM"
        product "2810XIV"
        selector "round-robin 0"
        path_grouping_policy multibus
        rr_min_io 15
        path_checker tur
        failback 15
        no_path_retry queue
        polling_interval 3
    }
}
```

The `user_friendly_names` parameter was addressed in 3.2.6, “Setting up Device Mapper Multipathing” on page 108. You can add it to file or leave it out. The values for `failback`, `no_path_retry`, `path_checker`, and `polling_interval` control the behavior of DM-MP in case of path failures. Normally, do not change them. If your situation requires a modification of these parameters, see the publications in 3.1.2, “Reference material” on page 88. The `rr_min_io` setting specifies the number of I/O requests that are sent to one path before switching to the next one. The value of 15 gives good load balancing results in most cases. However, you can adjust it as necessary.

Important: Upgrading or reinstalling the Host Attachment Kit does not change the `multipath.conf` file. Ensure that your settings match the values that were shown previously.

System z specific multipathing settings

Testing of Linux on System z with multipathing has shown that for best results, set the parameters as follows:

- ▶ `dev_loss_tmo` parameter to 90 seconds
- ▶ `fast_io_fail_tmo` parameter to 5 seconds

Modify the `/etc/multipath.conf` file and add the settings that are shown in Example 3-40.

Example 3-40 System z specific multipathing settings

```
...
defaults {
...
    dev_loss_tmo      90
    fast_io_fail_tmo  5
...
}
...
```

Make the changes effective by using the **reconfigure** command in the interactive **multipathd -k** prompt.

Disabling QLogic failover

The QLogic HBA kernel modules have limited built-in multipathing capabilities. Because multipathing is managed by DM-MP, make sure that the QLogic failover support is disabled. Use the **modinfo qla2xxx** command as shown in Example 3-41 to check.

Example 3-41 Checking for enabled QLogic failover

```
x36501ab9:~ # modinfo qla2xxx | grep version
version:      8.03.01.04.05-k
srcversion:   A2023F2884100228981F34F
```

If the version string ends with **-fo**, the failover capabilities are turned on and must be disabled. To do so, add a line to the `/etc/modprobe.conf` file of your Linux system as illustrated in Example 3-42.

Example 3-42 Disabling QLogic failover

```
x36501ab9:~ # cat /etc/modprobe.conf
...
options qla2xxx ql2xfailover=0
...
```

After you modify this file, run the **depmod -a** command to refresh the kernel driver dependencies. Then, reload the `qla2xxx` module to make the change effective. If you include the `qla2xxx` module in the `InitRAMFS`, you must create a new one.

3.3 Non-disruptive SCSI reconfiguration

This section highlights actions that can be taken on the attached host in a nondisruptive manner.

3.3.1 Adding and removing XIV volumes dynamically

Unloading and reloading the Fibre Channel HBA Adapter used to be the typical way to discover newly attached XIV volumes. However, this action is disruptive to all applications that use Fibre Channel-attached disks on this particular host.

With a modern Linux system, you can add newly attached LUNs without unloading the FC HBA driver. As shown in Example 3-43, you use a command interface that is provided by **sysfs**.

Example 3-43 Scanning for new Fibre Channel attached devices

```
x36501ab9:~ # ls /sys/class/fc_host/  
host0 host1  
x36501ab9:~ # echo "- - -" > /sys/class/scsi_host/host0/scan  
x36501ab9:~ # echo "- - -" > /sys/class/scsi_host/host1/scan
```

First, discover which SCSI instances your FC HBAs have, then issue a **scan** command to their **sysfs** representatives. The triple dashes “- - -” represent the Channel-Target-LUN combination to scan. A dash causes a scan through all possible values. A number would limit the scan to the provided value.

Tip: If you have the Host Attachment Kit installed, you can use the **xiv_fc_admin -R** command to scan for new XIV volumes.

New disk devices that are discovered this way automatically get device nodes and are added to DM-MP.

Tip: For some older Linux versions, you must force the FC HBA to run a port login to recognize the newly added devices. Use the following command, which must be issued to all FC HBAs:

```
echo 1 > /sys/class/fc_host/host<ID>/issue_lip
```

If you want to remove a disk device from Linux, follow this sequence to avoid system hangs because of incomplete I/O requests:

1. Stop all applications that use the device and make sure that all updates or writes are completed.
2. Unmount the file systems that use the device.
3. If the device is part of an LVM configuration, remove it from all logical volumes and volume groups.
4. Remove all paths to the device from the system (Example 3-44).

Example 3-44 Removing both paths to a disk device

```
x36501ab9:~ # echo 1 > /sys/class/scsi_disk/0\0\0\3/device/delete  
x36501ab9:~ # echo 1 > /sys/class/scsi_disk/1\0\0\3/device/delete
```

The device paths (or disk devices) are represented by their Linux SCSI address. For more information, see “Linux SCSI addressing explained” on page 104. Run the **multipathd -k“show topology”** command after you remove each path to monitor the progress.

DM-MP and udev recognize the removal automatically, and delete all corresponding disk and multipath device nodes. You must remove all paths that exist to the device before you detach the device on the storage system level.

You can use **watch** to run a command periodically for monitoring purposes. This example allows you to monitor the multipath topology with a period of one second:

```
watch -n 1 'multipathd -k"show top"'
```


3.3.2 Adding and removing XIV volumes in Linux on System z

The mechanisms to scan and attach new volumes do not work the same in Linux on System z. Commands are available that discover and show the devices that are connected to the FC HBAs. However, they do not do the logical attachment to the operating system automatically. In SLES10 SP3, use the `zfcplib_san_disc` command for discovery.

Example 3-45 shows how to discover and list the connected volumes, in this case one remote port or path, with the `zfcplib_san_disc` command. You must run this command for all available remote ports.

Example 3-45 Listing LUNs connected through a specific remote port

```
lnxvm01:~ # zfcplib_san_disc -L -p 0x5001738000cb0191 -b 0.0.0501
0x0001000000000000
0x0002000000000000
0x0003000000000000
0x0004000000000000
```

Remember: In more recent distributions, `zfcplib_san_disc` is no longer available because remote ports are automatically discovered. The attached volumes can be listed by using the `ls1uns` script.

After you discover the connected volumes, logically attach them using `sysfs` interfaces. Remote ports or device paths are represented in the `sysfs`. There is a directory for each local - remote port combination (path). It contains a representative of each attached volume and various meta files as interfaces for action. Example 3-46 shows such a `sysfs` structure for a specific XIV port.

Example 3-46 sysfs structure for a remote port

```
lnxvm01:~ # ls -l /sys/bus/ccw/devices/0.0.0501/0x5001738000cb0191/
total 0
drwxr-xr-x 2 root root    0 2010-12-03 13:26 0x0001000000000000
...
--w----- 1 root root 4096 2010-12-03 13:26 unit_add
--w----- 1 root root 4096 2010-12-03 13:26 unit_remove
```

Add LUN 0x0003000000000000 to both available paths by using the `unit_add` metafile as shown in Example 3-47.

Example 3-47 Adding a volume to all existing remote ports

```
lnxvm01:~ # echo 0x0003000000000000 > /sys/.../0.0.0501/0x5001738000cb0191/unit_add
lnxvm01:~ # echo 0x0003000000000000 > /sys/.../0.0.0501/0x5001738000cb0160/unit_add
```

Important: You must run discovery by using `zfcplib_san_disc` whenever new devices, remote ports, or volumes are attached. Otherwise, the system does not recognize them even if you do the logical configuration.

New disk devices that you attach this way automatically get device nodes and are added to DM-MP.

If you want to remove a volume from Linux on System z, complete the same steps as for the other platforms. These procedures avoid system hangs because of incomplete I/O requests:

1. Stop all applications that use the device, and make sure that all updates or writes are completed.
2. Unmount the file systems that use the device.
3. If the device is part of an LVM configuration, remove it from all logical volumes and volume groups.
4. Remove all paths to the device from the system.

Volumes can then be removed logically by using a method similar to attachment. Write the LUN of the volume into the `unit_remove` meta file for each remote port in `sysfs`.

Important: If you need the newly added devices to be persistent, use the methods in “Adding XIV volumes to a Linux on System z system” on page 106. Create the configuration files to be used at the next system start.

3.3.3 Adding new XIV host ports to Linux on System z

If you connect new XIV ports or a new XIV system to the Linux on System z system, you must logically attach the new remote ports. Discover the XIV ports that are connected to your HBAs as shown in Example 3-48.

Example 3-48 Showing connected remote ports

```
lnxvm01:~ # zfcplib_san_disc -W -b 0.0.0501
0x5001738000cb0191
0x5001738000cb0170
lnxvm01:~ # zfcplib_san_disc -W -b 0.0.0601
0x5001738000cb0160
0x5001738000cb0181
```

Attach the new XIV ports logically to the HBAs. In Example 3-49, a remote port is already attached to HBA 0.0.0501. Add the second connected XIV port to the HBA.

Example 3-49 Listing attached remote ports, attaching remote ports

```
lnxvm01:~ # ls /sys/bus/ccw/devices/0.0.0501/ | grep 0x
0x5001738000cb0191

lnxvm01:~ # echo 0x5001738000cb0170 > /sys/bus/ccw/devices/0.0.0501/port_add

lnxvm01:~ # ls /sys/bus/ccw/devices/0.0.0501/ | grep 0x
0x5001738000cb0191
0x5001738000cb0170
```

Add the second new port to the other HBA in the same way (Example 3-50).

Example 3-50 Attaching a remote port to the second HBA

```
lnxvm01:~ # echo 0x5001738000cb0181 > /sys/bus/ccw/devices/0.0.0601/port_add
lnxvm01:~ # ls /sys/bus/ccw/devices/0.0.0601/ | grep 0x
0x5001738000cb0160
0x5001738000cb0181
```

3.3.4 Resizing XIV volumes dynamically

At the time of writing, only SLES11 SP1 can use the additional capacity of dynamically enlarged XIV volumes. Reducing the size is not supported. To resize XIV volumes dynamically, complete the following steps:

1. Create an ext3 file system on one of the XIV multipath devices and mount it. The **df** command in Example 3-51 shows the available capacity.

Example 3-51 Checking the size and available space on a mounted file system

```
x3650lab9:~ # df -h /mnt/itso_0520/
file system          Size Used Avail Use% Mounted on
/dev/mapper/2001738000cb0520
                    16G 173M  15G   2% /mnt/itso_0520
```

2. Use the XIV GUI to increase the capacity of the volume from 17 to 51 GB (decimal, as shown by the XIV GUI). The Linux SCSI layer picks up the new capacity when you rescan each SCSI disk device (path) through **sysfs** (Example 3-52).

Example 3-52 Rescanning all disk devices (paths) of a XIV volume

```
x3650lab9:~ # echo 1 > /sys/class/scsi_disk/0\0:0\0:4/device/rescan
x3650lab9:~ # echo 1 > /sys/class/scsi_disk/1\0:0\0:4/device/rescan
```

The message log shown in Example 3-53 indicates the change in capacity.

Example 3-53 Linux message log indicating the capacity change of a SCSI device

```
x3650lab9:~ # tail /var/log/messages
...
Oct 13 16:52:25 lnxxvm01 kernel: [ 9927.105262] sd 0:0:0:4: [sdh] 100663296
512-byte logical blocks: (51.54 GB/48 GiB)
Oct 13 16:52:25 lnxxvm01 kernel: [ 9927.105902] sdh: detected capacity change
from 17179869184 to 51539607552
...
```

3. Indicate the device change to DM-MP by running the **resize_map** command of **multipathd**. The updated capacity is displayed in the output of **show topology** (Example 3-54).

Example 3-54 Resizing a multipath device

```
x3650lab9:~ # multipathd -k"resize map 2001738000cb0520"
ok
x3650lab9:~ # multipathd -k"show top map 2001738000cb0520"
2001738000cb0520 dm-4 IBM,2810XIV
[size=48G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=2][active]
  \_ 0:0:0:4 sdh 8:112 [active][ready]
  \_ 1:0:0:4 sdg 8:96 [active][ready]
```

4. Resize the file system and check the new capacity as shown in Example 3-55.

Example 3-55 Resizing file system and checking capacity

```
x3650lab9:~ # resize2fs /dev/mapper/2001738000cb0520
resize2fs 1.41.9 (22-Aug-2009)
file system at /dev/mapper/2001738000cb0520 is mounted on /mnt/itso_0520;
on-line resizing required
old desc_blocks = 4, new_desc_blocks = 7
```

Performing an on-line resize of /dev/mapper/2001738000cb0520 to 12582912 (4k) blocks.
The file system on /dev/mapper/2001738000cb0520 is now **12582912** blocks long.

```
x36501ab9:~ # df -h /mnt/itso_0520/
file system          Size Used Avail Use% Mounted on
/dev/mapper/2001738000cb0520
                     48G  181M   46G   1% /mnt/itso_0520
```

Restrictions: At the time of writing, the dynamic volume increase process has the following restrictions:

- ▶ Of the supported Linux distributions, only SLES11 SP1 has this capability. The upcoming RH-EL 6 will also have it.
- ▶ The sequence works only with unpartitioned volumes.
- ▶ The file system must be created directly on the DM-MP device.
- ▶ Only the modern file systems can be resized while they are mounted. The ext2 file system cannot.

3.3.5 Using snapshots and remote replication targets

The XIV snapshot and remote replication solutions create identical copies of the source volumes. The target has a unique identifier, which is made up from the XIV WWNN and volume serial number. Any metadata that is stored on the target, such as the file system identifier or LVM signature, however, is identical to that of the source. This metadata can lead to confusion and data integrity problems if you plan to use the target on the same Linux system as the source.

This section describes some methods to avoid integrity issues. It also highlights some potential traps that might lead to problems.

File system directly on a XIV volume

The copy of a file system that is created directly on a SCSI disk device or a DM-MP device can be used on the same host as the source without modification. However, it cannot have an extra virtualization layer such as RAID or LVM. If you follow the sequence carefully and avoid the highlighted traps, you can use a copy on the same host without problems. The procedure is described on an ext3 file system on a DM-MP device that is replicated with a snapshot.

1. Mount the original file system as shown in Example 3-56 using a device node that is bound to the unique identifier of the volume. The device node cannot be bound to any metadata that is stored on the device itself.

Example 3-56 Mounting the source volume

```
x36501ab9:~ # mount /dev/mapper/2001738000cb0520 /mnt/itso_0520/
x36501ab9:~ # mount
...
/dev/mapper/2001738000cb0520 on /mnt/itso_0520 type ext3 (rw)
```

2. Make sure that the data on the source volume is consistent by running the **sync** command.
3. Create the snapshot on the XIV, make it writeable, and map the target volume to the Linux host. In the example, the snapshot source has the volume ID 0x0520, and the target volume has ID 0x1f93.

4. Initiate a device scan on the Linux host. For more information, see 3.3.1, “Adding and removing XIV volumes dynamically” on page 117. DM-MP automatically integrates the snapshot target as shown in Example 3-57.

Example 3-57 Checking DM-MP topology for the target volume

```
x36501ab9:~ # multipathd -k"show top"
2001738000cb0520 dm-4 IBM,2810XIV
[size=48G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=2][active]
  \_ 0:0:0:4 sdh 8:112 [active][ready]
  \_ 1:0:0:4 sdg 8:96 [active][ready]
...
2001738000cb1f93 dm-7 IBM,2810XIV
[size=48G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=2][active]
  \_ 0:0:0:5 sdi 8:128 [active][ready]
  \_ 1:0:0:5 sdj 8:144 [active][ready]
...
```

5. Mount the target volume to a separate mount point by using a device node that is created from the unique identifier of the volume (Example 3-58).

Example 3-58 Mounting the target volume

```
x36501ab9:~ # mount /dev/mapper/2001738000cb1f93 /mnt/itso_fc/
x36501ab9:~ # mount
...
/dev/mapper/2001738000cb0520 on /mnt/itso_0520 type ext3 (rw)
/dev/mapper/2001738000cb1f93 on /mnt/itso_fc type ext3 (rw)
```

Now you can access both the original volume and the point-in-time copy through their respective mount points.

Important: udev also creates device nodes that relate to the file system Universally Unique Identifier (UUID) or label. These IDs are stored in the data area of the volume, and are identical on both source and target. Such device nodes are ambiguous if the source and target are mapped to the host at the same time. Using them in this situation can result in data loss.

File system in a logical volume managed by LVM

The Linux *Logical Volume Manager* (LVM) uses metadata that is written to the data area of the disk device to identify and address its objects. If you want to access a set of replicated volumes that are under LVM control, modify this metadata so it is unique. This process ensures data integrity. Otherwise, LVM might mix volumes from the source and the target sets.

A script called `vgimportclone.sh` is publicly available that automates the modification of the metadata. It can be downloaded from:

<http://sources.redhat.com/cgi-bin/cvsweb.cgi/LVM2/scripts/vgimportclone.sh?cvsroot=lvm2>

An online copy of the Linux man page for the script can be found at:

<http://www.cl.cam.ac.uk/cgi-bin/manpage?8+vgimportclone>

Tip: The `vgimportclone` script and commands are part of the standard LVM tools for RH-EL. The SLES 11 distribution does not contain the script by default.

Complete the following steps to ensure consistent data on the target volumes and avoid mixing up the source and target. In this example, a volume group contains a logical volume that is striped over two XIV volumes. Snapshots are used to create a point in time copy of both volumes. Both the original logical volume and the cloned one are then made available to the Linux system. The XIV serial numbers of the source volumes are 1fc5 and 1fc6, and the IDs of the target volumes are 1fe4 and 1fe5.

1. Mount the original file system by using the LVM logical volume device as shown in Example 3-59.

Example 3-59 Mounting the source volume

```
x3650lab9:~ # mount /dev/vg_xiv/lv_itso /mnt/lv_itso
x3650lab9:~ # mount
...
/dev/mapper/vg_xiv-lv_itso on /mnt/lv_itso type ext3 (rw)
```

2. Make sure that the data on the source volume is consistent by running the `sync` command.
3. Create the snapshots on the XIV, unlock them, and map the target volumes 1fe4 and 1fe5 to the Linux host.
4. Initiate a device scan on the Linux host. For more information, see 3.3.1, “Adding and removing XIV volumes dynamically” on page 117. DM-MP automatically integrates the snapshot targets as shown in Example 3-60.

Example 3-60 Checking DM-MP topology for target volume

```
x3650lab9:~ # multipathd -k"show topology"
...
2001738000cb1fe4 dm-9 IBM,2810XIV
[size=32G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=2][active]
  \_ 0:0:0:6 sdk 8:160 [active][ready]
  \_ 1:0:0:6 sdm 8:192 [active][ready]
2001738000cb1fe5 dm-10 IBM,2810XIV
[size=32G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=2][active]
  \_ 0:0:0:7 sdl 8:176 [active][ready]
  \_ 1:0:0:7 sdn 8:208 [active][ready]
```

Important: To avoid data integrity issues, it is important that no LVM configuration commands are issued until step 5 is complete.

5. Run the `vgimportclone.sh` script against the target volumes, and provide a new volume group name (Example 3-61).

Example 3-61 Adjusting the LVM metadata of the target volumes

```
x3650lab9:~ # ./vgimportclone.sh -n vg_itso_snap /dev/mapper/2001738000cb1fe4
/dev/mapper/2001738000cb1fe5
WARNING: Activation disabled. No device-mapper interaction will be attempted.
Physical volume "/tmp/snap.sHT13587/vgimport1" changed
1 physical volume changed / 0 physical volumes not changed
```

```
WARNING: Activation disabled. No device-mapper interaction will be attempted.
Physical volume "/tmp/snap.sHT13587/vgimport0" changed
1 physical volume changed / 0 physical volumes not changed
WARNING: Activation disabled. No device-mapper interaction will be attempted.
Volume group "vg_xiv" successfully changed
Volume group "vg_xiv" successfully renamed to "vg_itso_snap"
Reading all physical volumes. This may take a while...
Found volume group "vg_itso_snap" using metadata type lvm2
Found volume group "vg_xiv" using metadata type lvm2
```

6. Activate the volume group on the target devices and mount the logical volume as shown in Example 3-62.

Example 3-62 Activating volume group on target device and mounting the logical volume

```
x3650lab9:~ # vgchange -a y vg_itso_snap
  1 logical volume(s) in volume group "vg_itso_snap" now active
x3650lab9:~ # mount /dev/vg_itso_snap/lv_itso /mnt/lv_snap_itso/
x3650lab9:~ # mount
...
/dev/mapper/vg_xiv-lv_itso on /mnt/lv_itso type ext3 (rw)
/dev/mapper/vg_itso_snap-lv_itso on /mnt/lv_snap_itso type ext3 (rw)
```

3.4 Troubleshooting and monitoring

This section addresses topics that are related to troubleshooting and monitoring. As mentioned earlier, always check that the Host Attachment Kit is installed.

Afterward, key information can be found in the same directory the installation was started from, inside the `install.log` file.

3.4.1 Linux Host Attachment Kit utilities

The Host Attachment Kit now includes the following utilities:

► **xiv_devlist**

xiv_devlist is the command that validates the attachment configuration. This command generates a list of multipathed devices available to the operating system. Example 3-63 shows the available options.

Example 3-63 Options of xiv_devlist from Host Attachment Kit version 1.7

```
# xiv_devlist --help
Usage: xiv_devlist [options]
-h, --help                show this help message and exit
-t OUT, --out=OUT         Choose output method: tui, csv, xml (default: tui)
-o FIELDS, --options=FIELDS
                           Fields to display; Comma-separated, no spaces. Use -l
                           to see the list of fields
-f OUTFILE, --file=OUTFILE
                           File to output to (instead of STDOUT) - can be used
                           only with -t csv/xml
-H, --hex                 Display XIV volume and machine IDs in hexadecimal base
-u SIZE_UNIT, --size-unit=SIZE_UNIT
```

	The size unit to use (e.g. MB, GB, TB, MiB, GiB, TiB, ...)
-d, --debug	Enable debug logging
-l, --list-fields	List available fields for the -o option
-m MP_FRAMEWORK_STR, --multipath=MP_FRAMEWORK_STR	Enforce a multipathing framework <auto native veritas>
-x, --xiv-only	Print only XIV devices
-V, --version	Shows the version of the HostAttachmentKit framework

► **xiv_diag**

This utility gathers diagnostic information from the operating system. The resulting compressed file can then be sent to IBM-XIV support teams for review and analysis (Example 3-64).

Example 3-64 xiv_diag command

```
[/]# xiv_diag
Please type in a path to place the xiv_diag file in [default: /tmp]:
Creating archive xiv_diag-results_2010-9-27_13-24-54
...
INFO: Closing xiv_diag archive file                               DONE
Deleting temporary directory...                                  DONE
INFO: Gathering is now complete.
INFO: You can now send /tmp/xiv_diag-results_2010-9-27_13-24-54.tar.gz to IBM-XIV
for review.
INFO: Exiting.
```

3.4.2 Multipath diagnosis

Some key diagnostic information can be found from the following multipath commands.

To flush all multipath device maps:

```
multipath -F
```

To show the multipath topology (maximum information):

```
multipath -ll
```

For more detailed information, use the **multipath -v2 -d** as illustrated in Example 3-65.

Example 3-65 Linux command multipath output that shows the correct status

```
[root@bc-h-15-b7 ]# multipath -v2 -d
create: mpathc (20017380027950251) undef IBM,2810XIV
size=48G features='0' hwhandler='0' wp=undef
`-+- policy='round-robin 0' prio=1 status=undef
  |- 8:0:0:1 sdc 8:32 undef ready running
  `- 9:0:0:1 sde 8:64 undef ready running
create: mpathd (20017380027950252) undef IBM,2810XIV
size=48G features='0' hwhandler='0' wp=undef
`-+- policy='round-robin 0' prio=1 status=undef
  |- 8:0:0:2 sdd 8:48 undef ready running
  `- 9:0:0:2 sdf 8:80 undef ready running
```

Important: The `multipath` command sometimes finds errors in the `multipath.conf` file that do not exist. The following error messages can be ignored:

```
[root@b]# multipath -F
Sep 22 12:08:21 | multipath.conf line 30, invalid keyword: polling_interval
Sep 22 12:08:21 | multipath.conf line 41, invalid keyword: polling_interval
Sep 22 12:08:21 | multipath.conf line 53, invalid keyword: polling_interval
Sep 22 12:08:21 | multipath.conf line 54, invalid keyword: prio_callout
Sep 22 12:08:21 | multipath.conf line 64, invalid keyword: polling_interval
```

Another excellent command-line utility to use is the `xiv_devlist` command. Note that Example 3-66 shows paths that were not found in Figure 3-9 on page 128, and vice versa.

Example 3-66 Example of `xiv_devlist` showing multipath not correct

```
[root@bc-h-15-b7 ~]# xiv_devlist
```

XIV Devices

Device	Size (GB)	Paths	Vol Name	Vol Id	XIV Id	XIV Host
/dev/sdc	51.6	N/A	RedHat-Data_1	593	1310133	RedHat6.de.ibm.com
/dev/sdd	51.6	N/A	RedHat-Data_2	594	1310133	RedHat6.de.ibm.com
/dev/sde	51.6	N/A	RedHat-Data_1	593	1310133	RedHat6.de.ibm.com
/dev/sdf	51.6	N/A	RedHat-Data_2	594	1310133	RedHat6.de.ibm.com

Non-XIV Devices

Device	Size (GB)	Paths
/dev/sda	50.0	N/A
/dev/sdb	50.0	N/A

Figure 3-9 also shows paths that are not shown in Example 3-66 on page 127.

XIV Devices										
Device	Size (GB)	Serial	Lun	Paths	Vendor	Vol Name	Vol Id	XIV Id	XIV Host	
/dev/sdb	2044.4	N/A	1	N/A	N/A	prdjcl_sarcdb_03_001	9450	780251_7	prdjcl_sarcdb_03	
/dev/sdc	2044.4	N/A	2	N/A	N/A	prdjcl_sarcdb_03_002	9451	780251_7	prdjcl_sarcdb_03	
/dev/sdd	1030.8	N/A	3	N/A	N/A	prdjcl_sarcdb_03_003	9452	780251_7	prdjcl_sarcdb_03	
/dev/sde	2044.4	N/A	1	N/A	N/A	prdjcl_sarcdb_03_001	9450	780251_7	prdjcl_sarcdb_03	
/dev/sdf	2044.4	N/A	2	N/A	N/A	prdjcl_sarcdb_03_002	9451	780251_7	prdjcl_sarcdb_03	
/dev/sdg	1030.8	N/A	3	N/A	N/A	prdjcl_sarcdb_03_003	9452	780251_7	prdjcl_sarcdb_03	
/dev/sdh	2044.4	N/A	1	N/A	N/A	prdjcl_sarcdb_03_001	9450	780251_7	prdjcl_sarcdb_03	
/dev/sdi	2044.4	N/A	2	N/A	N/A	prdjcl_sarcdb_03_002	9451	780251_7	prdjcl_sarcdb_03	

Figure 3-9 Second example of `xiv_devlist` command, showing multipath not working properly

Important: When you are using the `xiv_devlist` command, note the number of paths that are indicated in the column for each device. You do not want the `xiv_devlist` output to show N/A in the paths column.

The expected output from the `multipath` and `xiv_devlist` commands is shown in Example 3-67.

Example 3-67 Example of `multipath` finding the XIV devices, and updating the paths correctly

```
[root@bc-h-15-b7 ~]#multipath
create: mpathc (20017380027950251) undef IBM,2810XIV
size=48G features='0' hwhandler='0' wp=undef
`-+- policy='round-robin 0' prio=1 status=undef
  |- 8:0:0:1 sdc 8:32 undef ready running
  `-- 9:0:0:1 sde 8:64 undef ready running
create: mpathd (20017380027950252) undef IBM,2810XIV
size=48G features='0' hwhandler='0' wp=undef
`-+- policy='round-robin 0' prio=1 status=undef
  |- 8:0:0:2 sdd 8:48 undef ready running
  `-- 9:0:0:2 sdf 8:80 undef ready running
```

```
[root@bc-h-15-b7 ~]# xiv_devlist
```

XIV Devices						
Device	Size (GB)	Paths	Vol Name	Vol Id	XIV Id	XIV Host

```

-----
/dev/mapper/m 51.6      2/2    RedHat-Data_1 593      1310133 RedHat6.de.ib
pathc                                             m.com
-----
/dev/mapper/m 51.6      2/2    RedHat-Data_2 594      1310133 RedHat6.de.ib
pathd                                             m.com
-----

```

Non-XIV Devices

```

-----
Device      Size (GB) Paths
-----
/dev/sda    50.0      N/A
-----
/dev/sdb    50.0      N/A
-----

```

3.4.3 Other ways to check SCSI devices

The Linux kernel maintains a list of all attached SCSI devices in the /proc pseudo file system as illustrated in Example 3-68. The /proc/scsi/scsi pseudo file system contains basically the same information (apart from the device node) as the `ls SCSI` output. It is always available, even if `ls SCSI` is not installed.

Example 3-68 Alternate list of attached SCSI devices

```

x3650lab9:~ # cat /proc/scsi/scsi
Attached devices:
Host: scsi0 Channel: 00 Id: 00 Lun: 01
  Vendor: IBM      Model: 2810XIV      Rev: 10.2
  Type:   Direct-Access      ANSI SCSI revision: 05
Host: scsi0 Channel: 00 Id: 00 Lun: 02
  Vendor: IBM      Model: 2810XIV      Rev: 10.2
  Type:   Direct-Access      ANSI SCSI revision: 05
Host: scsi0 Channel: 00 Id: 00 Lun: 03
  Vendor: IBM      Model: 2810XIV      Rev: 10.2
  Type:   Direct-Access      ANSI SCSI revision: 05
Host: scsi1 Channel: 00 Id: 00 Lun: 01
  Vendor: IBM      Model: 2810XIV      Rev: 10.2
  Type:   Direct-Access      ANSI SCSI revision: 05
Host: scsi1 Channel: 00 Id: 00 Lun: 02
  Vendor: IBM      Model: 2810XIV      Rev: 10.2
  Type:   Direct-Access      ANSI SCSI revision: 05
Host: scsi1 Channel: 00 Id: 00 Lun: 03
  Vendor: IBM      Model: 2810XIV      Rev: 10.2
  Type:   Direct-Access      ANSI SCSI revision: 05
...

```

The **fdisk -l** command that is shown in Example 3-69 can be used to list all block devices, including their partition information and capacity (Example 3-69). However, it does not include SCSI address, vendor, and model information.

Example 3-69 Output of fdisk -l

```
x36501ab9:~ # fdisk -l

Disk /dev/sda: 34.3 GB, 34359738368 bytes
255 heads, 63 sectors/track, 4177 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

   Device Boot      Start         End      Blocks   Id  System
/dev/sda1                1         2089     16779861   83  Linux
/dev/sda2           3501         4177      5438002+   82  Linux swap / Solaris

Disk /dev/sdb: 17.1 GB, 17179869184 bytes
64 heads, 32 sectors/track, 16384 cylinders
Units = cylinders of 2048 * 512 = 1048576 bytes

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

...
```

3.4.4 Performance monitoring with iostat

You can use the **iostat** command to monitor the performance of all attached disks. It is part of the **sysstat** package that ships with every major Linux distribution. However, it is not necessarily installed by default. The **iostat** command reads data that are provided by the kernel in `/proc/stats` and prints it in human readable format. For more information, see the man page of **iostat**.

3.4.5 Generic SCSI tools

For Linux, the *sg_tools* allow low-level access to SCSI devices. They communicate with SCSI devices through the generic SCSI layer. This layer is represented by special device files `/dev/sg0`, `/dev/sg1`, and so on. In recent Linux versions, the **sg_tools** can also access the block devices `/dev/sda`, and `/dev/sdb`. They can also access any other device node that represents a SCSI device directly.

The following are the most useful `sg_tools`:

- ▶ **sg_inq /dev/sgx** prints SCSI Inquiry data, such as the volume serial number.
- ▶ **sg_scan** prints the SCSI host, channel, target, and LUN mapping for all SCSI devices.
- ▶ **sg_map** prints the `/dev/sdx` to `/dev/sgy` mapping for all SCSI devices.
- ▶ **sg_readcap /dev/sgx** prints the block size and capacity (in blocks) of the device.
- ▶ **sginfo /dev/sgx** prints SCSI inquiry and mode page data. You can also use it to manipulate the mode pages.

3.5 Boot Linux from XIV volumes

This section describes how you can configure a system to load the Linux kernel and operating system from a SAN-attached XIV volume. This process is illustrated with an example based on SLES11 SP1 on an x86 server with QLogic FC HBAs. Other distributions and hardware platforms that have deviations from the example are noted. For more information about configuring the HBA BIOS to boot from SAN-attached XIV volume, see 1.2.5, “Boot from SAN on x86/x64 based architecture” on page 23.

3.5.1 The Linux boot process

To understand how to boot a Linux system from SAN-attached XIV volumes, you need a basic understanding of the Linux boot process. The following are the basic steps a Linux system goes through until it presents the login prompt:

1. OS loader

The system firmware provides functions for rudimentary input/output operations such as the BIOS of x86 servers. When a system is turned on, it runs the *power-on self-test (POST)* to check which hardware is available and whether everything is working. Then, it runs the operating system loader (OS loader). The OS loader uses those basic I/O routines to read a specific location on the defined system disk and starts running the code that it contains. This code is either part of the boot loader of the operating system, or it branches to the location where the boot loader is located.

If you want to boot from a SAN-attached disk, make sure that the OS loader can access that disk. FC HBAs provide an extension to the system firmware for this purpose. In many cases, it must be explicitly activated.

On x86 systems, this location is called the *Master Boot Record (MBR)*.

Remember: For Linux on System z under z/VM, the OS loader is not part of the firmware. Instead, it is part of the z/VM program `ip1`.

2. The boot loader

The boot loader starts the operating system kernel. It must know the physical location of the kernel image on the system disk. It then reads it in, unpacks it if it is compressed, and starts it. This process is still done by using the basic I/O routines that are provided by the firmware. The boot loader also can pass configuration options and the location of the `InitRAMFS` to the kernel.

The most common Linux boot loaders are

- GRUB (Grand Unified Bootloader) for x86 systems
- `zip1` for System z
- `yaboot` for Power Systems

3. The kernel and the InitRAMFS

After the kernel is unpacked and running, it takes control of the system hardware. It starts and configures the following systems:

- Memory management
- Interrupt handling
- The built-in hardware drivers for the hardware that is common on all systems such as MMU and clock

It reads and unpacks the InitRAMFS image, again by using the same basic I/O routines. The InitRAMFS contains more drivers and programs that are needed to set up the Linux file system tree (root file system). To be able to boot from a SAN-attached disk, the standard InitRAMFS must be extended with the FC HBA driver and the multipathing software. In modern Linux distributions, this process is done automatically by the tools that create the InitRAMFS image.

After the root file system is accessible, the kernel starts the `init()` process.

4. The `init()` process

The `init()` process brings up the operating system itself, including networking, services, and user interfaces. The hardware is already abstracted. Therefore, `init()` is not platform-dependent, nor are there any SAN-boot specifics.

For more information about the Linux boot process for x86 based systems, see the *IBM developerWorks*® at:

<http://www.ibm.com/developerworks/linux/library/l-linuxboot/>

3.5.2 Configuring the QLogic BIOS to boot from an XIV volume

The first step to configure the HBA is to load a BIOS extension that provides the basic input/output capabilities for a SAN-attached disk. For more information, see 1.2.5, “Boot from SAN on x86/x64 based architecture” on page 23.

Tip: Emulex HBAs also support booting from SAN disk devices. You can enable and configure the Emulex BIOS extension by pressing Alt+e or Ctrl+e when the HBAs are initialized during server startup. For more information, see the following Emulex publications:

- ▶ *Supercharge Booting Servers Directly from a Storage Area Network*

<http://www.emulex.com/artifacts/fc0b92e5-4e75-4f03-9f0b-763811f47823/bootingServersDirectly.pdf>

- ▶ *Enabling Emulex Boot from SAN on IBM BladeCenter*

http://www.emulex.com/artifacts/4f6391dc-32bd-43ae-bcf0-1f51cc863145/enabling_boot_ibm.pdf

3.5.3 OS loader considerations for other platforms

The BIOS is the x86 specific way to start loading an operating system. This section briefly describes how this loading is done on the other supported platforms.

IBM Power Systems

When you install Linux on an IBM Power System server or LPAR, the Linux installer sets the boot device in the firmware to the drive that you are installing on. No special precautions need be taken whether you install on a local disk, a SAN-attached XIV volume, or a virtual disk provided by the VIO server.

IBM System z

Linux on System z can be loaded from traditional CKD disk devices or from Fibre-Channel-attached Fixed-Block (SCSI) devices. To load from SCSI disks, the SCSI IPL feature (FC 9904) must be installed and activated on the System z server. The SCSI *initial program load (IPL)* is generally available on recent System z systems (IBM z10™ and later).

Important: Activating the SCSI IPL feature is disruptive. It requires a POR of the whole system.

Linux on System z can run in two configurations:

1. Linux on System z running natively in a System z LPAR

After you install Linux on System z, you must provide the device from which the LPAR runs the IPL in the LPAR start dialog on the System z *Support Element*. After it is registered there, the IPL device entry is permanent until changed.

2. Linux on System z running under z/VM

Within z/VM, you start an operating system with the IPL command. This command provides the z/VM device address of the device where the Linux boot loader and kernel are installed.

When you boot from SCSI disk, you do not have a z/VM device address for the disk itself. For more information, see 3.2.1, “Platform-specific remarks” on page 92, and “System z” on page 94. You must provide information about which LUN the machine loader uses to start the operating system separately. z/VM provides the **cp** commands **set Loaddev** and **query Loaddev** for this purpose. Their use is illustrated in Example 3-70.

Example 3-70 Setting and querying SCSI IPL device in z/VM

```
SET LOADDEV PORTNAME 50017380 00CB0191 LUN 00010000 00000000

CP QUERY LOADDEV
PORTNAME 50017380 00CB0191 LUN 00010000 00000000 BOOTPROG 0
BR_LBA 00000000 00000000
```

The port name is the XIV host port that is used to access the boot volume. After the load device is set, use the IPL program with the device number of the FCP device (HBA) that connects to the XIV port and LUN to boot from. You can automate the IPL by adding the required commands to the z/VM profile of the virtual machine.

3.5.4 Installing SLES11 SP1 on an XIV volume

With recent Linux distributions, installation on a XIV volume is as easy as installation on a local disk. The process has the following extra considerations:

- ▶ Identifying the correct XIV volumes to install on
- ▶ Enabling multipathing during installation

Tip: After the SLES11 installation program (YAST) is running, the installation is mostly hardware independent. It works the same when it runs on an x86, IBM Power System, or System z server.

To install SLES11 SP1 on an XIV volume, complete the following steps:

1. Boot from an installation DVD. Follow the installation configuration windows until you come to the Installation Settings window shown in Figure 3-10.

Remember: The Linux on System z installer does not automatically list the available disks for installation. Use the Configure Disks window to discover and attach the disks that are needed to install the system by using a graphical user interface. This window is displayed before you get to the Installation Settings window. At least one disk device is required to run the installation.

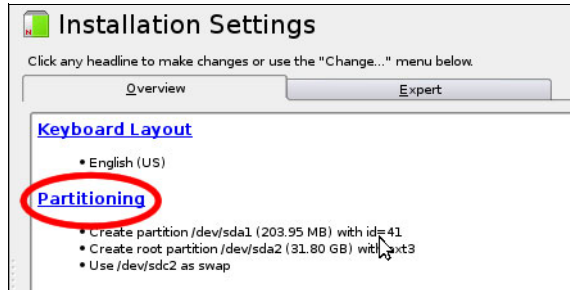


Figure 3-10 SLES11 SP1 installation settings

2. Click **Partitioning**.
3. In the Preparing Hard Disk: Step 1 window, make sure that **Custom Partitioning (for experts)** is selected and click **Next** (Figure 3-11). It does not matter which disk device is selected in the **Hard Disk** field.

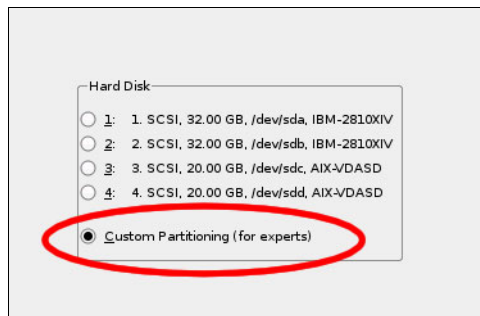


Figure 3-11 Preparing Hard Disk: Step 1 window

4. Enable multipathing in the Expert Partitioner window. Select **Hard disks** in the navigation section on the left side, then click **Configure** → **Configure Multipath** (Figure 3-12).

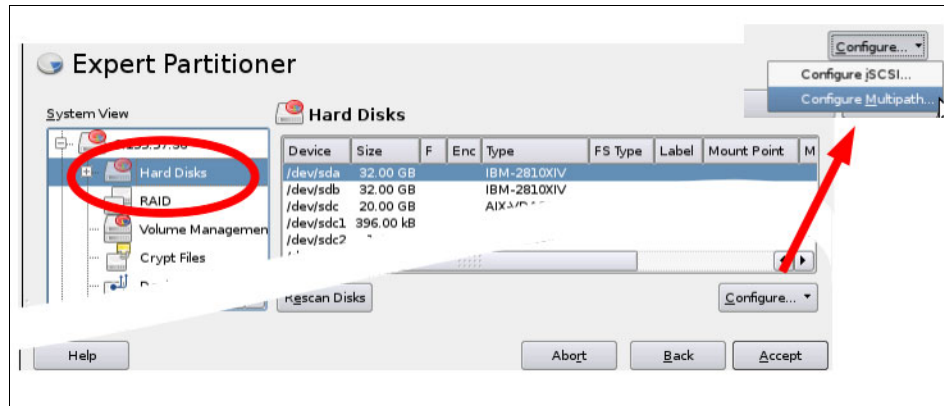


Figure 3-12 Enabling multipathing in the Expert Partitioner window

5. Confirm your selecting, and the tool rescans the disk devices. When finished, it presents an updated list of hard disks that also shows the multipath devices it found (Figure 3-13).

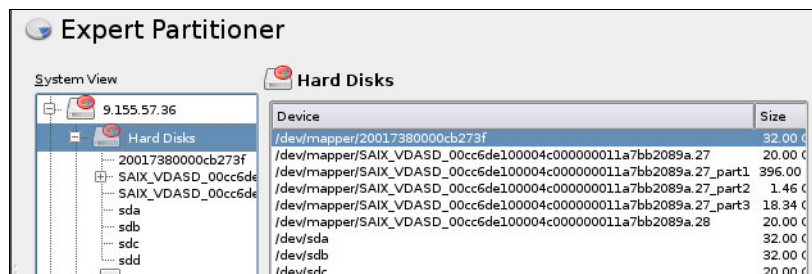


Figure 3-13 Selecting multipath device for installation

6. Select the multipath device (XIV volume) you want to install to and click **Accept**.
7. In the Partitioner window, create and configure the required partitions for your system the same way you would on a local disk.

You can also use the automatic partitioning capabilities of YAST after the multipath devices are detected in step 5. To do so, complete the following steps:

1. Click **Back** until you see the initial partitioning window again. It now shows the multipath devices instead of the disks, as illustrated in Figure 3-14.

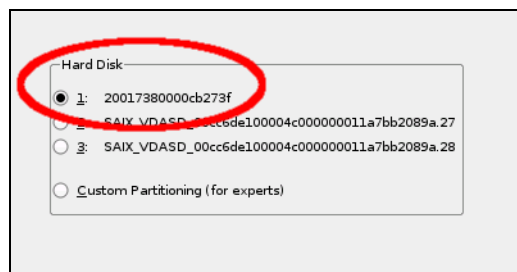


Figure 3-14 Preparing Hard Disk: Step 1 window with multipath devices

2. Select the multipath device that you want to install on and click **Next**.
3. Select the partitioning scheme that you want.

Important: All supported platforms can boot Linux from multipath devices. In some cases, however, the tools that install the boot loader can write only to simple disk devices. In these cases, install the boot loader with multipathing deactivated. For SLES10 and SLES11, add the parameter `multipath=off` to the boot command in the boot loader. The boot loader for IBM Power Systems and System z must be reinstalled whenever there is an update to the kernel or `InitRAMFS`. A separate entry in the boot menu allows you to switch between single and multipath mode when necessary. For more information, see the Linux distribution-specific documentation in 3.1.2, “Reference material” on page 88.

The installer does not implement any device-specific settings, such as creating the `/etc/multipath.conf` file. You must implement these settings manually after installation as explained in 3.2.7, “Special considerations for XIV attachment” on page 116. Because DM-MP is already started during the processing of the `InitRAMFS`, you also must build a new `InitRAMFS` image after you change the DM-MP configuration. For more information, see “Making the FC driver available early in the boot process” on page 97.

It is possible to add *Device Mapper* layers on top of DM-MP, such as software RAID or LVM. The Linux installers support these options.

Tip: RH-EL 5.1 and later support multipathing already. Turn on multipathing it by adding the option `mpath` to the kernel boot line of the installation system. Anaconda, the RH installer, then offers to install to multipath devices



XIV and AIX host connectivity

This chapter explains specific considerations and describes the host attachment-related tasks for the AIX operating system.

Important: The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, see the System Storage Interoperation Center (SSIC) at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

Host Attachment Kits and related publications can be downloaded from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

This chapter includes the following sections:

- ▶ Attaching XIV to AIX hosts
- ▶ SAN boot in AIX

4.1 Attaching XIV to AIX hosts

This section provides information and procedures for attaching the XIV Storage System to AIX on an IBM POWER® platform. Fibre Channel connectivity is addressed first, and then iSCSI attachment.

The AIX host attachment process with XIV is described in detail in the Host Attachment Guide for AIX, which is available from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

The XIV Storage System supports different versions of the AIX operating system through either Fibre Channel (FC) or iSCSI connectivity.

These general notes apply to all AIX releases:

- ▶ XIV Host Attachment Kit for AIX (current at the time of writing) supports all AIX releases except for AIX 5.2 and earlier
- ▶ Dynamic LUN expansion with LVM requires XIV firmware version 10.2 or later

4.1.1 Prerequisites

If the current AIX operating system level installed on your system is not compatible with XIV, you must upgrade before you attach the XIV storage. To determine the maintenance package or technology level that is currently installed on your system, use the `oslevel` command (Example 4-1).

Example 4-1 Determining current AIX version and maintenance level

```
# oslevel -s
7100-01-05-1228
```

In this example, the system is running AIX 7.1.0.0 technology level 1 (7.1TL1). Use this information with the SSIC to ensure that you have an IBM-supported configuration.

If AIX maintenance items are needed, consult the IBM Fix Central website. You can download fixes and updates for your systems software, hardware, and operating system at:

<http://www.ibm.com/eserver/support/fixes/fixcentral/main/pseries/aix>

Before further configuring your host system or the XIV Storage System, make sure that the physical connectivity between the XIV and the POWER system is properly established. Direct attachment of XIV to the host system is not supported. If you use FC switched connections, ensure that you have functioning zoning that uses the worldwide port name (WWPN) numbers of the AIX host.

4.1.2 AIX host FC configuration

Attaching the XIV Storage System to an AIX host using Fibre Channel involves the following activities from the host side:

1. Identifying the Fibre Channel host bus adapters (HBAs) and determine their WWPN values.
2. Installing the AIX Host Attachment Kit for XIV.
3. Configuring multipathing.

Identifying FC adapters and attributes

To allocate XIV volumes to an AIX host, identify the Fibre Channel adapters on the AIX server. Use the **lsdev** command to list all the FC adapter ports in your system as shown in Example 4-2.

Example 4-2 Listing FC adapters

```
# lsdev -Cc adapter|grep fcs
fcs0 Available 00-00 8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs1 Available 00-01 8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
```

In this example, there are two FC ports.

The **lsslot** command returns not just the ports, but also the PCI slot where the Fibre Channel adapters are in the system (Example 4-3). This command can be used to identify in what physical slot a specific adapter is placed.

Example 4-3 Locating FC adapters

```
# lsslot -c pci | grep fcs
U5802.001.00H3722-P1-C10 PCI-E capable, Rev 1 slot with 8x lanes fcs0 fcs1
```

To obtain the WWPN of each of the POWER system FC adapters, use the **lscfg** command as shown in Example 4-4.

Example 4-4 Finding Fibre Channel adapter WWPN

```
# lscfg -v1 fcs0
fcs0 U5802.001.00H3722-P1-C10-T1 8Gb PCI Express Dual Port FC
Adapter (df1000f114108a03)
```

```
Part Number.....10N9824
Serial Number.....1A113001FB
Manufacturer.....001A
EC Level.....D77040
Customer Card ID Number.....577D
FRU Number.....10N9824
Device Specific.(ZM).....3
Network Address.....10000000C9B7F27A
ROS Level and ID.....0278117B
Device Specific.(Z0).....31004549
Device Specific.(Z1).....00000000
Device Specific.(Z2).....00000000
Device Specific.(Z3).....09030909
Device Specific.(Z4).....FF781116
Device Specific.(Z5).....0278117B
Device Specific.(Z6).....0773117B
Device Specific.(Z7).....0B7C117B
Device Specific.(Z8).....20000000C9B7F27A
Device Specific.(Z9).....US1.11X11
Device Specific.(ZA).....U2D1.11X11
Device Specific.(ZB).....U3K1.11X11
Device Specific.(ZC).....00000000
Hardware Location Code.....U5802.001.00H3722-P1-C10-T1
```

You can also print the WWPN of an HBA directly by issuing this command:

```
lscfg -v1 <fcs#> | grep Network
```

where <fcs#> is the instance of an FC HBA to query.

Installing the XIV Host Attachment Kit for AIX

For AIX to correctly recognize the disks that are mapped from the XIV Storage System as *MPIO 2810 XIV Disk*, the *IBM XIV Host Attachment Kit for AIX* is required. This package also enables multipathing. At the time of writing, Host Attachment Kit 1.10.0 was used. The file set can be downloaded from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

Important: Although AIX now natively supports XIV using ODM changes that have been back-ported to several older AIX releases, install the XIV Host Attachment Kit. The kit provides support and access to the latest XIV utilities like `xiv_diag`. The output of these XIV utilities is mandatory for IBM support when you open an XIV-related service call.

To install the Host Attachment Kit, complete these steps:

1. Download or copy the downloaded Host Attachment Kit to your AIX system.
2. From the AIX prompt, change to the directory where your XIV package is located.
3. Run the `gunzip -c IBM_XIV_Host_Attachment_Kit_1.10.0-b1221_for_AIX.tar.gz | tar xvf -` command to extract the file.
4. Switch to the newly created directory and run the installation script (Example 4-5).

Example 4-5 Installing the AIX XIV Host Attachment Kit

```
# ./install.sh
Welcome to the XIV Host Attachment Kit installer.
Would you like to proceed and install the Host Attachment Kit? [Y/n]:
y
Please wait while the installer validates your existing configuration...
-----
Please wait, the Host Attachment Kit is being installed...
-----
Installation successful.
Please refer to the Host Attachment Guide for information on how to configure
this host.

-----
The IBM XIV Host Attachment Kit includes the following utilities:
xiv_attach: Interactive wizard that configures the host and verifies its
configuration for connectivity with the IBM XIV Storage System.
xiv_devlist: Lists of all XIV volumes that are mapped to the host, with general
info about non-XIV volumes.
xiv_syslist: Lists all XIV storage systems that are detected by the host.
xiv_diag: Performs complete diagnostics of the host and its connectivity with
the IBM XIV Storage System, and saves the information to a file.
xiv_fc_admin: Allows you to perform different administrative operations for
FC-connected hosts and XIV storage systems.
xiv_iscsi_admin: Allows you to perform different administrative operations for
iSCSI-connected hosts and XIV storage systems.
-----
```

5. Zone the host to XIV.
6. The Host Attachment Kit provides an interactive command-line utility to configure and connect the host to the XIV Storage System. Only Fibre Channel attachment is supported. If iSCSI attachment is needed, check the SSIC to find out which AIX versions are supported. The command `xiv_attach` starts a wizard that attaches the host to the XIV and create the host object on the XIV. Example 4-6 shows the `xiv_attach` command output.

Example 4-6 Attachment to XIV and host creation on XIV

```
# xiv_attach
-----
Welcome to the IBM XIV host attachment wizard, version 1.10.0.
This wizard will help you attach this host to the XIV storage system.

The wizard will now validate the host configuration for the XIV storage system.
Press [ENTER] to proceed.

-----
Only Fibre Channel connectivity is supported on this host.
Would you like to perform Fibre Channel attachment? [default: yes ]:
-----
Please wait while the wizard validates your existing configuration...
Verifying AIX packages                                     OK
This host is already configured for the XIV storage system.
-----
Please define zoning for this host and add its World Wide Port Names (WWPNs) to
the XIV storage system:
10:00:00:00:C9:B7:F2:7A: fcs0: [IBM]: N/A
10:00:00:00:C9:B7:F2:7B: fcs1: [IBM]: N/A
Press [ENTER] to proceed.

Would you like to rescan for new storage devices? [default: yes ]:
Please wait while rescanning for XIV storage devices...
-----
This host is connected to the following XIV storage arrays:
Serial   Version  Host Defined  Ports Defined  Protocol  Host Name(s)
1310114  11.1.1.0 No           None           FC         --
6000105  10.2.4.5 No           None           FC         --
1310133  11.1.1.0 No           None           FC         --
This host is not defined on some of the FC-attached XIV storage systems.
Do you want to define this host on these XIV systems now? [default: yes ]:
Please enter a name for this host [default: p7-770-02v21.mainz.de.ibm.com ]:
p770-02-lpar3
Please enter a username for system 1310114 [default: admin ]:  itso
Please enter the password of user itso for system 1310114:

Please enter a username for system 6000105 [default: admin ]:  itso
Please enter the password of user itso for system 6000105:

Please enter a username for system 1310133 [default: admin ]:  itso
Please enter the password of user itso for system 1310133:

Press [ENTER] to proceed.
```

The IBM XIV host attachment wizard has successfully configured this host.

Press [ENTER] to exit.

7. Create volumes on XIV and map these volumes (LUNs) to the host system that was configured by `xiv_attach`. You can use the XIV GUI for volume creation and mapping tasks, as illustrated in 1.4, “Logical configuration for host connectivity” on page 37. Use `cfgmgr` or `xiv_fc_admin -R` to rescan for the LUNs as shown in Example 4-7.

Example 4-7 XIV labeled FC disks

```
# xiv_fc_admin -R
# lsdev -Cc disk
hdisk1 Available          Virtual SCSI Disk Drive
hdisk2 Available 00-01-02 MPI0 2810 XIV Disk
hdisk3 Available 00-01-02 MPI0 2810 XIV Disk
```

8. Use `xiv_devlist` command to get more information about the mapped LUNs, as shown in Example 4-8.

Example 4-8 xiv_devlist command

```
# xiv_devlist -x
XIV Devices
-----
Device          Size (GB) Paths Vol Name          Vol Id  XIV Id  XIV Host
-----
/dev/hdisk2    103.2      12/12 p770_02_lpar3_1  6066   1310114 p770-02-lpar3
-----
/dev/hdisk3    103.2      12/12 p770_02_lpar3_2  6069   1310114 p770-02-lpar3
-----
```

To add disks to the system, complete the following steps:

1. Use the XIV GUI to map the new LUNs to the AIX server.
2. On the AIX system, run `xiv_fc_admin -R` to rescan for the new LUNs.
3. Use `xiv_devlist` to confirm that the new LUNs are present to the system.

Other AIX commands such as `cfgmgr` can also be used, but these commands are run within the XIV commands.

Portable XIV Host Attachment Kit Install and usage

The IBM XIV Host Attachment Kit is now offered in a portable format. With the portable package, you can use the Host Attachment Kit without having to install the utilities locally on the host. You can run all Host Attachment Kit utilities from a shared network drive or from a portable USB flash drive. This is the preferred method for deployment and management.

The `xiv_fc_admin` command can be used to confirm that the AIX server is running a supported configuration and ready to attach to the XIV storage. Use the `xiv_fc_admin -V` command to verify the configuration and be notified if any OS component is missing. The `xiv_attach` command must be run the first time that the server is attached to the XIV array. It is used to scan for new XIV LUNs and configure the server to work with XIV. Do not run the `xiv_attach` command more than once. If more LUNs are added in the future, use the

`xiv_fc_admin -R` command to scan for the new LUNs. All of these commands and others in the portable Host Attachment Kit are defined in 4.1.5, “Host Attachment Kit utilities” on page 159.

Using a network drive

To use the portable Host Attachment Kit package from a network drive:

1. Extract the files from `<HAK_build_name>_Portable.tar.gz` into a shared folder on a network drive.
2. Mount the shared folder to each host computer you intend to use the Host Attachment Kit on. The folder must be recognized and accessible as a network drive.

You can now use the IBM XIV Host Attachment Kit on any host to which the network drive is mounted.

To run commands from the portable Host Attachment Kit location, use `./` before every command.

Tip: Whenever a newer Host Attachment Kit version is installed on the network drive, all hosts to which that network drive was mounted have access to that version.

Using a portable USB flash drive

To use the portable Host Attachment Kit package from a USB flash drive, complete these steps:

1. Extract the files from `<HAK_build_name>_Portable.tar.gz` into a folder on the USB flash drive.
2. Plug the USB flash drive into any host on which you want to use the Host Attachment Kit.
3. Run any Host Attachment Kit utility from the drive.

For more information about setting up servers that use the portable Host Attachment Kit, see “AIX MPIO” on page 145.

Removing the Host Attachment Kit software

In some situation, you must remove the Host Attachment Kit. In most cases, when you are upgrading to a new version, the Host Attachment Kit can be installed without uninstalling the older version first. Check the release notes and instructions to determine the best procedure.

If the Host Attachment Kit is locally installed on the host, you can uninstall it without detaching the host from XIV.

The portable Host Attachment Kit packages do not require the uninstallation procedure. You can delete the portable Host Attachment Kit directory on the network drive or the USB flash drive to uninstall it. For more information about the portable Host Attachment Kit, see the previous section.

The regular uninstallation removes the locally installed Host Attachment Kit software without detaching the host. This process preserves all multipathing connections to the XIV Storage System.

Use the following command to uninstall the Host Attachment Kit software:

```
# /opt/xiv/host_attach/bin/uninstall
```

The **uninstall** command removes the following components:

- ▶ IBM Storage Solutions External Runtime Components
- ▶ IBM XIV Host Attachment Kit tools

If you get the message Please use the O/S package management services to remove the package, use the package management service to remove the Host Attachment Kit. The package name is `xiv.hostattachment.tools`. To remove the package, use the **installp -u xiv.hostattachment.tools** command as shown in Example 4-9.

Example 4-9 Uninstalling the Host Attach Kit

```
# installp -u xiv.hostattachment.tools
+-----+
                        Pre-deinstall Verification...
+-----+
Verifying selections...done
Verifying requisites...done
Results...

SUCSESSES
-----
  Filesets listed in this section passed pre-deinstall verification
  and will be removed.

  Selected Filesets
  -----
  xiv.hostattachment.tools 1.10.0.0          # Support tools for XIV connec...

  << End of Success Section >>

FILESET STATISTICS
-----
  1 Selected to be deinstalled, of which:
    1 Passed pre-deinstall verification
  ----
  1 Total to be deinstalled

+-----+
                        Deinstalling Software...
+-----+

installp: DEINSTALLING software for:
          xiv.hostattachment.tools 1.10.0.0

Removing dynamically created files from the system
Finished processing all filesets. (Total time: 3 secs).

+-----+
                        Summaries:
+-----+

Installation Summary
-----
```

Name	Level	Part	Event	Result
xiv.hostattachment.tools	1.10.0.0	USR	DEINSTALL	SUCCESS

AIX MPIO

AIX Multipath I/O (MPIO) is an enhancement to the base OS environment that provides native support for multi-path Fibre Channel storage attachment. MPIO automatically discovers, configures, and makes available every storage device path. The storage device paths provide high availability and load balancing for storage I/O. MPIO is part of the base AIX kernel, and is available with the current supported AIX levels.

The MPIO base functionality is limited. It provides an interface for vendor-specific path control modules (PCMs) that allow for implementation of advanced algorithms.

For more information, see the IBM System p® and AIX Information Center at:

<http://publib.boulder.ibm.com/infocenter/aix/v7r1/index.jsp>

Configuring XIV devices as MPIO or non-MPIO devices

Configuring XIV devices as MPIO provides the best solution. However, if you are using a third-party multipathing solution, you might want to manage the XIV 2810 device with the same solution. Using a non-IBM solution usually requires the XIV devices to be configured as non-MPIO devices.

AIX provides the `manage_disk_drivers` command to switch a device between MPIO and non-MPIO. This command can be used to change how the XIV device is configured. All XIV disks are converted.

Restriction: It is not possible to convert one XIV disk to MPIO and another XIV disk to non-MPIO.

To switch XIV 2810 devices from MPIO to non-MPIO, run the following command:

```
manage_disk_drivers -o AIX_non_MPIO -d 2810XIV
```

To switch XIV 2810 devices from non-MPIO to MPIO, run the following command:

```
manage_disk_drivers -o AIX_AAPCM -d 2810XIV
```

After you run either of these commands, the system must be rebooted for the configuration change to take effect.

To display the present settings, run the following command:

```
manage_disk_drivers -l
```

Disk behavior algorithms and queue depth settings

Using the XIV Storage System in a multipath environment, you can change the disk behavior algorithm between `round_robin` and `fail_over` mode. The default disk behavior mode is `round_robin`, with a queue depth setting of 40.

Check the disk behavior algorithm and queue depth settings as shown in Example 4-10.

Example 4-10 Viewing disk behavior and queue depth

```
# lsattr -El hdisk2 | grep -e algorithm -e queue_depth
algorithm          round_robin                Algorithm True
queue_depth        40                          Queue DEPTH True
```

If the application is I/O intensive and uses large block I/O, the `queue_depth` and the max transfer size might need to be adjusted. Such an environment typically needs a `queue_depth` of 64 - 256, and `max_transfer=0x100000`. Typical values are 40 - 64 as the queue depth per LUN, and 512-2048 per HBA in AIX.

Performance tuning

This section provides some performance considerations to help you adjust your AIX system to best fit your environment. If you boot from a SAN-attached LUN, create a `mksysb` image or a crash consistent snapshot of the boot LUN before you change the HBA settings. The following are performance considerations for AIX:

- ▶ Use multiple threads and asynchronous I/O to maximize performance on the XIV.
- ▶ Check with `iostat` on a per path basis for the LUNs to make sure that the load is balanced across all paths.
- ▶ Verify the HBA queue depth and per LUN queue depth for the host are sufficient to prevent queue waits. However, make sure that they are not so large that they overrun the XIV queues. The XIV queue limit is 1400 per XIV port and 256 per LUN per WWPN (host) per port. Do not submit more I/O per XIV port than the 1400 maximum it can handle. The limit for the number of queued I/O for an HBA on AIX systems with 8-Gb HBAs is 4096. This limit is controlled by the `num_cmd_elems` attribute for the HBA, which is the maximum number of commands that AIX queues. Increase it if necessary to the maximum value, which is 4096. The exception is if you have 1-Gbps, 2-Gbps, or 4-Gbps HBAs, in which cases the maximum is lower.
- ▶ The other setting to consider is the `max_xfer_size`. This setting controls the maximum I/O size the adapter can handle. The default is `0x100000`. If necessary, increase it to `0x200000` for large IOs, such as backups.

Check these values by using `lsattr -El fcsX` for each HBA as shown in Example 4-11.

Example 4-11 lsattr command

```
# lsattr -El fcs0
DIF_enabled      no          DIF (T10 protection) enabled      True
bus_intr_lvl     no          Bus interrupt level                 False
bus_io_addr      0xff800    Bus I/O address                     False
bus_mem_addr     0xffe76000 Bus memory address                  False
bus_mem_addr2    0xffe78000 Bus memory address                  False
init_link        auto       INIT Link flags                     False
intr_msi_1       135616     Bus interrupt level                 False
intr_priority    3          Interrupt priority                  False
lg_term_dma      0x800000   Long term DMA                       True
max_xfer_size    0x100000   Maximum Transfer Size               True
num_cmd_elems    500        Maximum number of COMMANDS to queue to the adapter True
pref_alpa        0x1        Preferred AL_PA                     True
sw_fc_class      2          FC Class for Fabric                 True
tme              no         Target Mode Enabled                 True

# lsattr -El fcs1
```

DIF_enabled	no	DIF (T10 protection) enabled	True
bus_intr_lvl		Bus interrupt level	False
bus_io_addr	0xffc00	Bus I/O address	False
bus_mem_addr	0xffe77000	Bus memory address	False
bus_mem_addr2	0xffe7c000	Bus memory address	False
init_link	auto	INIT Link flags	False
intr_msi_1	135617	Bus interrupt level	False
intr_priority	3	Interrupt priority	False
lg_term_dma	0x800000	Long term DMA	True
max_xfer_size	0x100000	Maximum Transfer Size	True
num_cmd_elems	500	Maximum number of COMMANDS to queue to the adapter	True
pref_alpa	0x1	Preferred AL_PA	True
sw_fc_class	2	FC Class for Fabric	True
tme	no	Target Mode Enabled	True

The maximum number of commands AIX queues to the adapter and the transfer size can be changed with the **chdev** command. Example 4-12 shows how to change these settings. The system must be rebooted for the changes to take effect.

Example 4-12 chdev command

```
# chdev -a 'num_cmd_elems=4096 max_xfer_size=0X200000' -l fcs0 -P
fcs0 changed

# chdev -a 'num_cmd_elems=4096 max_xfer_size=0X200000' -l fcs1 -P
fcs1 changed
```

The changes can be confirmed by running the **lsattr** command again (Example 4-13).

Example 4-13 lsattr command confirmation

```
# lsattr -El fcs0
...
max_xfer_size 0X200000 Maximum Transfer Size True
num_cmd_elems 4096 Maximum number of COMMANDS to queue to the adapter True
...
# lsattr -El fcs1
...
max_xfer_size 0X200000 Maximum Transfer Size True
num_cmd_elems 4096 Maximum number of COMMANDS to queue to the adapter True
...
```

To check the disk queue depth, periodically run **iostat -D 5**. If the `avgqsz` (average wait queue size) or `sqfull` are consistently greater zero, increase the disk queue depth. The maximum disk queue depth is 256. However, do not start at 256 and work down because you might flood the XIV with commands and waste memory on the AIX server. 64 is a good number for most environments. For more information about AIX disk queue depth tuning, see the following web page:

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TD105745>

See the following tables for the *minimum* level of service packs and the Host Attachment Kit version that are needed for each AIX version.

Table 4-1 shows the service packs and Host Attachment Kit version needed for AIX 5.3.

Table 4-1 AIX 5.3 minimum level service packs and Host Attachment Kit versions

AIX Release	Technology Level	Service pack	Host Attachment Kit Version
AIX 5.2	TL 10 ^a	SP 7	1.5.2
AIX 5.3	TL 7 ^a	SP 6	1.5.2
AIX 5.3	TL 8 ^a	SP 4	1.5.2
AIX 5.3	TL 9 ^a	SP 0	1.5.2
AIX 5.3	TL 10	SP 0	1.5.2
AIX 5.3	TL 11	SP 0	1.7.0
AIX 5.3	TL 12	SP 0	1.7.0

a. The queue depth is limited to 1 in round robin mode. Queue depth is limited to 256 when you use MPIO with the fail_over mode.

Table 4-2 shows the service packs and Host Attachment Kit version needed for AIX 6.1.

Table 4-2 AIX 6.1 minimum level service packs and Host Attachment Kit versions

AIX Release	Technology Level	Service pack	Host Attachment Kit Version
AIX 6.1	TL0 ^a	SP 6	1.5.2
AIX 6.1	TL1 ^a	SP 2	1.5.2
AIX 6.1	TL2 ^a	SP 0	1.5.2
AIX 6.1	TL3	SP 0	1.5.2
AIX 6.1	TL 4	SP0	1.5.2
AIX 6.1	TL 5	SP 0	1.7.0
AIX 6.1	TL 6	SP 0	1.7.0
AIX 6.1	TL 7	SP 0	1.8.0

a. The queue depth is limited to 1 in round robin mode. Queue depth is limited to 256 when you use MPIO with the fail_over mode.

Table 4-3 shows the service packs and Host Attachment Kit version needed for AIX 7.1.

Table 4-3 AIX 7.1 minimum level service pack and Host Attachment Kit version

AIX Release	Technology Level	Service pack	Host Attachment Kit Version
AIX 7.1	TL 0	SP 0	1.7.0
AIX 7.1	TL 1	SP 0	1.8.0

The default disk behavior algorithm is round_robin with a queue depth of 40. If the appropriate AIX technology level and service packs list are met, this queue depth restriction is lifted and the settings can be adjusted.

Example 4-14 shows how to adjust the disk behavior algorithm and queue depth setting.

Example 4-14 Changing disk behavior algorithm and queue depth command

```
# chdev -a algorithm=round_robin -a queue_depth=40 -l <hdisk#>
```

In the command, *<hdisk#>* stands for an instance of a hdisk.

If you want the fail_over disk behavior algorithm, load balance the I/O across the FC adapters and paths. Set the path priority attribute for each LUN so that $1/n^{\text{th}}$ of the LUNs are assigned to each of the *n* FC paths.

Useful MPIO commands

The following commands are used to change priority attributes for paths that can specify a preference for the path that is used for I/O. The effect of the priority attribute depends on whether the disk behavior algorithm attribute is set to fail_over or round_robin.

- ▶ For **algorithm=fail_over**, the path with the higher priority value handles all the I/O. If there is a path failure, the other path is used. After a path failure and recovery, if you have IY79741 installed, I/O will be redirected down the path with the highest priority. If you want the I/O to go down the primary path, use **chpath** to disable and then re-enable the secondary path. If the priority attribute is the same for all paths, the first path that is listed with **lspath -H1 <hdisk>** is the primary path. Set the primary path to priority value 1, the next path's priority (in case of path failure) to 2, and so on.
- ▶ For **algorithm=round_robin**, if the priority attributes are the same, I/O goes down each path equally. If you set pathA's priority to 1 and pathB's to 255, for every I/O going down pathA, 255 I/O are sent down pathB.

To change the path priority of an MPIO device, use the **chpath** command. An example of this process is shown in Example 4-17 on page 150.

Initially, use the **lspath** command to display the operational status for the paths to the devices as shown in Example 4-15.

Example 4-15 The lspath command shows the paths for hdisk2

```
# lspath -l hdisk2 -F status:name:parent:path_id:connection
Enabled:hdisk2:fscsi0:0:5001738027820170,1000000000000
Enabled:hdisk2:fscsi0:1:5001738027820160,1000000000000
Enabled:hdisk2:fscsi0:2:5001738027820150,1000000000000
Enabled:hdisk2:fscsi0:3:5001738027820140,1000000000000
Enabled:hdisk2:fscsi0:4:5001738027820180,1000000000000
Enabled:hdisk2:fscsi0:5:5001738027820190,1000000000000
Enabled:hdisk2:fscsi1:6:5001738027820162,1000000000000
Enabled:hdisk2:fscsi1:7:5001738027820152,1000000000000
Enabled:hdisk2:fscsi1:8:5001738027820142,1000000000000
Enabled:hdisk2:fscsi1:9:5001738027820172,1000000000000
Enabled:hdisk2:fscsi1:10:5001738027820182,1000000000000
Enabled:hdisk2:fscsi1:11:5001738027820192,1000000000000
```

The **lspath** command can also be used to read the attributes of a path to an MPIO-capable device as shown in Example 4-16. The *<connection>* information is either “<SCSI ID>, <LUN ID>” for SCSI (for example “5, 0”) or “<WWN>, <LUN ID>” for FC devices (Example 4-16).

Example 4-16 The lspath command reads attributes of the 0 path for hdisk2

```
# lspath -AHE -l hdisk2 -p fscsi0 -w "5001738027820170,10000000000000"
attribute value          description  user_settable

scsi_id  0x20ac00          SCSI ID    False
node_name 0x5001738027820000 FC Node Name False
priority  1              Priority    True
```

As noted, the **chpath** command is used to run change operations on a specific path. It can either change the operational status or tunable attributes that are associated with a path. It cannot run both types of operations in a single invocation.

Example 4-17 illustrates the use of the **chpath** command with an XIV Storage System. The command sets the primary path to fscsi0 using the first path listed. There are two paths from the switch to the storage for this adapter. For the next disk, set the priorities to 4, 1, 2, and 3. In fail-over mode, assuming the workload is relatively balanced across the hdisk. This setting balances the workload evenly across the paths.

Example 4-17 The chpath command

```
# chpath -l hdisk2 -p fscsi0 -w "5001738027820160,1000000000000000" -a priority=2
path Changed
# chpath -l hdisk2 -p fscsi1 -w "5001738027820162,1000000000000000" -a priority=3
path Changed
# chpath -l hdisk2 -p fscsi1 -w "5001738027820152,1000000000000000" -a priority=4
path Changed
```

The **rmppath** command unconfigures or undefines, or both, one or more paths to a target device. You cannot unconfigure (undefine) the last path to a target device by using the **rmppath** command. The only way to unconfigure (undefine) the last path to a target device is to unconfigure the device itself. Use the **rmdev** command to do so.

4.1.3 AIX host iSCSI configuration

To make sure that your AIX version is supported for iSCSI attachment (for iSCSI hardware or software initiator), check the IBM SSIC at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

For iSCSI, no Host Attachment Kit is required. Make sure that your system is equipped with the required file sets by running the **ls1pp** command as shown in Example 4-18.

Example 4-18 Verifying installed iSCSI file sets in AIX

```
# ls1pp -la "*.iscsi*"
Fileset          Level  State      Description
-----
Path: /usr/lib/objrepos
  devices.common.IBM.iscsi.rte
    6.1.0.0  COMMITTED  Common iSCSI Files
    6.1.3.0  COMMITTED  Common iSCSI Files
  devices.iscsi.disk.rte
    6.1.0.0  COMMITTED  iSCSI Disk Software
```


devices.iscsi.tape.rte	6.1.0.0	COMMITTED	iSCSI Tape Software
devices.iscsi_sw.rte	6.1.0.0	COMMITTED	iSCSI Software Device Driver
	6.1.3.0	COMMITTED	iSCSI Software Device Driver

Path: /etc/objrepos

devices.common.IBM.iscsi.rte	6.1.0.0	COMMITTED	Common iSCSI Files
	6.1.3.0	COMMITTED	Common iSCSI Files
devices.iscsi_sw.rte	6.1.0.0	COMMITTED	iSCSI Software Device Driver

Current limitations when using iSCSI

The code available at the time of writing has the following limitations when you are using the iSCSI software initiator in AIX:

- ▶ iSCSI is supported through a single path. No MPIIO support is provided.
- ▶ The `xiv_iscsi_admin` command does not discover new targets on AIX. You must manually add new targets.
- ▶ The `xiv_attach` wizard does not support iSCSI.

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 are created using iSCSI devices to be in an inactive state after reboot. After the iSCSI devices are configured, manually activate the iSCSI-backed Volume Groups. Then, mount any associated file systems.

Restriction: Volume Groups are activated during a different boot phase than the iSCSI software. For this reason, you cannot activate iSCSI Volume Groups during the boot process

- ▶ Do not span Volume Groups across non-iSCSI devices.

I/O failures

To avoid I/O failures, consider these recommendations:

- ▶ If connectivity to iSCSI target devices is lost, I/O failures occur. Before you do anything that causes long-term loss of connectivity to the active iSCSI targets, stop all I/O activity and unmount iSCSI-backed file systems.
- ▶ If a loss of connectivity occurs while applications are attempting I/O activities with iSCSI devices, I/O errors eventually occur. You might not be able to unmount iSCSI-backed file systems because the underlying iSCSI device remains busy.
- ▶ File system maintenance must be performed if I/O failures occur because of loss of connectivity to active iSCSI targets. To do file system maintenance, run the `fsck` command against the affected file systems.

Configuring the iSCSI software initiator and the server on XIV

To connect AIX to the XIV through iSCSI, complete the following steps:

1. Get the *iSCSI qualified name* (IQN) on the AIX server, and set the maximum number of targets by using the *System Management Interface Tool* (SMIT):
 - a. Select **Devices**.
 - b. Select **iSCSI**.
 - c. Select **iSCSI Protocol Device**.
 - d. Select **Change / Show Characteristics of an iSCSI Protocol Device**.
 - e. Select the device and verify the iSCSI Initiator Name value. The Initiator Name value is used by the iSCSI Target during login.

Tip: A default initiator name is assigned when the software is installed. This initiator name can be changed to match local network naming conventions.

You can also issue the `lsattr` command to verify the `initiator_name` parameter as shown in Example 4-19.

Example 4-19 Checking initiator name

```
# lsattr -El iscsi0
disc_filename /etc/iscsi/targets
Configuration file False
disc_policy file Discovery
Policy True
initiator_name iqn.com.ibm.de.mainz.p590-tic-1-8.hostid.099b5778 iSCSI
Initiator Name True
isns_srvnames auto iSNS
Servers IP Addresses True
isns_srvports iSNS
Servers Port Numbers True
max_targets 16 Maximum
Targets Allowed True
num_cmd_elems 200 Maximum
number of commands to queue to driver True
```

- f. The **Maximum Targets Allowed** field corresponds to the maximum number of iSCSI targets that can be configured. If you reduce this number, you also reduce the amount of network memory pre-allocated for the iSCSI protocol during configuration.

2. Define the AIX server on XIV with the host and cluster window (Figure 4-1).

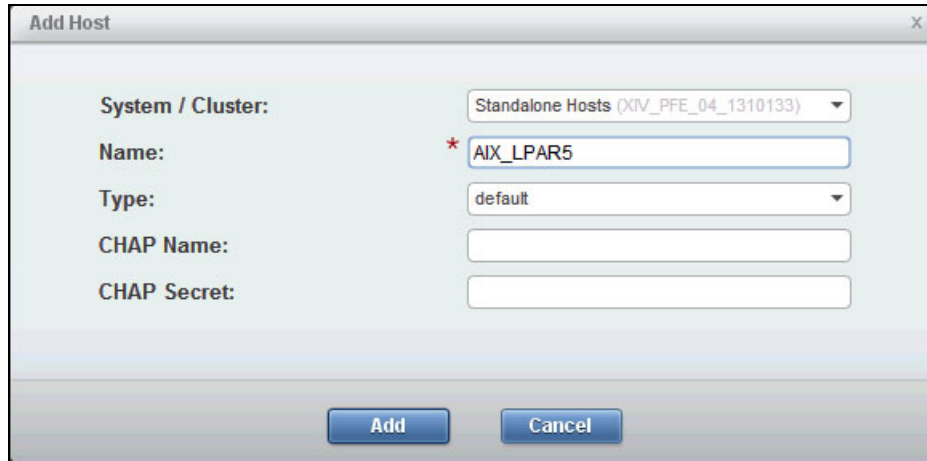


Figure 4-1 Adding the iSCSI host

3. Right-click the new host name and select **Add Port** (Figure 4-2).

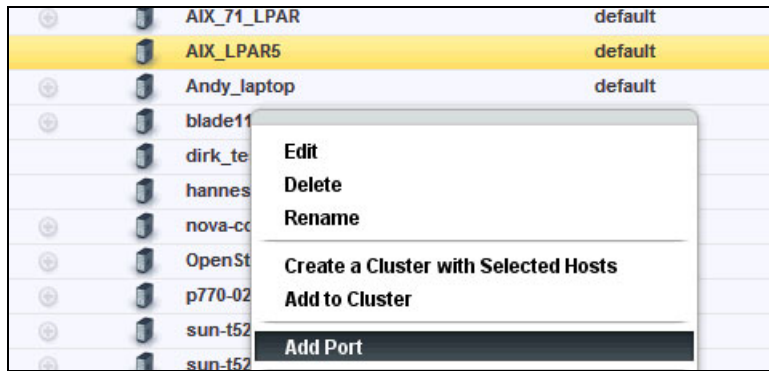


Figure 4-2 Adding port

4. Configure the port as an iSCSI port and enter the IQN name that you collected in Example 4-19. Add this value to the iSCSI Name as shown in Figure 4-3.



Figure 4-3 Configuring iSCSI port

5. Create the LUNs in XIV and map them to the AIX iSCSI server so the server can see them in the following steps.
6. Determine the iSCSI IP addresses in the XIV Storage System by selecting **iSCSI Connectivity** from the Host and LUNs menu (Figure 4-4).



Figure 4-4 iSCSI Connectivity

7. The iSCSI connectivity panel in Figure 4-5 shows all the available iSCSI ports. Set the MTU to 4500 if your network supports jumbo frames (Figure 4-5).

Name ▲	System	Address	Netmask	Gateway	MTU	Module
M4P1	XIV_PFE_04_1310...	9.155.50.11	255.255.255.0	9.155.50.1	4500	1:Module:4
M5P1	XIV_PFE_04_1310...	9.155.50.12	255.255.255.0	9.155.50.1	4500	1:Module:5
M6P1	XIV_PFE_04_1310...	9.155.50.13	255.255.255.0	9.155.50.1	4500	1:Module:6

Figure 4-5 XIV iSCSI ports

8. In the system view in the XIV GUI, right-click the XIV Storage box itself, and select **Settings** → **Parameters**.
9. Find the IQN of the XIV Storage System in the System Settings window (Figure 4-6).

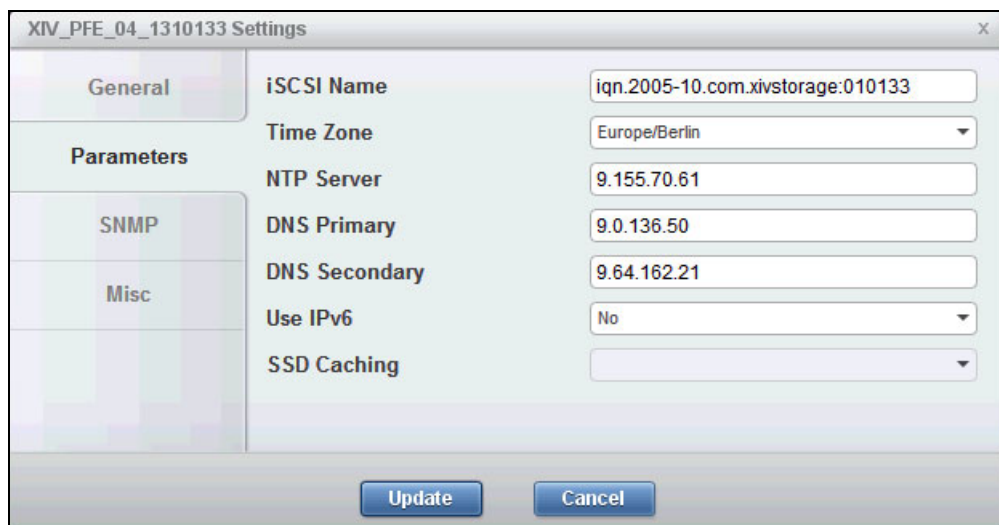


Figure 4-6 Verifying iSCSI name in XIV Storage System

If you are using XCLI, issue the `config_get` command as shown in Example 4-20.

Example 4-20 The config_get command in XCLI

```
XIV_PFE_04_1310133>>config_get
Name                Value
dns_primary         9.0.136.50
dns_secondary       9.64.162.21
system_name         XIV_PFE_04_1310133
snmp_location
snmp_contact
snmp_community      XIV
snmp_trap_community XIV
system_id           10133
machine_type        2810
machine_model        114
machine_serial_number 1310133
email_sender_address
email_reply_to_address
email_subject_format {severity}: {description}
internal_email_subject_format {machine_type}-{machine_model}:
{machine_serial_number}: {severity}: {description}
iscsi_name           iqn.2005-10.com.xivstorage:010133
timezone             -7200
ntp_server           9.155.70.61
ups_control          yes
support_center_port_type Management
isns_server
ipv6_state           disabled
ipsec_state          disabled
ipsec_track_tunnels no
```

10. Return to the AIX system and add the XIV iSCSI IP address, port name, and IQN to the `/etc/iscsi/targets` file. This file must include the iSCSI targets for the device configuration.

Tip: The iSCSI targets file defines the name and location of the iSCSI targets that the iSCSI software initiator attempts to access. This file is read every time that the iSCSI software initiator is loaded.

Each uncommented line in the file represents an iSCSI target. iSCSI device configuration requires that the iSCSI targets can be reached through a properly configured network interface. Although the iSCSI software initiator can work using a 10/100 Ethernet LAN, it is designed for use with a separate gigabit Ethernet network.

Include your specific connection information in the targets file as shown in Example 4-21.

Example 4-21 Inserting connection information into the /etc/iscsi/targets file in AIX

```
# cat /etc/iscsi/targets
...
9.155.50.11 3260 iqn.2005-10.com.xivstorage:010133
```

11. Enter the following command at the AIX prompt:

```
cfgmgr -l iscsi0
```

This command runs the following actions:

- Reconfigures the software initiator
- Causes the driver to attempt to communicate with the targets listed in the `/etc/iscsi/targets` file
- Defines a new hdisk for each LUN found on the targets

12. Run `lsdev -Cc disk` to view the new iSCSI devices. Example 4-22 shows one iSCSI disk.

Example 4-22 iSCSI confirmation

```
# lsdev -Cc disk
hdisk0 Available Virtual SCSI Disk Drive
hdisk1 Available Other iSCSI Disk Drive
```

Exception: If the appropriate disks are not defined, review the configuration of the initiator, the target, and any iSCSI gateways to ensure correctness. Then, rerun the `cfgmgr` command.

iSCSI performance considerations

To ensure the best performance, enable the following features of the AIX Gigabit Ethernet Adapter and the iSCSI Target interface:

- ▶ TCP Large Send
- ▶ TCP send and receive flow control
- ▶ Jumbo frame

The first step is to confirm that the network adapter supports jumbo frames. Jumbo frames are Ethernet frames that support more than 1500 bytes. Jumbo frames can carry up to 9000 bytes of payload, but some care must be taken when using the term. Many different Gigabit Ethernet switches and Gigabit Ethernet network cards can support jumbo frames. Check the network card specification or the vendor's support website to confirm that the network card supports this function.

You can use `lsattr` to list some of the current adapter device driver settings. Enter `lsattr -E -l ent0`, where `ent0` is the adapter name. Make sure that you are checking and modifying the correct adapter. A typical output is shown in Example 4-23.

Example 4-23 lsattr output that displays adapter settings

```
# lsattr -E -l ent0
alt_addr      0x000000000000    Alternate ethernet address      True
busintr       167              Bus interrupt level             False
busmem        0xe8120000       Bus memory address              False
chksum_offload yes              Enable hardware transmit and receive checksum True
compat_mode   no               Gigabit Backward compatability True
copy_bytes    2048            Copy packet if this many or less bytes True
delay_open    no               Enable delay of open until link state is known True
failback      yes              Enable auto failback to primary True
failback_delay 15              Failback to primary delay timer True
failover      disable          Enable failover mode            True
flow_ctrl     yes              Enable Transmit and Receive Flow Control True
intr_priority 3                Interrupt priority              False
intr_rate     10000           Max rate of interrupts generated by adapter True
jumbo_frames  no               Transmit jumbo frames           True
large_send    yes              Enable hardware TX TCP resegmentation True
media_speed   Auto_Negotiation Media speed                      True
rom_mem       0xe80c0000       ROM memory address              False
```

rx_hog	1000	Max rcv buffers processed per rcv interrupt	True
rxbuf_pool_sz	2048	Rcv buffer pool, make 2X rxdesc_que_sz	True
rxdesc_que_sz	1024	Rcv descriptor queue size	True
slih_hog	10	Max Interrupt events processed per interrupt	True
tx_que_sz	8192	Software transmit queue size	True
txdesc_que_sz	512	TX descriptor queue size	True
use_alt_addr	no	Enable alternate ethernet address	True

In the example, `jumbo_frames` are off. When this setting is not enabled, you cannot increase the network speed. Set up the `tcp_sendspace`, `tcp_recvspace`, `sb_max`, and `mtu_size` network adapter and network interface options to optimal values.

To see the current settings, use `lsattr` to list the settings for `tcp_sendspace`, `tcp_recvspace`, and `mtu_size` (Example 4-24).

Example 4-24 lsattr output that displays interface settings

```
lsattr -E -l en0
```

alias4		IPv4 Alias including Subnet Mask	True
alias6		IPv6 Alias including Prefix Length	True
arp	on	Address Resolution Protocol (ARP)	True
authority		Authorized Users	True
broadcast		Broadcast Address	True
mtu	1500	Maximum IP Packet Size for This Device	True
netaddr	9.155.87.120	Internet Address	True
netaddr6		IPv6 Internet Address	True
netmask	255.255.255.0	Subnet Mask	True
prefixlen		Prefix Length for IPv6 Internet Address	True
remmtu	576	Maximum IP Packet Size for REMOTE Networks	True
rfc1323		Enable/Disable TCP RFC 1323 Window Scaling	True
security	none	Security Level	True
state	up	Current Interface Status	True
tcp_mssdflt		Set TCP Maximum Segment Size	True
tcp_nodelay		Enable/Disable TCP_NODELAY Option	True
tcp_recvspace		Set Socket Buffer Space for Receiving	True
tcp_sendspace		Set Socket Buffer Space for Sending	True

Example 4-24 shows that all values are true, and that `mtu` is set to 1500.

To change the `mtu` setting, enable `jumbo_frames` on the adapter. Issue the following command:

```
chdev -l ent0 -a jumbo_frames=yes -P
```

Reboot the server by entering `shutdown -Fr`. Check the interface and adapter settings and confirm the changes (Example 4-25).

Example 4-25 The adapter settings after you make the changes

```
# lsattr -E -l ent0
```

...			
jumbo_frames	yes	Transmit jumbo frames	True
...			

Example 4-26 shows that the mtu value is changed to 9000. XIV only supports a mtu size of 4500.

Example 4-26 The interface settings after you make the changes

```
# lsattr -E -l en0
...
mtu          9000          Maximum IP Packet Size for This Device    True
...

```

Use the following command to change the mtu to 4500 on the AIX server:

```
chdev -l en0 -a mtu=4500
```

Confirm that the setting is changed. Use the `/usr/sbin/no -a` command to show the `sb_max`, `tcp_recvspace`, and `tcp_sendspace` values as shown in Example 4-27.

Example 4-27 Checking values by using the /usr/sbin/no -a command

```
# /usr/sbin/no -a
...
                sb_max = 1048576
...
                tcp_recvspace = 16384
                tcp_sendspace = 16384
...

```

There are three other settings to check:

- ▶ `tcp_sendspace`: This setting specifies how much data the sending application can buffer in the kernel before the application is blocked on a send call.
- ▶ `tcp_recvspace`: This setting specifies how many bytes of data the receiving system can buffer in the kernel on the receiving sockets queue.
- ▶ `sb_max`: Sets an upper limit on the number of socket buffers queued to an individual socket. It therefore controls how much buffer space is used by buffers that are queued to a sender socket or receiver socket.

Set these three settings as follows:

1. `tcp_sendspace`, `tcp_recvspace`, and `sb_max` network: The maximum transfer size of the iSCSI software initiator is 256 KB. Assuming that the system maximums for `tcp_sendspace` and `tcp_recvspace` are set to 262144 bytes, use the `ifconfig` command to configure a gigabit Ethernet interface by using the following command:

```
ifconfig en0 9.155.87.120 tcp_sendspace 262144 tcp_recvspace 262144
```
2. `sb_max`: Set this network option to at least 524288, and preferably 1048576. The `sb_max` sets an upper limit on the number of socket buffers queued. Set this limit with the command `/usr/sbin/no -o sb_max=1048576`.

4.1.4 Management volume LUN 0

According to the SCSI standard, XIV Storage System maps itself in every map to LUN 0 for inband Fibre Channel XIV management. This LUN serves as the “well known LUN” for that map. The host can then issue SCSI commands to that LUN that are not related to any specific volume. This device is displayed as a normal hdisk in the AIX operating system.

You might want to eliminate this management LUN on your system, or need to assign the LUN 0 number to a specific volume.

To convert LUN 0 to a real volume, complete the following steps:

1. Right-click LUN 0 and select **Enable** to allow mapping LUNs to LUN 0 (Figure 4-7).



Figure 4-7 Enabling LUN 0 mapping

2. Map your volume to LUN 0, and it replaces the management LUN to your volume.

4.1.5 Host Attachment Kit utilities

The Host Attachment Kit includes these useful utilities:

- ▶ **xiv_devlist**
- ▶ **xiv_diag**
- ▶ **xiv_attach**
- ▶ **xiv_fc_admin**
- ▶ **xiv_iscsi_admin** (xiv_iscsi_admin is not supported on AIX)
- ▶ **xiv_detach** (applicable to Windows Server only)

These utilities have the following functions”

- ▶ **xiv_devlist**

The **xiv_devlist** utility lists all volumes that are mapped to the AIX host. Example 4-28 shows the output of this command for two XIV disks that are attached over two Fibre Channel paths. The hdisk0 is a non-XIV device. The **xiv-devlist** command shows which hdisk represents which XIV volume.

Example 4-28 *xiv_devlist* output

```
# xiv_devlist
XIV Devices
-----
Device          Size (GB) Paths Vol Name          Vol Id  XIV Id  XIV Host
-----
/dev/hdisk2    103.2      12/12 p770_02_1par3_1  6066   1310114 p770-02-1par3
-----
/dev/hdisk3    103.2      12/12 p770_02_1par3_2  6069   1310114 p770-02-1par3
-----

Non-XIV Devices
-----
Device          Size (GB) Paths
-----
/dev/hdisk1    34.4       N/A
-----
```

The following options are available for the `xiv_devlist` command:

- `-h, --help`: Shows help
- `-t, xml`: Provides XML (default: tui)
- `-o`: Selects fields to display as comma-separated with no spaces.
- `-l`: Shows the list of fields
- `-f`: Shows file to output. Can be used only with `-t csv/xml`
- `-H, --hex`: Displays XIV volume and system IDs in hexadecimal base
- `-u --size-unit=SIZE_UNIT`: Selects size unit to use, such as MB, GB, TB, MiB, and GiB)
- `-d, --debug`: Enables debug logging
- `-l, --list-fields`: Lists available fields for the `-o` option
- `-m`: Enforces a multipathing framework <auto|natively|veritas>
- `-x, --xiv-only`: Displays only XIV devices
- `-V, --version`: Shows the version of the HostAttachmentKit framework

► **xiv_diag**

The `xiv_diag` utility gathers diagnostic data from the AIX operating system and saves it in a compressed file. This file can be sent to IBM support for analysis. Example 4-29 shows a sample output.

Example 4-29 xiv_diag output

```
# xiv_diag
Welcome to the XIV diagnostics tool, version 1.10.0.
This tool will gather essential support information from this host.
Please type in a path to place the xiv_diag file in [default: /tmp]:
Creating archive xiv_diag-results_2012-10-22_14-12-1
INFO: Gathering xiv_devlist logs... DONE
INFO: Gathering xiv_attach logs... DONE
INFO: Gathering build-revision file... DONE
INFO: Gathering HAK version... DONE
INFO: Gathering xiv_devlist... DONE
INFO: Gathering xiv_fc_admin -V... DONE
INFO: Gathering xiv_fc_admin -L... DONE
INFO: Gathering xiv_fc_admin -P... DONE
INFO: Gathering uname... DONE
INFO: Gathering snap: output... DONE
INFO: Gathering /tmp/ibmsupt.xiv directory... DONE
INFO: Gathering rm_cmd: output... DONE

INFO: Closing xiv_diag archive file DONE
Deleting temporary directory... DONE
INFO: Gathering is now complete.
INFO: You can now send /tmp/xiv_diag-results_2012-10-22_14-12-1.tar.gz to IBM-XIV
for review.
INFO: Exiting.
```

► **xiv_attach**

The `xiv_attach wizard` is a utility that assists with attaching the server to the XIV system. See Example 4-6 on page 141 to see the wizard and an example of what it does.

► **xiv_fc_admin**

The **xiv_fc_admin** utility is used to run administrative tasks and query Fibre Channel attachment-related information.

The following options are available for the **xiv_fc_admin** command:

- **-h, --help**: Shows this help message and exit
- **-v, --version**: Prints hostattachment kit version
- **-b, --build**: Prints hostattachment build number
- **-V, --verify**: Verifies host configuration tasks
- **-C, --configure**: Configures this host for attachment
- **-R, --rescan**: Rescans devices
- **-D, --define**: Defines this host on a system
- **-L, --list**: Lists attached XIV systems
- **-P, --print**: Prints WWPN of HBA devices

The following host definition options are available for the **xiv_fc_admin** command:

- **-u USERNAME, --user=USERNAME**: Sets username for XCLI
- **-p PASSWORD, --pass=PASSWORD**: Sets password for XCLI
- **-H HOSTNAME, --hostname=HOSTNAME**: Sets the optional hostname for this host. Unless specified, the os hostname is used
- **-S SERIAL, --serial=SERIAL**: Sets the serial number of the system. See the parameter **--list**.

► **xiv_iscsi_admin**

The **xiv_iscsi_admin** is not supported on AIX.

► **xiv_detach**

The **xiv_detach** command is applicable on Windows only.

For more information, see the IBM Storage Host Software Solutions link in the XIV Information Center at:

<http://publib.boulder.ibm.com/infocenter/ibmxiv/r2/index.jsp>

4.2 SAN boot in AIX

This section contains step-by-step instructions for SAN boot implementation for the IBM POWER System (formerly System p) in an AIX v6.1 environment. Similar steps can be followed for other AIX environments.

When you use AIX SAN boot with XIV, the default MPIO is used. During the boot sequence, AIX uses the bootlist to find valid paths to a LUN/hdisk that contains a valid boot logical volume (hd5). However, a maximum of five paths can be defined in the bootlist, while the XIV multipathing setup results in more than five paths to a hdisk. A fully redundant configuration establishes 12 paths (Figure 1-7 on page 16).

For example, consider two hdisks (hdisk0 and hdisk1) containing a valid boot logical volume, both having 12 paths to the XIV Storage System. To set the bootlist for hdisk0 and hdisk1, issue the following command:

```
/ > bootlist -m normal hdisk0 hdisk1
```

The `bootlist` command displays the list of boot devices as shown in Example 4-30.

Example 4-30 Displaying the bootlist

```
# bootlist -m normal -o
hdisk0 blv=hd5 pathid=0
```

Example 4-30 shows that `hdisk1` is not present in the bootlist. Therefore, the system cannot boot from `hdisk1` if the paths to `hdisk0` are lost.

There is a workaround in AIX 6.1 TL06 and AIX 7.1 to control the bootlist by using the `pathid` parameter as in the following command:

```
bootlist -m normal hdisk0 pathid=0 hdisk0 pathid=1 hdisk1 pathid=0 hdisk1 pathid=1
```

Implement SAN boot with AIX by using one of the following methods:

- ▶ For a system with an already installed AIX operating system, mirror the `rootvg` volume to the SAN disk.
- ▶ For a new system, start the AIX installation from a bootable AIX CD installation package or use Network Installation Management (NIM).

The *mirroring* method is simpler to implement than using the NIM.

4.2.1 Creating a SAN boot disk by mirroring

The mirroring method requires that you have access to an AIX system that is up and running. Locate an available system where you can install AIX on an internal SCSI disk.

To create a boot disk on the XIV system, complete the following steps:

1. Select a logical drive that is the same size or larger than the size of `rootvg` currently on the internal SCSI disk. Verify that your AIX system can see the new disk with the `lspv -L` command as shown in Example 4-31.

Example 4-31 lspv command

```
# lspv -L
hdisk0      00cc6de1b1d84ec9      rootvg      active
hdisk1      none                      None
hdisk2      none                      None
hdisk3      none                      None
hdisk4      none                      None
hdisk5      00cc6de1cfb8ea41      None
```

2. Verify the size with the `xiv_devlist` command to make sure that you are using an XIV (external) disk. Example 4-32 shows that `hdisk0` is 32 GB, `hdisks 1 through 5` are attached, and they are XIV LUNs. Notice that `hdisk1` is only 17 GB, so it is not large enough to create a mirror.

Example 4-32 xiv_devlist command

```
# ./xiv_devlist

XIV Devices
-----
Device      Size (GB) Paths Vol Name      Vol Id  XIV Id  XIV Host
-----
/dev/hdisk1 17.2      2/2  ITS0_Anthony_ 1018   1310114 AIX_P570_2_1p
```

		Blade1_Iomete		ar2	
/dev/hdisk2	1032.5	2/2	CUS_Jake	230	1310114 AIX_P570_2_1p ar2
/dev/hdisk3	34.4	2/2	CUS_Lisa_143	232	1310114 AIX_P570_2_1p ar2
/dev/hdisk4	1032.5	2/2	CUS_Zach	231	1310114 AIX_P570_2_1p ar2
/dev/hdisk5	32.2	2/2	LPAR2_boot_mi rror	7378	1310114 AIX_P570_2_1p ar2

Non-XIV Devices

Device	Size (GB)	Paths
/dev/hdisk0	32.2	1/1

3. Add the new disk to the rootvg volume group by clicking **smitty vg** → **Set Characteristics of a Volume Group** → **Add a Physical Volume to a Volume Group**.
4. Leave **Force the creation of volume group** set to no.
5. Enter the **Volume Group name** (in this example, rootvg) and **Physical Volume name** that you want to add to the volume group (Figure 4-8).

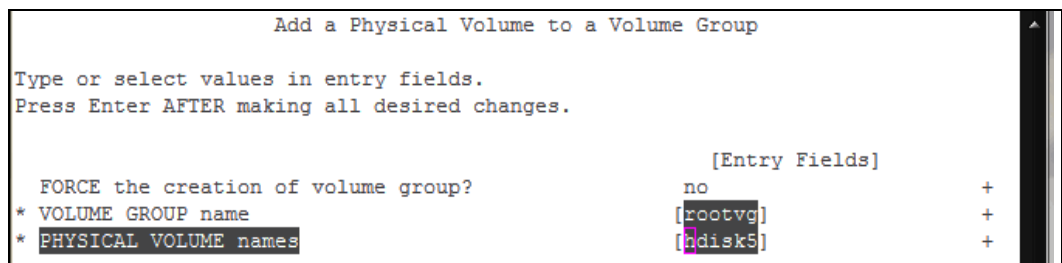


Figure 4-8 Adding the disk to the rootvg

Figure 4-9 shows the settings confirmation.

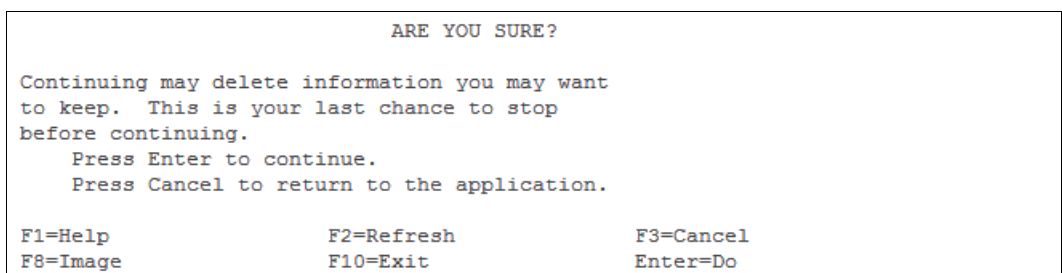


Figure 4-9 Adding disk confirmation

6. Create the mirror of rootvg. If the rootvg is already mirrored, create a third copy on the new disk by clicking **smitty vg** → **Mirror a Volume Group** (Figure 4-10).

```

Mirror a Volume Group

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

                                [Entry Fields]
* VOLUME GROUP name                rootvg
Mirror Sync Mode                    [Foreground]          +
PHYSICAL VOLUME names              [hdisk5]              +
Number of COPIES of each logical   2                    +
partition
Keep Quorum Checking On?           no                    +
Create Exact LV Mapping?           no                    +

```

Figure 4-10 Creating a rootvg mirror

Enter the volume group name that you want to mirror (rootvg, in this example).

7. Select the one of the following mirror sync modes:
 - **Foreground:** This option causes the command to run until the mirror copy synchronization completes. The synchronization can take a long time. The amount of time depends mainly on the speed of your network and how much data you have.
 - **Background:** This option causes the command to complete immediately, and mirror copy synchronization occurs in the background. With this option, it is not obvious when the mirrors complete their synchronization.
 - **No Sync:** This option causes the command to complete immediately without running any type of mirror synchronization. If this option is used, the new remote mirror copy exists but is marked as stale until it is synchronized with the **syncvg** command.
8. Select the Physical Volume name. You added this drive to your disk group in Figure 4-8 on page 163. The number of copies of each logical volume is the number of physical partitions that are allocated for each logical partition. The value can be one to three. A value of two or three indicates a mirrored logical volume. Leave the **Keep Quorum Checking on** and **Create Exact LV Mapping** settings at no.

After the volume is mirrored, you see confirmation that the mirror was successful as shown in Figure 4-11.

```

COMMAND STATUS

Command: OK          stdout: yes          stderr: no

Before command completion, additional instructions may appear below.

0516-1804 chvg: The quorum change takes effect immediately.
0516-1126 mirrorvg: rootvg successfully mirrored, user should perform
                  bosboot of system to initialize boot records. Then, user must modify
                  bootlist to include: hdisk0 hdisk5.

```

Figure 4-11 Mirror completed

- Verify that all partitions are mirrored with `lsvg -l rootvg` (Figure 4-12). The Physical Volume (PVs) column displays as two or three, depending on the number you chose when you created the mirror.

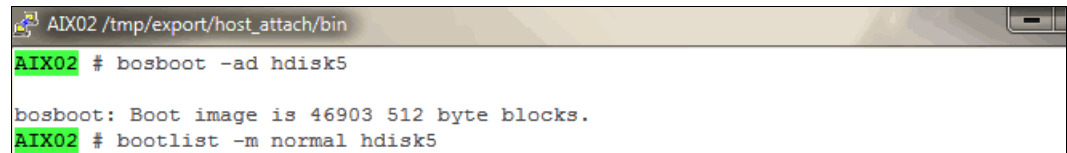
```
AIX02 # lsvg -l rootvg
rootvg:
LV NAME          TYPE      LPs      PPs      PVs  LV STATE  MOUNT POINT
hd5              boot      1         2         2    closed/syncd  N/A
hd6              paging    16        32        2    open/syncd    N/A
hd8              jfs2log   1          2         2    open/syncd    N/A
hd4              jfs2      28         56        2    open/syncd    /
hd2              jfs2     320        640        2    open/syncd    /usr
hd9var           jfs2      79        158        2    open/syncd    /var
hd3              jfs2      32         64        2    open/syncd    /tmp
hd1              jfs2      32         64        2    open/syncd    /home
hd10opt          jfs2      14         28        2    open/syncd    /opt
hd11admin        jfs2       4          8         2    open/syncd    /admin
livedump         jfs2       8         16        2    open/syncd    /var/adm/ras/livedu
```

Figure 4-12 Verifying that all partitions are mirrored

- Re-create the boot logical drive, and change the normal boot list with the following commands:

```
bosboot -ad hdiskx
bootlist -m normal hdiskx
```

Figure 4-13 shows the output after you run the commands.



```
AIX02 /tmp/export/host_attach/bin
AIX02 # bosboot -ad hdisk5
bosboot: Boot image is 46903 512 byte blocks.
AIX02 # bootlist -m normal hdisk5
```

Figure 4-13 Relocating boot volume

- Select the rootvg volume group and the original hdisk that you want to remove, then click **smitty vg** → **Unmirror a Volume Group**.
- Select rootvg for the volume group name ROOTVG and the internal SCSI disk you want to remove.
- Click **smitty vg** → **Set Characteristics of a Volume Group** → **Remove a Physical Volume from a Volume Group**.
- Run the following commands again:

```
bosboot -ad hdiskx
bootlist -m normal hdiskx
```

At this stage, the creation of a bootable disk on the XIV is completed. Restarting the system makes it boot from the SAN (XIV) disk.

After the system reboots, use the `lspv -L` command to confirm that the server is booting from the XIV hdisk as shown in Figure 4-14.

```
Terminal Edit Font Encoding Options
Console login: root
root's Password:
*****
*
*
* Welcome to AIX Version 6.1!
*
*
* Please see the README file in /usr/lpp/bos for information pertinent to
* this release of the AIX Operating System.
*
*
*****
Last unsuccessful login: Wed Oct  5 13:59:32 2011 on /dev/vty0
Last login: Wed Oct  5 13:59:50 2011 on /dev/vty0

# bash
AIX02 # lspv -L
hdisk0          00cc6de1b1d84ec9          None
hdisk1          none                          None
hdisk2          none                          None
hdisk3          00cc6de1bb94fd29          None
hdisk4          none                          None
hdisk5          00cc6de1cfb8ea41          rootvg          active
AIX02 #
```

Figure 4-14 XIV SAN boot disk confirmation

4.2.2 Installation on external storage from bootable AIX CD-ROM

To install AIX on XIV System disks, complete the following preparations:

1. Update the Fibre Channel (FC) adapter (HBA) microcode to the latest supported level.
2. Make sure that you have an appropriate SAN configuration, and the host is properly connected to the SAN
3. Make sure that the zoning configuration is updated, and at least one LUN is mapped to the host.

Tip: If the system cannot see the SAN fabric at login, configure the HBAs at the server open firmware prompt.

Because a SAN allows access to many devices, identifying the hdisk to install to can be difficult. Use the following method to facilitate the discovery of the lun_id to hdisk correlation:

1. If possible, zone the switch or disk array such that the system being installed can discover only the disks to be installed to. After the installation completes, you can reopen the zoning so the system can discover all necessary devices.
2. If more than one disk is assigned to the host, make sure that you are using the correct one using one of the following methods:
 - Assign Physical Volume Identifiers (PVIDs) to all disks from an already installed AIX system that can access the disks. Assign PWVIDS by using the following command:
`chdev -a pv=yes -l hdiskX`

where *X* is the appropriate disk number. Create a table mapping PVIDs to physical disks. Make the PVIDs visible in the installation menus by selecting option **77 display more disk info**. You can also use the PVIDs to do an unprompted NIM installation.

- Another way to ensure that the selection of the correct disk is to use Object Data Manager (ODM) commands:
 - i. Boot from the AIX installation CD-ROM.
 - ii. From the main installation menu, click **Start Maintenance Mode for System Recovery** → **Access Advanced Maintenance Functions** → **Enter the Limited Function Maintenance Shell**.
 - iii. At the prompt, issue one of the following commands:

```
odmget -q "attribute=lun_id AND value=0xNN..N" CuAt
odmget -q "attribute=lun_id" CuAt (list every stanza with lun_id
attribute)
```

where *OxNN..N* is the *lun_id* that you are looking for. This command prints the ODM stanzas for the *hdisks* that have that *lun_id*.
 - iv. Enter **Exit** to return to the installation menus.

The Open Firmware implementation can only boot from *lun_ids* 0 through 7. The firmware on the Fibre Channel adapter (HBA) promotes this *lun_id* to an 8-byte FC *lun-id*. The firmware does this promotion by adding a byte of zeros to the front and 6 bytes of zeros to the end. For example, *lun_id* 2 becomes 0x0002000000000000. The *lun_id* is normally displayed without the leading zeros. Take care when you are installing because the procedure allows installation to *lun_ids* outside of this range.

To install on external storage, complete these steps:

1. Insert an AIX CD that has a bootable image into the CD-ROM drive.
2. Select **CD-ROM** as the installation device to make the system boot from the CD. The way to change the bootlist varies model by model. In most System p models, you use the system management services (SMS) menu. For more information, see the user's guide for your model.
3. Allows the system to boot from the AIX CD image after you leave the SMS menu.
4. After a few minutes, the console displays a window that directs you to press a key to use the device as the system console.
5. A window prompts you to select an installation language.
6. The Welcome to the Base Operating System Installation and Maintenance window is displayed. Change the installation and system settings for this system to select a Fibre Channel-attached disk as a target disk. Enter **2** to continue.
7. On the Installation and Settings window, enter **1** to change the system settings and select the **New and Complete Overwrite** option.
8. On the Change (the destination) Disk window, select the Fibre Channel disks that are mapped to your system. To see more information, enter **77** to display the detailed information window that includes the PVID. Enter **77** again to show WWPN and LUN_ID information. Type the number, but do not press **Enter**, for each disk that you choose. Typing the number of a selected disk clears the device. Be sure to include an XIV disk.
9. After you select the Fibre Channel-attached disks, the Installation and Settings window is displayed with the selected disks. Verify the installation settings, and then enter **0** to begin the installation process.

Important: Verify that you made the correct selection for root volume group. The existing data in the destination root volume group is deleted during Base Operating System (BOS) installation.

When the system reboots, a window displays the address of the device from which the system is reading the boot image.

4.2.3 AIX SAN installation with NIM

NIM is a client/server infrastructure and service that allows remote installation of the operating system. It manages software updates, and can be configured to install and update third-party applications. The NIM server and client file sets are part of the operating system. A separate NIM server must be configured to keep the configuration data and the installable product file sets.

Deploy the NIM environment, and ensure that the following configurations have been completed:

- ▶ The NIM server is properly configured as the NIM master and the basic NIM resources are defined.
- ▶ The Fibre Channel adapters are already installed on the system onto which AIX is to be installed.
- ▶ The Fibre Channel adapters are connected to a SAN, and on the XIV system have at least one logical volume (LUN) mapped to the host.
- ▶ The target system (NIM client) currently has no operating system installed, and is configured to boot from the NIM server.

For more information about how to configure a NIM server, see *AIX 5L Version 5.3: Installing AIX*, SC23-4887-02.

Before the installation, modify the `bosinst.data` file, where the installation control is stored. Insert your appropriate values at the following stanza:

```
SAN_DISKID
```

This stanza specifies the worldwide port name and a logical unit ID for Fibre Channel-attached disks. The worldwide port name and logical unit ID are in the format that is returned by the `lsattr` command (that is, 0x followed by 1–16 hexadecimal digits). The `ww_name` and `lun_id` are separated by two slashes (`//`).

```
SAN_DISKID = <worldwide_portname//lun_id>
```

For example:

```
SAN_DISKID = 0x0123456789FEDCBA//0x2000000000000
```

Or you can specify PVID (example with internal disk):

```
target_disk_data:  
PVID = 000c224a004a07fa  
SAN_DISKID =  
CONNECTION = scsi0//10,0  
LOCATION = 10-60-00-10,0  
SIZE_MB = 34715  
HDISKNAME = hdisk0
```

To install AIX SAN with NIM, complete the following steps:

1. Enter the # `smit nim_bosinst` command.
2. Select the **lpp_source** resource for the BOS installation.
3. Select the **SPOT** resource for the BOS installation.
4. Select the **BOSINST_DATA to use during installation** option, and select a `bosinst_data` resource that can run a non-prompted BOS installation.
5. Select the **RESOLV_CONF to use for network configuration** option, and select a `resolv_conf` resource.
6. Click the **Accept New License Agreements** option, and select **Yes**. Accept the default values for the remaining menu options.
7. Press Enter to confirm and begin the NIM client installation.
8. To check the status of the NIM client installation, enter the following command:

```
# lsnim -l va09
```




XIV and HP-UX host connectivity

This chapter addresses specific considerations for attaching the XIV system to an HP-UX host.

For the latest information, see the Host Attachment Kit publications from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

HP-UX manuals are available at the HP Business Support Center at:

<http://www.hp.com/go/hpux-core-docs>

Important: The procedures and instructions are based on code that was available at the time of writing. For the latest support information and instructions, see the System Storage Interoperation Center (SSIC) at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

This chapter contains the following sections:

- ▶ Attaching XIV to an HP-UX host
- ▶ HP-UX multi-pathing solutions
- ▶ Veritas Volume Manager on HP-UX
- ▶ HP-UX SAN boot

5.1 Attaching XIV to an HP-UX host

At the time of writing, XIV Storage System Software release 11.0.0 supports Fibre Channel attachment to HP servers that run HP-UX 11iv2 (11.23), and HP-UX 11iv3 (11.31). For more information, see the SSIC at:

<http://www.ibm.com/systems/support/storage/config/ssic>

The HP-UX host attachment process with XIV is described in the *Host Attachment Guide for HP-UX*, which is available from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

The attachment process includes completing the following steps:

1. Getting the worldwide names (WWNs) of the host Fibre Channel adapters
2. Completing the SAN zoning
3. Defining volumes and host objects on the XIV Storage System
4. Mapping the volumes to the host
5. Installing the XIV Host Attachment Kit, which can be downloaded from Fix Central

This section focuses on the HP-UX specific steps. The steps that are not specific to HP-UX are described in Chapter 1, “Host connectivity” on page 1.

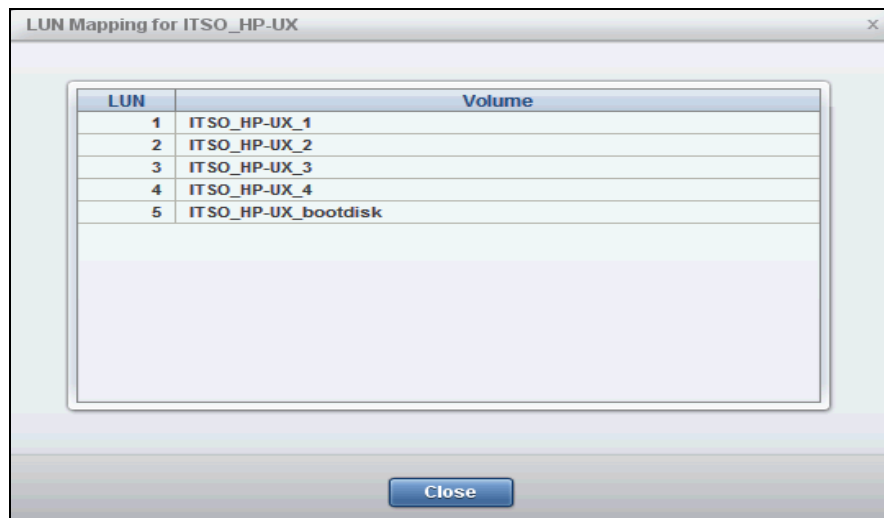
Figure 5-1 shows the host object that was defined for the HP-UX server that is used for the examples.



Name ▲	
ITSO_HP-UX	default
500110A000101960	FC
50060B000039DDE0	FC

Figure 5-1 XIV host object for the HP-UX server

Figure 5-2 shows the volumes that were defined for the HP-UX server.



LUN	Volume
1	ITSO_HP-UX_1
2	ITSO_HP-UX_2
3	ITSO_HP-UX_3
4	ITSO_HP-UX_4
5	ITSO_HP-UX_bootdisk

Figure 5-2 XIV volumes that are mapped to the HP-UX server

The HP-UX utility `ioscan` displays the Fibre Channel adapters of the host. The `fcmsutil` utility displays details of these adapters, including the worldwide name (WWN), as shown in Example 5-1.

Example 5-1 HP Fibre Channel adapter properties

```
# ioscan -fnk|grep fcd
fc          0 0/2/1/0      fcd          CLAIMED    INTERFACE  HP A6826-60001 2Gb Dual
Port PCI/PCI-X Fibre Channel Adapter (FC Port 1)
/dev/fcd0
fc          1 0/2/1/1      fcd          CLAIMED    INTERFACE  HP A6826-60001 2Gb Dual
Port PCI/PCI-X Fibre Channel Adapter (FC Port 2)
/dev/fcd1
fc          2 0/5/1/0      fcd          CLAIMED    INTERFACE  HP A6826-60001 2Gb Dual
Port PCI/PCI-X Fibre Channel Adapter (FC Port 1)
/dev/fcd2
fc          3 0/5/1/1      fcd          CLAIMED    INTERFACE  HP A6826-60001 2Gb Dual
Port PCI/PCI-X Fibre Channel Adapter (FC Port 2)
/dev/fcd3

# fcmsutil /dev/fcd0

        Vendor ID is = 0x1077
        Device ID is = 0x2312
    PCI Sub-system Vendor ID is = 0x103C
        PCI Sub-system ID is = 0x12BA
            PCI Mode = PCI-X 133 MHz
        ISP Code version = 3.3.166
        ISP Chip version = 3
            Topology = PTTOPT_FABRIC
        Link Speed = 2Gb
    Local N_Port_id is = 0x0b3400
    Previous N_Port_id is = None
    N_Port Node World Wide Name = 0x50060b000039dde1
    N_Port Port World Wide Name = 0x50060b000039dde0
    Switch Port World Wide Name = 0x203200053353e557
    Switch Node World Wide Name = 0x100000053353e557
        Driver state = ONLINE
        Hardware Path is = 0/2/1/0
        Maximum Frame Size = 2048
    Driver-Firmware Dump Available = NO
    Driver-Firmware Dump Timestamp = N/A
        Driver Version = @(#) fcd B.11.31.01 Jan  7 2007
```

The XIV Host Attachment Kit version 1.10.0 supports HP-UX 11iv3 on HP Integrity servers. The Host Attachment Kit includes scripts to facilitate HP-UX attachment to XIV. For example, the `xiv_attach` script runs the following tasks (Example 5-2):

- ▶ Identifies the Fibre Channel adapters of the hosts that are connected to XIV storage systems.
- ▶ Identifies the name of the host object that is defined on the XIV Storage System for this host (if already created).
- ▶ Supports rescanning for new storage devices.

Example 5-2 xiv_attach script output

```
# xiv_attach
-----
Welcome to the XIV Host Attachment wizard, version 1.10.0.
```

This wizard will assist you to attach this host to the XIV system.

The wizard will now validate host configuration for the XIV system.
Press [ENTER] to proceed.

Only fibre-channel is supported on this host.
Would you like to set up an FC attachment? [default: yes]:

Please wait while the wizard validates your existing configuration...
This host is already configured for the XIV system

Please zone this host and add its WWPNs with the XIV storage system:
50060b000039dde0: /dev/fcd0: []:
50060b000039dde2: /dev/fcd1: []:
500110a000101960: /dev/fcd2: []:
500110a000101962: /dev/fcd3: []:
Press [ENTER] to proceed.

Would you like to rescan for new storage devices now? [default: yes]:
Please wait while rescanning for storage devices...

The host is connected to the following XIV storage arrays:
Serial Version Host Defined Ports Defined Protocol Host Name(s)
1310114 11.0.0.0 Yes All FC ITS0_HP-UX
This host is defined on all FC-attached XIV storage arrays

Press [ENTER] to proceed.

The XIV host attachment wizard successfully configured this host

Press [ENTER] to exit.

5.2 HP-UX multi-pathing solutions

HP introduced HP Native Multi-Pathing with HP-UX 11iv3. The earlier *pvl* multi-pathing is still available, but use Native Multi-Pathing. HP Native Multi-Pathing provides I/O load balancing across the available I/O paths, whereas *pvl* provides path failover and failback, but no load balancing. Both multi-pathing methods can be used for HP-UX attachment to XIV.

HP Native Multi-Pathing uses the so-called Agile View Device Addressing which addresses a device by its worldwide ID (WWID) as an object. The device can be discovered by its WWID regardless of the hardware controllers, adapters, or paths between the HP-UX server and the device itself. Therefore, this addressing method creates only one device file for each device.

Example 5-3 shows the HP-UX view of five XIV volumes using agile addressing and the conversion from agile to legacy view.

Example 5-3 HP-UX agile and legacy views

```
# ioscan -fnNkC disk
Class    I  H/W Path  Driver S/W State  H/W Type  Description
=====
disk     3  64000/0xfa00/0x0  esdisk CLAIMED  DEVICE  HP 146 GMAT3147NC
                /dev/disk/disk3  /dev/disk/disk3_p2  /dev/rdisk/disk3
/dev/rdisk/disk3_p2
```



```

/dev/disk/disk3_p1 /dev/disk/disk3_p3 /dev/rdisk/disk3_p1
/dev/rdisk/disk3_p3
disk 4 64000/0xfa00/0x1 esdisk CLAIMED DEVICE HP 146 GMAT3147NC
/dev/disk/disk4 /dev/disk/disk4_p2 /dev/rdisk/disk4
/dev/rdisk/disk4_p2
/dev/disk/disk4_p1 /dev/disk/disk4_p3 /dev/rdisk/disk4_p1
/dev/rdisk/disk4_p3
disk 5 64000/0xfa00/0x2 esdisk CLAIMED DEVICE TEAC DV-28E-C
/dev/disk/disk5 /dev/rdisk/disk5
disk 16 64000/0xfa00/0xa2 esdisk CLAIMED DEVICE IBM 2810XIV
/dev/disk/disk16 /dev/rdisk/disk16
disk 17 64000/0xfa00/0xa3 esdisk CLAIMED DEVICE IBM 2810XIV
/dev/disk/disk17 /dev/rdisk/disk17
disk 18 64000/0xfa00/0xa4 esdisk CLAIMED DEVICE IBM 2810XIV
/dev/disk/disk18 /dev/rdisk/disk18
disk 19 64000/0xfa00/0xa5 esdisk CLAIMED DEVICE IBM 2810XIV
/dev/disk/disk19 /dev/rdisk/disk19
disk 20 64000/0xfa00/0xa6 esdisk CLAIMED DEVICE IBM 2810XIV
/dev/disk/disk20 /dev/rdisk/disk20
# ioscan -m dsf /dev/disk/disk16
Persistent DSF Legacy DSF(s)
=====
/dev/disk/disk16 /dev/dsk/c5t0d1
/dev/dsk/c7t0d1

```

If device special files are missing on the HP-UX server, you can create them in two ways. The first option is rebooting the host, which is disruptive. The other option is to run the command **insf -eC disk**, which reinstalls the special device files for all devices of the class disk.

Volume groups, logical volumes, and file systems can be created on the HP-UX host. Example 5-4 shows the HP-UX commands to initialize the physical volumes and create a volume group in an LVM environment. The rest is standard HP-UX system administration that is not specific to XIV, and therefore is not addressed.

HP Native Multi-Pathing is used to automatically specify the Agile View device files, for example `/dev/(r)disk/disk1299`. To use *pvl*links, specify the Legacy View device files of all available hardware paths to a disk device, for example `/dev/(r)dsk/c153t0d1` and `c155t0d1`.

Example 5-4 Volume group creation

```

# pvcreate /dev/rdisk/disk16
Physical volume "/dev/rdisk/disk16" has been successfully created.
# pvcreate /dev/rdisk/disk17
Physical volume "/dev/rdisk/disk17" has been successfully created.
# mkdir /dev/vg02
# mknod /dev/vg02/group c 64 0x020000
# vgcreate vg02 /dev/disk/disk16 /dev/disk/disk17
Increased the number of physical extents per physical volume to 8205.
Volume group "/dev/vg02" has been successfully created.
Volume Group configuration for /dev/vg02 has been saved in /etc/lvmconf/vg02.conf

```

5.3 Veritas Volume Manager on HP-UX

With HP-UX 11i 3, you can use one of two volume managers:

- ▶ The HP Logical Volume Manager (LVM).
- ▶ The Veritas Volume Manager (VxVM). With this manager, any I/O is handled in pass-through mode and therefore run by Native Multipathing, not by Dynamic Multipathing (DMP).

According to the *HP-UX System Administrator's Guide*, both volume managers can coexist on an HP-UX server. For more information, see HP-UX System Administration at:

<http://www.hp.com/go/hpux-core-docs>

You can use both simultaneously (on separate physical disks), but usually you choose to use one or the other exclusively.

The configuration of XIV volumes on HP-UX with LVM is described in 5.2, “HP-UX multi-pathing solutions” on page 174. Example 5-5 shows the initialization of disks for VxVM use and the creation of a disk group with the `vxdiskadm` utility.

Example 5-5 Disk initialization and disk group creation with vxdiskadm

```
# vxctl enable
# vxdisk list
DEVICE      TYPE          DISK          GROUP          STATUS
Disk_0s2    auto:LVM      -             -             LVM
Disk_1s2    auto:LVM      -             -             LVM
XIV2_0      auto:none     -             -             online invalid
XIV2_1      auto:none     -             -             online invalid
XIV2_2      auto:none     -             -             online invalid
XIV2_3      auto:LVM      -             -             LVM
XIV2_4      auto:LVM      -             -             LVM
```

```
# vxdiskadm
```

Volume Manager Support Operations

Menu: VolumeManager/Disk

- 1 Add or initialize one or more disks
- 2 Remove a disk
- 3 Remove a disk for replacement
- 4 Replace a failed or removed disk
- 5 Mirror volumes on a disk
- 6 Move volumes from a disk
- 7 Enable access to (import) a disk group
- 8 Remove access to (deport) a disk group
- 9 Enable (online) a disk device
- 10 Disable (offline) a disk device
- 11 Mark a disk as a spare for a disk group
- 12 Turn off the spare flag on a disk
- 13 Remove (deport) and destroy a disk group
- 14 Unrelocate subdisks back to a disk
- 15 Exclude a disk from hot-relocation use
- 16 Make a disk available for hot-relocation use
- 17 Prevent multipathing/Suppress devices from VxVM's view

18 Allow multipathing/Unsuppress devices from VxVM's view
19 List currently suppressed/non-multipathed devices
20 Change the disk naming scheme
21 Change/Display the default disk layouts
22 Mark a disk as allocator-reserved for a disk group
23 Turn off the allocator-reserved flag on a disk
list List disk information

? Display help about menu
?? Display help about the menuing system
q Exit from menus

Select an operation to perform: 1

Add or initialize disks
Menu: VolumeManager/Disk/AddDisks

Use this operation to add one or more disks to a disk group. You can add the selected disks to an existing disk group or to a new disk group that will be created as a part of the operation. The selected disks may also be added to a disk group as spares. Or they may be added as nohotuses to be excluded from hot-relocation use. The selected disks may also be initialized without adding them to a disk group leaving the disks available for use as replacement disks.

More than one disk or pattern may be entered at the prompt. Here are some disk selection examples:

all: all disks
c3 c4t2: all disks on both controller 3 and controller 4, target 2
c3t4d2: a single disk (in the c#t#d# naming scheme)
xyz_0: a single disk (in the enclosure based naming scheme)
xyz_: all disks on the enclosure whose name is xyz

disk#: a single disk (in the new naming scheme)

Select disk devices to add: [<pattern-list>,all,list,q,?] XIV2_1 XIV2_2

Here are the disks selected. Output format: [Device_Name]

XIV2_1 XIV2_2

Continue operation? [y,n,q,?] (default: y) y

You can choose to add these disks to an existing disk group, a new disk group, or you can leave these disks available for use by future add or replacement operations. To create a new disk group, select a disk group name that does not yet exist. To leave the disks available for future use, specify a disk group name of "none".

Which disk group [<group>,none,list,q,?] (default: none) dg01

Create a new group named dg01? [y,n,q,?] (default: y)

Create the disk group as a CDS disk group? [y,n,q,?] (default: y) n

Use default disk names for these disks? [y,n,q,?] (default: y)

Add disks as spare disks for dg01? [y,n,q,?] (default: n) n

Exclude disks from hot-relocation use? [y,n,q,?] (default: n)

Add site tag to disks? [y,n,q,?] (default: n)

A new disk group will be created named dg01 and the selected disks will be added to the disk group with default disk names.

XIV2_1 XIV2_2

Continue with operation? [y,n,q,?] (default: y)

Do you want to use the default layout for all disks being initialized? [y,n,q,?] (default: y) n

Do you want to use the same layout for all disks being initialized? [y,n,q,?] (default: y)

Enter the desired format [cdsdisk,hpdisk,q,?] (default: cdsdisk) hpdisk

Enter desired private region length [<privlen>,q,?] (default: 32768)

Initializing device XIV2_1.

Initializing device XIV2_2.

VxVM NOTICE V-5-2-120

Creating a new disk group named dg01 containing the disk device XIV2_1 with the name dg0101.

VxVM NOTICE V-5-2-88

Adding disk device XIV2_2 to disk group dg01 with disk name dg0102.

Add or initialize other disks? [y,n,q,?] (default: n) n

vxdisk list

DEVICE	TYPE	DISK	GROUP	STATUS
Disk_0s2	auto:LVM	-	-	LVM
Disk_1s2	auto:LVM	-	-	LVM
XIV2_0	auto:none	-	-	online invalid
XIV2_1	auto:hpdisk	dg0101	dg01	online
XIV2_2	auto:hpdisk	dg0102	dg01	online
XIV2_3	auto:LVM	-	-	LVM
XIV2_4	auto:LVM	-	-	LVM

The graphical equivalent for the `vxdiskadm` utility is the Veritas Enterprise Administrator (VEA). Figure 5-3 shows disks as they are displayed in this graphical user interface.

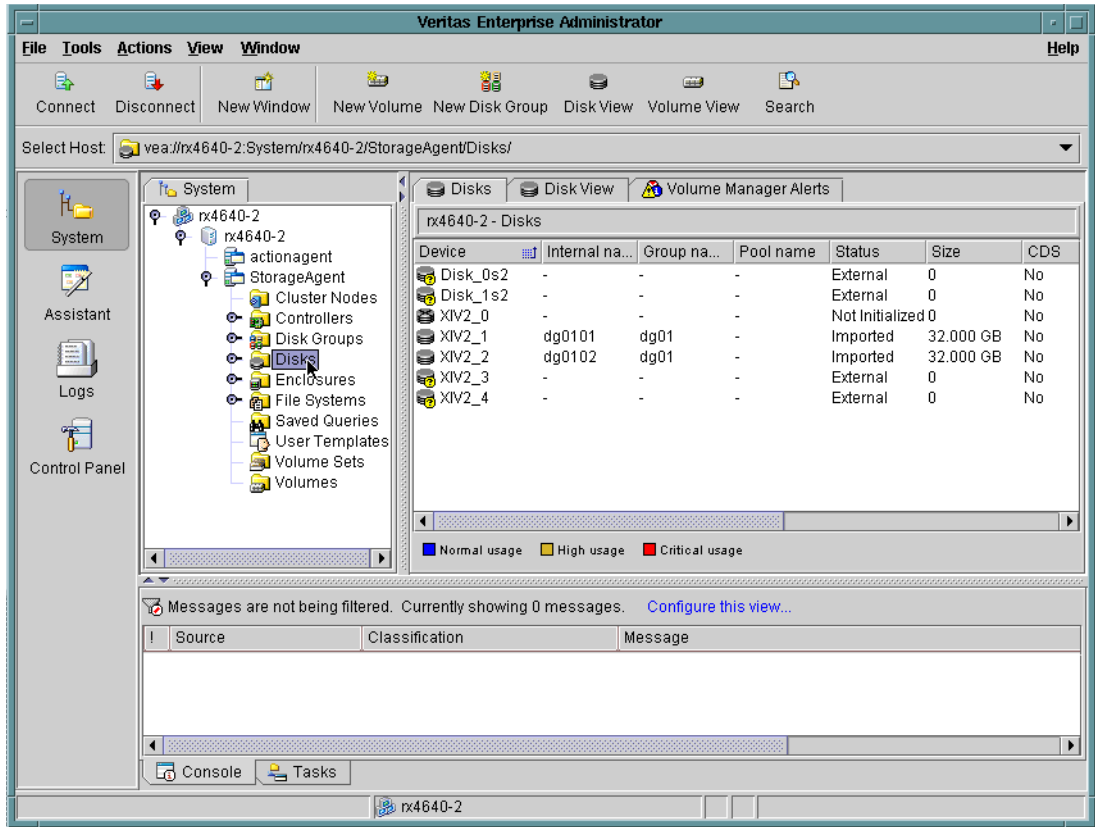


Figure 5-3 Disk presentation by Veritas Enterprise Administrator

In this example, after you create the disk groups and the VxVM disks, you must create file systems and mount them.

5.3.1 Array Support Library for an IBM XIV Storage System

Veritas Volume Manager (VxVM) offers a device discovery service that is implemented in the so-called Device Discovery Layer (DDL). For a specific storage system, this service is provided by an Array Support Library (ASL). The ASL can be downloaded from the Symantec website. An ASL can be dynamically added to or removed from VxVM.

On a host system, the VxVM command `vxddladm listsupport` displays a list of storage systems that are supported by the VxVM version installed on the operating system (Example 5-6).

Example 5-6 VxVM command to list Array Support Libraries

```
# vxddladm listsupport
LIBNAME          VID
=====
...
libvxxiv.s1     XIV, IBM

# vxddladm listsupport libname=libvxxiv.s1
ATTR_NAME       ATTR_VALUE
```

```

=====
LIBNAME          libvxxiv.sl
VID              XIV, IBM
PID              NEXTRA, 2810XIV
ARRAY_TYPE       A/A
ARRAY_NAME       Nextra, XIV
=====

```

On a host system, ASLs allow easier identification of the attached disk storage devices. The ASL serially numbers the attached storage systems of the same type and the volumes of a single storage system that are assigned to this host.

Example 5-7 shows that five volumes of one XIV system are assigned to that HP-UX host. VxVM controls the devices XIV2_1 and XIV2_2, and the disk group name is dg01. The HP LVM controls the remaining XIV devices, except for XIV2_0.

Example 5-7 VxVM disk list

```

# vxdisk list
DEVICE      TYPE          DISK          GROUP         STATUS
Disk_0s2    auto:LVM      -             -             LVM
Disk_1s2    auto:LVM      -             -             LVM
XIV2_0      auto:none     -             -             online invalid
XIV2_1      auto:hpdisk   dg0101       dg01          online
XIV2_2      auto:hpdisk   dg0102       dg01          online
XIV2_3      auto:LVM      -             -             LVM
XIV2_4      auto:LVM      -             -             LVM

```

More information about ASL is available at:

<http://www.symantec.com/business/support/index?page=content&id=TECH21351>

ASL packages for XIV and HP-UX 11iv3 are available for download at:

<http://www.symantec.com/business/support/index?page=content&id=TECH63130>

5.4 HP-UX SAN boot

The IBM XIV Storage System provides Fibre Channel boot from SAN capabilities for HP-UX. This section describes the SAN boot implementation for HP Integrity server that is running HP-UX 11iv3 (11.31). Boot management is provided by the Extensible Firmware Interface (EFI). Earlier systems ran another boot manager, so the SAN boot process might differ.

There are various possible implementations of SAN boot with HP-UX:

- ▶ To implement SAN boot for a new system, start the HP-UX installation from a bootable HP-UX CD or DVD installation package. You can also use a network-based installation such as Ignite-UX.
- ▶ To implement SAN boot on a system with an already installed HP-UX operating system, mirror the system disk volume to the SAN disk.

5.4.1 Installing HP-UX on external storage

To install HP-UX on XIV system volumes, make sure that you have an appropriate SAN configuration. The host must be properly connected to the SAN, the zoning configuration must be updated, and at least one LUN must be mapped to the host.

Discovering the LUN ID

Because a SAN allows access to many devices, identifying the volume to install can be difficult. To discover the *lun_id* to HP-UX device file correlation, complete these steps:

1. If possible, zone the switch and change the LUN mapping on the XIV Storage System so that the system being installed can discover only the disks to be installed to. After the installation completes, reopen the zoning so the system can discover all necessary devices.
2. If possible, temporarily attach the volumes to an already installed HP-UX system. Write down the hardware paths of the volumes so you can later compare them to the other system's hardware paths. Example 5-8 shows the output of the **ioscan** command that creates a hardware path list.
3. Write down the LUN identifiers on the XIV system to identify the volumes to install to during HP-UX installation. For example, LUN Id 5 matches to the disk named 64000/0xfa00/0x68 in the **ioscan** list that is shown in Example 5-8. This disk's hardware path name includes the string 0x5.

Example 5-8 HP-UX disk view (ioscan)

```
# ioscan -m hwdisk
Lun H/W Path      Lunpath H/W Path      Legacy H/W Path
=====
64000/0xfa00/0x0
           0/4/1/0.0x5000c500062ac7c9.0x0  0/4/1/0.0.0.0.0
64000/0xfa00/0x1
           0/4/1/0.0x5000c500062ad205.0x0  0/4/1/0.0.0.1.0
64000/0xfa00/0x5
           0/3/1/0.0x5001738000cb0140.0x0  0/3/1/0.19.6.0.0.0
           0/3/1/0.19.6.255.0.0.0
           0/3/1/0.0x5001738000cb0170.0x0  0/3/1/0.19.1.0.0.0
           0/3/1/0.19.1.255.0.0.0
           0/7/1/0.0x5001738000cb0182.0x0  0/7/1/0.19.54.0.0.0
           0/7/1/0.19.54.255.0.0.0
           0/7/1/0.0x5001738000cb0192.0x0  0/7/1/0.19.14.0.0.0
           0/7/1/0.19.14.255.0.0.0
64000/0xfa00/0x63
           0/3/1/0.0x5001738000690160.0x0  0/3/1/0.19.62.0.0.0
           0/3/1/0.19.62.255.0.0.0
           0/7/1/0.0x5001738000690190.0x0  0/7/1/0.19.55.0.0.0
           0/7/1/0.19.55.255.0.0.0
64000/0xfa00/0x64
           0/3/1/0.0x5001738000690160.0x1000000000000
0/3/1/0.19.62.0.0.0.1
           0/7/1/0.0x5001738000690190.0x1000000000000
0/7/1/0.19.55.0.0.0.1
64000/0xfa00/0x65
           0/3/1/0.0x5001738000690160.0x20000000000000
0/3/1/0.19.62.0.0.0.2
```

```
0/7/1/0.0x5001738000690190.0x2000000000000
0/7/1/0.19.55.0.0.0.2
64000/0xfa00/0x66
0/3/1/0.0x5001738000690160.0x3000000000000
0/3/1/0.19.62.0.0.0.3
0/7/1/0.0x5001738000690190.0x3000000000000
0/7/1/0.19.55.0.0.0.3
64000/0xfa00/0x67
0/3/1/0.0x5001738000690160.0x4000000000000
0/3/1/0.19.62.0.0.0.4
0/7/1/0.0x5001738000690190.0x4000000000000
0/7/1/0.19.55.0.0.0.4
64000/0xfa00/0x68
0/3/1/0.0x5001738000690160.0x5000000000000
0/3/1/0.19.62.0.0.0.5
0/7/1/0.0x5001738000690190.0x5000000000000
0/7/1/0.19.55.0.0.0.5
```

Installing HP-UX

The examples in this chapter involve an HP-UX installation on HP Itanium-based Integrity systems. On older HP PA-RISC systems, the processes to boot the server and select disks to install HP-UX to are different. A complete description of the HP-UX installation processes on HP Integrity and PA-RISC systems is provided in the HP manual *HP-UX 11iv3 Installation and Update Guide*. Click **HP-UX 11iv3** at:

<http://www.hp.com/go/hpux-core-docs>

To install HP-UX 11iv3 on an XIV volume from DVD on an HP Integrity system, complete these steps:

1. Insert the first HP-UX Operating Environment DVD into the DVD drive.
2. Reboot or power on the system and wait for the EFI panel.

3. Select **Boot from DVD** and continue as shown in Figure 5-4.

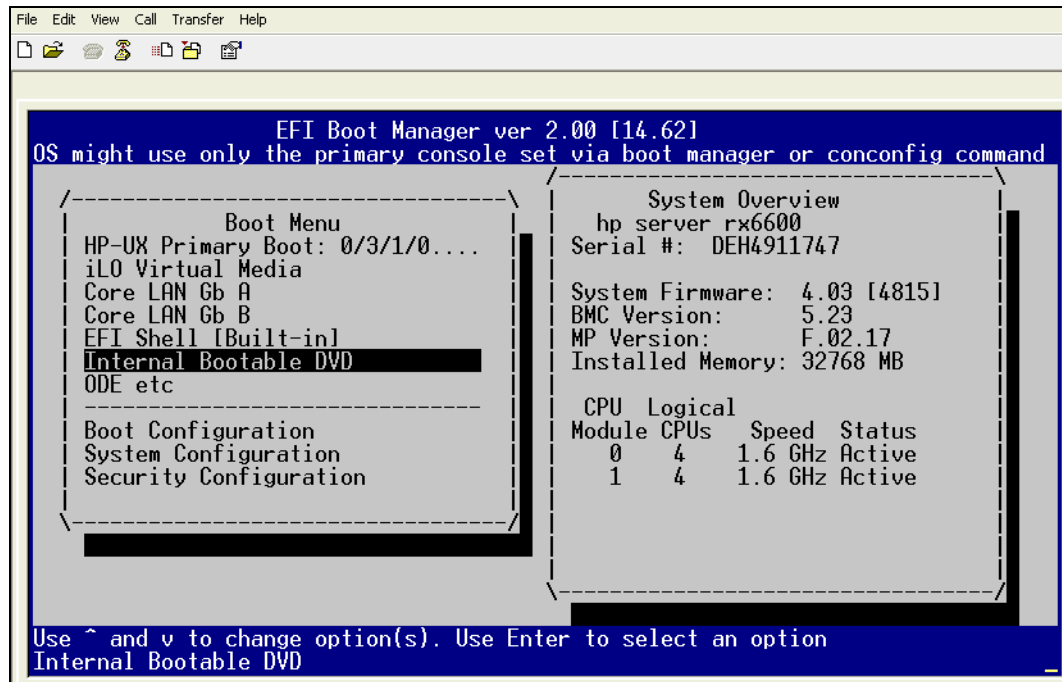


Figure 5-4 Boot device selection with EFI Boot Manager

4. The server boots from the installation media. On the HP-UX installation and recovery process window, select **Install HP-UX** (Figure 5-5).

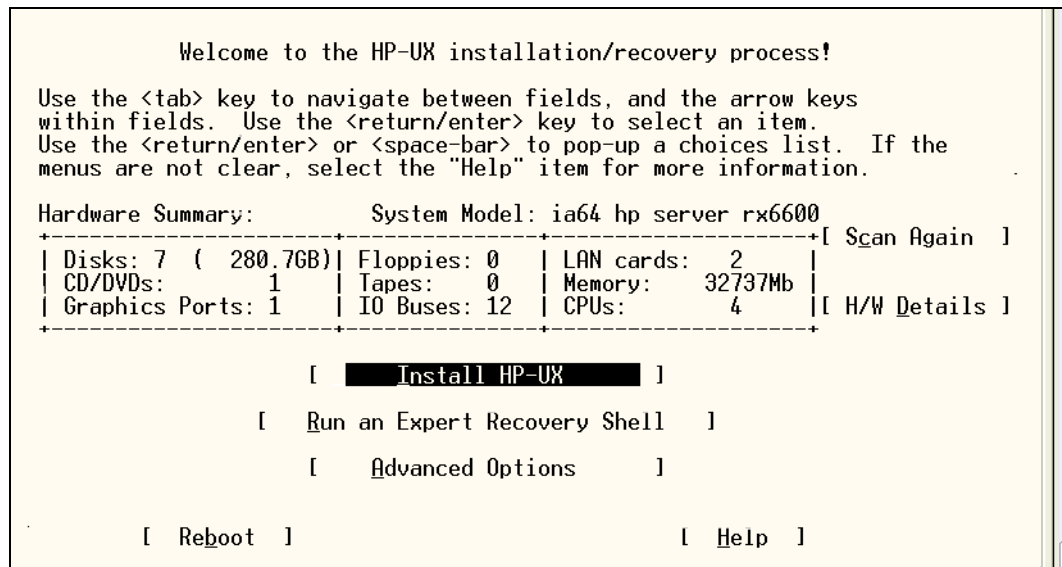


Figure 5-5 HP-UX installation window: Starting OS installation

- The HP-UX installation procedure displays the disks that are suitable for operating system installation. Identify and select the XIV volume to install HP-UX to as shown in Figure 5-6.

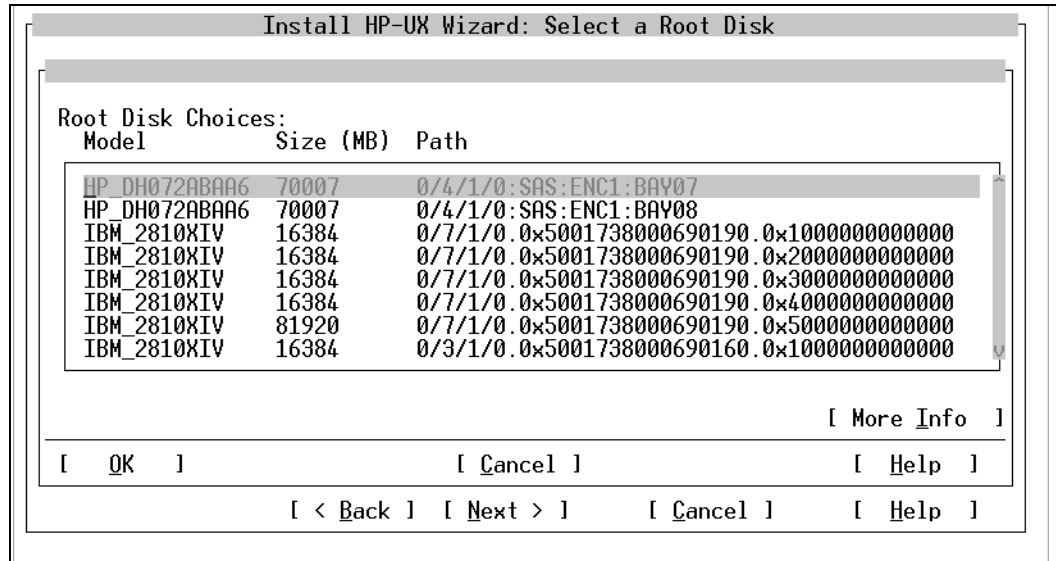


Figure 5-6 HP-UX installation panel: Selecting a root disk

- The remaining steps of an HP-UX installation on a SAN disk do not differ from installation on an internal disk.

5.4.2 Creating a SAN boot disk by mirroring

The “Mirroring the Boot Disk” section of the *HP-UX System Administrator's Guide: Logical Volume Management HP-UX 11i V3* includes a detailed description of the boot disk mirroring process. Click **HP-UX 11i Volume Management (LVM/VxVM) Software** at:

<http://www.hp.com/go/hpux-core-docs>

The storage-specific part is the identification of the XIV volume to install to on HP-UX. For more information, see 5.4.1, “Installing HP-UX on external storage” on page 181.



XIV and Solaris host connectivity

This chapter explains specific considerations for attaching the XIV system to a Solaris host.

Important: The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, ALWAYS see the System Storage Interoperation Center (SSIC) at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

You can find the latest Host Attachment Kit software and User Guide at:

http://www.ibm.com/support/fixcentral/swg/selectFixes?parent=Enterprise+Storage+Servers&product=ibm/Storage_Disk/XIV+Storage+System+%282810,+2812%29&release=A11&platform=All&function=all#IBM%20XIV%20Host%20Attachment%20Kit

This chapter includes the following sections:

- ▶ Attaching a Solaris host to XIV
- ▶ Solaris host configuration for Fibre Channel
- ▶ Solaris host configuration for iSCSI
- ▶ Solaris Host Attachment Kit utilities
- ▶ Creating partitions and file systems with UFS

6.1 Attaching a Solaris host to XIV

Before you start the configuration, set the network and establish the connection in the SAN for the FC connectivity. For the iSCSI connection, the iSCSI ports must be configured first. For more information, see 1.3, “iSCSI connectivity” on page 28.

Tip: You can use both Fibre Channel and iSCSI connections to attach hosts. However, do not use both connections for the same LUN on the same host.

6.2 Solaris host configuration for Fibre Channel

This section describes attaching a Solaris host to XIV over Fibre Channel. It provides detailed descriptions and installation instructions for the various software components required. To make sure that your HBA and the firmware are supported, check the IBM SSIC at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

The environment in the examples in this chapter consists of a SUN Sparc T5220 running with Solaris 10 U10.

6.2.1 Obtaining WWPN for XIV volume mapping

To map the volumes to the Solaris host, you need the worldwide port names (WWPNs) of the HBAs. WWPNs can be found by using the `fcinfo` command as shown in Example 6-1.

Example 6-1 WWPNs of the HBAs

```
# fcinfo hba-port | grep HBA
HBA Port WWN: 2100001b32919ab1
HBA Port WWN: 2101001b32b19ab1
HBA Port WWN: 2100001b3291d4b1
HBA Port WWN: 2101001b32b1d4b1
```

6.2.2 Installing the Host Attachment Kit

To install the Host Attachment Kit, complete the following steps:

1. Open a terminal session and go to the directory where the package is.
2. Run the command that is shown in Example 6-2 to extract the archive.

Example 6-2 Extracting the Host Attachment Kit

```
# gunzip -c IBM_XIV_Host_Attachment_Kit-<version>-<os>-<arch>.tar.gz | tar xvf -
```

3. Change to the newly created directory and start the Host Attachment Kit installer as shown in Example 6-3.

Example 6-3 Starting the installation

```
# cd HAK_<version>
# /bin/sh ./install.sh
Welcome to the XIV Host Attachment Kit installer.
Would you like to proceed and install the Host Attachment Kit? [Y/n]:
```

```

y
Please wait while the installer validates your existing configuration...
-----
Please wait, the Host Attachment Kit is being installed...
-----
Installation successful.
Please refer to the Host Attachment Guide for information on how to configure
this host.

-----
The IBM XIV Host Attachment Kit includes the following utilities:
xiv_attach: Interactive wizard that configures the host and verifies its
configuration for connectivity with the IBM XIV Storage System.
xiv_devlist: Lists of all XIV volumes that are mapped to the host, with general
info about non-XIV volumes.
xiv_syslist: Lists all XIV storage systems that are detected by the host.
xiv_diag: Performs complete diagnostics of the host and its connectivity with
the IBM XIV Storage System, and saves the information to a file.
xiv_fc_admin: Allows you to perform different administrative operations for
FC-connected hosts and XIV storage systems.
xiv_iscsi_admin: Allows you to perform different administrative operations for
iSCSI-connected hosts and XIV storage systems.
-----

```

4. Follow the prompts to install the Host Attachment Kit.
5. After you run the installation script, review the installation log file `install.log` in the same directory.

6.2.3 Configuring the host

Use the utilities that are provided in the Host Attachment Kit to configure the Solaris host. Host Attachment Kit packages are installed in the `/opt/xiv/host_attach` directory.

Restriction: You must be logged in as root or have root privileges to use the Host Attachment Kit.

The main executable files are installed in the folder that is shown in Example 6-4.

Example 6-4 Location of main executable files

```

/opt/xiv/host_attach/bin/xiv_attach

```

They can also be used from every working directory.

To configure your system in and for the XIV, set up your SAN zoning first so that the XIV is visible for the host. To start the configuration, complete the following steps:

1. Run the `xiv_attach` command, which is mandatory for support. Example 6-5 shows an example of host configuration with this command.

Example 6-5 Sample results of the xiv_attach command

```

# xiv_attach

```

```

-----
Welcome to the IBM XIV host attachment wizard, version 1.10.0.

```

This wizard will help you attach this host to the XIV storage system.

The wizard will now validate the host configuration for the XIV storage system.
Press [ENTER] to proceed.

Please specify the connectivity type: [f]c / [i]scsi : f

Please wait while the wizard validates your existing configuration...
Verifying for mpxio DMP solution
Verifying Boot From SAN (BFS) device configuration NOT OK
Verifying SCSI driver NOT OK
Verifying scsi_vhci.conf settings NOT OK
Verifying fp.conf settings NOT OK

The wizard needs to configure this host for the XIV storage system.
Do you want to proceed? [default: yes]:
Please wait while the host is being configured...

Configuring for mpxio DMP solution
Configuring Boot From SAN (BFS) device configuration REBOOT
Configuring SCSI driver OK
Configuring scsi_vhci.conf settings REBOOT
Configuring fp.conf settings REBOOT

A reboot is required.
Please reboot this host and then restart the wizard.

Press [ENTER] to exit.

Remember: After you run the **xiv_attach** command for the first time, you must reboot the server.

2. After the system reboot, start **xiv_attach** again to finish the system to XIV configuration for the Solaris host, and to define the host object on XIVs, as shown in Example 6-6.

Example 6-6 Fibre Channel host attachment configuration after reboot

xiv_attach

Welcome to the IBM XIV host attachment wizard, version 1.10.0.
This wizard will help you attach this host to the XIV storage system.

The wizard will now validate the host configuration for the XIV storage system.
Press [ENTER] to proceed.

Please specify the connectivity type: [f]c / [i]scsi : f

Please wait while the wizard validates your existing configuration...
Verifying for mpxio DMP solution
Verifying Boot From SAN (BFS) device configuration OK
Verifying SCSI driver OK
Verifying scsi_vhci.conf settings OK
Verifying fp.conf settings OK

This host is already configured for the XIV storage system.

Please define zoning for this host and add its World Wide Port Names (WWPNs) to the XIV storage system:

2101001b32b1d4b1: /dev/cfg/c5: [QLogic Corp.]: 371-4325-01

2101001b32b19ab1: /dev/cfg/c3: [QLogic Corp.]: 371-4325-01

2100001b3291d4b1: /dev/cfg/c4: [QLogic Corp.]: 371-4325-01

2100001b32919ab1: /dev/cfg/c2: [QLogic Corp.]: 371-4325-01

Press [ENTER] to proceed.

Would you like to rescan for new storage devices? [default: yes]:

Please wait while rescanning for XIV storage devices...

This host is connected to the following XIV storage arrays:

Serial	Version	Host Defined	Ports Defined	Protocol	Host Name(s)
--------	---------	--------------	---------------	----------	--------------

1310114	11.1.1.0	No	None	FC	--
---------	----------	----	------	----	----

1310133	11.1.1.0	No	None	FC	--
---------	----------	----	------	----	----

6000105	10.2.4.5	No	None	FC	--
---------	----------	----	------	----	----

This host is not defined on some of the FC-attached XIV storage systems.

Do you want to define this host on these XIV systems now? [default: yes]:

Please enter a name for this host [default: sun-t5220-01-1]:

Please enter a username for system 1310114 [default: admin]: itso

Please enter the password of user itso for system 1310114:

Please enter a username for system 1310133 [default: admin]: itso

Please enter the password of user itso for system 1310133:

Please enter a username for system 6000105 [default: admin]: itso

Please enter the password of user itso for system 6000105:

Press [ENTER] to proceed.

The IBM XIV host attachment wizard has successfully configured this host.

Press [ENTER] to exit.

The XIV host attachment wizard successfully configured this host

Press [ENTER] to exit.

xiv_attach detected connectivity to three XIVs (zoning to XIVs is already completed) and checked whether a valid host definition exists. Provide a user (with storageadmin rights) and password for each detected XIV. The system then connects to each remote XIV and defines the host and ports on the XIVs.

Tip: A rescan of for new XIV LUNs can be done with `xiv_fc_admin -R`.

- Run the `/opt/xiv/host_attach/bin/xiv_devlist` or `xiv_devlist` command from each working directory. These commands display the mapped volumes and the number of paths to the IBM XIV Storage System as shown in Example 6-7.

Example 6-7 Showing mapped volumes and available paths

```
# xiv_devlist -x

XIV Devices
-----
Device          Size (GB) Paths Vol Name   Vol Id   XIV Id   XIV Host
-----
/dev/dsk/c6t0  103.2      12/12  T5220_01_1 2900     1310114  sun-t5220-01-
017380027820B
54d0s2
-----
/dev/dsk/c6t0  103.2      12/12  T5220_01_2 2902     1310114  sun-t5220-01-
017380027820B
56d0s2
-----
```

6.3 Solaris host configuration for iSCSI

This section explains how to connect an iSCSI volume to the server. The example environment consists of a SUN Sparc T5220 running with Solaris 10 U10. To configure the Solaris host for iSCSI, complete these steps:

- Run the command `xiv_attach` as shown in Example 6-8.

Example 6-8 xiv_attach for iSCSI

```
# xiv_attach
-----
Welcome to the IBM XIV host attachment wizard, version 1.10.0.
This wizard will help you attach this host to the XIV storage system.

The wizard will now validate the host configuration for the XIV storage system.
Press [ENTER] to proceed.

-----
Please specify the connectivity type: [f]c / [i]scsi : i
-----
Please wait while the wizard validates your existing configuration...
Verifying for mpxio DMP solution
Verifying Boot From SAN (BFS) device configuration                                OK
Verifying SCSI driver                                                            OK
Verifying scsi_vhci.conf settings                                               OK
Verifying iscsi.conf settings                                                  OK
This host is already configured for the XIV storage system.

-----
Would you like to discover a new iSCSI target? [default: yes ]:
Please enter an XIV iSCSI discovery address (iSCSI interface): 9.155.116.64
Is this host defined to use CHAP authentication with the XIV storage system?
[default: no ]:
```



```

Would you like to discover a new iSCSI target? [default: yes ]:
Please enter an XIV iSCSI discovery address (iSCSI interface): 9.155.50.11
Is this host defined to use CHAP authentication with the XIV storage system?
[default: no ]:
Would you like to discover a new iSCSI target? [default: yes ]: no
Would you like to rescan for new storage devices? [default: yes ]:

```

```

-----
This host is connected to the following XIV storage arrays:
Serial    Version  Host Defined  Ports Defined  Protocol  Host Name(s)
1310114   11.1.1.0 No            None           iSCSI      --
1310133   11.1.1.0 No            None           iSCSI      --
This host is not defined on some of the iSCSI-attached XIV storage systems.
Do you want to define this host on these XIV systems now? [default: yes ]:
Please enter a name for this host [default: sun-t5220-01-1 ]:
sun-t5220-01-1-iscsi
Please enter a username for system 1310114 [default: admin ]:  itso
Please enter the password of user itso for system 1310114:

Please enter a username for system 1310133 [default: admin ]:  itso
Please enter the password of user itso for system 1310133:

```

Press [ENTER] to proceed.

```

-----
The IBM XIV host attachment wizard has successfully configured this host.

```

Press [ENTER] to exit.

2. Define the host and iSCSI port when it rescans for storage devices as seen in Example 6-8 on page 190. You need a valid storageadmin ID to do so. The host and iSCSI port can also be defined on the GUI.
3. Discover the iSCSI qualified name (IQN) of your server with the **xiv_iscsi_admin -P** command as shown in Example 6-9.

Example 6-9 Display IQN

```

# xiv_iscsi_admin -P
iqn.1986-03.com.sun:01:0021284fe446.508551e0

```

4. Define and map volumes on the XIV system.
5. Rescan the iSCSI using the **xiv_iscsi_admin -R** command. You see all XIV devices that are mapped to the host as shown in Example 6-10.

Example 6-10 xiv_devlist

```

# xiv_devlist -x
XIV Devices
-----
Device          Size (GB)  Paths  Vol Name      Vol Id  XIV Id  XIV Host
-----
/dev/dsk/c6t0  51.6      3/3    T5220-01-iSCS  6329   1310133 sun-t5220-01-
0173800279518          I                               1-iscsi
B9d0s2
-----

```

6.4 Solaris Host Attachment Kit utilities

The Host Attachment Kit now includes the following utilities:

► **xiv_devlist**

This utility allows validation of the attachment configuration. It generates a list of multipathed devices available to the operating system. Example 6-11 shows the options of the `xiv_devlist` commands.

Example 6-11 xiv_devlist

```
# xiv_devlist --help
Usage: xiv_devlist [options]

Options:
  -h, --help           show this help message and exit
  -t OUT, --out=OUT    Choose output method: tui, csv, xml (default: tui)
  -f OUTFILE, --file=OUTFILE
                        File name to output to (instead of STDOUT). Can be
                        used only with -t csv/xml
  -o FIELDS, --options=FIELDS
                        Choose which fields to display in the output; Comma-
                        separated, no spaces. Use -l to see the list of fields
  -l, --list-fields    List available fields for the -o option
  -H, --hex            Display XIV volume and machine IDs in hexadecimal base
  -u SIZE_UNIT, --size-unit=SIZE_UNIT
                        The size unit to use (e.g. MB, GB, TB, MiB, GiB, TiB,
                        ...)
  -m MP_FRAMEWORK_STR, --multipath=MP_FRAMEWORK_STR
                        Enforce a multipathing framework <auto|native|veritas>
  -x, --xiv-only       Print only XIV devices
  -d, --debug          Enable debug logging
  -V, --version        Display the version number
```

► **xiv_diag**

This utility gathers diagnostic information from the operating system. The resulting compressed file can then be sent to IBM-XIV support teams for review and analysis. Example results are shown in Example 6-12.

Example 6-12 xiv_diag command

```
# xiv_diag
Welcome to the XIV diagnostics tool, version 1.10.0.
This tool will gather essential support information from this host.
Please type in a path to place the xiv_diag file in [default: /tmp]:
Creating archive xiv_diag-results_2012-10-23_14-39-57
INFO: Gathering HAK version... DONE
INFO: Gathering uname... DONE
INFO: Gathering cfgadm... DONE
INFO: Gathering find /dev... DONE
INFO: Gathering Package list... DONE
INFO: Gathering xiv_devlist... DONE
INFO: Gathering xiv_fc_admin -V... DONE
INFO: Gathering xiv_iscsi_admin -V... DONE
INFO: Gathering xiv_fc_admin -L... DONE
```

```

INFO: Gathering xiv_fc_admin -P... DONE
INFO: Gathering xiv_iscsi_admin -L... DONE
INFO: Gathering xiv_iscsi_admin -P... DONE
INFO: Gathering SCSI Inquiries... DONE
INFO: Gathering scsi_vhci.conf... DONE
INFO: Gathering release... DONE
INFO: Gathering fp.conf... DONE
INFO: Gathering /var/adm directory... DONE
INFO: Gathering /var/log directory... DONE
INFO: Gathering build-revision file... DONE

INFO: Closing xiv_diag archive file DONE
Deleting temporary directory... DONE
INFO: Gathering is now complete.
INFO: You can now send /tmp/xiv_diag-results_2012-10-23_14-39-57.tar.gz to IBM-XIV
for review.
INFO: Exiting.

```

6.5 Creating partitions and file systems with UFS

This section describes how to create a partition and file systems with UFS on mapped XIV volumes. A system with Solaris 10 on a Sparc is used to illustrate the process as shown in Example 6-13.

Example 6-13 Mapped XIV volumes

```

# xiv_devlist -x
XIV Devices
-----
Device          Size (GB) Paths Vol Name   Vol Id  XIV Id  XIV Host
-----
/dev/dsk/c6t0  103.2      10/10 T5220_01_1 2900    1310114 sun-t5220-01-
017380027820B 1
54d0s2
-----
/dev/dsk/c6t0  103.2      10/10 T5220_01_2 2902    1310114 sun-t5220-01-
017380027820B 1
56d0s2
-----

```

Example 6-14 shows how to use the command **format** on the Solaris system.

Example 6-14 Solaris format tool

```

# format
Searching for disks...done

c6t0017380027820B54d0: configured with capacity of 96.14GB
c6t0017380027820B56d0: configured with capacity of 96.14GB

AVAILABLE DISK SELECTIONS:
  0. c1t0d0 <SUN146G cyl 14087 alt 2 hd 24 sec 848>

```

```

    /pci@0/pci@0/pci@2/scsi@0/sd@0,0
1. c1t1d0 <SUN146G cyl 14087 alt 2 hd 24 sec 848>
    /pci@0/pci@0/pci@2/scsi@0/sd@1,0
2. c1t2d0 <LSILOGIC-Logical Volume-3000-136.67GB>
    /pci@0/pci@0/pci@2/scsi@0/sd@2,0
3. c1t4d0 <LSILOGIC-Logical Volume-3000-136.67GB>
    /pci@0/pci@0/pci@2/scsi@0/sd@4,0
4. c6t0017380027820B54d0 <IBM-2810XIV-0000 cyl 12306 alt 2 hd 128 sec 128>
    /scsi_vhci/ssd@g0017380027820b54
5. c6t0017380027820B56d0 <IBM-2810XIV-0000 cyl 12306 alt 2 hd 128 sec 128>
    /scsi_vhci/ssd@g0017380027820b56

```

```

Specify disk (enter its number): 4
selecting c6t0017380027820B54d0
[disk formatted]
Disk not labeled. Label it now? yes

```

FORMAT MENU:

```

disk      - select a disk
type      - select (define) a disk type
partition - select (define) a partition table
current   - describe the current disk
format    - format and analyze the disk
repair    - repair a defective sector
label     - write label to the disk
analyze   - surface analysis
defect    - defect list management
backup    - search for backup labels
verify    - read and display labels
save      - save new disk/partition definitions
inquiry   - show vendor, product and revision
volname   - set 8-character volume name
!<cmd>   - execute <cmd>, then return
quit

```

The standard partition table can be used, but you can also define a user-specific table. Use the **partition** command in the format tool to change the partition table. You can see the newly defined table by using the **print** command as shown in Example 6-15.

Example 6-15 Solaris format/partition tool

```
format> partition
```

PARTITION MENU:

```

0      - change `0' partition
1      - change `1' partition
2      - change `2' partition
3      - change `3' partition
4      - change `4' partition
5      - change `5' partition
6      - change `6' partition
7      - change `7' partition
select - select a predefined table
modify - modify a predefined partition table

```

```

    name - name the current table
    print - display the current table
    label - write partition map and label to the disk
    !<cmd> - execute <cmd>, then return
    quit
partition> print
Current partition table (default):
Total disk cylinders available: 12306 + 2 (reserved cylinders)

```

Part	Tag	Flag	Cylinders	Size	Blocks
0	root	wm	0 - 15	128.00MB	(16/0/0) 262144
1	swap	wu	16 - 31	128.00MB	(16/0/0) 262144
2	backup	wu	0 - 12305	96.14GB	(12306/0/0) 201621504
3	unassigned	wm	0	0	(0/0/0) 0
4	unassigned	wm	0	0	(0/0/0) 0
5	unassigned	wm	0	0	(0/0/0) 0
6	usr	wm	32 - 12305	95.89GB	(12274/0/0) 201097216
7	unassigned	wm	0	0	(0/0/0) 0

```

partition> label
Ready to label disk, continue? yes

```

```

partition> quit

```

FORMAT MENU:

```

    disk - select a disk
    type - select (define) a disk type
    partition - select (define) a partition table
    current - describe the current disk
    format - format and analyze the disk
    repair - repair a defective sector
    label - write label to the disk
    analyze - surface analysis
    defect - defect list management
    backup - search for backup labels
    verify - read and display labels
    save - save new disk/partition definitions
    inquiry - show vendor, product and revision
    volname - set 8-character volume name
    !<cmd> - execute <cmd>, then return
    quit
format> quit

```

Verify the new table as shown in Example 6-16.

Example 6-16 Verifying the table

```

# prtvtoc /dev/rdisk/c6t0017380027820B54d0s0
* /dev/rdisk/c6t0017380027820B54d0s0 partition map
*
* Dimensions:
*   512 bytes/sector
*   128 sectors/track
*   128 tracks/cylinder
*   16384 sectors/cylinder

```

```

* 12308 cylinders
* 12306 accessible cylinders
*
* Flags:
* 1: unmountable
* 10: read-only
*
*
* Partition Tag Flags First Sector Last Sector Mount Directory
* 0 2 00 0 262144 262143
* 1 3 01 262144 262144 524287
* 2 5 01 0 201621504 201621503
* 6 4 00 524288 201097216 201621503

```

Create file systems on the partition/volume as shown in Example 6-17.

Example 6-17 Making new file systems

```

# newfs /dev/rdisk/c6t0017380027820B54d0s2
newfs: construct a new file system /dev/rdisk/c6t0017380027820B54d0s2: (y/n)? y
/dev/rdisk/c6t0017380027820B54d0s2: 201621504 sectors in 32816 cylinders of 48
tracks, 128 sectors
98448.0MB in 2051 cyl groups (16 c/g, 48.00MB/g, 5824 i/g)
super-block backups (for fsck -F ufs -o b=#) at:
32, 98464, 196896, 295328, 393760, 492192, 590624, 689056, 787488, 885920,
Initializing cylinder groups:
.....
super-block backups for last 10 cylinder groups at:
200645792, 200744224, 200842656, 200941088, 201039520, 201137952, 201236384,
201326624, 201425056, 201523488

```

You can optionally check the file systems as seen in Example 6-18.

Example 6-18 Checking the file systems

```

# fsck /dev/rdisk/c6t0017380027820B54d0s2
** /dev/rdisk/c6t0017380027820B54d0s2
** Last Mounted on
** Phase 1 - Check Blocks and Sizes
** Phase 2 - Check Pathnames
** Phase 3a - Check Connectivity
** Phase 3b - Verify Shadows/ACLs
** Phase 4 - Check Reference Counts
** Phase 5 - Check Cylinder Groups
2 files, 9 used, 99284750 free (14 frags, 12410592 blocks, 0.0% fragmentation)

```

After creation of the mount point and mounting the volume as shown in Example 6-19, you can start using the volume with ufs file systems.

Example 6-19 Mounting the volume to Solaris

```

# mkdir /XIV_Vol
# mount /dev/dsk/c6t0017380027820B54d0s2 /XIV_Vol
# df -h
Filesystem size used avail capacity Mounted on
/dev/dsk/c1t0d0s0 35G 4.2G 31G 13% /

```

/devices	OK	OK	OK	0%	/devices
ctfs	OK	OK	OK	0%	/system/contract
proc	OK	OK	OK	0%	/proc
mnttab	OK	OK	OK	0%	/etc/mnttab
swap	13G	1.6M	13G	1%	/etc/svc/volatile
objfs	OK	OK	OK	0%	/system/object
sharefs	OK	OK	OK	0%	/etc/dfs/sharetab
/platform/SUNW,SPARC-Enterprise-T5220/lib/libc_psr/libc_psr_hwcap2.so.1					
	35G	4.2G	31G	13%	
/platform/sun4v/lib/libc_psr.so.1					
/platform/SUNW,SPARC-Enterprise-T5220/lib/sparcv9/libc_psr/libc_psr_hwcap2.so.1					
	35G	4.2G	31G	13%	
/platform/sun4v/lib/sparcv9/libc_psr.so.1					
fd	OK	OK	OK	0%	/dev/fd
swap	13G	32K	13G	1%	/tmp
swap	13G	40K	13G	1%	/var/run
/dev/dsk/c1t0d0s7	99G	142M	98G	1%	/export/home
/dev/dsk/c6t0017380027820B54d0s2					
	95G	96M	94G	1%	/XIV_Vol



XIV and Symantec Storage Foundation

This chapter addresses specific considerations for host connectivity. It describes host attachment-related tasks for the operating systems that use Symantec Storage Foundation instead of their built-in functions.

Important: The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, see the System Storage Interoperation Center (SSIC) at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

You can find the latest Host Attachment Kit software and User Guide at:

http://www-933.ibm.com/support/fixcentral/swg/selectFixes?parent=Enterprise+Storage+Servers&product=ibm/Storage_Disk/XIV+Storage+System+%282810,+2812%29&release=All&platform=All&function=all#IBM%20XIV%20Host%20Attachment%20Kit

This chapter includes the following sections:

- ▶ Introduction
- ▶ Prerequisites
- ▶ Placing XIV LUNs under VxVM control
- ▶ Configuring multipathing with DMP
- ▶ Working with snapshots

7.1 Introduction

The Symantec Storage Foundation is available as a unified method of volume management at the OS level. It was formerly known as the Veritas Volume Manager (VxVM) and Veritas Dynamic Multipathing (DMP).

At the time of writing, XIV supports the use of VxVM and DMP with the following operating systems:

- ▶ HP-UX
- ▶ AIX
- ▶ Red Hat Enterprise Linux
- ▶ SUSE Linux
- ▶ Linux on Power
- ▶ Solaris

Depending on the OS version and hardware, only specific versions and releases of Veritas Volume Manager are supported when connected to XIV. In general, IBM supports VxVM versions 5.0 and 5.1.

For most of the OS and VxVM versions, IBM supports space reclamation on thin provisioned volumes.

For more information about the operating systems and VxVM versions that are supported, see the SSIC at:

<http://www.ibm.com/systems/support/storage/config/ssic>

For more information about attaching the IBM XIV Storage System to hosts with VxVM and DMP, see the Symantec website at:

<https://sort.symantec.com/asl>

7.2 Prerequisites

Common prerequisites such as cabling, defining SAN zoning, and creating volumes and mapping them to the host, must be completed. In addition, the following tasks must be completed to successfully attach XIV to host systems by using VxVM with DMP:

- ▶ Check Array Support Library (ASL) availability for XIV Storage System on your Symantec Storage Foundation installation.
- ▶ Place the XIV volumes under VxVM control.
- ▶ Set up DMP multipathing with IBM XIV.

Make sure that you install all the patches and updates available for your Symantec Storage Foundation installation. For more information, see your Symantec Storage Foundation documentation.

7.2.1 Checking ASL availability and installation

The examples that are used to illustrate attachment to XIV and configuration for hosts by using VxVM with DMP as logical volume manager use Solaris version 10 on SPARC. The steps are similar for most UNIX and Linux hosts.

To check for the presence of the XIV ASL on your host system, log on to the host as root and run the command that is shown in Example 7-1.

Example 7-1 Checking the availability of ASL for IBM XIV Storage System

```
# vxddladm listversion|grep xiv
libvxxiv.so                vm-5.1.100-rev-1 5.1
```

If the command output does not show that the required ASL is already installed, locate the installation package. The installation package for the ASL is available at:

<https://sort.symantec.com/asl>

Specify the vendor of your storage system, your operating system, and the version of your Symantec Storage Foundation. You are then redirected to a page from which you can download the ASL package for your environment. Installation instructions are available on the same page.

Install the required XIV Host Attachment Kit for your platform. You can check the Host Attachment Kit availability for your platform at:

<http://www.ibm.com/support/fixcentral>

7.2.2 Installing the XIV Host Attachment Kit

You must install the XIV Host Attachment Kit for your system to get support. To install the Host Attachment Kit in the Solaris/SPARC experimentation scenario, complete the following steps:

1. Open a terminal session and go to the directory where the package was downloaded.
2. Extract files from the archive by running the commands that are shown in Example 7-2.

Example 7-2 Extracting the Host Attachment Kit

```
# gunzip IBM_XIV_Host_Attachment_Kit_<version>-<os>-<arch>.tar.gz
# tar -xvf IBM_XIV_Host_Attachment_Kit_<version>-<os>-<arch>.tar
```

3. Change to the newly created directory and start the Host Attachment Kit installer, as seen in Example 7-3.

Example 7-3 Starting the installation

```
# cd IBMxivhak-<version>-<os>-<arch>
# /bin/sh ./install.sh
```

4. Follow the prompts.
5. Review the installation log file `install.log` in the same directory.

7.2.3 Configuring the host

Use the utilities that are provided in the Host Attachment Kit to configure the host. The Host Attachment Kit packages are installed in /opt/xiv/host_attach directory.

Requirement: You must be logged in as root or have root privileges to use the Host Attachment Kit.

1. Run the `xiv_attach` utility as shown in Example 7-4. The command can also be started from any working directory.

Example 7-4 Starting `xiv_attach`

```
# /opt/xiv/host_attach/bin/xiv_attach
```

```
-----  
Welcome to the IBM XIV host attachment wizard, version 1.10.0.  
This wizard will help you attach this host to the XIV storage system.
```

```
The wizard will now validate the host configuration for the XIV storage system.  
Press [ENTER] to proceed.
```

```
-----  
Please specify the connectivity type: [f]c / [i]scsi : f  
Notice: VxDMP is available and will be used as the DMP software  
Press [ENTER] to proceed.
```

```
-----  
Please wait while the wizard validates your existing configuration...  
Verifying for vxdmp DMP solution  
Verifying VxDMP configuration OK  
Verifying SCSI driver NOT OK  
Verifying scsi_vhci.conf settings NOT OK  
Verifying fp.conf settings NOT OK
```

```
-----  
The wizard needs to configure this host for the XIV storage system.  
Do you want to proceed? [default: yes ]:  
Please wait while the host is being configured...
```

```
-----  
Configuring for vxdmp DMP solution  
Configuring VxDMP configuration OK  
Configuring SCSI driver  
devfsadm: driver failed to attach: sgen  
Warning: Driver (sgen) successfully added to system but failed to attach  
OK  
Configuring scsi_vhci.conf settings REBOOT  
Configuring fp.conf settings REBOOT
```

```
-----  
A reboot is required.  
Please reboot this host and then restart the wizard.
```

```
Press [ENTER] to exit.
```

2. For the Solaris on SUN server that is used in this example, you must reboot the host before you proceed to the next step. Other systems can vary.
3. Zone SUN server to XIVs.

4. After the system reboot, start `xiv_attach` again to complete the host system configuration for XIV attachment (Example 7-5).

Example 7-5 Fibre Channel host attachment configuration after reboot

```
# /opt/xiv/host_attach/bin/xiv_attach
-----
Welcome to the IBM XIV host attachment wizard, version 1.10.0.
This wizard will help you attach this host to the XIV storage system.

The wizard will now validate the host configuration for the XIV storage system.
Press [ENTER] to proceed.

-----
Please specify the connectivity type: [f]c / [i]scsi : f
Notice: VxDMP is available and will be used as the DMP software
Press [ENTER] to proceed.

-----
Please wait while the wizard validates your existing configuration...
Verifying for vxdmp DMP solution
Verifying VxDMP configuration                                OK
Verifying SCSI driver                                       OK
Verifying scsi_vhci.conf settings                           OK
Verifying fp.conf settings                                  OK
This host is already configured for the XIV storage system.

-----
Please define zoning for this host and add its World Wide Port Names (WWPNs) to
the XIV storage system:
2101001b32b1b0b1: /dev/cfg/c3: [QLogic Corp.]: 371-4325-01
2101001b32b1beb1: /dev/cfg/c5: [QLogic Corp.]: 371-4325-01
2100001b3291b0b1: /dev/cfg/c2: [QLogic Corp.]: 371-4325-01
2100001b3291beb1: /dev/cfg/c4: [QLogic Corp.]: 371-4325-01
Press [ENTER] to proceed.

Would you like to rescan for new storage devices? [default: yes ]:
Please wait while rescanning for XIV storage devices...

-----
This host is connected to the following XIV storage arrays:
Serial    Version  Host Defined  Ports Defined  Protocol  Host Name(s)
1310114   11.1.1.0 No           None           FC        --
6000105   10.2.4.5 No           None           FC        --
1310133   11.1.1.0 No           None           FC        --
This host is not defined on some of the FC-attached XIV storage systems.
Do you want to define this host on these XIV systems now? [default: yes ]:
Please enter a name for this host [default: sun-t5220-02-1 ]:
Please enter a username for system 1310114 [default: admin ]: itso
Please enter the password of user itso for system 1310114:

Please enter a username for system 6000105 [default: admin ]: itso
Please enter the password of user itso for system 6000105:

Please enter a username for system 1310133 [default: admin ]: itso
Please enter the password of user itso for system 1310133:
```

Press [ENTER] to proceed.

The IBM XIV host attachment wizard has successfully configured this host.

Press [ENTER] to exit.

-
5. Create volumes on XIV and map these volumes (LUNs) to the host system, which was configured by `xiv_attach`. You can use the XIV GUI for volume creation and mapping tasks, as illustrated in 1.4, “Logical configuration for host connectivity” on page 37.
 6. After the LUN mapping is completed, discover the mapped LUNs on your host by running the `xiv_fc_admin -R` command.
 7. Use the command `/opt/xiv/host_attach/bin/xiv_devlist` to check the mapped volumes and the number of paths to the XIV Storage System (Example 7-6).

Example 7-6 Showing mapped volumes and available paths

```
# xiv_devlist -x
XIV Devices
-----
Device          Size (GB) Paths Vol Name   Vol Id  XIV Id  XIV Host
-----
/dev/vx/dmp/x   103.2     12/12  T5220_02_1  9498   1310114 sun-t5220-02-iv0_251a
-----
/dev/vx/dmp/x   103.2     12/12  T5220_02_2  9499   1310114 sun-t5220-02-iv0_251b
-----
```

7.3 Placing XIV LUNs under VxVM control

To place XIV LUNs under VxVM control, complete the following steps:

1. Label the disks with the `format` command.
2. Discover new devices on your hosts by using either the `vxdiskconfig` or `vxdisk -f scandisks` command.
3. Check for new devices that were discovered by using the `vxdisk list` command as illustrated in Example 7-7.

Example 7-7 Discovering and checking new disks on your host

```
# vxdisk -f scandisks
# vxdisk list
DEVICE      TYPE          DISK      GROUP      STATUS
disk_0      auto:ZFS      -         -          ZFS
disk_1      auto:ZFS      -         -          ZFS
disk_2      auto:ZFS      -         -          ZFS
disk_3      auto:ZFS      -         -          ZFS
xiv0_251a   auto:none     -         -          online invalid
xiv0_251b   auto:none     -         -          online invalid
```

4. After you discover the new disks on the host, you might need to format the disks. For more information, see your OS-specific Symantec Storage Foundation documentation. In this

example, the disks must be formatted. Run the `vxdi skadm` command as shown in Example 7-8. Select option 1 and then follow the instructions, accepting all defaults except for the questions Encapsulate this device? (answer no), and Instead of encapsulating, initialize? (answer yes).

Example 7-8 Configuring disks for VxVM

```
# vxdiskadm
```

```
Volume Manager Support Operations
```

```
Menu: VolumeManager/Disk
```

- 1 Add or initialize one or more disks
 - 2 Encapsulate one or more disks
 - 3 Remove a disk
 - 4 Remove a disk for replacement
 - 5 Replace a failed or removed disk
 - 6 Mirror volumes on a disk
 - 7 Move volumes from a disk
 - 8 Enable access to (import) a disk group
 - 9 Remove access to (deport) a disk group
 - 10 Enable (online) a disk device
 - 11 Disable (offline) a disk device
 - 12 Mark a disk as a spare for a disk group
 - 13 Turn off the spare flag on a disk
 - 14 Unrelocate subdisks back to a disk
 - 15 Exclude a disk from hot-relocation use
 - 16 Make a disk available for hot-relocation use
 - 17 Prevent multipathing/Suppress devices from VxVM's view
 - 18 Allow multipathing/Unsuppress devices from VxVM's view
 - 19 List currently suppressed/non-multipathed devices
 - 20 Change the disk naming scheme
 - 21 Get the newly connected/zoned disks in VxVM view
 - 22 Change/Display the default disk layouts
 - list List disk information
-
- ? Display help about menu
 - ?? Display help about the menuing system
 - q Exit from menus

```
Select an operation to perform: 1
```

```
Add or initialize disks
```

```
Menu: VolumeManager/Disk/AddDisks
```

Use this operation to add one or more disks to a disk group. You can add the selected disks to an existing disk group or to a new disk group that will be created as a part of the operation. The selected disks may also be added to a disk group as spares. Or they may be added as nohotuses to be excluded from hot-relocation use. The selected disks may also be initialized without adding them to a disk group leaving the disks available for use as replacement disks.

More than one disk or pattern may be entered at the prompt. Here are some disk selection examples:

all: all disks
c3 c4t2: all disks on both controller 3 and controller 4, target 2
c3t4d2: a single disk (in the c#t#d# naming scheme)
xyz_0 : a single disk (in the enclosure based naming scheme)
xyz_ : all disks on the enclosure whose name is xyz

Select disk devices to add: [<pattern-list>,all,list,q,?] xiv0_251a xiv0_251b
Here are the disks selected. Output format: [Device_Name]

xiv0_251a xiv0_251b

Continue operation? [y,n,q,?] (default: y)
You can choose to add these disks to an existing disk group, a new disk group, or you can leave these disks available for use by future add or replacement operations. To create a new disk group, select a disk group name that does not yet exist. To leave the disks available for future use, specify a disk group name of "none".

Which disk group [<group>,none,list,q,?] (default: none) XIV_DG

Create a new group named XIV_DG? [y,n,q,?] (default: y)

Create the disk group as a CDS disk group? [y,n,q,?] (default: y)

Use default disk names for these disks? [y,n,q,?] (default: y)

Add disks as spare disks for XIV_DG? [y,n,q,?] (default: n)

Exclude disks from hot-relocation use? [y,n,q,?] (default: n)

Add site tag to disks? [y,n,q,?] (default: n)
A new disk group will be created named XIV_DG and the selected disks will be added to the disk group with default disk names.

xiv0_251a xiv0_251b

Continue with operation? [y,n,q,?] (default: y)
The following disk devices have a valid VTOC, but do not appear to have been initialized for the Volume Manager. If there is data on the disks that should NOT be destroyed you should encapsulate the existing disk partitions as volumes instead of adding the disks as new disks.
Output format: [Device_Name]

xiv0_251a xiv0_251b

Encapsulate these devices? [Y,N,S(elect),q,?] (default: Y) N

xiv0_251a xiv0_251b

Instead of encapsulating, initialize?
[Y,N,S(elect),q,?] (default: N) Y

Do you want to use the default layout for all disks being initialized?
[y,n,q,?] (default: y)


```

Initializing device xiv0_251a.
Initializing device xiv0_251b.
VxVM NOTICE V-5-2-120
Creating a new disk group named XIV_DG containing the disk
device xiv0_251a with the name XIV_DG01.
VxVM NOTICE V-5-2-88
Adding disk device xiv0_251b to disk group XIV_DG with disk
name XIV_DG02.

Add or initialize other disks? [y,n,q,?] (default: n)

```

Tip: If the `vxdi skadm` initialization function complains the disk is offline, you might need to initialize it using the default OS-specific utility. For example, use the `format` command in Solaris.

5. Check the results by using the `vxdisk list` and `vx dg list` commands as shown in Example 7-9.

Example 7-9 Showing the results of putting XIV LUNs under VxVM control

```

# vxdisk list
DEVICE      TYPE          DISK          GROUP          STATUS
disk_0      auto:ZFS      -             -             ZFS
disk_1      auto:ZFS      -             -             ZFS
disk_2      auto:ZFS      -             -             ZFS
disk_3      auto:ZFS      -             -             ZFS
xiv0_251a   auto:cdsdisk  XIV_DG01     XIV_DG         online thinrc1m
xiv0_251b   auto:cdsdisk  XIV_DG02     XIV_DG         online thinrc1m
# vx dg list XIV_DG
Group:      XIV_DG
dgid:       1349985305.13.sun-t5220-02-1
import-id:  1024.12
flags:      cds
version:    160
alignment:  8192 (bytes)
ssb:        on
autotagging: on
detach-policy: global
dg-fail-policy: dgdisable
copies:     nconfig=default nlog=default
config:     seqno=0.1029 perm1en=48144 free=48140 templen=2 loglen=7296
config disk xiv0_251a copy 1 len=48144 state=clean online
config disk xiv0_251b copy 1 len=48144 state=clean online
log disk xiv0_251a copy 1 len=7296
log disk xiv0_251b copy 1 len=7296

```

The XIV LUNs that were added are now available for volume creation and data storage. The status `thinrc1m` means that the volumes from the XIV are thin-provisioned, and the XIV storage has the Veritas thin reclamation API implemented.

- Use the **vxdisk reclaim <diskgroup> | <disk>** command to free up any space that can be reclaimed. For more information about thin reclamation on XIV, see *Thin Reclamation Using Veritas Storage Foundation Enterprise HA from Symantec and the IBM XIV Storage System*, TSL03051USEN. This publication can be found at:
<ftp://public.dhe.ibm.com/common/ssi/ecm/en/ts103051usen/TSL03051USEN.PDF>
- Check that you get adequate performance, and, if required, configure DMP multipathing settings.

7.4 Configuring multipathing with DMP

The Symantec Storage Foundation version 5.1 uses MinimumQ *iopolicy* by default for the enclosures on Active/Active storage systems. For most typical SAN workloads the default DMP policy of Minimum Queue (MinQ) is optimal. Other **iopolicy** algorithms can be selected based on specific application or operating system needs. For example, to set the **iopolicy** parameter to **round-robin** and enable the use of all paths, complete the following steps:

- Identify names of enclosures on the XIV Storage System.
- Log on to the host as root user, and run the **vxddmpadm listenclosure all** command. Examine the results to determine which enclosure names belong to an XIV Storage System. In Example 7-10, the enclosure name is xiv0.

Example 7-10 Identifying names of enclosures on an IBM XIV Storage System

```
# vxddmpadm listenclosure all
```

ENCLR_NAME	ENCLR_TYPE	ENCLR_SNO	STATUS	ARRAY_TYPE	LUN_
disk	Disk	DISKS	CONNECTED	Disk	4
xiv0	XIV	130210114	CONNECTED	A/A	2

- Change the **iopolicy** parameter for the identified enclosures by running the command **vxddmpadm setattr enclosure <identified enclosure name> iopolicy=round-robin** for each identified enclosure.
- Check the results of the change by running the command **vxddmpadm getattr enclosure <identified enclosure name>** as shown in Example 7-11.

Example 7-11 Changing DMP settings by using the iopolicy parameter

```
# vxddmpadm setattr enclosure xiv0 iopolicy=round-robin
# vxddmpadm getattr enclosure xiv0
```

ENCLR_NAME	ATTR_NAME	DEFAULT	CURRENT
xiv0	iopolicy	MinimumQ	Round-Robin
xiv0	partitionsizes	512	512
xiv0	use_all_paths	-	-
xiv0	failover_policy	Global	Global
xiv0	recoveryoption[throttle]	Nothrottle[0]	Nothrottle[0]
xiv0	recoveryoption[errorretry]	Timebound[300]	Timebound[300]
xiv0	redundancy	0	0
xiv0	dmp_lun_retry_timeout	0	0
xiv0	failovermode	explicit	explicit

- For heavy workloads, increase the queue depth parameter to 64. The queue depth can be set as high as 128 if needed. Run the command **vxddmpadm gettune dmp_queue_depth** to get information about current settings and run **vxddmpadm settune**

dmp_queue_depth=<new queue depth value> to adjust them as shown in Example 7-12.

Example 7-12 Changing the queue depth parameter

```
# vxdmpadm gettune dmp_queue_depth
      Tunable                Current Value  Default Value
-----
dmp_queue_depth                32             32
# vxdmpadm settune dmp_queue_depth=64
Tunable value will be changed immediately
# vxdmpadm gettune dmp_queue_depth
      Tunable                Current Value  Default Value
-----
dmp_queue_depth                64             32
```

7.5 Working with snapshots

Version 5.0 of Symantec Storage Foundation introduced a new function to work with hardware cloned or snapshot target devices. Starting with version 5.0, VxVM stores the unique disk identifier (UDID) in the disk private region. The UDID is stored when the disk is initialized or when the disk is imported into a disk group.

Whenever a disk is brought online, the current UDID value is compared to the UDID stored in the private region of the disk. If the UDID does not match, the `udid_mismatch` flag is set on the disk. This flag allows LUN snapshots to be imported on the same host as the original LUN. It also allows multiple snapshots of the same LUN to be concurrently imported on a single server. These snapshots can then be used for the offline backup or processing.

After you create XIV snapshots for LUNs used on a host under VxVM control, unlock (enable writing) those snapshots and map them to your host. When this process is complete, the snapshot LUNs can be imported on the host by using the following steps:

1. Check that the created snapshots are visible for your host by running the commands **vxdisk scandisks** and **vxdisk list** as shown in Example 7-13.

Example 7-13 Identifying created snapshots on host side

```
# vxdisk scandisks
# vxdisk list
DEVICE      TYPE          DISK          GROUP          STATUS
disk_0      auto:ZFS      -             -              ZFS
disk_1      auto:ZFS      -             -              ZFS
disk_2      auto:ZFS      -             -              ZFS
disk_3      auto:ZFS      -             -              ZFS
xiv0_0b70   auto:cdsdisk -             -              online udid_mismatch
xiv0_0b71   auto:cdsdisk -             -              online udid_mismatch
xiv0_251a   auto:cdsdisk XIV_DG01      XIV_DG         online thinrc1m
xiv0_251b   auto:cdsdisk XIV_DG02      XIV_DG         online thinrc1m
```

2. Import the created snapshot on your host by running the command **vxdbg -n <name for new volume group> -o useclonedev=on,updateid -C import <name of original volume group>**.

3. Run the **vxdisk list** command to ensure that the LUNs were imported as shown in Example 7-14.

Example 7-14 Importing snapshots onto your host

```
# vxdg -n XIV_DG_SNAP -o useclonedev=on,updateid -C import XIV_DG
# vxdisk list
```

DEVICE	TYPE	DISK	GROUP	STATUS
disk_0	auto:ZFS	-	-	ZFS
disk_1	auto:ZFS	-	-	ZFS
disk_2	auto:ZFS	-	-	ZFS
disk_3	auto:ZFS	-	-	ZFS
xiv0_0b70	auto:cdsdisk	XIV_DG01	XIV_DG_SNAP	online clone_disk
xiv0_0b71	auto:cdsdisk	XIV_DG02	XIV_DG_SNAP	online clone_disk
xiv0_251a	auto:cdsdisk	XIV_DG01	XIV_DG	online thinrclm
xiv0_251b	auto:cdsdisk	XIV_DG02	XIV_DG	online thinrclm



IBM i and AIX clients connecting to XIV through VIOS

This chapter explains XIV connectivity with Virtual I/O Server (VIOS) clients, including AIX, Linux on Power and IBM i. VIOS is a component of PowerVM that provides the ability for logical partitions (LPARs) that are VIOS clients to share resources.

Important: The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, see the System Storage Interoperation Center (SSIC) at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

Host Attachment Kits and documentation can be downloaded from Fix Central at:

<http://www.ibm.com/support/fixcentral/>

This chapter includes the following sections:

- ▶ Introduction to IBM PowerVM
- ▶ Planning for VIOS and IBM i
- ▶ Connecting an PowerVM IBM i client to XIV
- ▶ Mapping XIV volumes in the Virtual I/O Server
- ▶ Matching XIV volume to IBM i disk unit
- ▶ Performance considerations for IBM i with XIV

8.1 Introduction to IBM PowerVM

Virtualization on IBM Power Systems servers provides a rapid and cost-effective response to many business needs. Virtualization capabilities have become an important element in planning for IT floor space and servers. Growing commercial and environmental concerns create pressure to reduce the power footprint of servers. With the escalating cost of powering and cooling servers, consolidation and efficient utilization of the servers is becoming critical.

Virtualization on Power Systems servers allows an efficient utilization of servers by reducing the following needs:

- ▶ Server management and administration costs because there are fewer physical servers
- ▶ Power and cooling costs with increased utilization of existing servers
- ▶ Time to market because virtual resources can be deployed immediately

8.1.1 IBM PowerVM overview

IBM PowerVM is a virtualization technology for AIX, IBM i, and Linux environments on IBM POWER® processor-based systems. It is a special software appliance that is tied to IBM Power Systems, which are the converged IBM i and IBM p server platforms. It is licensed on a POWER processor basis.

PowerVM offers a secure virtualization environment with the following features and benefits:

- ▶ Consolidates diverse sets of applications that are built for multiple operating systems (AIX, IBM i, and Linux) on a single server
- ▶ Virtualizes processor, memory, and I/O resources to increase asset utilization and reduce infrastructure costs
- ▶ Dynamically adjusts server capability to meet changing workload demands
- ▶ Moves running workloads between servers to maximize availability and avoid planned downtime

Virtualization technology is offered in three editions on Power Systems:

- ▶ PowerVM Express Edition
- ▶ PowerVM Standard Edition
- ▶ PowerVM Enterprise Edition

PowerVM provides logical partitioning technology by using the following features:

- ▶ Either the Hardware Management Console (HMC) or the Integrated Virtualization Manager (IVM)
- ▶ Dynamic logical partition (LPAR) operations
- ▶ IBM Micro-Partitioning® and VIOS capabilities
- ▶ N_Port ID Virtualization (NPIV)

PowerVM Express Edition

PowerVM Express Edition is available only on the IBM Power 520 and Power 550 servers. It is designed for clients who want an introduction to advanced virtualization features at an affordable price.

With PowerVM Express Edition, you can create up to three partitions on a server (two client partitions and one for the VIOS and IVM). You can use virtualized disk and optical devices,

and try the shared processor pool. All virtualization features can be managed by using the IVM, including:

- ▶ Micro-Partitioning
- ▶ Shared processor pool
- ▶ VIOS
- ▶ PowerVM LX86
- ▶ Shared dedicated capacity
- ▶ NPIV
- ▶ Virtual tape

PowerVM Standard Edition

For clients who are ready to gain the full value from their server, IBM offers the PowerVM Standard Edition. This edition provides the most complete virtualization functionality for UNIX and Linux in the industry, and is available for all IBM Power Systems servers.

With PowerVM Standard Edition, you can create up to 254 partitions on a server. You can use virtualized disk and optical devices, and try out the shared processor pool. All virtualization features can be managed by using a Hardware Management Console or the IVM. These features include Micro-Partitioning, shared processor pool, Virtual I/O Server, PowerVM Lx86, shared dedicated capacity, NPIV, and virtual tape.

PowerVM Enterprise Edition

PowerVM Enterprise Edition is offered exclusively on IBM POWER6® and IBM POWER7® servers. It includes all the features of the PowerVM Standard Edition, plus the *PowerVM Live Partition Mobility* capability.

With PowerVM Live Partition Mobility, you can move a running partition from one POWER6 or POWER7 technology-based server to another with no application downtime. This capability results in better system utilization, improved application availability, and energy savings. With PowerVM Live Partition Mobility, planned application downtime because of regular server maintenance is no longer necessary.

8.1.2 Virtual I/O Server

Virtual I/O Server (VIOS) is virtualization software that runs in a separate partition of the POWER system. VIOS provides virtual storage and networking resources to one or more client partitions.

VIOS owns physical I/O resources such as Ethernet and SCSI/FC adapters. It virtualizes those resources for its client LPARs to share them remotely using the built-in hypervisor services. These client LPARs can be created quickly, and typically own only real memory and shares of processors without any physical disks or physical Ethernet adapters.

With Virtual SCSI support, VIOS client partitions can share disk storage that is physically assigned to the VIOS LPAR. This virtual SCSI support of VIOS can be used with storage devices that do not support the IBM i proprietary 520-byte/sectors format available to IBM i clients of VIOS. These storage devices include IBM XIV Storage System server.

VIOS owns the physical adapters, such as the Fibre Channel storage adapters, that are connected to the XIV system. The logical unit numbers (LUNs) of the physical storage devices that are detected by VIOS are mapped to VIOS virtual SCSI (VSCSI) server adapters. The VSCSI adapters are created as part of its partition profile.

The client partition connects to the VIOS VSCSI server adapters by using the hypervisor. The corresponding VSCSI client is adapters that are defined in its partition profile. VIOS runs SCSI emulation and acts as the SCSI target for the IBM i operating system.

Figure 8-1 shows an example of the VIOS owning the physical disk devices, and their virtual SCSI connections to two client partitions.

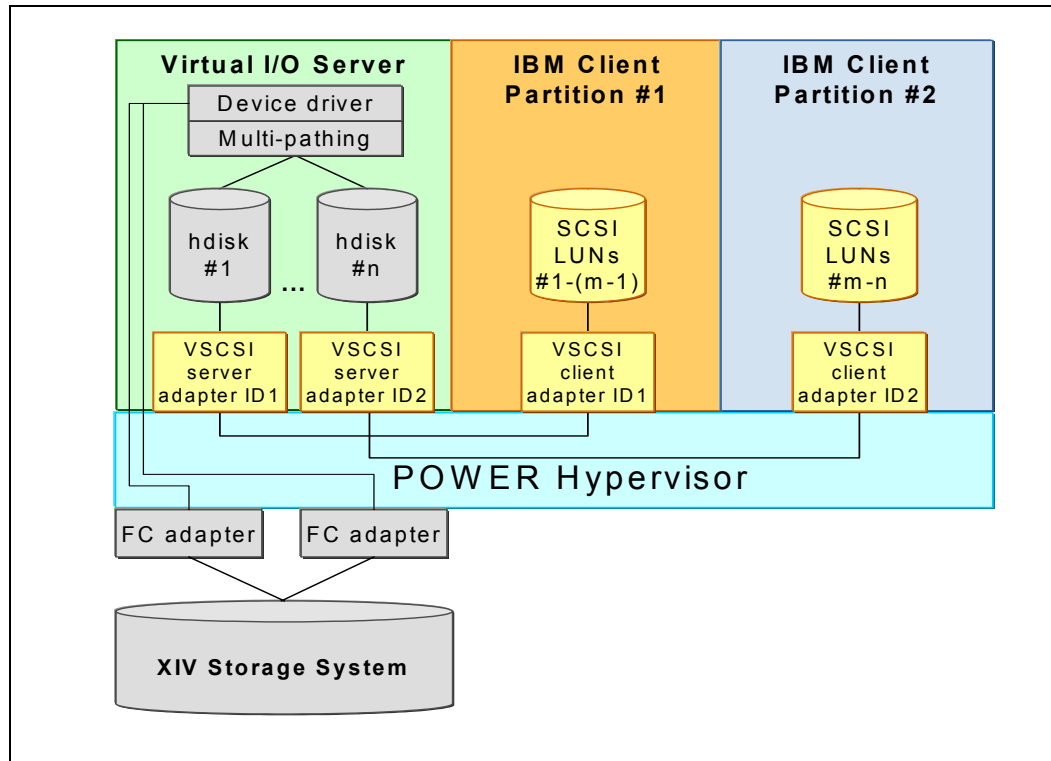


Figure 8-1 VIOS virtual SCSI support

8.1.3 Node Port ID Virtualization

The VIOS technology has been enhanced to boost the flexibility of IBM Power Systems servers with support for NPIV. NPIV simplifies the management and improves performance of Fibre Channel SAN environments. It does so by standardizing a method for Fibre Channel ports to virtualize a physical node port ID into multiple virtual node port IDs. The VIOS takes advantage of this feature, and can export the virtual node port IDs to multiple virtual clients. The virtual clients see this node port ID and can discover devices as though the physical port was attached to the virtual client.

The VIOS does not do any device discovery on ports that use NPIV. Therefore, no devices are shown in the VIOS connected to NPIV adapters. The discovery is left for the virtual client, and all the devices that are found during discovery are detected only by the virtual client. This way, the virtual client can use FC SAN storage-specific multipathing software on the client to discover and manage devices.

For more information about PowerVM virtualization management, see *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590.

Restriction: Connection through VIOS NPIV to an IBM i client is possible only for storage devices that can attach natively to the IBM i operating system. These devices include the IBM System Storage DS8000 and DS5000. To connect to other storage devices, such as XIV Storage Systems, use VIOS with virtual SCSI adapters.

8.2 Planning for VIOS and IBM i

The XIV system can be connected to an IBM i partition through VIOS. PowerVM and VIOS themselves are supported on the POWER5, POWER6, and IBM POWER7 Systems™. However, IBM i, being a client of VIOS, is supported only on POWER6 and POWER7 Systems.

Important: The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, see the SSIC at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

You can find Host Attachment publications at:

<http://www.ibm.com/support/fixcentral/>

8.2.1 Requirements

The following are general requirements, current at the time of writing, to fulfill when you attach an XIV Storage System to an IBM i VIOS client (Table 8-1).

These requirements serve as an orientation to the required hardware and software levels for XIV Storage System with IBM i. For current information, see the SSIC at:

<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

Table 8-1 IBM i and VIOS requirements for XIV attachment

XIV attach	Server	VIOS level	i 6.1	i 7.1
VIOS VSCSI	POWER7	2.2 or later	yes (i 6.1.1)	yes
VIOS VSCSI	Blade servers that are based on POWER7 and IBM BladeCenter® Chassis H	2.2 or later	yes (i 6.1.1)	yes
VIOS VSCSI	IBM POWER6+™	2.1.1 or later	yes	yes
VIOS VSCSI	Blade servers that are based on POWER6 and BladeCenter Chassis H	2.1.1 or later	yes	yes
VIOS VSCSI	POWER6	2.1.1 or later	yes	yes

The following websites provide up-to-date information about the environments that are used when connecting XIV Storage System to IBM i:

1. IBM System i® storage solutions
<http://www.ibm.com/systems/i/hardware/storage/index.html>
2. Virtualization with IBM i, PowerVM, and Power Systems
<http://www.ibm.com/systems/i/os/>
3. IBM Power Systems Hardware information center
http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/iphdx/550_m50_landing.htm
4. IBM Power Blade servers
<http://www.ibm.com/systems/power/hardware/blades/index.html>
5. IBM i and System i Information Center
<http://publib.boulder.ibm.com/iseries/>
6. IBM Support portal
<http://www.ibm.com/support/entry/portal/>
7. System Storage Interoperation Center (SSIC)
<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>

8.2.2 Supported SAN switches

For the list of supported SAN switches when connecting the XIV Storage System to the IBM i operating system, see the SSIC at:

http://www.ibm.com/systems/support/storage/config/ssic/displayesssearchwithoutjs.wss?start_over=yes

8.2.3 Physical Fibre Channel adapters and virtual SCSI adapters

You can connect up to 4,095 LUNs per target, and up to 510 targets per port on a VIOS physical FC adapter. You can assign up to 16 LUNs to one VSCSI adapter. Therefore, you can use the number of LUNs to determine the number of virtual adapters that you need.

Restriction: When the IBM i operating system and VIOS are on an IBM Power Architecture blade server, you can define only one VSCSI adapter to assign to an IBM i client. Therefore, the number of LUNs that can connect to the IBM i operating system is limited to 16.

8.2.4 Queue depth in the IBM i operating system and Virtual I/O Server

When you connect the IBM XIV Storage System server to an IBM i client through the VIOS, consider the following types of queue depths:

- ▶ The IBM i queue depth to a virtual LUN
The SCSI command tag queuing in the IBM i operating system allows up to 32 I/O operations to one LUN at the same time.

- ▶ The queue depth per physical disk (hdisk) in the VIOS
This queue depth indicates the maximum number of I/O requests that can be outstanding on a physical disk in the VIOS at a time.
- ▶ The queue depth per physical adapter in the VIOS
This queue depth indicates the maximum number of I/O requests that can be outstanding on a physical adapter in the VIOS at the same time.

The IBM i operating system has a fixed queue depth of 32, which is *not* changeable. However, the queue depths in the VIOS can be set up by a user. The default setting in the VIOS varies based on these factors:

- ▶ Type of connected storage
- ▶ Type of physical adapter
- ▶ Type of multipath driver or Host Attachment Kit

The XIV Storage System typically has the following characteristics:

- ▶ The queue depth per physical disk is 40
- ▶ The queue depth per 4-Gbps FC adapter is 200
- ▶ The queue depth per 8-Gbps FC adapter is 500

Check the queue depth on physical disks by entering the following VIOS command:

```
lsdev -dev hdiskxx -attr queue_depth
```

If needed, set the queue depth to 32 using the following command:

```
chdev -dev hdiskxx -attr queue_depth=32
```

This last command ensures that the queue depth in the VIOS matches the IBM i queue depth for an XIV LUN.

8.2.5 Multipath with two Virtual I/O Servers

The IBM XIV Storage System server is connected to an IBM i client partition through the VIOS. For redundancy, connect the XIV Storage System to an IBM i client with two or more VIOS partitions. Assign one VSCSI adapter in the IBM i operating system to a VSCSI adapter in each VIOS. The IBM i operating system then establishes multipath to an XIV LUN, with each path using one separate VIOS. For XIV attachment to VIOS, the VIOS integrated native MPIO multipath driver is used. Up to eight VIOS partitions can be used in such a multipath connection. However, most installations use multipath by using two VIOS partitions.

For more information, see 8.3.3, “IBM i multipath capability with two Virtual I/O Servers” on page 222.

8.2.6 General guidelines

This section presents general guidelines for IBM XIV Storage System servers that are connected to a host server. These practices also apply to the IBM i operating system.

With the grid architecture and massive parallelism inherent to XIV system, the general approach is to always maximize the use of all XIV resources.

Distributing connectivity

The goal for host connectivity is to create a balance of the resources in the IBM XIV Storage System server. Balance is achieved by distributing the physical connections across the

interface modules. A host usually manages multiple physical connections to the storage device for redundancy purposes by using a SAN connected switch. The ideal is to distribute these connections across each of the interface modules. This way, the host uses the full resources of each module to which it connects for maximum performance.

You do not need to connect each host instance to each interface module. However, when the host has more than one physical connection, have the connections (cabling) spread across separate interface modules.

Similarly, if multiple hosts have multiple connections, you must distribute the connections evenly across the interface modules.

Zoning SAN switches

To maximize balancing and distribution of host connections to an IBM XIV Storage System server, create a zone for the SAN switches. In this zone, have each host adapter connect to each XIV interface module and through each SAN switch. For more information, see 1.2.2, “Fibre Channel configurations” on page 15 and 1.2.3, “Zoning” on page 19.

Appropriate zoning: Use a separate zone for each host adapter (initiator). For each zone that contains the host adapter, add all switch port connections from the XIV Storage System (targets).

Queue depth

SCSI command tag queuing for LUNs on the IBM XIV Storage System server allows multiple I/O operations to one LUN at the same time. The LUN queue depth indicates the number of I/O operations that can be done simultaneously to a LUN.

The XIV architecture eliminates the storage concept of a large central cache. Instead, each module in the XIV grid has its own dedicated cache. The XIV algorithms that stage data between disk and cache work most efficiently when multiple I/O requests are coming in parallel. This process is where the host queue depth becomes an important factor in maximizing XIV I/O performance. Therefore, configure the host HBA queue depths as large as possible.

Number of application threads

The overall design of the IBM XIV Storage System grid architecture excels with applications that employ threads to handle the parallel execution of I/O. The multi-threaded applications profit the most from XIV performance.

8.3 Connecting an PowerVM IBM i client to XIV

The XIV system can be connected to an IBM i partition through the VIOS. This section explains how to set up a POWER6 system to connect the XIV Storage System to an IBM i client with multipath through two VIOS partitions. Setting up a POWER7 system to an XIV Storage System is similar.

8.3.1 Creating the Virtual I/O Server and IBM i partitions

This section describes the steps to complete the following tasks

- ▶ Create a VIOS partition and an IBM i partition through the POWER6 HMC
- ▶ Create VSCSI adapters in the VIOS and the IBM i partition
- ▶ Assign VSCSI adapters so that the IBM i partition can work as a client of the VIOS

For more information, see 6.2.1, “Creating the VIOS LPAR,” and 6.2.2, “Creating the IBM i LPAR,” in *IBM i and Midrange External Storage*, SG24-7668.

Creating a Virtual I/O Server partition in a POWER6 server

To create a POWER6 LPAR for VIOS, complete these steps:

1. Insert the PowerVM activation code in the HMC. Click **Tasks** → **Capacity on Demand (CoD)** → **Advanced POWER Virtualization** → **Enter Activation Code**.
2. Create the partition by clicking **Systems Management** → **Servers**.
3. Select the server to use for creating the VIOS partition, and click **Tasks** → **Configuration** → **Create Logical Partition** → **VIO Server** (Figure 8-2).

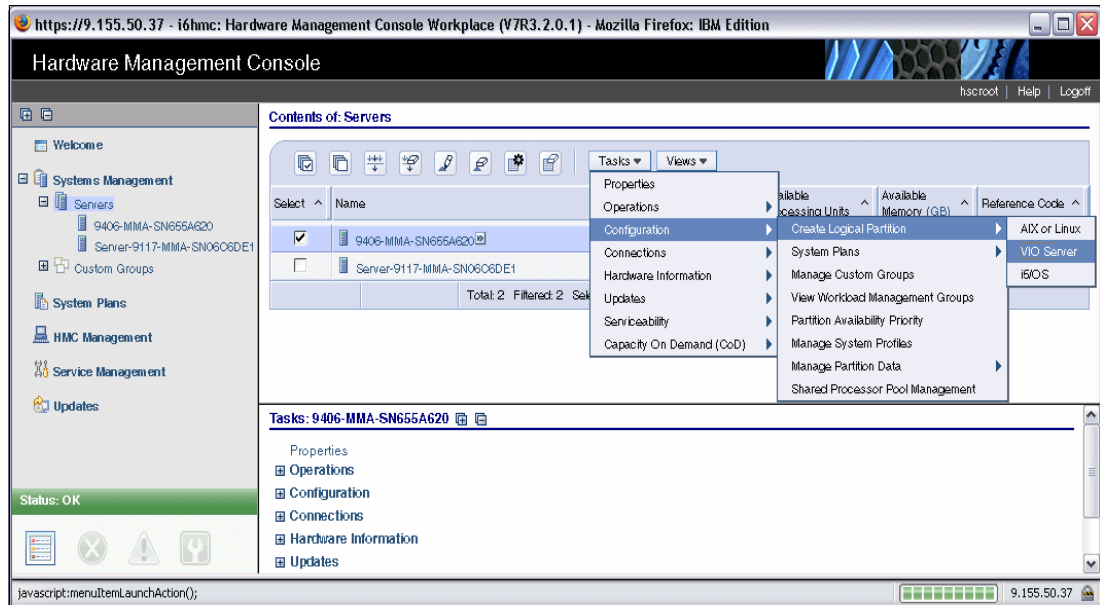


Figure 8-2 Creating the VIOS partition

4. In the Create LPAR wizard:
 - a. Enter the partition ID and name.
 - b. Enter the partition profile name.
 - c. Select whether the processors in the LPAR are dedicated or shared. Whenever possible with your environment, select **Dedicated**.
 - d. Select the minimum, wanted, and maximum number of processors for the partition.
 - e. Select the minimum, wanted, and maximum amount of memory in the partition.

- In the I/O window, select the I/O devices to include in the new LPAR. In this example, the RAID controller is included to attach the internal SAS drive for the VIOS boot disk and DVD_RAM drive. The physical Fibre Channel adapters are included to connect to the XIV server. As shown in Figure 8-3, they are added as **Required**.

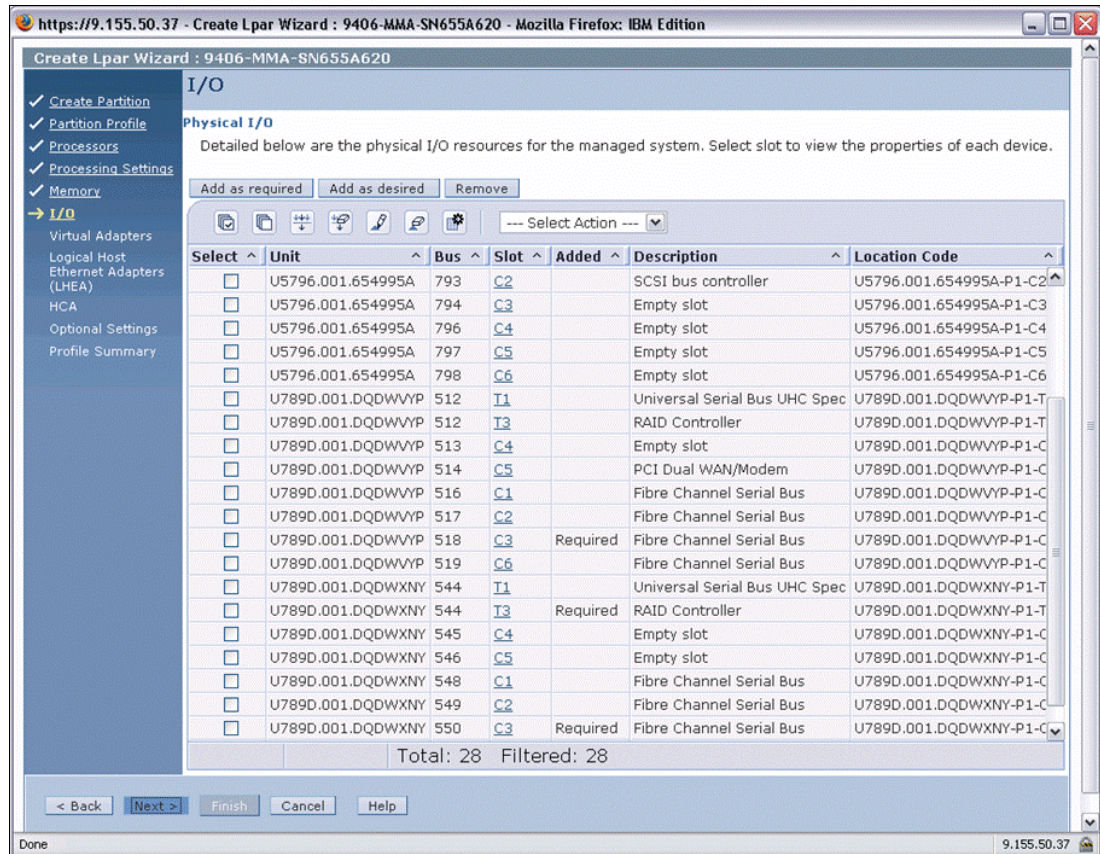


Figure 8-3 Adding the I/O devices to the VIOS partition

- In the Virtual Adapters window, create an Ethernet adapter by clicking **Actions** → **Create** → **Ethernet Adapter**. Mark it as **Required**.
- Create the VSCSI adapters to assign to the virtual adapters in the IBM i client:
 - Click **Actions** → **Create** → **SCSI Adapter**.
 - In the next window, either leave the **Any Client partition can connect** selected or limit the adapter to a particular client.

If DVD-RAM is virtualized to the IBM i client, you might want to create another VSCSI adapter for DVD-RAM.
- Configure the logical host Ethernet adapter:
 - Select the logical host Ethernet adapter from the list.
 - Click **Configure**.
 - Verify that the selected logical host Ethernet adapter is not selected by any other partitions, and select **Allow all VLAN IDs**.
- In the Profile Summary window, review the information, and click **Finish** to create the LPAR.

Creating an IBM i partition in the POWER6 processor-based server

To create an IBM i partition to be the VIOS client, complete these steps:

1. From the HMC, click **Systems Management** → **Servers**.
2. In the right pane, select the server in which you want to create the partition. Then, click **Tasks** → **Configuration** → **Create Logical Partition** → **i5/OS™**.
3. In the Create LPAR wizard:
 - a. Enter the Partition ID and name.
 - b. Enter the partition Profile name.
 - c. Select whether the processors in the LPAR are dedicated or shared. Whenever possible in your environment, select **Dedicated**.
 - d. Select the minimum, wanted, and maximum number of processors for the partition.
 - e. Select the minimum, wanted, and maximum amount of memory in the partition.
 - f. In the I/O window, if the IBM i client partition is not supposed to own any physical I/O hardware, click **Next**.
4. In the Virtual Adapters window, click **Actions** → **Create** → **Ethernet Adapter** to create a virtual Ethernet adapter.
5. In the Create Virtual Ethernet Adapter window, accept the suggested adapter ID and the VLAN ID. Select **This adapter is required for partition activation** and click **OK**.
6. In the Virtual Adapters window, click **Actions** → **Create** → **SCSI Adapter**. This sequence creates the VSCSI client adapters on the IBM i client partition that is used for connecting to the corresponding VIOS.
7. For the VSCSI client adapter ID, specify the ID of the adapter:
 - a. For the type of adapter, select **Client**.
 - b. Select **Mark this adapter is required for partition activation**.
 - c. Select the VIOS partition for the IBM i client.
 - d. Enter the server adapter ID to which you want to connect the client adapter.
 - e. Click **OK**.

If necessary, repeat this step to create another VSCSI client adapter. Use the second adapter to connect to the VIOS VSCSI server adapter that is used for virtualizing the DVD-RAM.
8. Configure the logical host Ethernet adapter:
 - a. Select the logical host Ethernet adapter from the menu and click **Configure**.
 - b. In the next window, ensure that no other partitions have selected the adapter, and select **Allow all VLAN IDs**.
9. In the OptiConnect Settings window, if OptiConnect is not used in IBM i, click **Next**.
10. If the connected XIV system is used to boot from a storage area network (SAN), select the virtual adapter that connects to the VIOS.

Tip: The IBM i Load Source device is on an XIV volume.

11. In the Alternate Restart Device window, if the virtual DVD-RAM device is used in the IBM i client, select the corresponding virtual adapter.
12. In the Console Selection window, select the default of HMC for the console device and click **OK**.

13. Depending on the planned configuration, click **Next** in the three windows that follow until you reach the Profile Summary window.
14. In the Profile Summary window, check the specified configuration and click **Finish** to create the IBM i LPAR.

8.3.2 Installing the Virtual I/O Server

For more information about how to install the VIOS in a partition of the POWER6 processor-based server, see *IBM i and Midrange External Storage, SG24-7668*.

Using LVM mirroring for the Virtual I/O Server

Set up LVM mirroring to mirror the VIOS root volume group (rootvg). The example is mirrored across two RAID0 arrays (hdisk0 and hdisk1) to help protect the VIOS from POWER6 server internal SAS disk drive failures.

Configuring Virtual I/O Server network connectivity

To set up network connectivity in the VIOS, complete these steps:

1. Log in to the HMC terminal window as padmin, and enter the following command:

```
lsdev -type adapter | grep ent
```

Look for the logical host Ethernet adapter resources. In this example, it is ent1 as shown in Figure 8-4.

```
$ lsdev -type adapter | grep ent
ent0          Available   Virtual I/O Ethernet Adapter (1-lan)
ent1        Available   Logical Host Ethernet Port (1p-hea)
```

Figure 8-4 Available logical host Ethernet port

2. Configure TCP/IP for the logical Ethernet adapter entX by using the **mktcpip** command. Specify the corresponding interface resource enX.
3. Verify the created TCP/IP connection by pinging the IP address that you specified in the **mktcpip** command.

Upgrading the Virtual I/O Server to the latest fix pack

As the last step of the installation, upgrade the VIOS to the latest fix pack.

8.3.3 IBM i multipath capability with two Virtual I/O Servers

The IBM i operating system provides multipath capability, allowing access to an XIV volume (LUN) through multiple connections. One path is established through each connection. Up to eight paths to the same LUN or set of LUNs are supported. Multipath provides redundancy in case a connection fails, and it increases performance by using all available paths for I/O operations to the LUNs.

With Virtual I/O Server release 2.1.2 or later, and IBM i release 6.1.1 or later, you can establish multipath to a set of LUNs. Each path uses a connection through a separate VIOS. This topology provides redundancy if a connection or the VIOS fails. Up to eight multipath connections can be implemented to the same set of LUNs, each through a separate VIOS. However, most IT centers establish no more than two such connections.

8.3.4 Virtual SCSI adapters in multipath with two Virtual I/O Servers

In the example setup, two VIOS and two VSCSI adapters are used in the IBM i partition. Each adapter is assigned to a virtual adapter in one VIOS. The same XIV LUNs are connected to each VIOS through two physical FC adapters in the VIOS for multipath. They are also mapped to VSCSI adapters that serve the IBM i partition. This way, the IBM i partition sees the LUNs through two paths, each path using one VIOS. Figure 8-5 shows the configuration.

For testing purposes, separate switches were not used. Instead, separate blades in the same SAN Director were used. In a production environment, use separate switches for redundancy as shown in Figure 8-5.

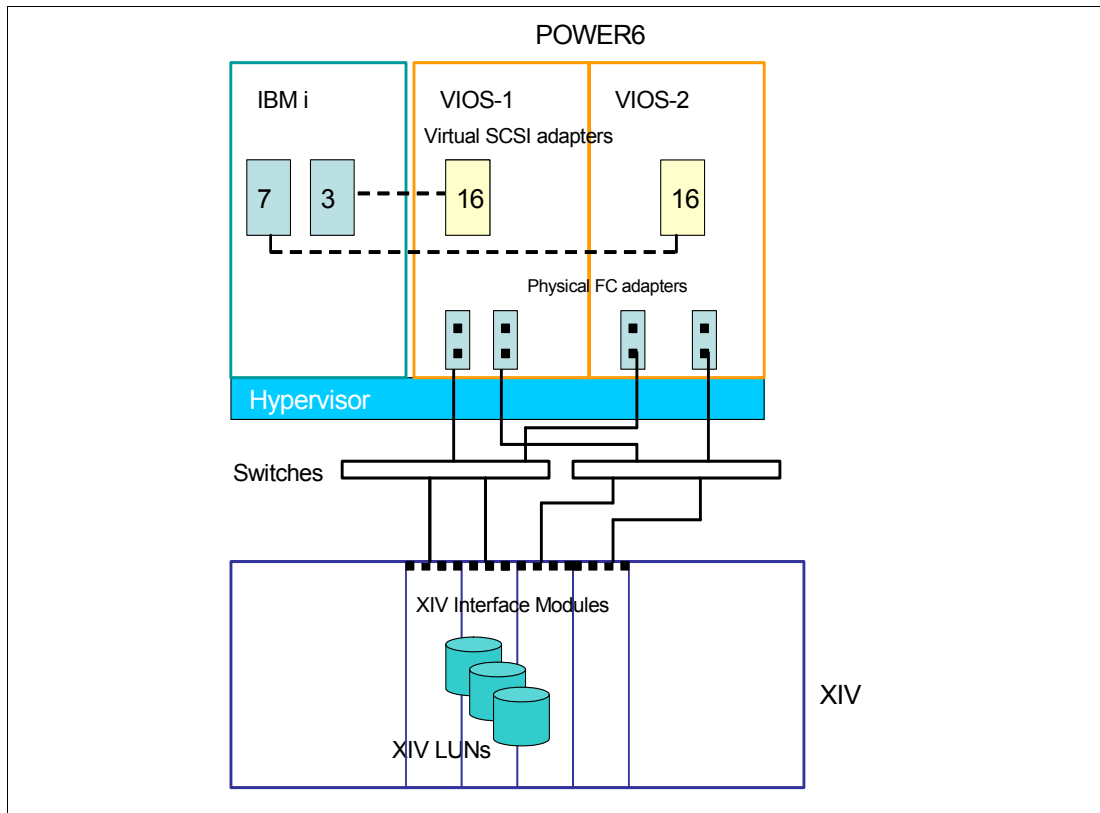


Figure 8-5 Setup for multipath with two VIOS

To connect XIV LUNs to an IBM i client partition in multipath with two VIOS, complete these steps:

Important: Perform steps 1 through 5 for each of the VIOS partitions.

1. After the LUNs are created in the XIV system, map the LUNs to the VIOS host as shown in 8.4, “Mapping XIV volumes in the Virtual I/O Server” on page 225. You can use the XIV Storage Management GUI or Extended Command Line Interface (XCLI).
2. Log on to VIOS as administrator. The example uses PuTTY to log in as described in 6.5, “Configuring VIOS virtual devices” in *IBM i and Midrange External Storage, SG24-7668*.
3. Run the `cfgdev` command so that the VIOS can recognize newly attached LUNs.

- In the VIOS, remove the SCSI reservation attribute from the LUNs (hdisks) that are to be connected through two VIOS. Enter the following command for each hdisk that connects to the IBM i operating system in multipath:

```
chdev -dev hdiskX -attr reserve_policy=no_reserve
```

- To get more bandwidth by using multiple paths, enter the following command for each hdisk (hdiskX):

```
chdev -dev hdiskX -perm -attr algorithm=round_robin
```

- Set the queue depth in VIOS for IBM i to 32 or higher. The default for XIV Storage Systems is 40 in VIOS. Higher values use more memory, so 40 is the usual value for AIX, Linux, and IBM i under VIOS.

- Verify the attributes by using the following command:

```
lsdev -dev hdiskX -attr
```

The command is illustrated in Figure 8-6.

```
$ lsdev -dev hdisk94 -attr
attribute      value                description
user_settable

PCM            PCM/friend/fcpothor Path Control Module      False
algorithm    round_robin        Algorithm                 True
clr_q          no                   Device CLEARS its Queue on error True
dist_err_pcmt 0                     Distributed Error Percentage True
dist_tw_width 50                    Distributed Error Sample Time True
hcheck_cmd     inquiry              Health Check Command      True
hcheck_interval 60                    Health Check Interval     True
hcheck_mode    nonactive             Health Check Mode         True
location       Location Label       Location Label            True
lun_id         0x1000000000000000 Logical Unit Number ID    False
lun_reset_spt  yes                  LUN Reset Supported       True
max_retry_delay 60                    Maximum Quiesce Time      True
max_transfer   0x40000              Maximum TRANSFER Size     True
node_name      0x5001738000cb0000 FC Node Name              False
pvid           none                  Physical volume identifier False
q_err          yes                   Use QERR bit              True
q_type         simple                Queuing TYPE              True
queue_depth  40                  Queue DEPTH               True
reassign_to    120                   REASSIGN time out value   True
reserve_policy no_reserve        Reserve Policy            True
rw_timeout     30                    READ/WRITE time out value True
scsi_id        0xa1400              SCSI ID                   False
start_timeout  60                    START unit time out value True
unique_id      261120017380000CB1797072810XIV03IBMfcp Unique device identifier  False
ww_name        0x5001738000cb0150 FC World Wide Name       False
$
```

Figure 8-6 'lsdev -dev hdiskX -attr' output

8. Map the disks that correspond to the XIV LUNs to the VSCSI adapters that are assigned to the IBM i client:
 - a. Check the IDs of assigned virtual adapters.
 - b. In the HMC, open the partition profile of the IBM i LPAR, click the Virtual Adapters tab, and observe the corresponding VSCSI adapters in the VIOS.
 - c. In the VIOS, look for the device name of the virtual adapter that is connected to the IBM i client. You can use the command `lsmmap -a11` to view the virtual adapters.
 - d. Map the disk devices to the SCSI virtual adapter that is assigned to the SCSI virtual adapter in the IBM i partition by entering the following command:

```
mkvdev -vdev hdiskxx -vadapter vhostx
```

Upon completing these steps, in each VIOS partition, the XIV LUNs report in the IBM i client partition by using two paths. The resource name of disk unit that represents the XIV LUN starts with DMPxxx, which indicates that the LUN is connected in multipath.

8.4 Mapping XIV volumes in the Virtual I/O Server

The XIV volumes must be added to both VIOS partitions. To make them available for the IBM i client, complete the following tasks in each VIOS:

1. Connect to the VIOS partition. The example uses a PuTTY session to connect.
2. In the VIOS, enter the `cfgdev` command to discover the newly added XIV LUNs. This command makes the LUNs available as disk devices (hdisks) in the VIOS. In the example, the LUNs added correspond to hdisk132 - hdisk140, as shown in Figure 8-7.

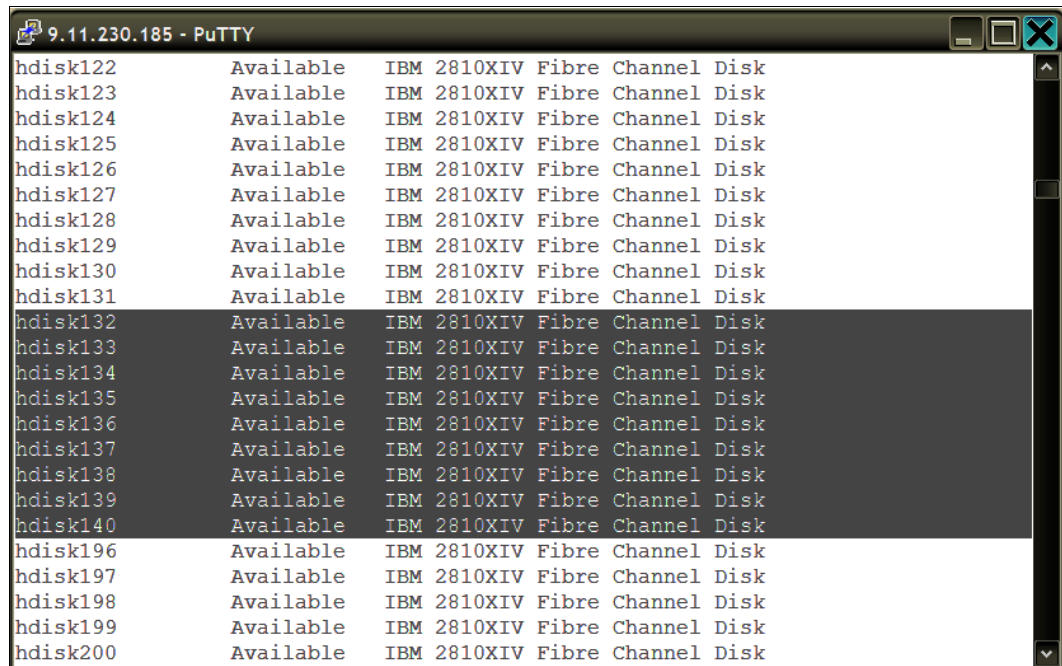


Figure 8-7 The hdisks of the added XIV volumes

For a multipath setup for IBM i, each XIV LUN is connected to both VIOS partitions. Before you assign these LUNs (from any of the VIOS partitions) to the IBM i client, make sure that the volume is not SCSI reserved.

3. Because a SCSI reservation is the default in the VIOS, change the reservation attribute of the LUNs to *non-reserved*. First, check the current reserve policy by entering the following command:

```
lsdev -dev hdiskx -attr reserve_policy
```

where *hdiskx* represents the XIV LUN.

If the reserve policy is not *no_reserve*, change it to *no_reserve* by entering the following command:

```
chdev -dev hdiskX -attr reserve_policy=no_reserve
```

4. Before mapping hdisks to a VSCSI adapter, check whether the adapter is assigned to the client VSCSI adapter in IBM i. Also, check whether any other devices are mapped to it.
 - a. Enter the following command to display the virtual slot of the adapter and see whether any other devices are assigned to it:

```
lsmmap -vadapter <name>
```

In the example setup, no other devices are assigned to the adapter, and the relevant slot is C16 as seen in Figure 8-8.

```

$ lsmmap -vadapter vhost5
SVSA          Physloc          C1
-----
vhost5        U9117.MMA.10AF384-V2-C16  02
VTD
                NO VIRTUAL TARGET DEVICE FOUND

```

Figure 8-8 Virtual SCSI adapter in the VIOS

- b. From the HMC, select the partition and click **Configuration** → **Manage Profiles**.
 - c. Select the profile and click **Actions** → **Edit**.

- d. In the partition profile, click the Virtual Adapters tab. Make sure that a client VSCSI adapter is assigned to the server adapter with the same ID as the virtual slot number. In the example, client adapter 3 is assigned to server adapter 16 (thus matching the virtual slot C16) as shown in Figure 8-9.

Type	Adapter ID	Connecting Partition	Connecting Adapter	Rec
Client SCSI	10	xivios-2(3)	19	No
Client SCSI	11	xivios-1(2)	15	Yes
Client SCSI	2	xivios-1(2)	11	Yes
Client SCSI	3	xivios-1(2)	16	Yes
Client SCSI	4	xivios-1(2)	17	No
Client SCSI	5	xivios-1(2)	18	No
Client SCSI	6	xivios-1(2)	19	No
Client SCSI	7	xivios-2(3)	16	Yes
Client SCSI	8	xivios-2(3)	17	No
Client SCSI	9	xivios-2(3)	18	No
Server Serial	0	Any Partition	Any Partition Slot	Yes

Figure 8-9 Assigned virtual adapters

5. Map the relevant hdisks to the VSCSI adapter by entering the following command:

```
mkvdev -vdev hdiskx -vadapter <name>
```

In this example, the XIV LUNs are mapped to the adapter vhost5. Each LUN is given a virtual device name by using the **-dev** parameter as shown in Figure 8-10.

```
$ mkvdev -vdev hdisk132 -vadapter vhost5 -dev vadamaboot132
vadamaboot132 Available
```

Figure 8-10 Mapping the LUNs in the VIOS

After you complete these steps for each VIOS, the LUNs are available to the IBM i client in multipath (one path through each VIOS).

8.5 Matching XIV volume to IBM i disk unit

To identify which IBM i disk unit is a particular XIV volume, complete these steps:

1. In VIOS, run the following commands to list the VIOS disk devices and their associated XIV volumes:
 - **eom_setup_env**: Initiates the OEM software installation and setup environment
 - **# XIV_devlist**: Lists the hdisks and corresponding XIV volumes
 - **# exit**: Returns to the VIOS prompt

The output of `XIV_devlist` command in one of the VIO servers in the example setup is shown in Figure 8-11. In this example, `hdisk5` corresponds to the XIV volumes `ITSO_i_1` with serial number 4353.

XIV Devices						
Device	Size	Paths	Vol Name	Vol Id	XIV Id	XIV Host
<code>/dev/hdisk5</code>	154.6GB	2/2	ITSO_i_1	4353	1300203	VIOS_1
<code>/dev/hdisk6</code>	154.6GB	2/2	ITSO_i_CG.snap _group_00001.I TSO_i_4	4497	1300203	VIOS_1
<code>/dev/hdisk7</code>	154.6GB	2/2	ITSO_i_3	4355	1300203	VIOS_1
<code>/dev/hdisk8</code>	154.6GB	2/2	ITSO_i_CG.snap _group_00001.I TSO_i_6	4499	1300203	VIOS_1
<code>/dev/hdisk9</code>	154.6GB	2/2	ITSO_i_5	4357	1300203	VIOS_1
<code>/dev/hdisk10</code>	154.6GB	2/2	ITSO_i_CG.snap _group_00001.I TSO_i_2	4495	1300203	VIOS_1
<code>/dev/hdisk11</code>	154.6GB	2/2	ITSO_i_7	4359	1300203	VIOS_1
<code>/dev/hdisk12</code>	154.6GB	2/2	ITSO_i_8	4360	1300203	VIOS_1

Figure 8-11 VIOS devices and matching XIV volumes

2. In VIOS, run **lsmmap -vadapter vhostx** for the virtual adapter that connects your disk devices to observe which virtual SCSI device corresponds to which hdisk. This process is illustrated in Figure 8-12.

```

$ lsmmap -vadapter vhost0
SVSA          Physloc                               Client Partition
ID
-----
vhost0        U9117.MMA.06C6DE1-V15-C20                    0x00000013

VTD           vtscsi0
Status        Available
LUN           0x8100000000000000
Backing device hdisk5
Physloc
U789D.001.DQD904G-P1-C1-T1-W5001738000CB0160-L1000000000000
Mirrored      false

VTD           vtscsi1
Status        Available
LUN           0x8200000000000000
Backing device hdisk6
Physloc
U789D.001.DQD904G-P1-C1-T1-W5001738000CB0160-L20000000000000
Mirrored      false

```

Figure 8-12 Hdisk to vscsi device mapping

3. For a particular virtual SCSI device, observe the corresponding LUN ID by using VIOS command **lsdev -dev vtscsi0 -vpd**. In the example, the virtual LUN id of device vtscsi0, is 1, as can be seen in Figure 8-13.

```

$
$ lsdev -dev vtscsi0 -vpd
vtscsi0      U9117.MMA.06C6DE1-V15-C20-L1  Virtual Target Device - Disk
$

```

Figure 8-13 LUN ID of a virtual SCSI device

4. In IBM i, use the command **STRSST** to start the use system service tools (SST). You need the SST user ID and password to sign in. After you are in SST, complete these steps:
 - a. Select **Option 3. Work with disk units** → **Option 1. Display disk configuration** → **Option 1. Display disk configuration status**.
 - b. In the Disk Configuration Status panel, press F9 to display disk unit details.

- c. In the Display Disk Unit Details panel, the columns Ct1 specifies which LUN ID belongs to which disk unit (Figure 8-14). In this example, the LUN ID 1 corresponds to IBM i disk unit 5 in ASP 1.

Display Disk Unit Details										
Type option, press Enter.										
5=Display hardware resource information details										
OPT	ASP	Unit	Serial Number	Sys Bus	Sys Card	Sys Board	I/O Adapter	I/O Bus	Ct1	Dev
	1	1	Y37DQDZREGE6	255	20	128		0	8	0
	1	2	Y33PKSV4ZE6A	255	21	128		0	7	0
	1	3	YQ2MN79SN934	255	21	128		0	3	0
	1	4	YGAZV3SLRQCM	255	21	128		0	5	0
	1	5	YS9NR8ZRT74M	255	21	128		0	1	0
33	4001		Y8NMB8T2W85D	255	21	128		0	2	0
33	4002		YH733AETK3YL	255	21	128		0	6	0
33	4003		YS7L4Z75EUEW	255	21	128		0	4	0

F3=Exit F9=Display disk units F12=Cancel

Figure 8-14 LUN ids of IBM i disk units

8.6 Performance considerations for IBM i with XIV

One purpose of experimenting with IBM i and XIV is to show the performance difference between using a few XIV volumes, and XIV volumes on IBM i.

During experimentation, the same capacity was always used. For some experiments, a few large volumes were used. For others, a larger number of smaller volumes were used. Specifically, a 6-TB capacity was used. The capacity was divided into 6 * 1-TB volumes, or into 42 * 154-GB volumes.

The experimentation is intended to show the performance improvement for an IBM i workload when it runs on XIV Gen 3 compared to an XIV generation 2 system. Tests with large and small numbers of LUNs was run on both XIV Gen 3 and XIV generation 2 systems. Both systems are equipped with 15 modules.

Remember: The purpose of the tests is to show the difference in IBM i performance between using a few large LUNs and many smaller LUNs. They also compare IBM i performance between XIV Gen 3 and XIV generation 2 systems.

The goal is *not* to make an overall configuration and setup recommendation for XIV to handle a specific IBM i workload.

8.6.1 Testing environment

The testing environment used the following configuration:

- ▶ IBM POWER7 system, model 770.
- ▶ Two IBM i LPARs, named LPAR2 and LPAR3, each of them running with six processing units and 80 GB of memory.
- ▶ IBM i software level V7.R1 with cumulative PTF package C1116710 and Hyper group PTF SF99709 level 40 installed in each LPAR.
- ▶ Two Virtual IO servers in the POWER7 system.
- ▶ VIOS software level 2.2.0.11, Fix pack 24, service pack 01, was installed in each VIOS system.
- ▶ XIV Storage System generation 2 with 15 modules / 1-TB disk drives code level 10.2.4.
- ▶ XIV Gen 3 Storage system with 15 modules / 2-TB disk drives, code level 11.0.0.
- ▶ Each VIOS uses two ports in an 8-Gb Fibre Channel adapter to connect to XIV. Each port is connected to three interface modules in XIV Storage System.
- ▶ The XIV Storage System has six volumes of 1-TB size and 42 volumes of 154 GB defined. These volumes are assigned to both VIOS.
- ▶ Each VIOS has the XIV volumes that are mapped as follows:
 - 6 * 1-TB volumes to the LPAR2 using 2 virtual SCSI adapters, three volumes to each virtual SCSI adapter
 - 42 * 154 volumes to LPAR3 using 3 virtual SCSI adapters, 16 volumes to each virtual SCSI adapter

In each of these configurations, the number of LUNs is a multiple of six. For a fully configured XIV System with six Interface Modules, this configuration equally distributes the workload (I/O traffic) across the Interface Modules.

This environment is used to connect to both the XIV Storage System generation 2 and XIV Gen 3.

Remember: In all the tests except the test with combined double workload, the XIV is exclusively used by one IBM i LPAR. In other words, no other applications or server I/O are running on the XIV.

In the test with combined double workload, the XIV is used only by the two IBM i LPARs. Again, no other workloads are on the XIV.

The testing scenario is illustrated in Figure 8-15.

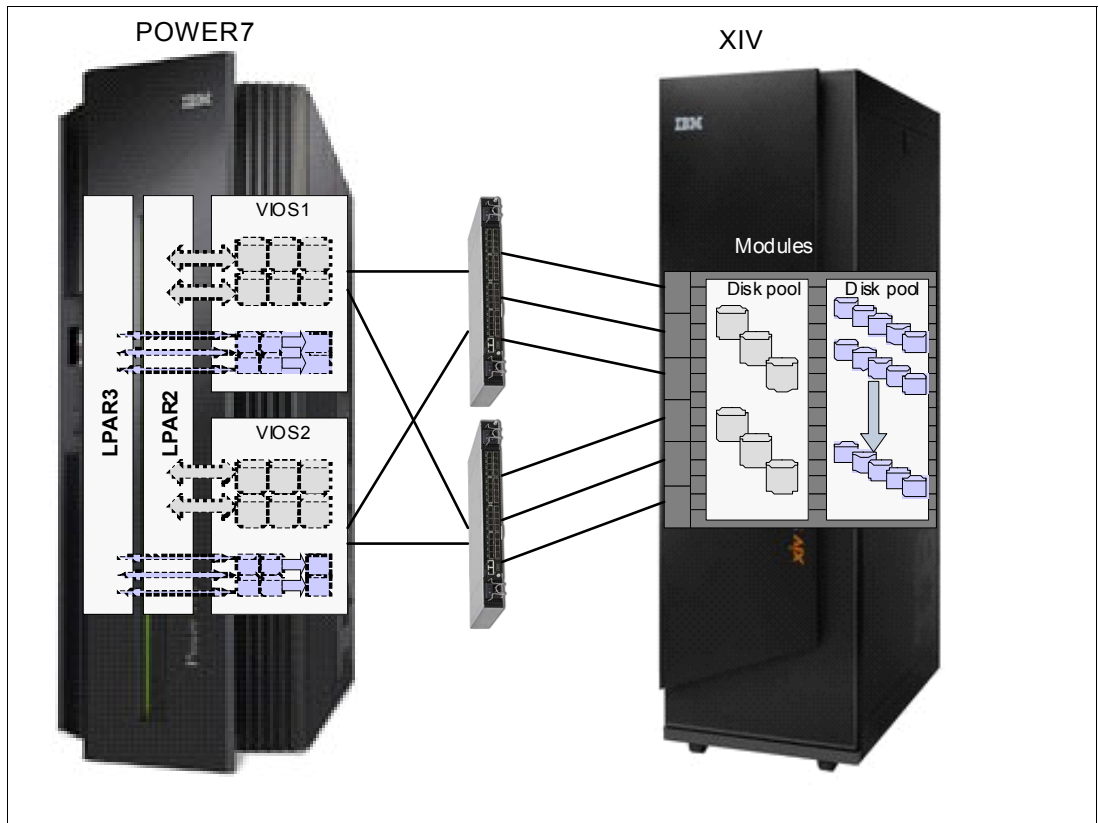


Figure 8-15 Testing environment

The IBM i LUNs defined on XIV Storage System for each LPAR are shown in Figure 8-16.

ITSO_Lpar2			
ITSO_Lpar2_1	vol		1
ITSO_Lpar2_2	vol		2
ITSO_Lpar2_3	vol		3
ITSO_Lpar2_4	vol		4
ITSO_Lpar2_5	vol		5
ITSO_Lpar2_6	vol		6

ITSO_Lpar3			
ITSO_Lpar3_01	vol		10
ITSO_Lpar3_02	vol		11
ITSO_Lpar3_03	vol		12
ITSO_Lpar3_04	vol		13
ITSO_Lpar3_05	vol		14
ITSO_Lpar3_06	vol		15
ITSO_Lpar3_07	vol		16
ITSO_Lpar3_08	vol		17
ITSO_Lpar3_09	vol		18
ITSO_Lpar3_10	vol		19
ITSO_Lpar3_11	vol		20
ITSO_Lpar3_33	vol		42
ITSO_Lpar3_34	vol		43
ITSO_Lpar3_35	vol		44
ITSO_Lpar3_36	vol		45
ITSO_Lpar3_37	vol		46
ITSO_Lpar3_38	vol		47
ITSO_Lpar3_39	vol		48
ITSO_Lpar3_40	vol		49
ITSO_Lpar3_41	vol		50
ITSO_Lpar3_42	vol		51

Figure 8-16 The LUNs for IBM i LPARs

Figure 8-17 shows the XIV volumes reporting in IBM i SST for the 6 * 1-TB LUNs configuration.

Display Disk Configuration Status							
ASP	Unit	Serial Number	Type	Model	Resource Name	Status	Hot Spare Protection
1	1	Y7WKQ2FQGW5N	6B22	050	DMP001	Configured	N
	2	Y7Y24LBTSUJJ	6B22	050	DMP026	Configured	N
	3	Y22QKZEEUB7B	6B22	050	DMP013	Configured	N
	4	YFVJ4STNADU5	6B22	050	DMP023	Configured	N
	5	YTXL2478XA3	6B22	050	DMP027	Configured	N
	6	YZLEQY7AB82C	6B22	050	DMP024	Configured	N

Figure 8-17 6 * 1-TB LUNs reporting in IBM i

Figure 8-18 shows the XIV volumes reporting in IBM i SST for the 42 * 154-GB LUNs configuration.

Display Disk Configuration Status						
ASP Unit	Serial Number	Type	Model	Resource Name	Status	Hot Spare Protection
1					Unprotected	
	1 Y9DY6HCARYRB	6B22	050	DMP001	Configured	N
	2 YR657TNBKKY4	6B22	050	DMP003	Configured	N
	3 YB9HSWBCJZRD	6B22	050	DMP006	Configured	N
	4 Y3U8YL3WVABW	6B22	050	DMP008	Configured	N
	5 Y58LXN6E3T8L	6B22	050	DMP010	Configured	N
	6 YUYBRDN3597T	6B22	050	DMP011	Configured	N
.....						
	35 YEES6NPSR6MJ	6B22	050	DMP050	Configured	N
	36 YP5QPYTA89DP	6B22	050	DMP051	Configured	N
	37 YNTD9ER85M4F	6B22	050	DMP076	Configured	N
	38 YGLUSQJXUMGP	6B22	050	DMP079	Configured	N
	39 Y6G7F38HSGQQ	6B22	050	DMP069	Configured	N
	40 YKGF2RZWDJXA	6B22	050	DMP078	Configured	N
	41 YG7PPW6KG58B	6B22	050	DMP074	Configured	N
	42 YP9P768LTLLM	6B22	050	DMP083	Configured	N

Figure 8-18 42 * 154-GB LUNs reporting in IBM i

8.6.2 Testing workload

The tests use the commercial processing workload (CPW). CPW is designed to evaluate a computer system and associated software in a commercial environment. It is maintained internally within the IBM i Systems Performance group.

The CPW application simulates the database server of an online transaction processing (OLTP) environment. These transactions are all handled by batch server jobs. They represent the type of transactions that might be done interactively in a client environment. Each of the transactions interacts with three to eight of the nine database files that are defined for the workload. Database functions and file sizes vary.

Functions that are exercised are single and multiple row retrieval, single and multiple row insert, single row update, single row delete, journal, and commitment control. These operations are run against files that vary from hundreds of rows to hundreds of millions of rows. Some files have multiple indexes, whereas some have only one. Some accesses are to the actual data and some take advantage of advanced functions such as index-only access.

CPW is considered a reasonable approximation of a steady-state, database-oriented commercial application.

After the workload is started, it generates the jobs in the CPW subsystems. Each job generates transactions. The CPW transactions are grouped by regions, warehouses, and users. Each region represents 1000 users or 100 warehouses: Each warehouse runs 10 users. CPW generates commercial types of transactions such as orders, payments, delivery, end stock level.

For the tests, the CPW is run with 96000 users, or 9600 warehouses. After you start the transaction workload, there is 50-minute delay, and a performance collection that lasts for one hour is started. After the performance collection is finished, several other IBM i analyzing tools, such as PEX, are run. At the end, the CPW database is restored. The entire CPW run lasts for about five hours.

8.6.3 Test with 154-GB volumes on XIV generation 2

The first test is with the 154-GB volumes defined on an XIV generation 2 system.

Table 8-2 shows the number of different transaction types, the percentage of each type of transaction, the average response time, and the maximal response time. The average response time for most of the transactions is between 0.03 and 0.04 seconds, and the maximum response time is 11.4 seconds.

Table 8-2 CPW transaction response times

Transaction ID	Count	Percentage	Average Resp. time (sec)	Maximal Resp. time (sec)
Neworder	7013230	44.33	0.038	2.210
Ordersts	648538	4.10	0.042	2.550
Payment	6846381	43.27	0.033	11.350
Delivery	658587	4.16	0.000	0.250
Stocklvl	655281	4.14	0.083	2.340

The I/O rate and the disk service time during the collection period are shown in Figure 8-19.

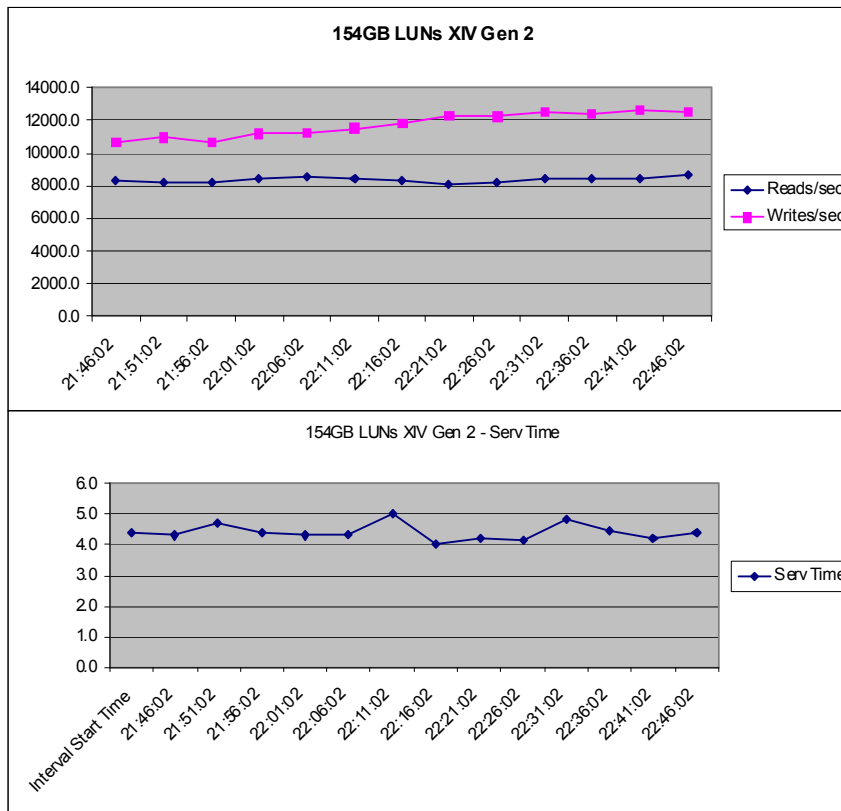


Figure 8-19 I/O rate and disk service time

During this collection period, CPW experienced an average of 8355 reads/sec and 11745 writes/sec. The average service time was 4.4 ms.

Tip: Because the reported disk *wait time* in IBM i collection services reports was 0 in all the tests, it is not included in the graphs.

The restore of the CPW database took 23 minutes.

Figure 8-20 on page 237, Figure 8-21 on page 238, and Figure 8-22 on page 239 show the following values that were reported in XIV:

- ▶ I/O rate
- ▶ Latency
- ▶ Bandwidth in MBps
- ▶ Read/write ratio
- ▶ Cache hits

Figure 8-20 shows these values during the whole CPW run.

Tip: For readers who cannot see the colors, the various data and scales are labeled in Figure 8-20. The other graphs are similar.

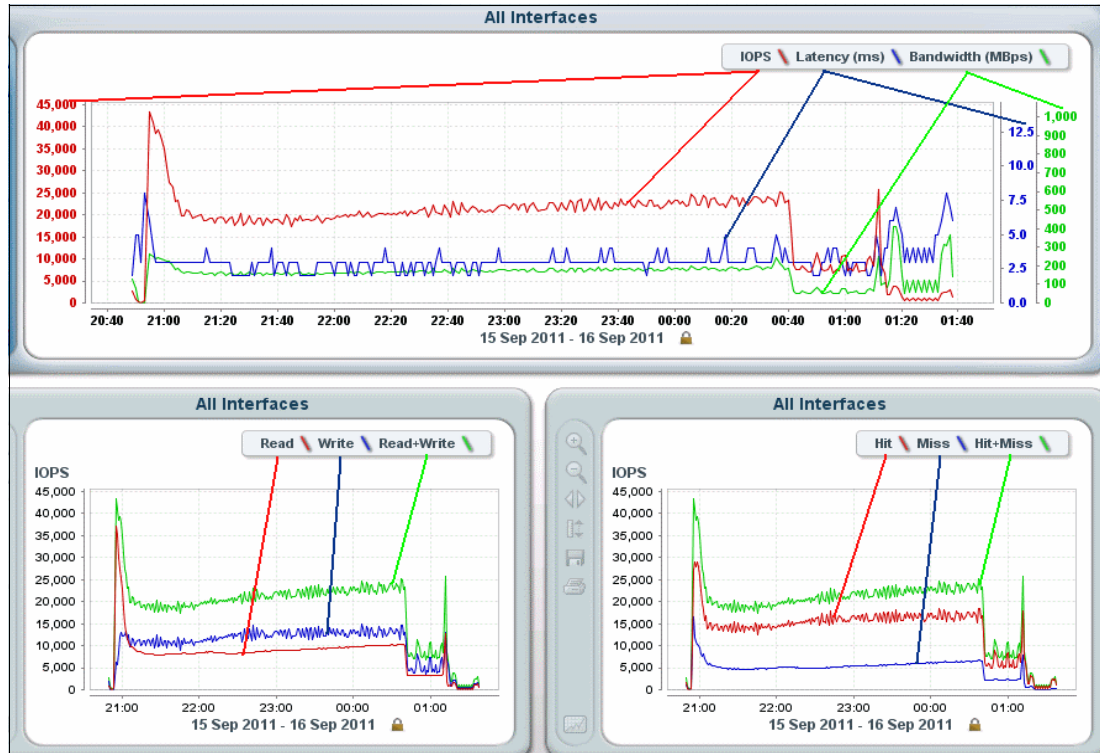


Figure 8-20 XIV values during the entire CPW run

Figure 8-21 shows the system during the IBM i collection period. The average latency during the collection period was 3 ms, and the cache hit percentage was 75%.

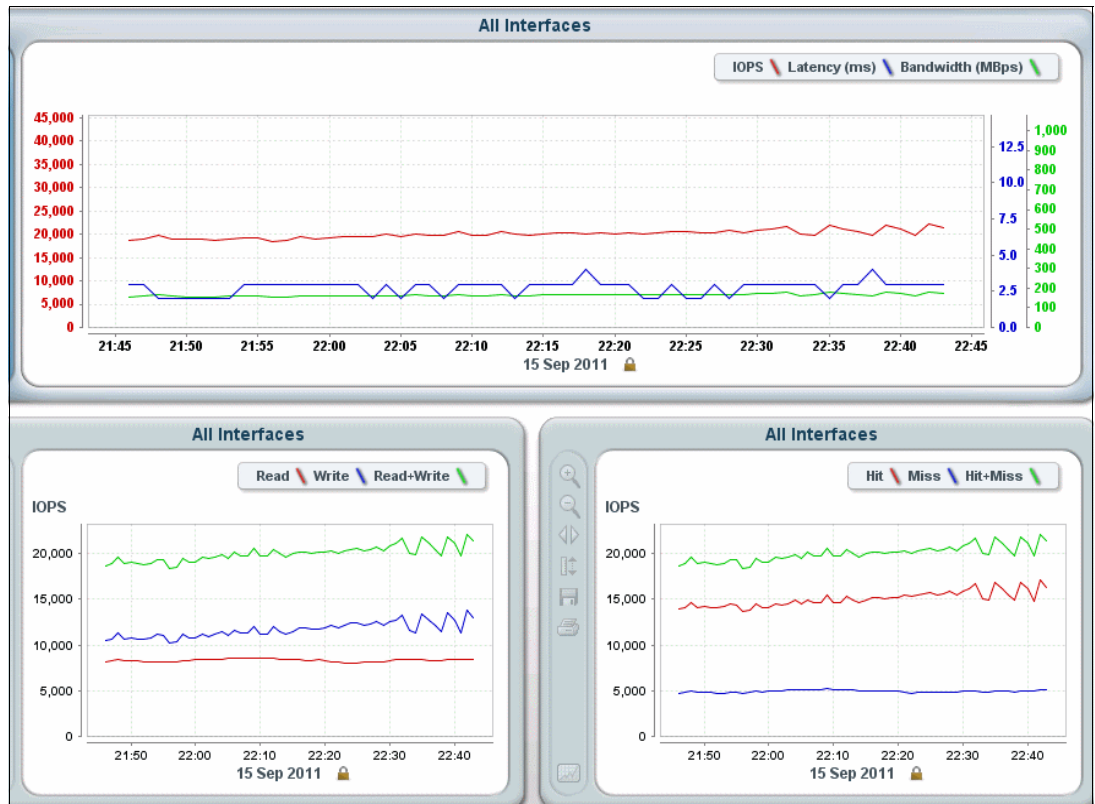


Figure 8-21 XIV values during the collection period

Figure 8-22 shows the system during the CPW database restore. The average latency was 6 ms, and the percentage of cache hits was 90% - 100%.



Figure 8-22 XIV values during restore of the database

8.6.4 Test with 1-TB volumes on XIV generation 2

The second test is with 1-TB volumes on the XIV generation 2 system.

Table 8-3 shows the number of different transaction types, the percentage of each type of transaction, the average response time, and the maximal response time. The average response time for most of the transactions is between 0.3 and 10 seconds. The maximal response time is 984.2 seconds.

Table 8-3 CPW transaction response times

Transaction ID	Count	Percentage	Average Resp. time (sec)	Maximal Resp. time (sec)
Neworder	3197534	46.21	10.219	984.170
Ordersts	271553	3.92	0.422	21.170
Payment	2900103	41.92	0.324	796.140
Delivery	275252	3.98	0.000	0.940
Stocklvl	274522	3.97	1.351	418.640

The I/O rate and the disk service time during the collection period is shown in the Figure 8-23. During this period, CPW experienced average 3949 reads/sec and 3907 writes/sec. The average service time was 12.6 ms.

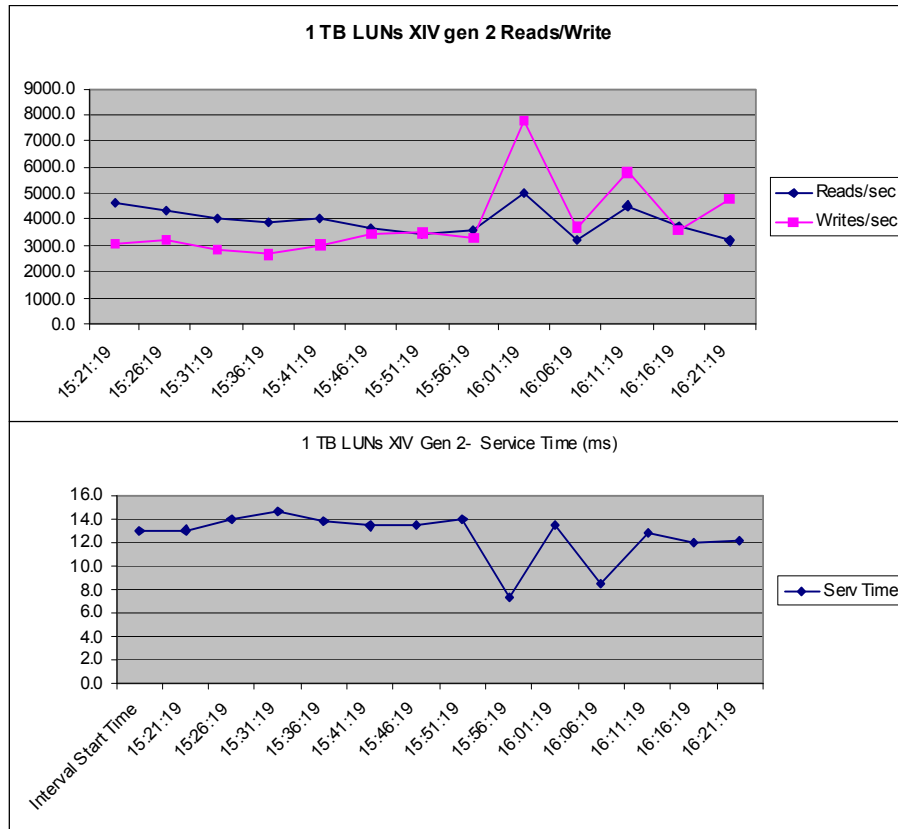


Figure 8-23 I/O rate and disk service time

The restore of CPW database took 24 minutes.

Figure 8-24 on page 241, Figure 8-25 on page 242, and Figure 8-26 on page 243 show the following values that were reported in XIV:

- ▶ I/O rate
- ▶ Latency
- ▶ Bandwidth in MBps
- ▶ Read/write ratio
- ▶ Cache hits

Figure 8-24 shows these values during the whole CPW run.



Figure 8-24 XIV values during entire CPW run

Figure 8-25 shows these values during the IBM i collection period. The average latency during the collection period was 20 ms, and the approximate percentage of cache hits was 50%.



Figure 8-25 XIV values during collection period

Figure 8-26 shows these values while restoring the CPW database. The average latency during the restore was 2.5 ms, and the approximate cache hit percentage was between 90% to 100%.

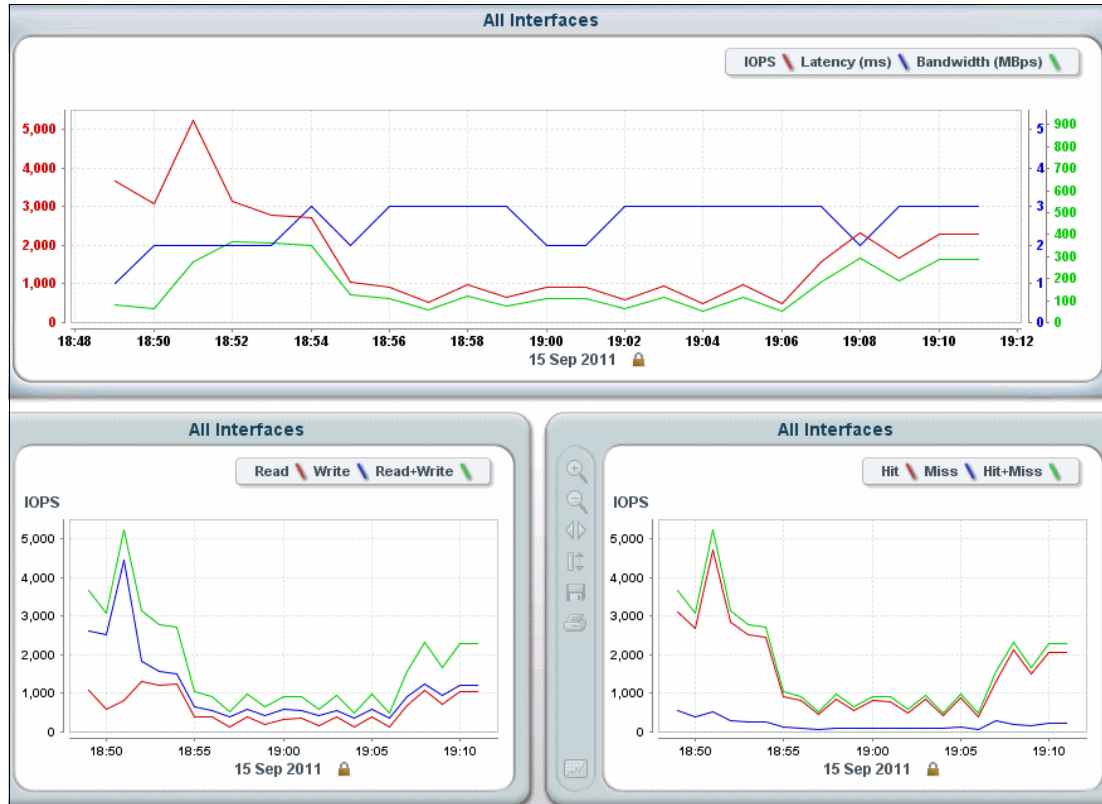


Figure 8-26 XIV values during restore of the database

8.6.5 Test with 154-GB volumes on XIV Gen 3

Table 8-4 shows the number of different transaction types, the percentage of each type of transaction, the average response time, and the maximal response time. The average response time for most of the transactions varies from 0.003 to 0.006 seconds. The maximal response time is 2.5 seconds.

Table 8-4 CPW transaction response time

Transaction ID	Count	Percentage	Average Resp. time (sec)	Maximal Resp. time (sec)
Neworder	7031508	44.32	0.006	0.540
Ordersts	650366	4.10	0.004	0.390
Payment	6864817	43.27	0.003	2.460
Delivery	660231	4.16	0.000	0.010
Stocklvl	656972	4.14	0.031	0.710

The disk service time response time during the collection period is shown in Figure 8-27. The average service time of the one hour collection period was 0.5 ms.

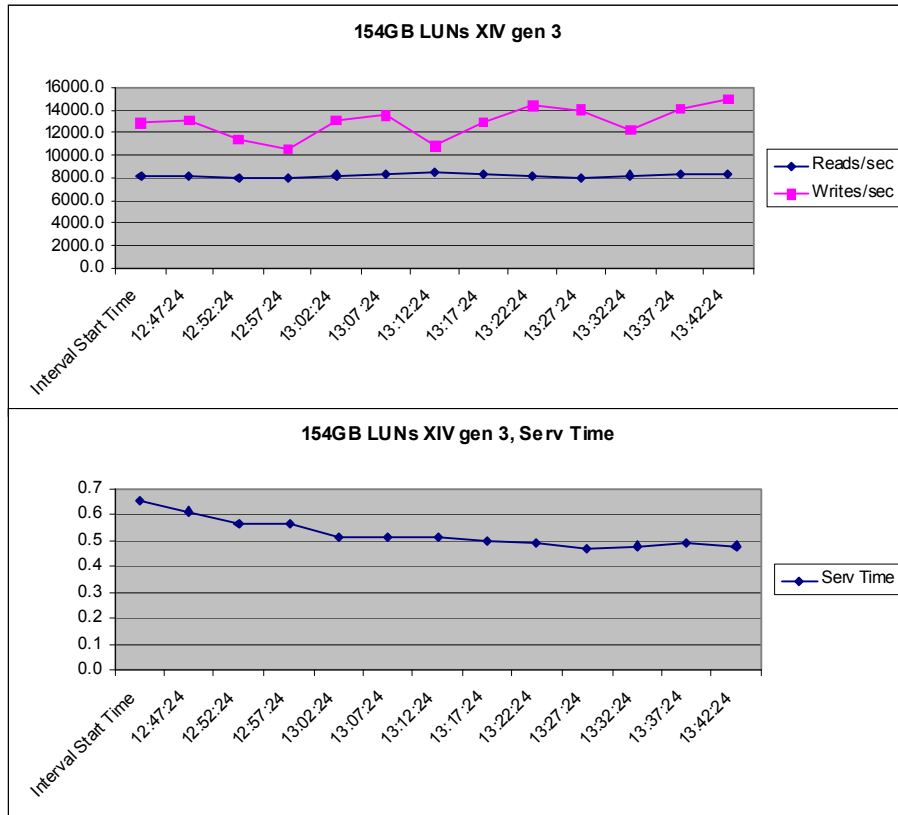


Figure 8-27 I/O rate and disk service times

Figure 8-28, Figure 8-29 on page 246, and Figure 8-30 on page 247 show the following values that were reported in XIV:

- ▶ I/O rate
- ▶ Latency
- ▶ Bandwidth in MBps
- ▶ Read/write ratio
- ▶ Cache hits

Figure 8-28 shows these values during the whole CPW run.



Figure 8-28 XIV values during the entire CPW run

Figure 8-29 shows these values during the IBM i collection period. The average latency during the collection period was close to 0. The average percentage of cache hits was close to 100%.



Figure 8-29 XIV values during collection period

Figure 8-30 shows these values during CPW database restore. The average latency was close to 0, and the percentage of cache was close to 100%.



Figure 8-30 XIV values during restore of the database

8.6.6 Test with 1-TB volumes on XIV Gen 3

Table 8-5 shows the number of different transaction types, the percentage of each type of transaction, the average response time, and the maximal response time. The average response time of most of the transactions varies from 0.003 seconds to 0.006 seconds. The maximal response time is 2.6 seconds.

Table 8-5 CPW transaction response times

Transaction ID	Count	Percentage	Average Resp. time (sec)	Maximal Resp. time (sec)
Neworder	7032182	44.33	0.006	0.390
Ordersts	650306	4.10	0.005	0.310
Payment	6864866	43.27	0.003	2.620
Delivery	660280	4.16	0.000	0.040
Stocklvl	657016	4.14	0.025	0.400

The disk service time during the collection period is shown in the Figure 8-31. The average service time for a one hour collection period was 0.5 ms.

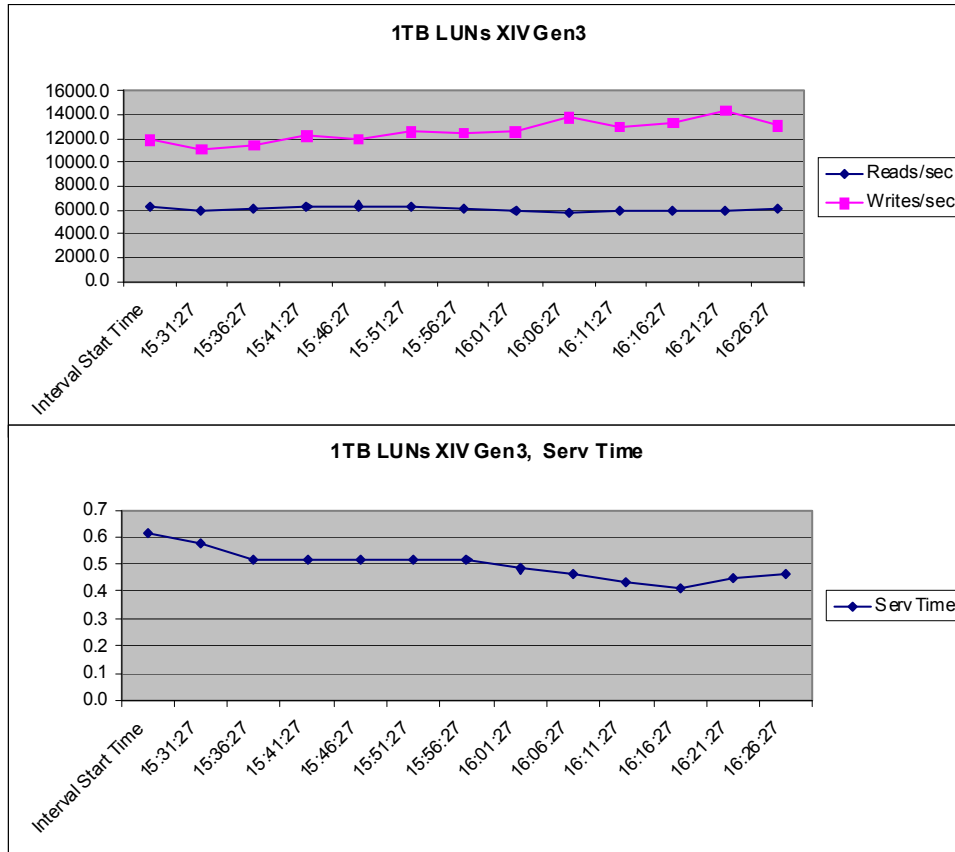


Figure 8-31 I/O rate and disk service time

Figure 8-32, Figure 8-33 on page 250, and Figure 8-34 on page 251 show the following values that were reported in XIV:

- ▶ I/O rate
- ▶ Latency
- ▶ Bandwidth in MBps
- ▶ Read/write ratio
- ▶ Cache hits

Figure 8-32 shows these values during the whole CPW run.

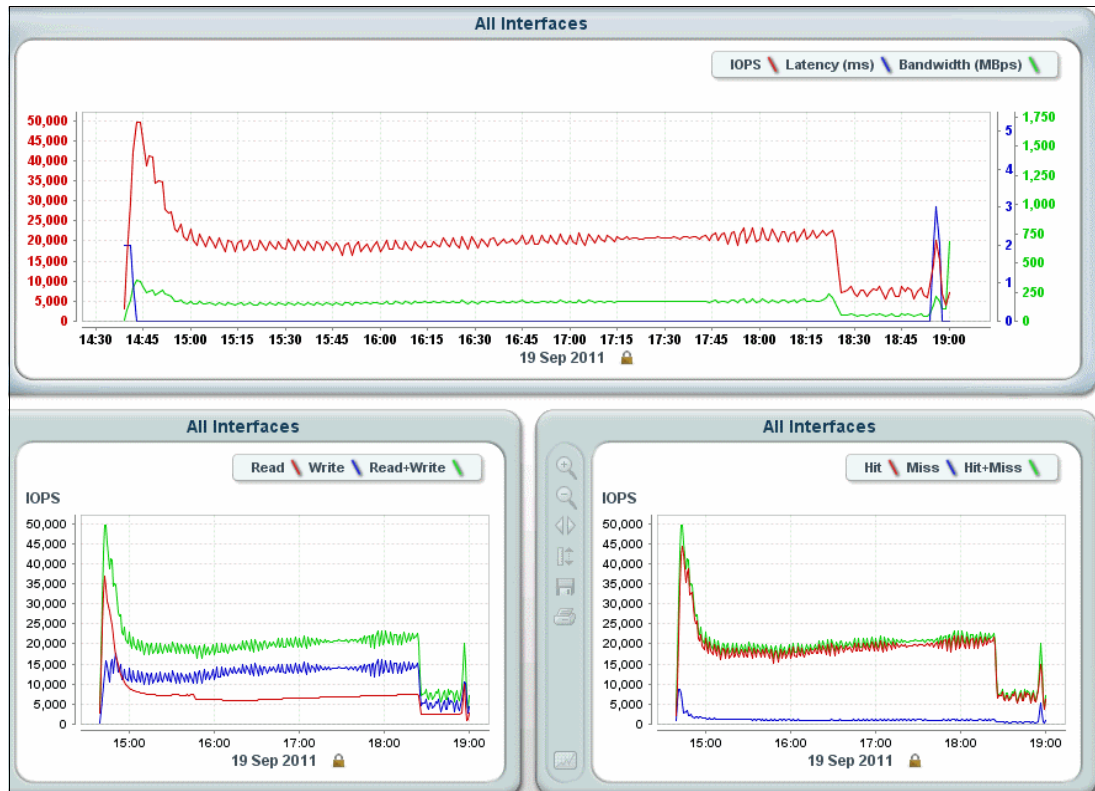


Figure 8-32 XIV values during the entire CPW run

Figure 8-33 shows these values during the IBM i collection period. The average latency during the collection period was 0.2 ms, and the cache hit percentage was almost 100%.



Figure 8-33 XIV values during collection period

Figure 8-34 shows these values during the CPW database restore. The latency during the database restore was close to 0 and the percentage of cache hits was close to 100%.



Figure 8-34 XIV values during restore of the database

8.6.7 Test with doubled workload on XIV Gen 3

In the tests that were run on XIV Storage System Gen 3, the workload experienced cache hits close to 100%. Response time did not differ between environments with 42 * 154-GB LUNs and 6 * 1-TB LUNs. To better show the performance difference between the two different LUN sizes on XIV Gen 3, the I/O rate on XIV was increased. The CPW was run with 192,000 users on each IBM i LPAR, and the workload was run in both LPARs at the same time.

Table 8-6 shows the number of different transaction types, the percentage of each type of transaction, the average response time, and the maximal response time. The average response time for most of the transactions varies from 0.6 to 42 seconds. The maximal response time is 321 seconds.

Table 8-6 1-TB LUNs

Transaction ID	Count	Percentage	Average Resp. time (sec)	Maximal Resp. time (sec)
Neworder	1423884	44.33	42.181	321.330
Ordersts	131548	4.10	0.733	30.150
Payment	1389705	43.27	0.558	38.550
Delivery	133612	4.16	0.000	0.150
Stocklvl	133113	4.14	9.560	44.920

The disk service time response time during the collection period is shown in Figure 8-35. The average disk service time of one hour collection period was 8.2 ms. The average LUN utilization was 91%.

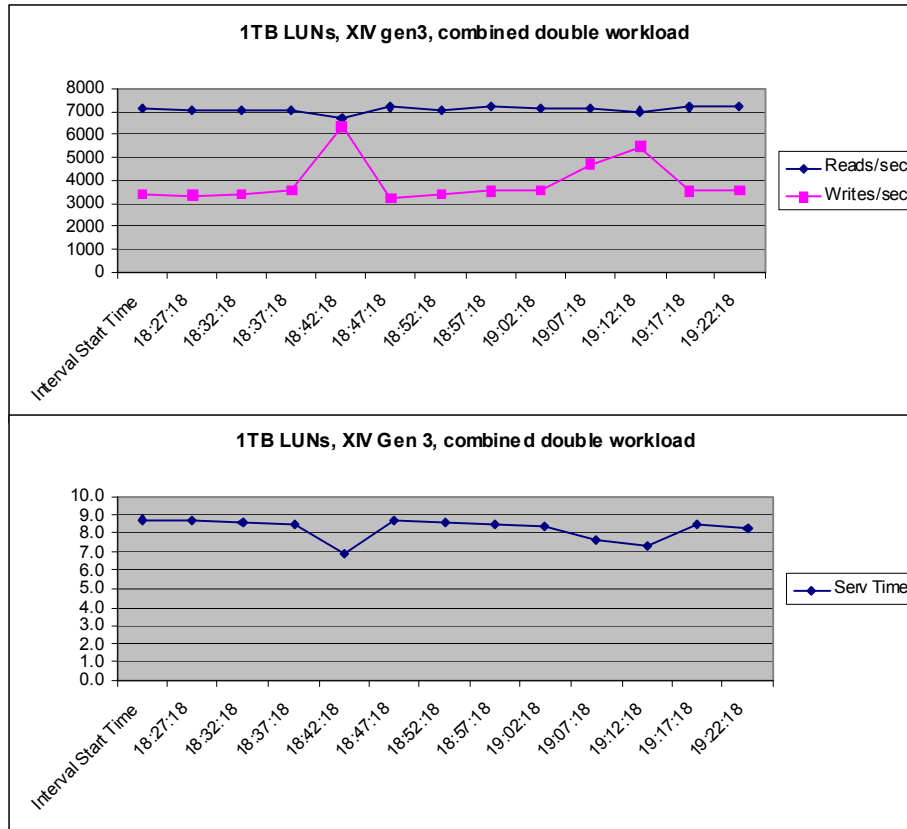


Figure 8-35 1-TB LUNs, combined double workload, I/O rate, and service time

The restore of CPW database took 16 minutes.

Table 8-7 shows the transaction response time for the CPW run on the 154-GB LUNs. Average response time for most of the transactions is between 1.6 and 12 seconds.

Table 8-7 154-GB LUNs

Transaction ID	Count	Percentage	Average Resp. time (sec)	Maximal Resp. time (sec)
Neworder	6885794	47.39	12.404	626.260
Ordersts	553639	3.81	0.511	16.850
Payment	5968925	41.08	1.545	178.690
Delivery	560421	3.86	0.042	6.210
Stocklvl	560005	3.85	2.574	21.810

The disk service time response time during the collection period is shown in Figure 8-36. The average disk service time of one hour collection period was 4.6 ms. The average LUN utilization was 78.3%.

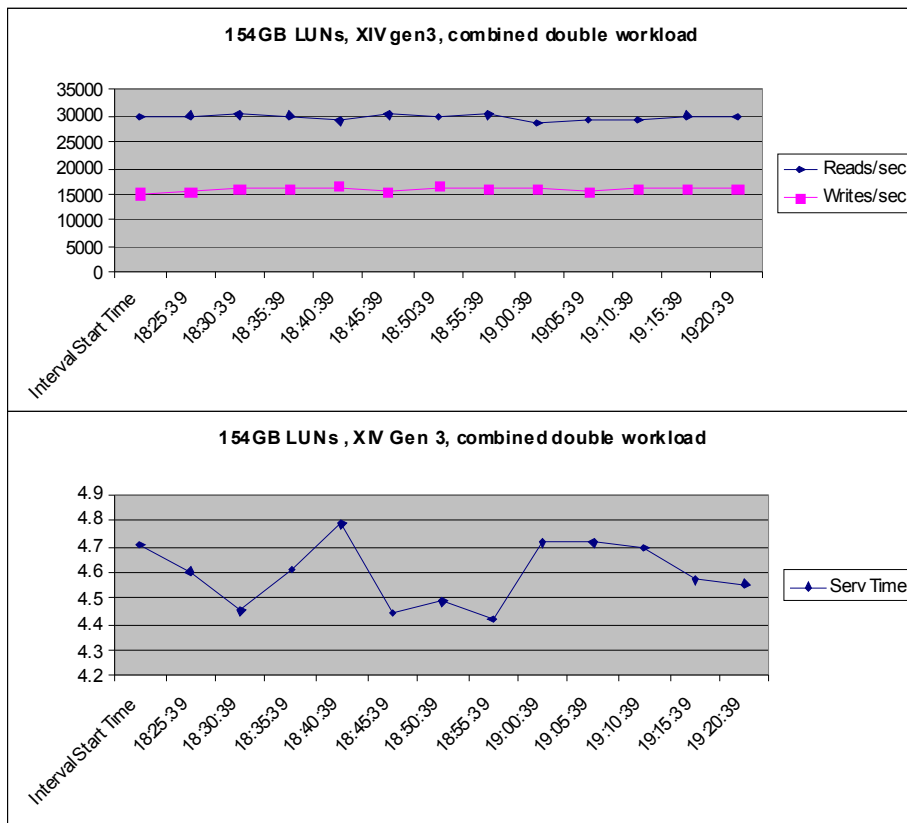


Figure 8-36 154-GB LUNs, combined double workload, I/O rate, and service time

The CPW database restore took 13 minutes.

In this test, workloads were run in both IBM i LPARs at the same time. Figure 8-37 on page 254, Figure 8-38 on page 255, and Figure 8-39 on page 256 show the following values in XIV:

- ▶ I/O rate
- ▶ Latency
- ▶ Bandwidth in MBps
- ▶ Cache hits

Figure 8-37 shows these values during the whole CPW run. The pictures show the XIV values of one LUN. Multiply the I/O rate and MBps by the number of LUNs to get the overall rates. The latency and cache hits that are shown for one LUN are about the same as the average across all LUNs in the LPAR.

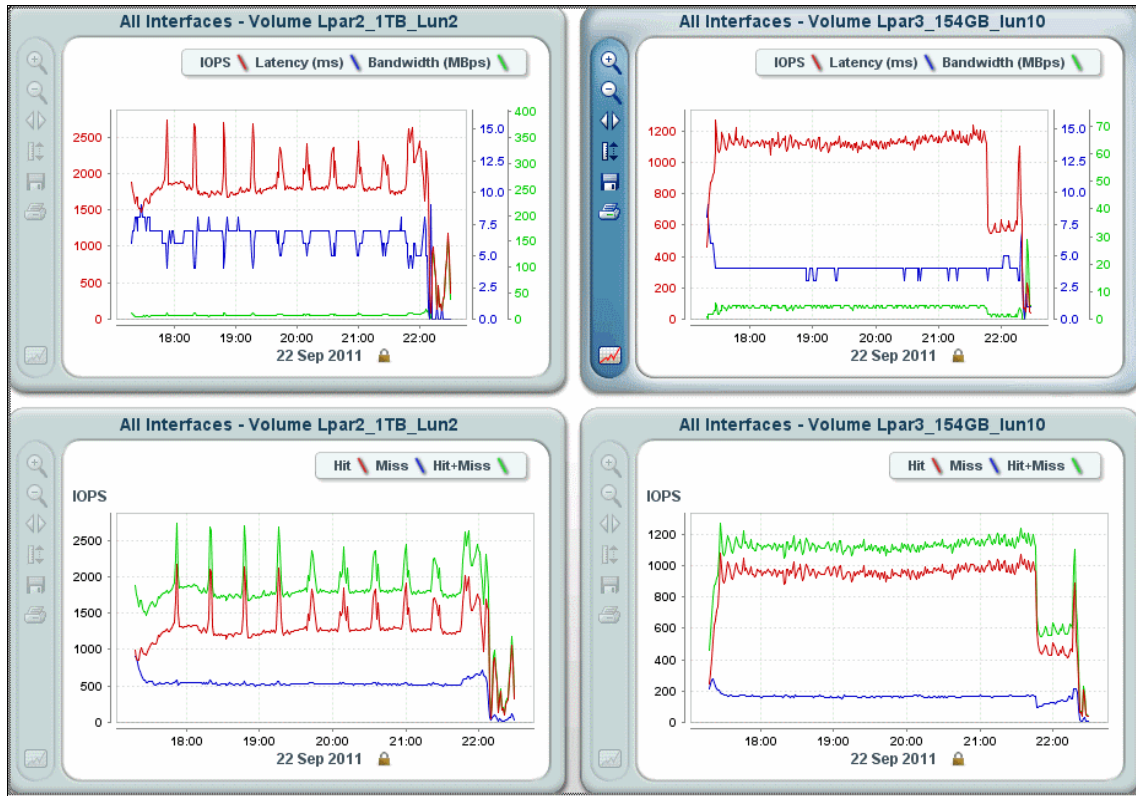


Figure 8-37 XIV values for entire run of both CPW workloads

Figure 8-38 shows these values during the IBM i collection period. The average latency of 1-TB LUNs was about 7 ms, whereas the latency of 154-GB LUNs was close to 4 ms. On 1-TB LUNs, the workload experiences about 60% cache hit, whereas on 154-GB LUNs the cache hits were about 80%.



Figure 8-38 XIV values for data collection of both CPW workloads

Figure 8-39 shows these values during the CPW database restore. During the database restore, the 1-TB LUNs experienced almost no latency. The latency on 15-4 GB LUNs was about 1 ms. Cache hits on both LUNs were close to 100%.

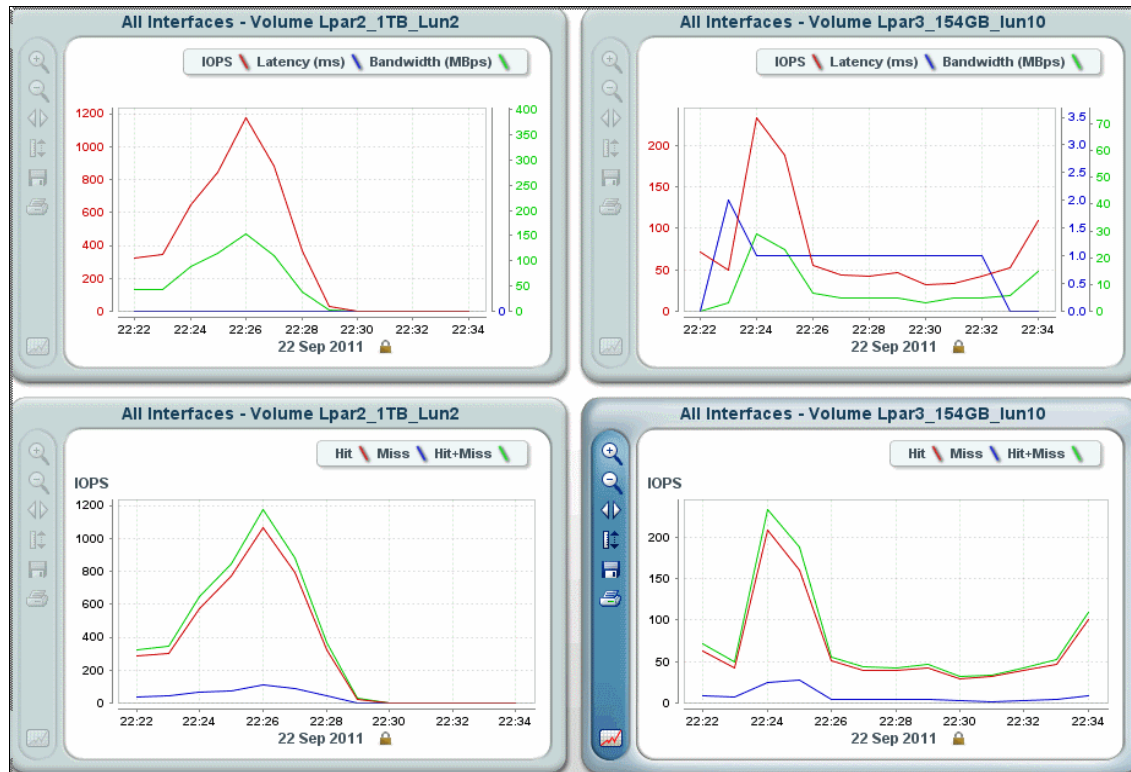


Figure 8-39 XIV values for database restore of both CPW workloads

8.6.8 Testing conclusions

Table 8-8 shows the results of the tests. The average transaction response time and disk service time were reported by IBM i Performance Tools during the CPW data collection period. In addition, the latency and cache hits were reported by XIV statistics during the same period.

Table 8-8 Test results

	Average transactions response time (seconds)	Average disk service time (milliseconds)	App. average latency on XIV (milliseconds)	App. percentage of average cache hits on XIV
XIV Gen 2				
42 * 154-GB LUNs	0.036	4.4	3	75
6 * 1-TB LUNs	4.9	12.6	20	50
XIV Gen 3				
42 * 154-GB LUNs	0.005	0.5	Near 0	Near 100
6 * 1-TB LUNs	0.005	0.5	Near 0	Near 100

	Average transactions response time (seconds)	Average disk service time (milliseconds)	App. average latency on XIV (milliseconds)	App. percentage of average cache hits on XIV
XIV Gen 3 Concurrent double workload				
42 * 154-GB LUNs	6.6	4.6	4	80
6 * 1-TB LUNs	19.3	8.2	7	60

Comparing many small LUNs to a few large LUNs

On an XIV Generation 2, the workload experiences much better response times when you use many smaller LUNs compared to using a few large LUNs.

Whether using small LUNs or few large LUNs on an XIV Gen 3 system, the performance is good in both cases. There is no significant difference between the response times in the two environments. However, when the XIV Gen 3 is stressed by running double workload in both LPARs, a large difference develops in response times between the two environments. The advantage goes to the configuration with many small LUNs.

The better performance with many small LUNs is for the following reasons:

- ▶ Queue-depth is the number of I/O operations that can be done concurrently to a volume. The queue-depth for an IBM i volume in VIOS is a maximum of 32. This maximum is modest comparing to the maximal queue-depths for other open servers. Therefore, in IBM i, define a larger number of small LUNs than for other open system servers. This configuration provides a comparable number of concurrent IO operations for the disk space available.
- ▶ The more LUNs that are available to an IBM i system, the more server tasks IBM i storage management uses to manage the I/O operations to the disk space. Therefore, better I/O performance is achieved.

Comparing XIV Storage System generation 2 and Gen 3

When you run the same LUN configuration (size and number) on XIV generation 2 and XIV Gen 3, XIV Gen 3 has better response times. This difference is caused by the bigger cache and enhanced storage architecture of the XIV Gen 3. The CPW workload with 96,000 users that ran on XIV Gen 3 experienced almost 100% cache hits and a disk response time below 1 ms. Running two such workloads at the same time did not stress XIV much more. The cache hits were 90% - 100%, and the service times were 1.2 to 1.6 ms.

When the increased workload of 192,000 users was run concurrently in the two LPARs, cache hits fell to 60% - 80%. Disk service times increased to 4 - 8 ms.

Conclusion about XIV LUN size for IBM i

Comparing the six 1-TB LUNs configuration against 42 LUNs of 154-GB (equal disk capacity in each environment), the 42 LUN configuration had the better performance. To keep the number of LUNs reasonable for ease of management in XIV, VIOS, and IBM i, generally a LUN size of 100 GB to 150 GB is appropriate.

In each configuration, the number of LUNs is a multiple of 6. For a fully configured XIV System with six Interface Modules, this configuration equally distributes the workload (I/O traffic) across the Interface Modules.



XIV Storage System and VMware connectivity

The IBM XIV Storage System is an excellent choice for your VMware storage requirements. XIV achieves consistent high performance by balancing the workload across physical resources. This chapter addresses OS-specific considerations for host connectivity and describes the host attachment-related tasks for VMware ESX version 3.5, ESX/ESXi version 4.x, and ESXi 5.0/5.1.

Note: For information about the XIV Storage System and VMware Integration concepts and implementation, see *XIV Storage System in a VMware Environment*, REDP-4965 at:

<http://www.redbooks.ibm.com/abstracts/redp4965.html?Open>

9.1 Integration concepts and implementation guidelines

This section is geared toward IT decision makers, storage administrators, and VMware administrators. It offers an overview of XIV storage and VMware integration concepts, and general implementation guidelines.

At a fundamental level, the goal of both the XIV Storage System and VMware's storage features is to significantly reduce the complexity of deploying and managing storage resources. With XIV, storage administrators can provide consistent tier-1 storage performance and quick change-request cycles. This support is possible because they need perform little planning and maintenance to keep performance levels high and storage optimally provisioned.

The following underlying strategies are built into the vSphere storage framework to insulate administrators from complex storage management tasks, and non-optimal performance and capacity resource utilization:

- ▶ Make storage objects much larger and more scalable, reducing the number that must be managed by the administrator
- ▶ Extend specific storage resource-awareness by attaching features and profiling attributes to the storage objects
- ▶ Help administrators make the correct storage provisioning decision for each virtual machine or even fully automate the intelligent deployment of virtual machine storage.
- ▶ Remove many time-consuming and repetitive storage-related tasks, including the need for repetitive physical capacity provisioning.

Clearly, vCenter relies upon the storage subsystem to fully support several key integration features to effectively implement these strategies. Appropriately compatible storage, such as XIV, is essential.

9.1.1 vSphere storage architectural overview

The vSphere storage architecture, including physical and logical storage elements, is depicted in Figure 9-1 on page 261. Although not intended to thoroughly explore vSphere storage concepts and terminology, the essential components and their relationships provide the foundational framework necessary to understand the upcoming integration principles.

The VMware file system (VMFS) is the central abstraction layer that acts as a medium between the storage and the hypervisor layers. The current generation of VMFS includes the following distinguishing attributes, among others:

- ▶ Clustered file system: Purpose-built, high performance clustered file system for storing virtual machine files on shared storage (Fibre Channel and iSCSI). The primary goal of VMFS's design is as an abstraction layer between the VMs and the storage to efficiently pool and manage storage as a unified, multi-tenant resource.
- ▶ Shared data file system: Enable multiple vSphere hosts to read and write from the same datastore concurrently.
- ▶ Online insertion or deletion of nodes: Add or remove vSphere hosts from VMFS volume with no impact to adjacent hosts or VMs.
- ▶ On-disk file locking: Ensure that the same virtual machine is not accessed by multiple vSphere hosts concurrently.

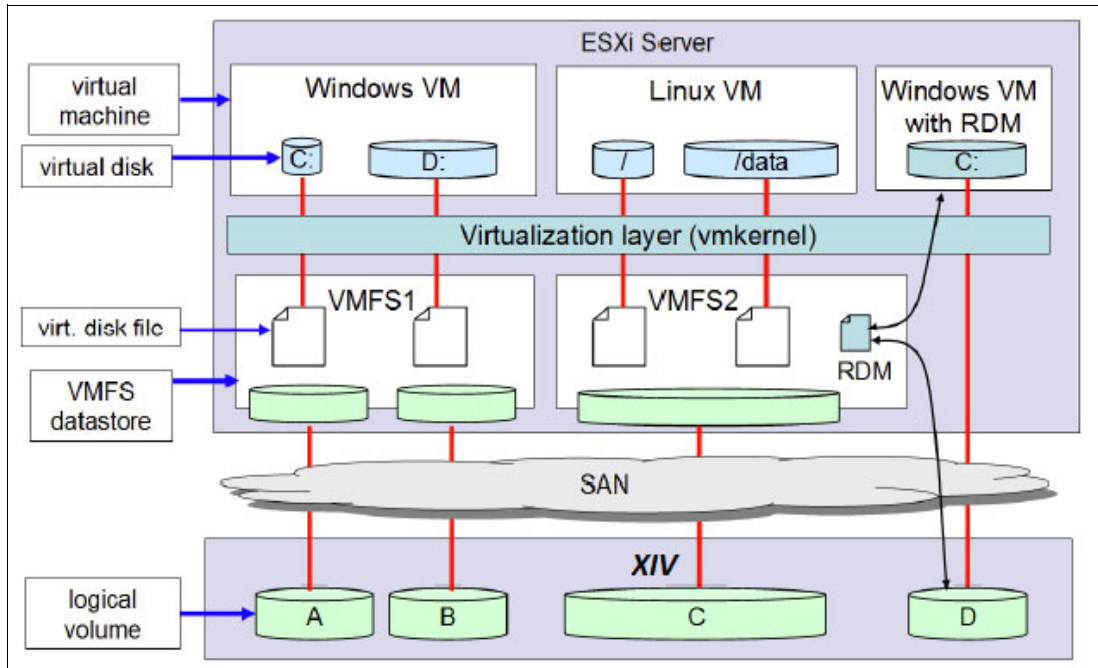


Figure 9-1 ESX/ESXi Basic Storage Elements in the vSphere Infrastructure

This chapter examines concepts and guidelines crucial to building an adaptable, efficient, high performance vSphere infrastructure with the XIV Storage System's inherently cloud-optimized design and deep vSphere integration capabilities at its foundation.

9.1.2 XIV and VMware general connectivity guidelines

When you implement Fibre Channel connectivity for the XIV Storage System in a vSphere environment, adhere to the following practices:

- ▶ Use XIV host cluster groups for LUN assignment
- ▶ Configure single initiator zones
- ▶ At the time of this writing, the VMware specifies that there can be a maximum of 1024 paths and 256 LUNs per ESX/ESXi host, as shown in Figure 9-2 on page 262. The following conditions must be simultaneously satisfied to achieve the optimal storage configuration:
 - Effectively balance paths across these objects:
 - Host HBA ports
 - XIV Interface Modules
 - Ensure that the wanted minimum number of host paths per LUN and the wanted minimum number of LUNs per host can be simultaneously met.
- ▶ Configure the Path Selection plug-in (PSP) multipathing based on vSphere version:
 - Use Round Robin policy if the vSphere version is vSphere 4.0 or higher.
 - Use Fixed Path policy if the vSphere version is lower than vSphere 4.0.
 - Do *not* use the Most Recently Used (MRU) policy.

Refer to Figure 9-2 and Figure 9-3 on page 263 for suggested configurations to satisfy these criteria.

When you implement iSCSI connectivity for the XIV Storage Systems in a vSphere environment, adhere to the following practices:

- ▶ One VMkernel port group per physical network interface card (NIC):
 - VMkernel port is bound to physical NIC port in vSwitch, which creates a “path”
 - Creates 1-to-1 “path” for VMware NMP
 - Use the same PSP as for FC connectivity
- ▶ Enable jumbo frames for throughput intensive workloads (must be done at all layers).
- ▶ Use Round Robin PSP to enable load balancing across all XIV Interface Modules. Each initiator should see a target port on each module.
- ▶ Queue depth can also be changed on the iSCSI software initiator. If more bandwidth is needed, the LUN queue depth can be modified.

Figure 9-2 shows a suggested configuration that uses two HBAs.

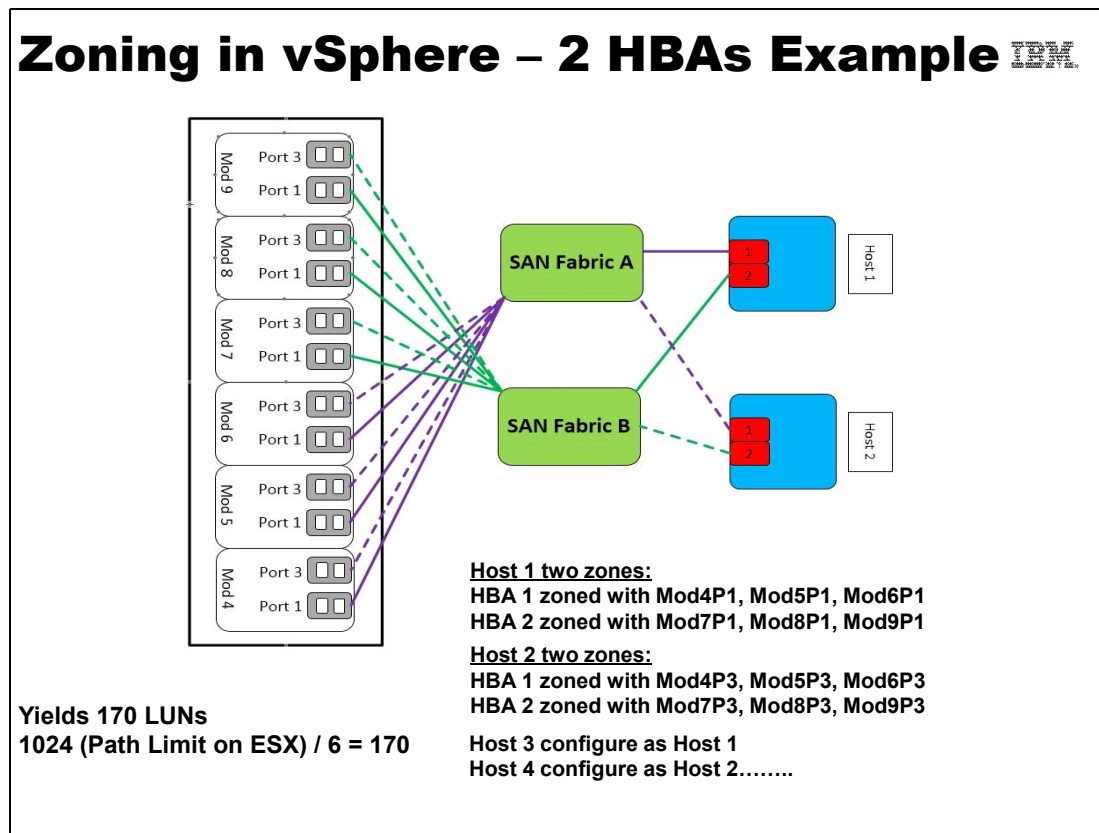


Figure 9-2 Zoning guidelines: 2 HBAs per ESX Host

Figure 9-3 shows a suggested configuration that uses four HBAs.

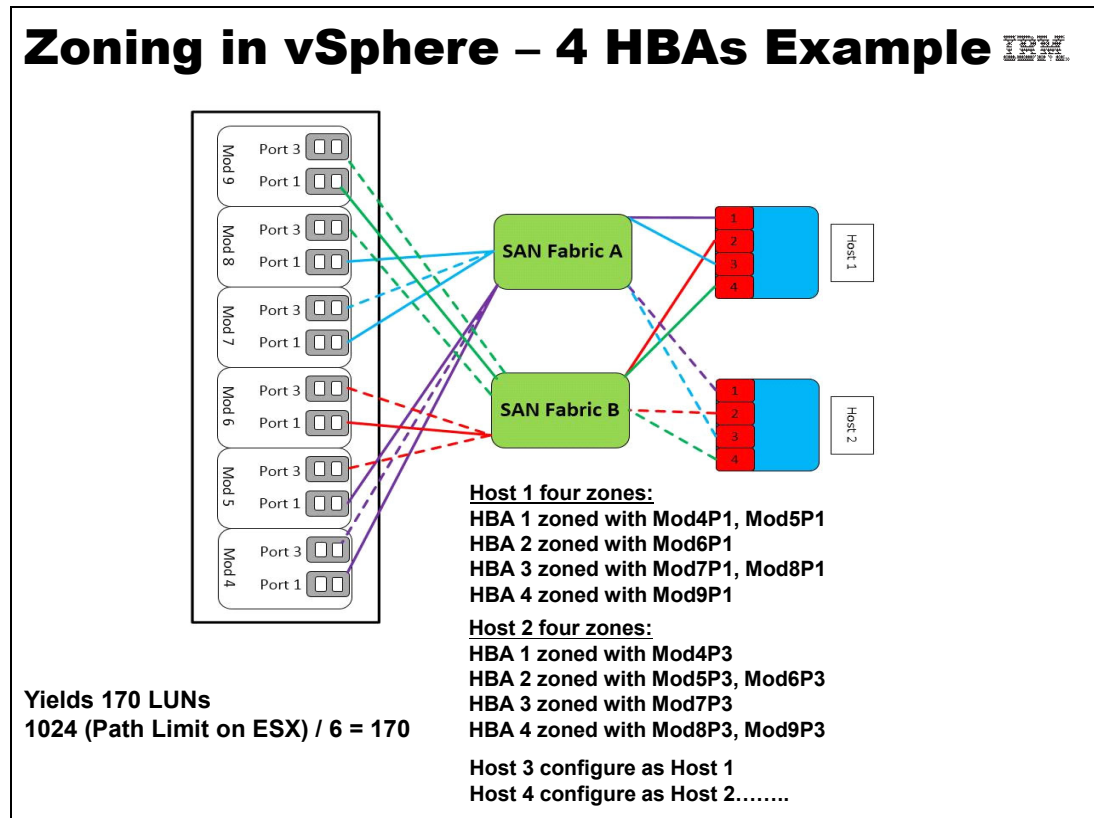


Figure 9-3 Zoning guidelines: 4 HBAs per ESX Host

Table 9-1 lists maximum values for various storage elements in VSphere 5.0/5.1.

Table 9-1 Notable storage maximums in vSphere 5.0/5.1

Storage Element Limit	Maximum
Virtual Disk Size	2 TB minus 512 bytes
Virtual Disks per Host	2048
LUNs per Host	256
Total Number of Paths per Host	1024
Total Number of Paths to per LUN	32
LUN Size	64 TB
Concurrent Storage vMotions per Datastore	8
Concurrent Storage vMotions per Host	2



XIV and N series Gateway connectivity

This chapter addresses specific considerations for attaching an IBM System Storage N series Gateway to an IBM XIV Storage System.

This chapter includes the following sections:

- ▶ Overview of N series Gateway
- ▶ Attaching N series Gateway to XIV
- ▶ Cabling
- ▶ Zoning
- ▶ Configuring the XIV for N series Gateway
- ▶ Installing Data ONTAP

10.1 Overview of N series Gateway

The IBM System Storage N series Gateway can be used to provide network-attached storage (NAS) functionality with XIV. For example, it can be used for Network File System (NFS) exports and Common Internet File System (CIFS) shares. N series Gateway is supported by software level 10.1 and later. Exact details about currently supported levels can be found in the N series interoperability matrix at:

<http://www-01.ibm.com/support/docview.wss?uid=ssg1S7003656>

Figure 10-1 illustrates attachment of the XIV Storage System with the N Series Gateway.

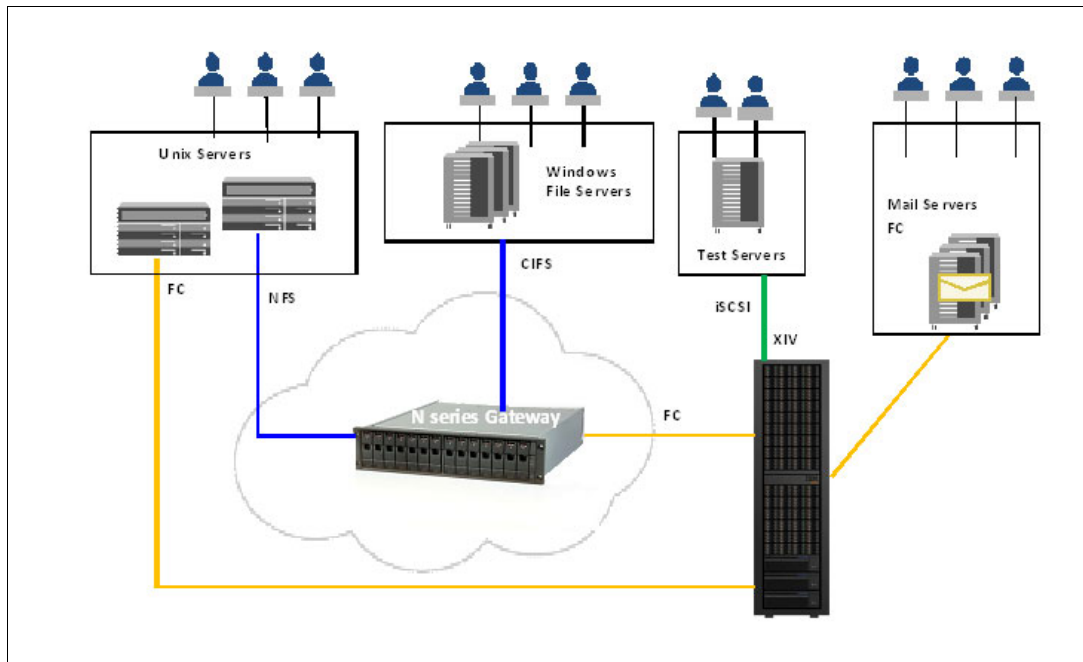


Figure 10-1 N series Gateway with IBM XIV Storage System

10.2 Attaching N series Gateway to XIV

When you attach the N series Gateway to an XIV, the following considerations apply. These considerations are in addition to the connectivity guidelines in Chapter 1, “Host connectivity” on page 1.

- ▶ Check for supported XIV and N series Operating System versions
- ▶ Plan and install the appropriate N series cabling
- ▶ Define SAN zoning on the fabric for XIV and N series entities
- ▶ Create XIV host definitions for the N series array
- ▶ Create XIV volumes, and optionally create a pool for these volumes
- ▶ Map XIV volumes to corresponding N series hosts
- ▶ Install Data ONTAP and upgrades onto the N series root volume on XIV

10.2.1 Supported versions

At the time of writing, the configurations that are shown in Table 10-1 are supported by XIV:

Table 10-1 Currently supported N series models and Data ONTAP versions

XIV model	Level	N series	ONTAP OS
XIV Gen2 2810-A14 XIV Gen2 2812-A14	10.2.0.a 10.2.1 10.2.1b 10.2.2 10.2.2a 10.2.4 10.2.4.a 10.2.4.b 10.2.4.c 10.2.4.e	N7900 N7700 N6070 N6060 N6040 N5600 N5300	Data ONTAP 7.3.3; Data ONTAP 7.3.4
XIV Gen2 2810-A14 XIV Gen2 2812-A14	10.2.0.a; 10.2.1; 10.2.1b; 10.2.2; 10.2.2a; 10.2.4; 10.2.4.a; 10.2.4.b; 10.2.4.c; 10.2.4.e	N5300 N5600 N6040 N6060 N6070 N6210 N6240 N6270 N7700 N7900	Data ONTAP 7.3.5.1 Data ONTAP 7.3.6 Data ONTAP 7.3.7
XIV Gen2 2810-A14 XIV Gen2 2812-A14	10.2.2 10.2.2a 10.2.4 10.2.4.a 10.2.4.b 10.2.4.c 10.2.4.e	N5300 N5600 N6040 N6060 N6070 N6210 N6240 N6270 N7700 N7900 N7950T	Data ONTAP 8.0.1 7-Mode Data ONTAP 8.0.2 7-Mode Data ONTAP 8.0.3 7-Mode Data ONTAP 8.0.4 7-Mode Data ONTAP 8.1 7-Mode Data ONTAP 8.1.1 7-Mode
XIV Gen3 2810-114; XIV Gen3 2812-114	11.0.0.a; 11.1.0; 11.1.0.a; 11.1.0.b; 11.1.1	N6040 N6060 N6070 N6210 N6240 N6270 N7700 N7900 N7950T	Data ONTAP 8.1 7-Mode; Data ONTAP 8.1.1 7-Mode;

For the latest information and supported versions, always verify your configuration in the N series Gateway interoperability matrix at:

<http://www-01.ibm.com/support/docview.wss?uid=ssg1S7003656>

Additionally, Gateway MetroClusters and Stretch MetroClusters are also supported as listed in the interoperability matrix.

10.2.2 Other considerations

Also, take into account the following considerations when you attach N series Gateway to XIV:

- ▶ Only FC connections between N series Gateway and an XIV system are allowed.
- ▶ Direct attach is not supported as of this writing.
- ▶ Do not map any volume using LUN 0. For more information, see *IBM System Storage N series Hardware Guide*, SG24-7840, available at:
<http://www.redbooks.ibm.com/redbooks/pdfs/sg247840.pdf>
- ▶ N series can handle only two paths per LUN. For more information, see 10.4, “Zoning” on page 269.
- ▶ N series can handle only up to 2-TB LUNs. For more information, see 10.6.4, “Adding data LUNs to N series Gateway” on page 280.

10.3 Cabling

This section addresses how to cable when you connect the XIV Storage System, either to a single N series Gateway or to an N series cluster Gateway.

10.3.1 Cabling example for single N series Gateway with XIV

Cable the N series Gateway so that one fiber port connects to each of the switch fabrics. You can use any of the fiber ports on the N series Gateway, but make sure that they are set as initiators. The example in Figure 10-2 uses 0a and 0c because they are on separate Fibre Channel chips, thus providing better resiliency.

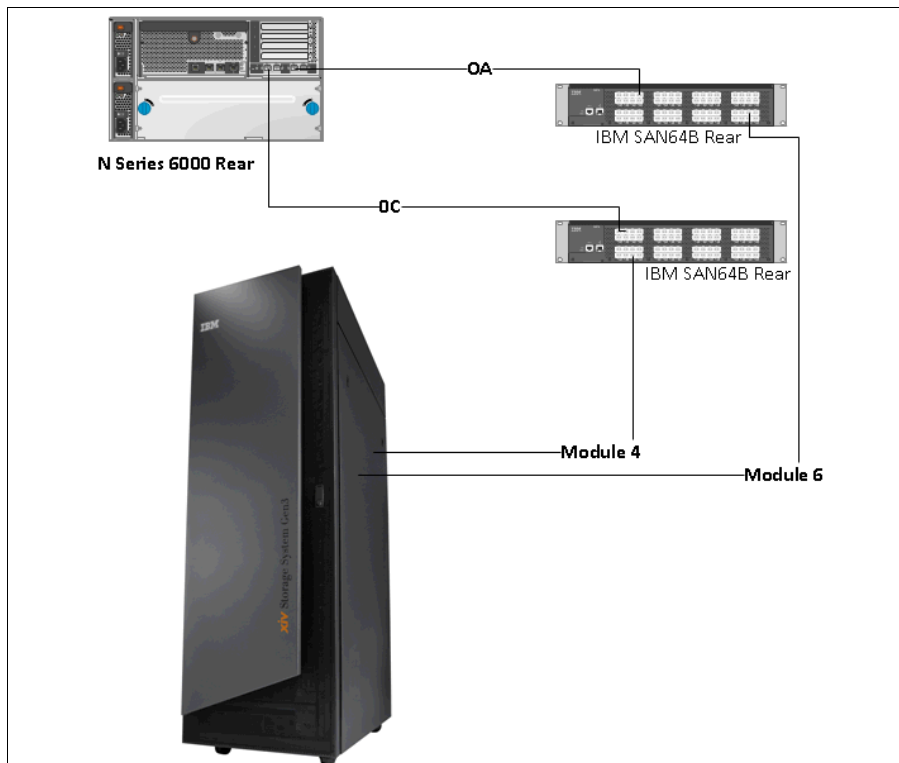


Figure 10-2 Single N series to XIV cabling overview

10.3.2 Cabling example for N series Gateway cluster with XIV

Cable an N series Gateway cluster so that one fiber port connects to each of the switch fabrics. You can use any of the fiber ports on the N series Gateway, but make sure that they are set as initiators. The example uses 0a and 0c because they are on separate Fibre Channel chips, which provides better resiliency.

The link between the N series Gateways is the cluster interconnect as shown in Figure 10-3.

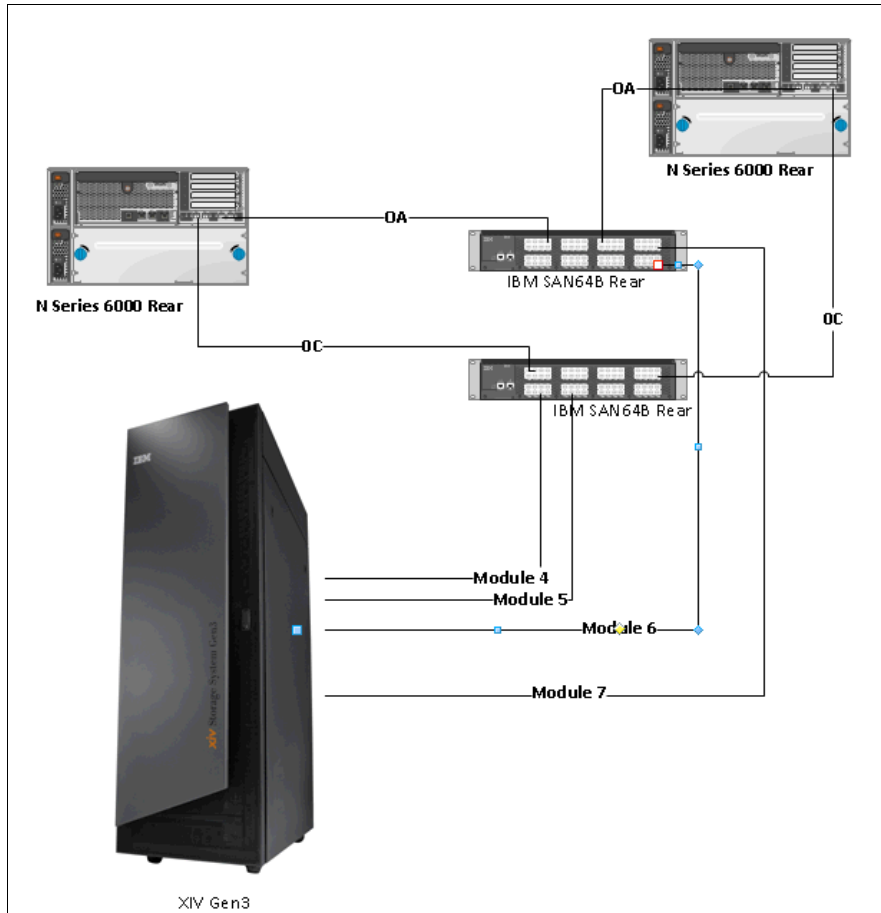


Figure 10-3 Clustered N series to XIV cabling overview

10.4 Zoning

Create zones so that there is only one initiator in each zone. Using a single initiator per zone ensures that every LUN presented to the N series Gateway has only two paths. It also limits the registered state change notification (RSCN) traffic in the switch.

10.4.1 Zoning example for single N series Gateway attachment to XIV

The following is an example of zoning definition for a single N series Gateway:

- ▶ Switch 1
 - Zone 1
 - NSeries_port_0a, XIV_module4_port1
- ▶ Switch 2
 - Zone 1
 - NSeries_port_0c, XIV_module6_port1

10.4.2 Zoning example for clustered N series Gateway attachment to XIV

The following is an example of zoning definition for a clustered N series Gateway:

- ▶ Switch 1
 - Zone 1
 - NSeries1_port_0a, XIV_module4_port1
 - Zone 2
 - Nseries2_port_0a, XIV_module5_port1
- ▶ Switch 2
 - Zone 1
 - NSeries1_port_0c, XIV_module6_port1
 - Zone 2
 - Nseries2_port_0c, XIV_module7_port1

10.5 Configuring the XIV for N series Gateway

N series Gateway boots from an XIV volume. Before you can configure an XIV for an N series Gateway, the correct root sizes must be chosen. Figure 10-4 shows the minimum root volume sizes from the N series Gateway interoperability matrix.

	Root Volume Minimums	
Nseries Model	Prior to 8.0	Starting in 8.0.x
N5500	25 GB	n/a
N5300	25 GB	256 GB
N5600	37 GB	368 GB
N6040	25 GB	256 GB
N6060	37 GB	368 GB
N6070	60 GB	400 GB
N6210	16 GB	160 GB
N6240	24 GB	240 GB
N6270	48 GB	480 GB
N7600	60 GB	400 GB
N7800	111 GB	400 GB
N7700	60 GB	400 GB
N7900	111 GB	400 GB

Note: The minimum size of the array LUN needed for the root volume, **is larger** than the Data ONTAP minimum root volume size. It ensures that there is sufficient space in the root volume for system files, log files, and core files, especially in the event of a crash or core dump.

Figure 10-4 Minimum root volume sizes on different N series hardware

The volumes that you present from an XIV round up to the nearest increment of 17 GB.

Important: XIV reports capacity in GB (decimal) and N Series reports in GiB (Binary). For more information, see 10.5.2, “Creating the root volume in XIV” on page 272.

N Series imports, by default, use Block Checksums (BCS). These imports use one block of every nine for checksum, which uses 12.5% of total capacity. Alternatively, you can import LUNs by using Zone checksum (ZCS). ZCS uses one block of every 64 for checksums. However, using ZCS negatively affects performance, especially on random read intensive workloads. Consider using Zone checksum on LUNs designated for backups.

The following space concerns also reduce usable capacity:

- ▶ N series itself uses approximately 1.5% of the capacity for metadata and metadata snapshots.
- ▶ N series Write Anywhere File Layout (WAFL) file system uses approximately 10% of the capacity for formatting.

10.5.1 Creating a Storage Pool in XIV

When N series Gateway is attached to XIV, it does not support XIV snapshots, synchronous mirror, asynchronous mirror, or thin provisioning features. If you need these features, they must be used from the corresponding functions that N series Data ONTAP natively offers.

To prepare the XIV Storage System for use with N series Gateway, first create a Storage Pool by using, for instance, the XIV GUI as shown in Figure 10-5.

Tip: No Snapshot space reservation is needed because the XIV snapshots are not supported by N series Gateways.

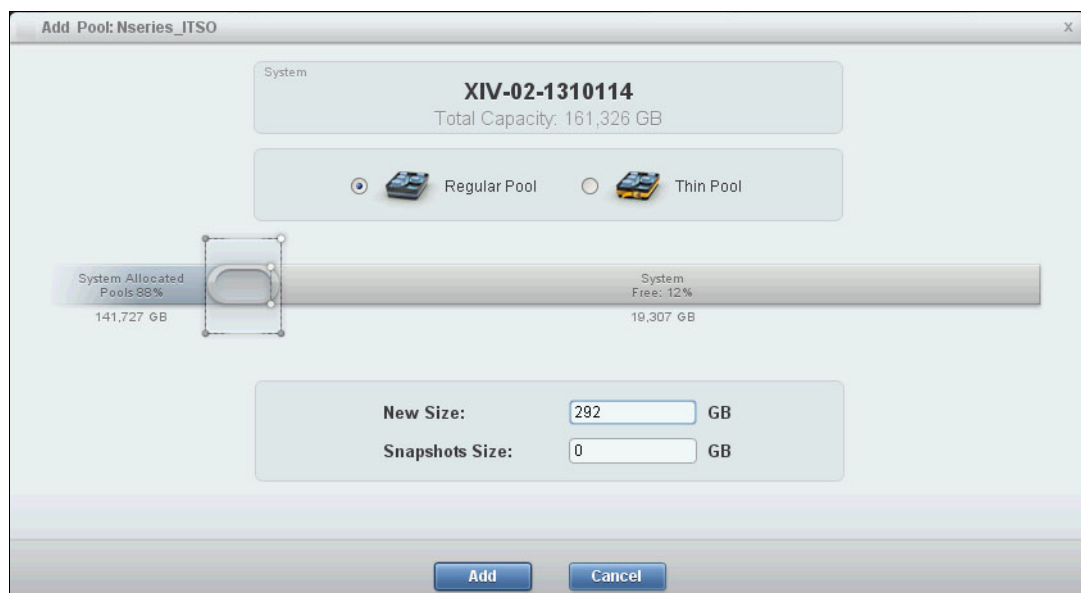


Figure 10-5 Creating a regular storage pool for N series Gateway

10.5.2 Creating the root volume in XIV

The N series interoperability matrix displayed (in part) in Figure 10-4 on page 270 shows the correct minimum sizing for the supported N series models.

N series calculates capacity differently than XIV, and you must make adjustments to get the correct size. N series GB are expressed as 1000 x 1024 x 1024 bytes, whereas XIV GB are expressed as either 1000 x 1000 x 1000 bytes or 1024x1024x1024 bytes.

The N series formula is not the same as GB or GiB. Figure 10-6 lists XIV GUI options that help ensure that you create the correct volume size for other operating systems and hosts.

Consideration: Inside the XIV GUI, you have several choices in allocation definitions:

- ▶ If GB units are chosen, a single storage unit is regarded as 10^9 (1,000,000,000) bytes.
- ▶ If GiB units are chosen, a single storage unit is regarded as 2^{30} bytes. This is known as *binary notation*.

To calculate the size for a minimum N series gateway root volume, use this formula:

$$\langle \text{min_size} \rangle \text{ GB} \times (1000 \times 1024 \times 1024) / (1000 \times 1000 \times 1000) = \langle \text{XIV_size_in_GB} \rangle$$

Because XIV is using capacity increments of about 17 GB, it will automatic set the size to the nearest increment of 17 GB.

As shown in Figure 10-6, create a volume with the correct size for the root volume in the XIV pool previously created. Also, create an extra 17-GB dummy volume that can be mapped as LUN 0. This additional volume might not be needed depending on the specific environment.

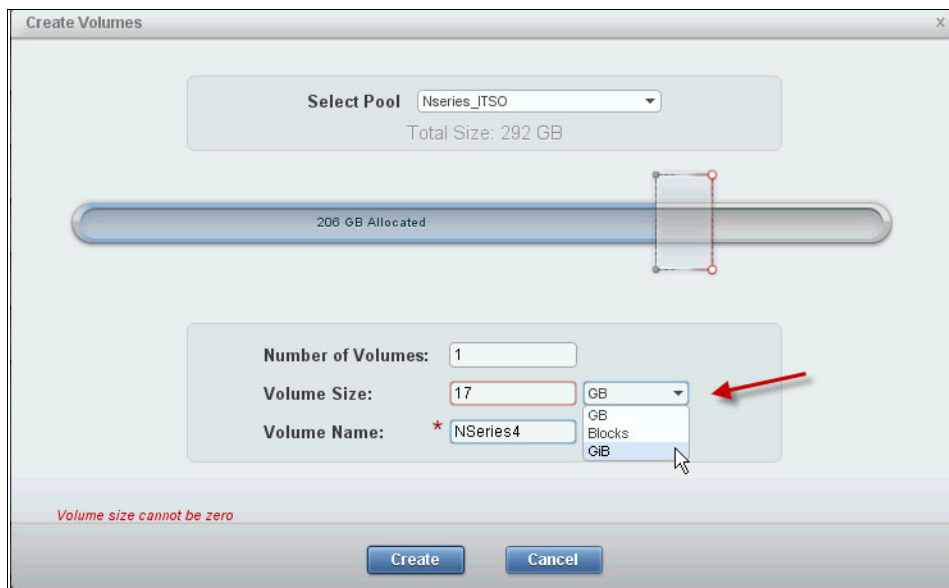


Figure 10-6 Creating a volume

10.5.3 Creating the N series Gateway host in XIV

Create the host definitions in XIV. The example that is shown in Figure 10-7 is for a single N series Gateway. You can just create a host. For a two node cluster Gateway, you must create a cluster in XIV first, and then add the corresponding hosts to the cluster.

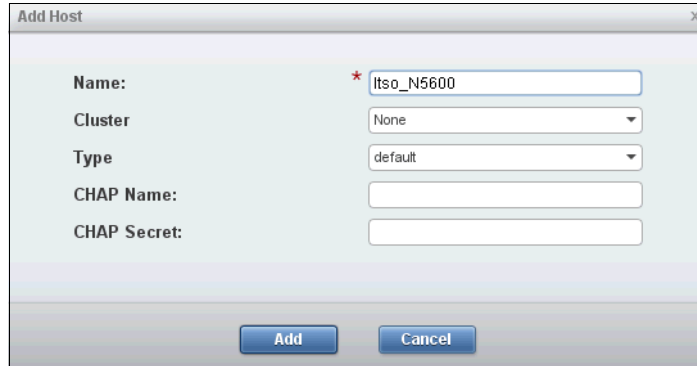


Figure 10-7 Creating a single N series Gateway host

10.5.4 Adding the WWPN to the host in XIV

Obtain the worldwide port name (WWPN) of the N series Gateway. You can do so by starting the N series Gateway in Maintenance mode. Maintenance mode makes the Gateway log in to the switches. To get the N series into Maintenance mode, you must access the N series console. Use the null modem cable that came with the system or the Remote LAN Module (RLM) interface.

To find the WWPN by using the RLM method, complete these steps:

1. Power on your N series Gateway.
2. Connect to RLM ip by using ssh, and log in as naroot.
3. Enter system console.
4. Observe the boot process as illustrated in Example 10-1, and when you see Press CTRL-C for special boot menu, immediately press Ctrl+C.

Example 10-1 N series Gateway booting by using SSH

```
Phoenix TrustedCore(tm) Server
Copyright 1985-2004 Phoenix Technologies Ltd.
All Rights Reserved
BIOS version: 2.4.0
Portions Copyright (c) 2006-2009 NetApp All Rights Reserved
CPU= Dual Core AMD Opteron(tm) Processor 265 X 2
Testing RAM
512MB RAM tested
8192MB RAM installed
Fixed Disk 0: STEC    NACF1GM1U-B11
```

```
Boot Loader version 1.7
Copyright (C) 2000-2003 Broadcom Corporation.
Portions Copyright (C) 2002-2009 NetApp
```

```
CPU Type: Dual Core AMD Opteron(tm) Processor 265
```

```
Starting AUTOBOOT press Ctrl-C to abort...
Loading x86_64/kernel/primary.krn:.....0x200000/46415944
0x2e44048/18105280 0x3f88408/6178149 0x456c96d/3 Entry at 0x00202018
Starting program at 0x00202018
```

Press CTRL-C for special boot menu

Special boot options menu will be available.

Tue Oct 5 17:20:23 GMT [nvram.battery.state:info]: The NVRAM battery is currently ON.

Tue Oct 5 17:20:24 GMT [fci.nserr.noDevices:error]: The Fibre Channel fabric attached to adapter 0c reports the presence of no Fibre Channel devices.

Tue Oct 5 17:20:25 GMT [fci.nserr.noDevices:error]: The Fibre Channel fabric attached to adapter 0a reports the presence of no Fibre Channel devices.

Tue Oct 5 17:20:33 GMT [fci.initialization.failed:error]: Initialization failed on Fibre Channel adapter 0d.

Data ONTAP Release 7.3.3: Thu Mar 11 23:02:12 PST 2010 (IBM)

Copyright (c) 1992-2009 NetApp.

Starting boot on Tue Oct 5 17:20:16 GMT 2010

Tue Oct 5 17:20:33 GMT [fci.initialization.failed:error]: Initialization failed on Fibre Channel adapter 0b.

Tue Oct 5 17:20:39 GMT [diskown.isEnabled:info]: software ownership has been enabled for this system

Tue Oct 5 17:20:39 GMT [config.noPartnerDisks:CRITICAL]: No disks were detected for the partner; this node will be unable to takeover correctly

- (1) Normal boot.
- (2) Boot without /etc/rc.
- (3) Change password.
- (4) No disks assigned (use 'disk assign' from the Maintenance Mode).
- (4a) Same as option 4, but create a flexible root volume.
- (5) Maintenance mode boot.

Selection (1-5)?

- 5. Select 5 for Maintenance mode.
- 6. Enter storage show adapter to find which WWPN belongs to 0a and 0c. Verify the WWPN in the switch and check that the N series Gateway is logged in. See Figure 10-8.

18(0x12)	55	123700	N	50:0a:09:82:00:02:43:4a	50:0a:09:82:00:02:43:4a
18(0x12)	51	123300	N	50:0a:09:80:00:02:43:4a	50:0a:09:80:00:02:43:4a

Figure 10-8 N series Gateway logged in to switch as network appliance

7. Add the WWPN to the host in the XIV GUI as depicted in Figure 10-9.

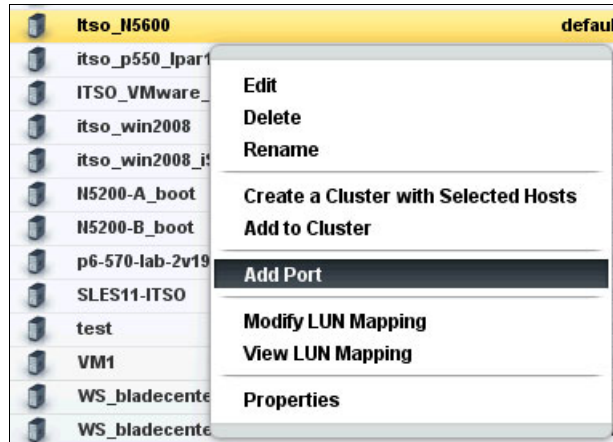


Figure 10-9 Adding port to the host

Make sure that you add both ports as shown in Figure 10-10. If your zoning is correct, they are displayed in the list. If they do not show up, check your zoning.

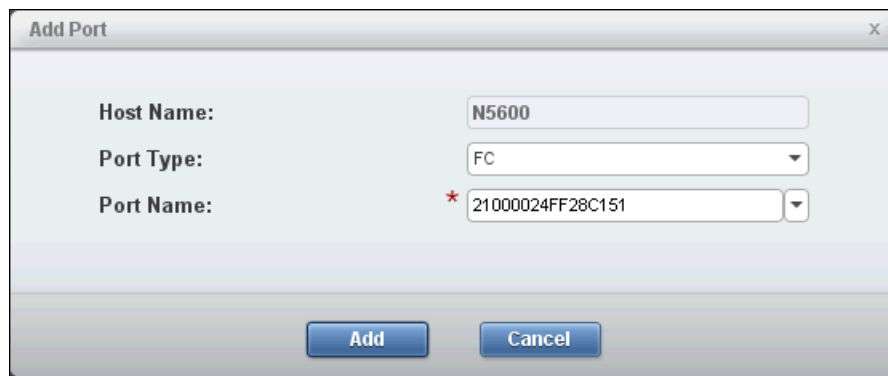


Figure 10-10 Adding both ports: 0a and 0c

8. Verify that both ports are connected to XIV by checking the Host Connectivity view in the XIV GUI as shown in Figure 10-11.

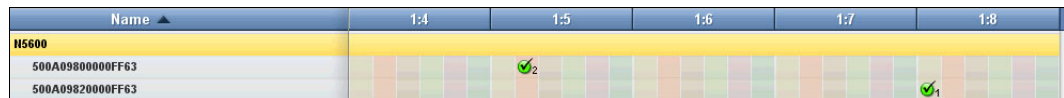


Figure 10-11 Verifying that ports are connected

10.5.5 Mapping the root volume to the N series host in XIV GUI

To map the root volume as LUN 0, complete these additional steps. This procedure is only needed for N series firmware 7.3.5 and earlier. In most N series environments, map the root volume as LUN 1, and skip the steps for LUN 0.

1. In the XIV GUI host view, right-click the host name and select **Modify LUN Mapping** as shown in Figure 10-12.

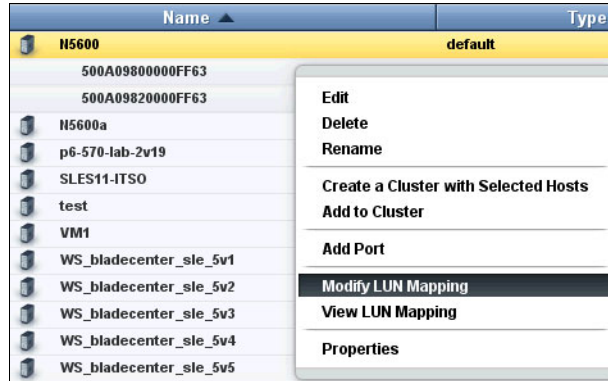


Figure 10-12 Selecting Modify LUN Mapping

2. Right-click **LUN 0** and select **Enable** as shown in Figure 10-13.

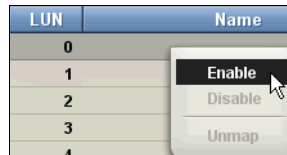


Figure 10-13 Enabling LUN 0

3. Click the 17-GB dummy volume for LUN 0, then map it to LUN 0 by clicking **Map** as illustrated in Figure 10-14.
4. Use steps 1-3 to map your N series root volume as LUN 1, also shown in Figure 10-14.



Figure 10-14 XIV Host Mapping view: LUN 0 and LUN 1 mapped correctly

Tip: If you have any problems, map the dummy XIV volume to LUN 0 and the N series root volume to LUN 1.

Fibre Channel configurations must adhere to SCSI-3 storage standards. In correctly configured storage arrays, LUN 0 is assigned to the controller (not to a disk device) and is accessible to all servers. This LUN 0 assignment is part of the SCSI-3 standard because many operating systems do not boot unless the controller is assigned as LUN 0. Assigning LUN 0 to the controller allows it to assume the critical role in discovering and reporting a list of all other LUNs available through that adapter.

In Windows, these LUNs are reported back to the kernel in response to the SCSI REPORT LUNS command. Unfortunately, not all vendor storage arrays comply with the standard of assigning LUN 0 to the controller. Failure to comply with that standard means that the boot process might not proceed correctly. In certain cases, even with LUN 0 correctly assigned, the boot LUN cannot be found, and the operating system fails to load. In these cases (without HBA LUN remapping), the kernel finds LUN 0, but might not be successful in enumerating the LUNs correctly.

If you are deploying an N series Gateway cluster, you must map both N series Gateway root volumes to the XIV cluster group.

10.6 Installing Data ONTAP

Follow the procedures in this section to install Data ONTAP on the XIV volume.

10.6.1 Assigning the root volume to N series Gateway

In the N series Gateway ssh shell, enter `disk show -v` to see the mapped disk as illustrated in Example 10-2.

Example 10-2 Running the disk show -v command

```
*> disk show -v
Local System ID: 118054991
```

DISK	OWNER	POOL	SERIAL NUMBER	CHKSUM
-----	-----	----	-----	-----
Primary_SW2:6.126L0	Not Owned	NONE	13000CB11A4	Block
Primary_SW2:6.126L1	Not Owned	NONE	13000CB11A4	Block

```
*>
```

Tip: If you do not see any disks, make sure that you have Data ONTAP 7.3.3 or later. If you must upgrade, follow the N series documentation to run a netboot update.

Assign the root LUN to the N series Gateway with `disk assign <disk name>` as shown in Example 10-3.

Example 10-3 Running the disk assign all command

```
*> disk assign Primary_SW2:6.126L1
Wed Ocdisk assign: Disk assigned but unable to obtain owner name. Re-run 'disk
assign' with -o option to specify name.t
  6 14:03:07 GMT [diskown.changingOwner:info]: changing ownership for disk
Primary_SW2:6.126L1 (S/N 13000CB11A4) from unowned (ID -1) to (ID 118054991)
*>
```

Verify the newly assigned disk by entering the `disk show` command as shown in Example 10-4.

Example 10-4 Running the disk show command

```
*> disk show
Local System ID: 118054991
```

DISK	OWNER	POOL	SERIAL NUMBER
Primary_SW2:6.126L1	(118054991)	Poo10	13000CB11A4

10.6.2 Installing Data ONTAP

To proceed with the Data ONTAP installation, complete these steps:

1. Stop Maintenance mode with `halt` as illustrated in Example 10-5.

Example 10-5 Stopping maintenance mode

```
*> halt
```

Phoenix TrustedCore(tm) Server
Copyright 1985-2004 Phoenix Technologies Ltd.
All Rights Reserved
BIOS version: 2.4.0
Portions Copyright (c) 2006-2009 NetApp All Rights Reserved
CPU= Dual Core AMD Opteron(tm) Processor 265 X 2
Testing RAM
512MB RAM tested
8192MB RAM installed
Fixed Disk 0: STEC NACF1GM1U-B11

Boot Loader version 1.7
Copyright (C) 2000-2003 Broadcom Corporation.
Portions Copyright (C) 2002-2009 NetApp

CPU Type: Dual Core AMD Opteron(tm) Processor 265
LOADER>

2. Enter **boot_ONTAP** and then press the Ctrl+C to get to the special boot menu, as shown in Example 10-6.

Example 10-6 Special boot menu

```
LOADER> boot_ontap
Loading x86_64/kernel/primary.krn:.....0x200000/46415944
0x2e44048/18105280 0x3f88408/6178149 0x456c96d/3 Entry at 0x00202018
Starting program at 0x00202018
Press CTRL-C for special boot menu
Special boot options menu will be available.
Wed Oct 6 14:27:24 GMT [nvram.battery.state:info]: The NVRAM battery is
currently ON.
Wed Oct 6 14:27:33 GMT [fci.initialization.failed:error]: Initialization
failed on Fibre Channel adapter 0d.

Data ONTAP Release 7.3.3: Thu Mar 11 23:02:12 PST 2010 (IBM)
Copyright (c) 1992-2009 NetApp.
Starting boot on Wed Oct 6 14:27:17 GMT 2010
Wed Oct 6 14:27:34 GMT [fci.initialization.failed:error]: Initialization
failed on Fibre Channel adapter 0b.
Wed Oct 6 14:27:37 GMT [diskown.isEnabled:info]: software ownership has been
enabled for this system

(1) Normal boot.
(2) Boot without /etc/rc.
(3) Change password.
(4) Initialize owned disk (1 disk is owned by this filer).
(4a) Same as option 4, but create a flexible root volume.
(5) Maintenance mode boot.
Selection (1-5)? 4a
```

3. Select option 4a to install Data ONTAP.

Remember: Use (4a) flexible root volumes because this option is far more flexible, and allows more expansion and configuration options than option 4.

4. The N series installs Data ONTAP, and also prompts for environment questions such as IP address and netmask.

10.6.3 Updating Data ONTAP

After Data ONTAP installation is finished and you enter all the relevant information, update Data ONTAP. An update is needed because the installation from special boot menu is a limited installation. Follow normal N series update procedures to update Data ONTAP to run a full installation.

Transfer the correct code package to the root volume in directory /etc/software. To transfer the package from Windows, complete these steps:

1. Start cifs and map c\$ of the N series Gateway.
2. Go to the /etc directory and create a folder called software.

3. Copy the code package to the software folder.
4. When the copy is finished, run **software update <package name>** from the N series Gateway shell.

Tip: Always assign a second LUN to use as the core dump LUN. The size that you need depends on the hardware. Consult the interoperability matrix to find the appropriate size.

10.6.4 Adding data LUNs to N series Gateway

Adding data LUNs to N series Gateway is same procedure as adding the root LUN. However, the maximum LUN size that Data ONTAP can handle is 2 TB. To reach the maximum of 2 TB, you must consider the following calculation.

As previously mentioned, N series expresses GB differently than XIV. A transformation is required to determine the exact size for the XIV LUN. N series expresses GB as 1000 x 1024 x 1024 bytes, whereas XIV uses GB as 1000 x 1000 x 1000 bytes.

For Data ONTAP 2-TB LUN, the XIV size expressed in GB must be $2000 \times (1000 \times 1024 \times 1024) / (1000 \times 1000 \times 1000) = 2097$ GB

However, the largest LUN size that can effectively be used in XIV is 2095 GB because XIV capacity is based on 17-GB increments.



ProtecTIER Deduplication Gateway connectivity

This chapter addresses specific considerations for using the IBM XIV Storage System as back-end storage for a TS7650G ProtecTIER Deduplication Gateway (3958-DD3).

For more information about TS7650G ProtecTIER Deduplication Gateway (3958-DD3), see *IBM System Storage TS7650, TS7650G, and TS7610, SG24-7652*.

This chapter includes the following sections:

- ▶ Overview
- ▶ Preparing an XIV for ProtecTIER Deduplication Gateway
- ▶ IBM SSR installs the ProtecTIER software

11.1 Overview

The ProtecTIER Deduplication Gateway is used to provide virtual tape library functions with deduplication features. Deduplication means that only the unique data blocks are stored on the attached storage.

Data deduplication is a technology that is used to reduce the amount of space that is required to store data on disk. It is achieved by storing a single copy of data that is backed up repetitively. IBM data deduplication can provide greater data reduction than previous technologies, such as Lempel-Ziv (LZ) compression and differencing, which is used for differential backups. The ProtecTIER presents virtual tapes to the backup software, making the process invisible to the backup software. The backup software runs backups as usual, but the backups are deduplicated before they are stored on the attached storage.

The effectiveness of data deduplication is dependent upon many variables, including the rate of data change, the number of backups, and the data retention period. For example, if you back up the exact same data once a week for six months, you save the first copy and do not save the next 24. This process provides a 25 to 1 data deduplication ratio. If you back up an incompressible file on week one, back up the exact same file again on week two and never back it up again, you have a 2 to 1 deduplication ratio.

In Figure 11-1, you can see ProtecTIER in a backup solution with XIV Storage System as the backup storage device. Fibre Channel attachment over switched fabric is the only supported connection mode.

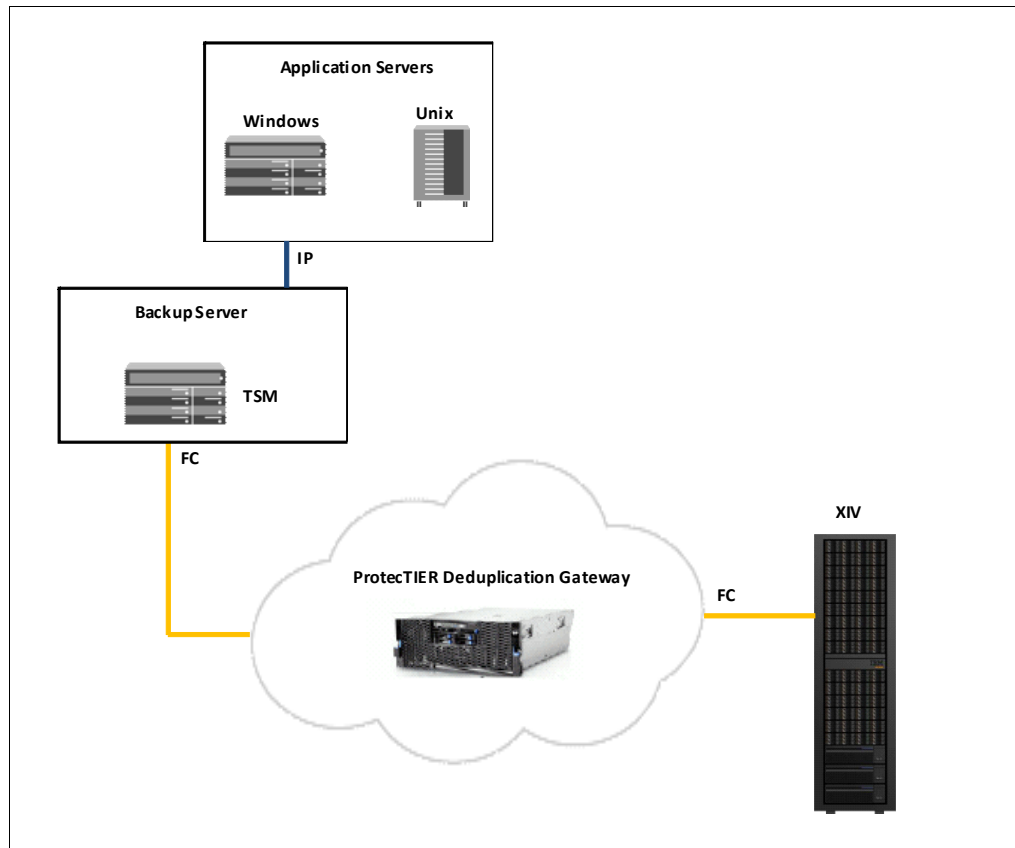


Figure 11-1 Single ProtecTIER Deduplication Gateway

TS7650G ProtecTIER Deduplication Gateway (3958-DD3) combined with IBM System Storage ProtecTIER Enterprise Edition software is designed to address the data protection needs of enterprise data centers. The solution offers high performance, high capacity, scalability, and a choice of disk-based targets for backup and archive data. TS7650G ProtecTIER Deduplication Gateway (3958-DD3) can also be ordered as a High Availability cluster, which includes two ProtecTIER nodes. The TS7650G ProtecTIER Deduplication Gateway offers the following benefits:

- ▶ Inline data deduplication that is powered by IBM HyperFactor® technology
- ▶ Multi-core virtualization and deduplication engine
- ▶ Clustering support for higher performance and availability
- ▶ Fibre Channel ports for host and server connectivity
- ▶ Performance of up to 1000 MBps or more sustained inline deduplication (two node clusters)
- ▶ Virtual tape emulation of up to 16 virtual tape libraries per single node or two-node cluster configuration
- ▶ Up to 512 virtual tape drives per two-node cluster or 256 virtual tape drives per TS7650G node
- ▶ Emulation of the IBM TS3500 tape library with IBM Ultrium 2 or Ultrium 3 tape drives
- ▶ Emulation of the Quantum P3000 tape library with DLT tape drives
- ▶ Scales to 1 PB of physical storage over 25 PB of user data

For more information about ProtecTIER, see *IBM System Storage TS7650, TS7650G, and TS7610*, SG24-7652, at:

<http://www.redbooks.ibm.com/redbooks/pdfs/sg247652.pdf>

11.2 Preparing an XIV for ProtecTIER Deduplication Gateway

When you attach the TS7650G ProtecTIER Deduplication Gateway to IBM XIV Storage System, preliminary conditions must be met. Preparation tasks must be run with these connectivity guidelines already presented in Chapter 1, “Host connectivity” on page 1:

- ▶ Check supported versions and other prerequisites
- ▶ Physical cabling in place
- ▶ Define appropriate zoning
- ▶ Create XIV pool and then volumes
- ▶ Make XIV host definitions for the ProtecTier Gateway
- ▶ Map XIV volumes to corresponding ProtecTier Gateway

The TS7650G ProtecTIER Deduplication Gateway is ordered together with ProtecTIER Software, but the ProtecTIER Software is shipped separately.

11.2.1 Supported versions and prerequisites

A TS7650G ProtecTIER Deduplication Gateway works with IBM XIV Storage System when the following prerequisites are fulfilled:

- ▶ The TS7650G ProtecTIER Deduplication Gateway (3958-DD3) and (3958-DD4) are supported.
- ▶ XIV Storage System Software is at code level 10.0.0.b or later. At the time of writing, this update code version is at 11.1.1.
- ▶ XIV Storage System must be functional before you install the TS7650G ProtecTIER Deduplication Gateway.
- ▶ The fiber connectivity must be installed, zoned, and working. Although one switch can provide basic functions, use two switches for redundancy.
- ▶ The Fibre Channel switches must be in the list of Fibre Channel switches supported by the IBM XIV Storage System. For more information, see the IBM System Storage Interoperation Center at:

<http://www.ibm.com/systems/support/storage/config/ssic>

Restriction: Direct attachment between TS7650G ProtecTIER Deduplication Gateway and IBM XIV Storage System is not supported.

11.2.2 Fibre Channel switch cabling

For maximum performance with an IBM XIV Storage System, connect all available XIV Interface Modules and use all of the back-end ProtecTier ports. For redundancy, connect Fibre Channel cables from TS7650G ProtecTIER Deduplication Gateway to two Fibre Channel (FC) switched fabrics.

If a single IBM XIV Storage System is being connected, each Fibre Channel switched fabric must have six available ports for Fibre Channel cable attachment to IBM XIV Storage System. Generally, use two connections for each interface module in XIV.

Typically, XIV interface module port 1 is used for Fibre Channel switch 1, and port 3 for switch 2 (Figure 11-2). When using a partially configured XIV rack, see Figure 1-1 on page 4 to locate the available FC ports.

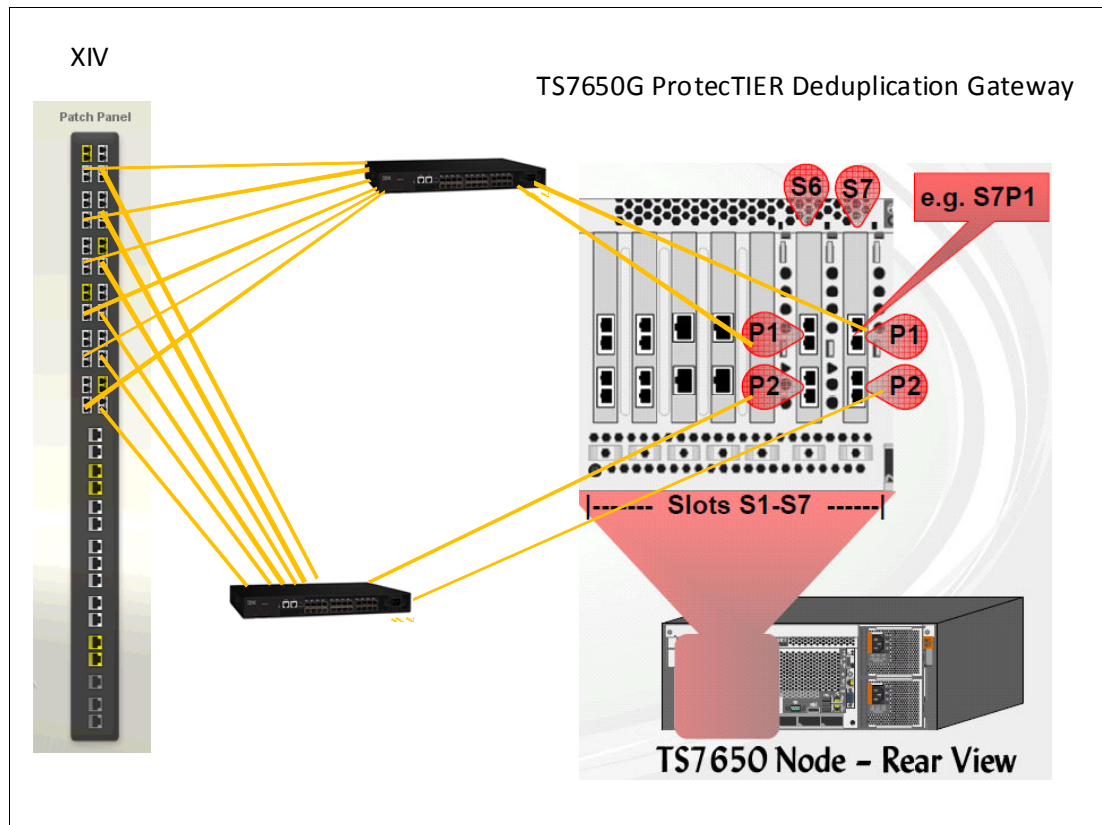


Figure 11-2 Cable diagram for connecting a TS7650G to IBM XIV Storage System

11.2.3 Zoning configuration

For each TS7650G disk attachment port, multiple XIV host ports are configured into separate isolated zone pairing in a 1:1 manner:

- ▶ All XIV Interface Modules on port 1 are zoned to the ProtecTIER host bus adapters (HBAs) in slot 6 port 1 and slot 7 port 1
- ▶ All XIV Interface Modules in port 3 are zoned to the ProtecTIER HBAs in slot 6 port 2 and slot 7 port 2

Each interface module in IBM XIV Storage System has connection with both TS7650G HBAs. Typical ProtecTIER configuration uses 1:1 zoning (one initiator and one target in each zone) to create zones. These zones allow the connection of a single ProtecTIER node with a 15 module IBM XIV Storage System with all six Interface Modules. See Example 11-1.

Example 11-1 Zoning example for an XIV Storage System attach

Switch 1:

- Zone 01: PT_S6P1, XIV_Module4Port1
- Zone 02: PT_S6P1, XIV_Module6Port1
- Zone 03: PT_S6P1, XIV_Module8Port1
- Zone 04: PT_S7P1, XIV_Module5Port1
- Zone 05: PT_S7P1, XIV_Module7Port1
- Zone 06: PT_S7P1, XIV_Module9Port1

Switch 02:

- Zone 01: PT_S6P2, XIV_Module4Port3
- Zone 02: PT_S6P2, XIV_Module6Port3
- Zone 03: PT_S6P2, XIV_Module8Port3
- Zone 04: PT_S7P2, XIV_Module5Port3
- Zone 05: PT_S7P2, XIV_Module7Port3
- Zone 06: PT_S7P2, XIV_Module9Port3

This example has the following characteristics:

- ▶ Each ProtecTIER Gateway back-end HBA port sees three XIV interface modules.
- ▶ Each XIV interface module is connected redundantly to two separate ProtecTIER back-end HBA ports.
- ▶ There are 12 paths (4 x 3) to one volume from a single ProtecTIER Gateway node.

11.2.4 Configuring XIV Storage System for ProtecTIER Deduplication Gateway

An IBM System Service Representative (SSR) uses the ProtecTIER Capacity Planning Tool to size the ProtecTIER repository metadata and user data. Capacity planning is always different because it depends heavily on your type of data and expected deduplication ratio. The planning tool output includes the detailed information about all volume sizes and capacities for your specific ProtecTIER installation. If you do not have this information, contact your IBM SSR to get it.

An example for XIV is shown in Figure 11-3.

The screenshot shows the 'Meta Data Planner' interface. At the top, it displays 'Company: IBM' and 'Workload: IBM Best Practices'. On the right, it shows 'Created: 01-Sep-11' and 'Updated: ' with a 'By:' field. The main configuration area includes: 'Model: TS7650G', 'Repository Size: 79', 'Raid Type: FC-15K8+8', 'Include Growth: []', 'Release: 2.5.1.0', 'Factoring Ratio: 12.1', 'Drive Capacity: 600', 'Planner: 2.5.1', 'Max Throughput: 1,000', and 'Emulation: VTL'. Below this is a 'Meta Data Configuration' section with a 'Results: Meta Data:' tab. It shows 'Capacity (GB): 2,888' and 'Spindles: 32'. A 'File Systems (GB)' table is also present with two rows: '1-15:' and '16-30:'. The '1-15:' row has values: 1, 1,316, 1,571, and then seven 0s. The '16-30:' row has values: 0, 0, 0, and then seven 0s.

Figure 11-3 ProtecTIER Capacity Planning Tool example

Tip: In the capacity planning tool for metadata, the fields RAID Type and Drive capacity show the most optimal choice for an XIV Storage System. The Factoring Ratio number is directly related to the size of the metadata volumes, and can be estimated by using the IBM ProtecTIER Performance Calculator.

Be sure to take the Max Throughput and Repository Size values into account during the calculations for both the initial install and future growth.

You must configure the IBM XIV Storage System before the ProtecTIER Deduplication Gateway is installed on it by an IBM SSR. Complete these steps:

- ▶ Configure one storage pool for ProtecTIER Deduplication Gateway. Set snapshot space to zero because creating snapshots on IBM XIV Storage System is not supported by ProtecTIER Deduplication Gateway.
- ▶ Configure the IBM XIV Storage System into volumes. Follow the ProtecTIER Capacity Planning Tool output. The capacity planning tool output gives you the metadata volume size and the size of the 32 data volumes. Configure a Quorum volume with a minimum of 1 GB as well, in case the solution needs more ProtecTIER nodes in the future.
- ▶ Map the volumes to ProtecTIER Deduplication Gateway, or, if you have a ProtecTIER Deduplication Gateway cluster, map the volumes to the cluster.

Example of configuring an IBM XIV Storage System

Create a Storage pool for the capacity you want to use for ProtecTIER Deduplication Gateway with the XIV GUI as shown in Figure 11-4.



Figure 11-4 Creating a storage pool in the XIV GUI

Tip: Use a Regular Pool and zero the snapshot reserve space. Snapshots and thin provisioning are not supported when XIV is used with ProtecTIER Deduplication Gateway.

In the example in Figure 11-3 on page 286 with a 79-TB XIV Storage System and a deduplication Factoring Ratio of 12, the volumes sizes are as follows:

- ▶ 2x 1571-GB volumes for metadata. Make these volumes equal to each other, and nearest to XIV allocation size, in this case 1583.
- ▶ 1x 17 G volume for Quorum (it must be 17 GB because that is the XIV min size).
- ▶ 32 x *<Remaining Pool Space available>*, which is 75440. Dividing 75440 by 32 means that user data LUNs on the XIV are 2357 GB each. The XIV 3.0 client GUI makes this calculation easy for you. Enter the number of Volumes to create, then drag the slider to the right to fill the entire pool. The GUI automatically calculates the appropriate equivalent amount.

Figure 11-5 shows the creation of the metadata volumes.

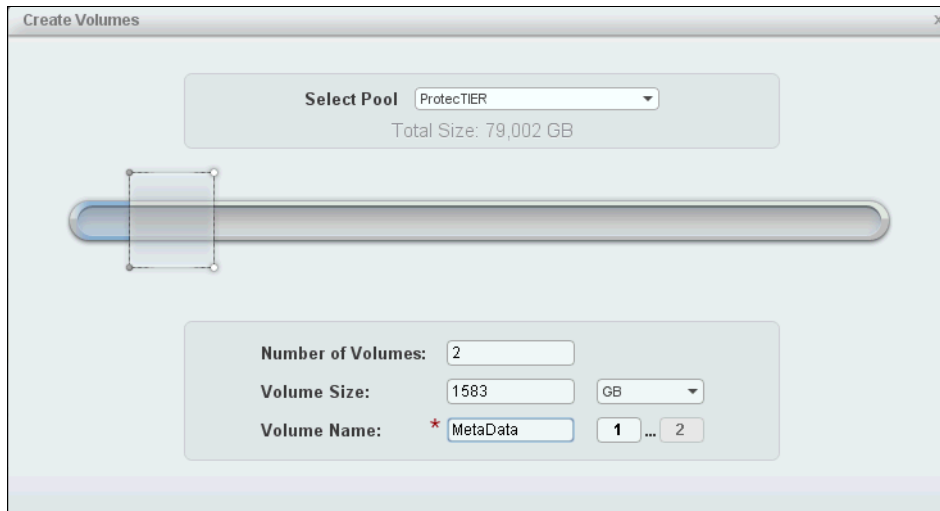


Figure 11-5 Creating metadata volumes

Figure 11-6 shows the creation of the Quorum volume.

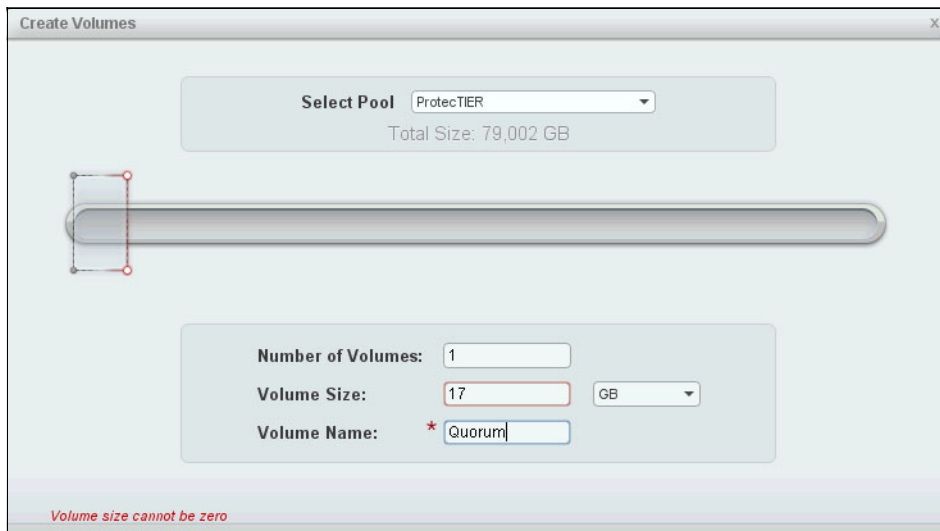


Figure 11-6 Creating a Quorum volume

Figure 11-7 shows the creation of volumes for user data. The arrows show dragging the slider to use all of the pool. This action automatically calculates the appropriate size for all volumes.

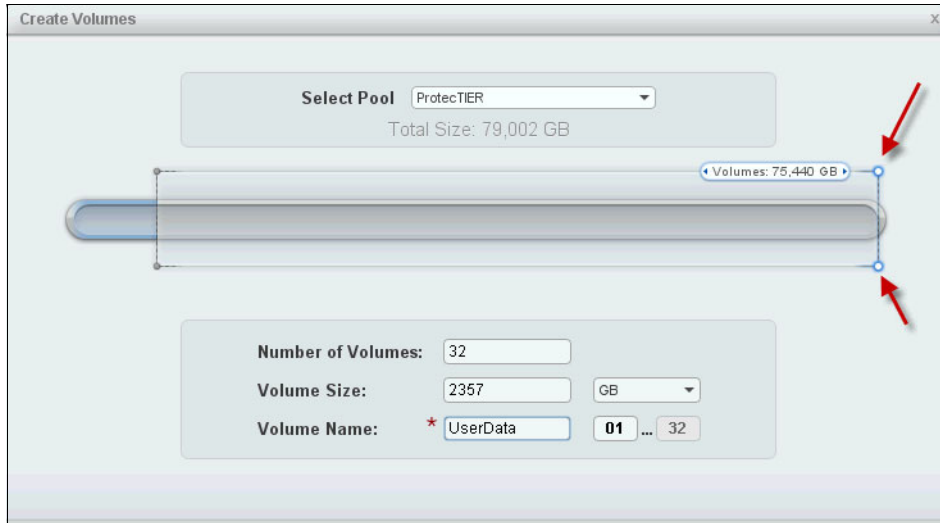


Figure 11-7 Creating User Data volumes

If you have a ProtectTIER Gateway cluster (two ProtectTIER nodes in a High Availability solution), complete these steps:

1. Create a cluster group (Figure 11-8).

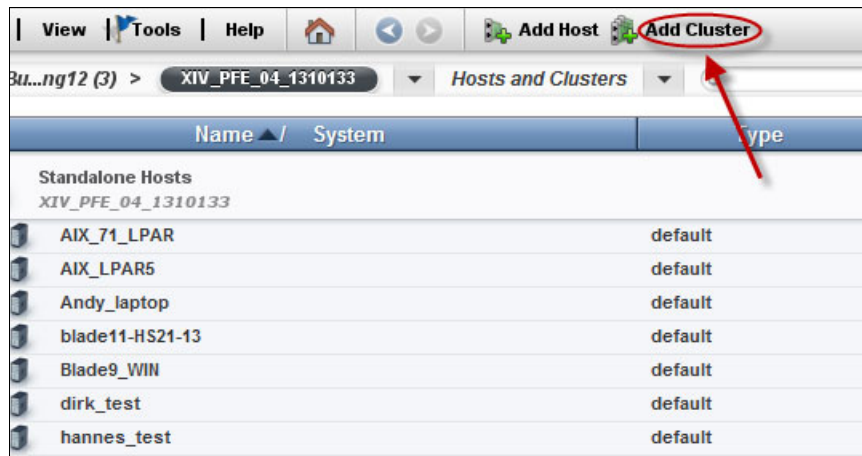


Figure 11-8 Adding Cluster ProtectTIER

2. Add a host that is defined for each node to that cluster group.

3. Create a cluster definition for the high available ProtecTIER cluster (Figure 11-9).

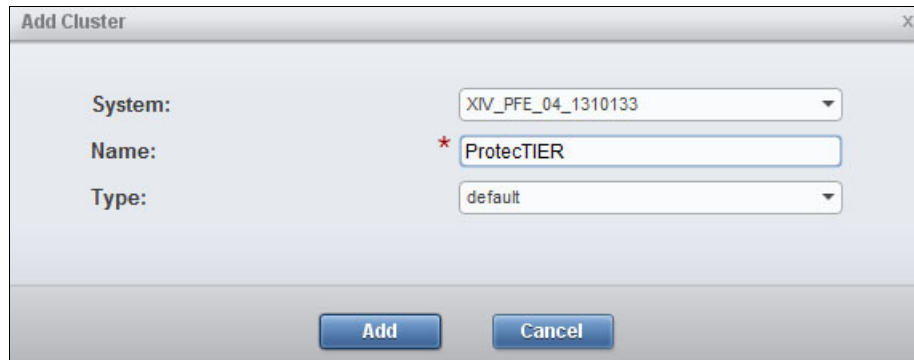


Figure 11-9 Defining cluster ProtecTIER

4. Highlight the cluster and select **Add Host** (Figure 11-10).

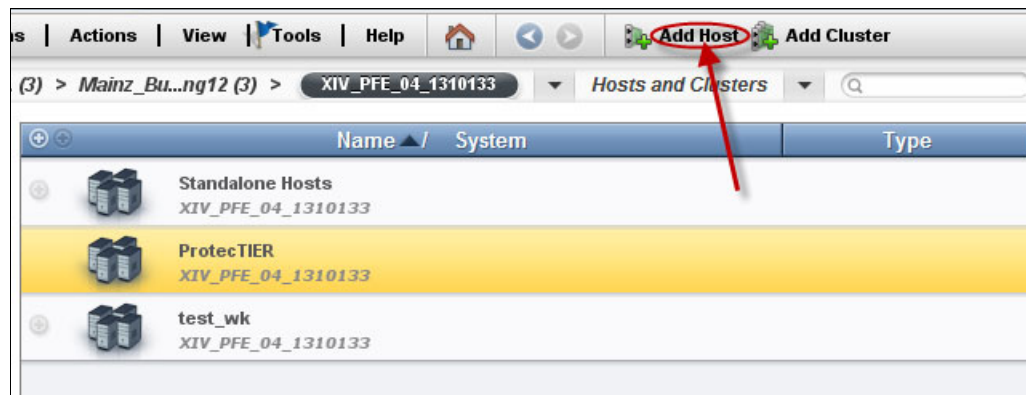


Figure 11-10 Adding Host

5. Enter the information for the new ProtecTIER host and click **Add** (Figure 11-11).

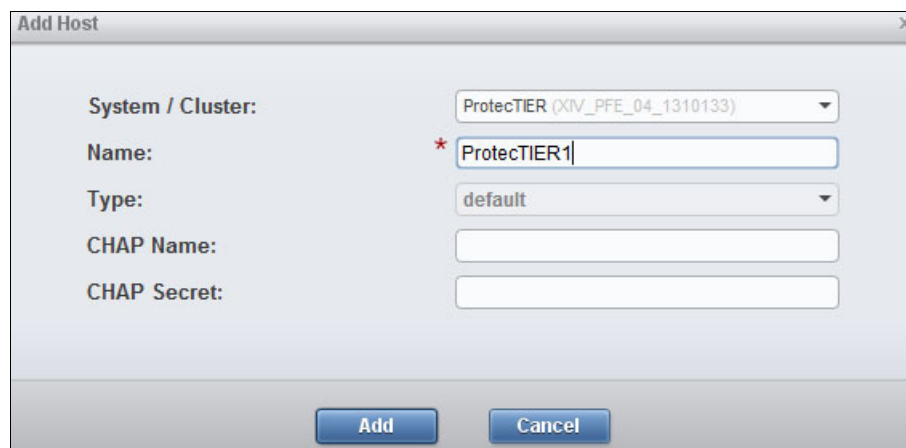


Figure 11-11 Adding the Host ProtecTIER Node 1

6. Add the additional ProtecTIER Host Node2 by following the previous steps.
7. Find the WWPNs of the ProtecTIER nodes. WWPNs can be found in the name server of the Fibre Channel switch. If zoning is in place, they are selectable from the menu.

Alternatively they can also be found in the BIOS of the HBA cards and then entered by hand in the XUICV GUI.

- Click **Add Port** to add the WWPNs of both nodes to the ProtecTIER Gateway hosts as shown in Figure 11-12.

Figure 11-12 Add Port Name window

Figure 11-13 shows the Cluster with Hosts and WWPNs.

ProtectTIER XIV_PFE_04_1310133			
ProtectTIER Node 1	default	ProtectTIER	
10000000C9660C20	FC		
ProtectTIER Node 2	default	ProtectTIER	
10000000C9660C21	FC		

Figure 11-13 Host view of the ProtecTIER cluster

- Map the volumes to the ProtecTIER Gateway cluster. In the XIV GUI, right-click the cluster name, or on the host if you have only one ProtecTIER node, and select **Modify LUN Mapping** (Figure 11-14).

Name	Size (GB)	LUN	Name	Size (GB)	Serial
volume-0000001	86.0	0			
volume-0000002	120.0	1	volume-0000001	86	10,681
volume-0000003	51.0	2	volume-0000002	120	10,823
volume-0000005	51.0	3	volume-0000003	51	8,412
volume-0000007	68.0	4	volume-0000005	51	10,516
WK_mirror_test_1	103.0	5	volume-0000007	68	6,170
WK_mirror_test_2	103.0	6			
WK_mirror_test_3	103.0	7			
WK_mirror_test_4	103.0	8			
WK_mirror_test_5	103.0	9			
WK_mirror_test_6	103.0	10			
		11			
		12			

Figure 11-14 Mapping LUNs to ProtecTIER cluster

Tip: If you have only one ProtecTIER Gateway node, map the volumes directly to the ProtecTIER gateway node.

11.3 IBM SSR installs the ProtecTIER software

The IBM SSR can now install ProtecTIER software on the ProtecTIER Gateway nodes. The repository setup window is shown in Figure 11-15.

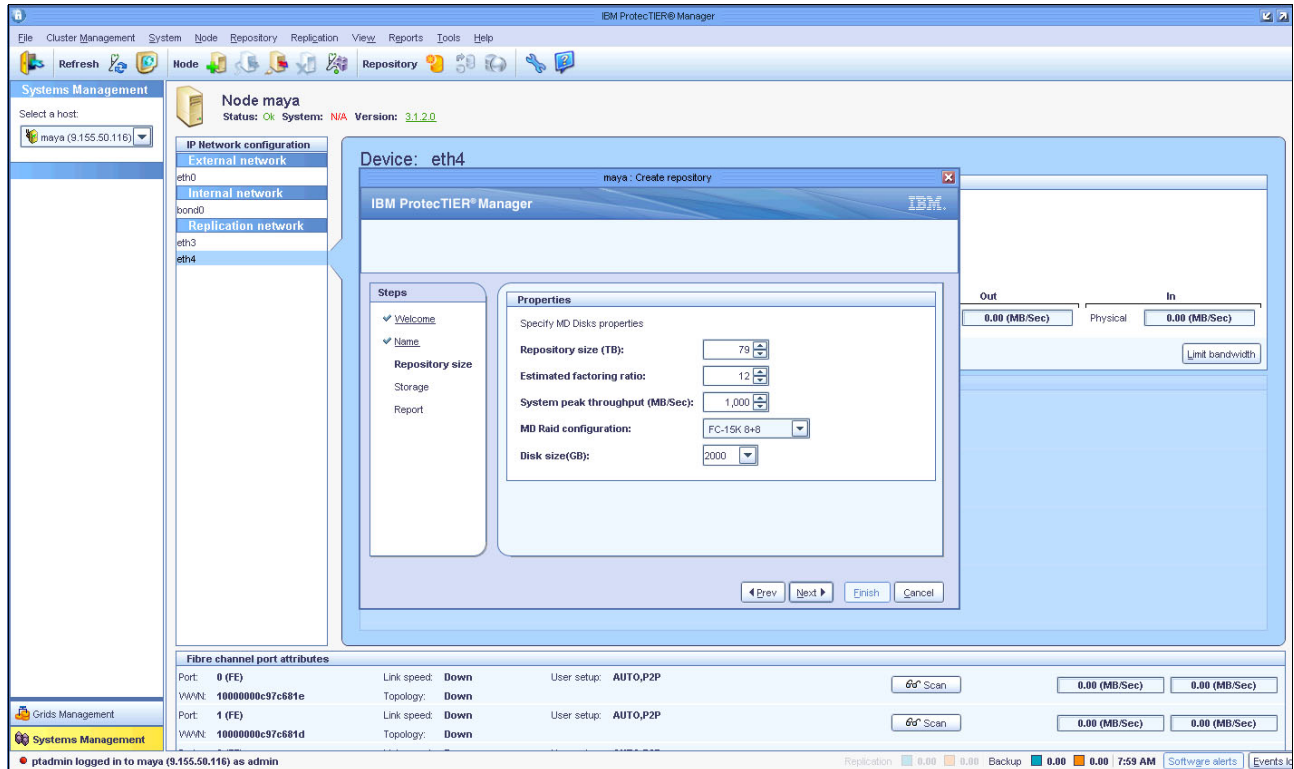


Figure 11-15 ProtecTIER Administration window during the setup of the Repository, on XIV



XIV in database and SAP application environments

This chapter provides guidelines on how to use the IBM XIV Storage System, Microsoft SQL Server, Oracle, and DB2 database application environments. It also includes guidelines for SAP environments.

The chapter focuses on the storage-specific aspects of space allocation for database environments. If I/O bottlenecks show up in a database environment, a performance analysis of the complete environment might be necessary. Look at the database engine, file systems, operating system, and storage. The chapter also gives hints, tips, and web links for more information about the non-storage components of the environment.

This chapter contains the following sections:

- ▶ XIV volume layout for database applications
- ▶ Guidelines for SAP
- ▶ Database Snapshot backup considerations

12.1 XIV volume layout for database applications

The XIV Storage System uses a unique process to balance data and I/O across all disk drives within the storage system. This data distribution method is fundamentally different from conventional storage subsystems, and significantly simplifies database management considerations. Conventional storage systems require detailed volume layout requirements to allocate database space for optimal performance. This effort is not required for the XIV Storage System.

Most storage vendors publish recommendations on how to distribute data across the storage system resources to achieve optimal I/O performance. Unfortunately, the original setup and tuning cannot usually be kept over the lifetime of an application environment. Because applications change and storage landscapes grow, traditional storage systems must be constantly *retuned* to maintain optimal performance. One common, less-than-optimal solution is providing more storage capacity on a best effort level, which tends to cause I/O performance to deteriorate.

This aging process that affects application environments on many storage systems does not occur with the XIV architecture because of these advantages:

- ▶ Volumes are always distributed across all disk drives.
- ▶ Volumes can be added or resized without downtime.
- ▶ Applications get maximized system and I/O power regardless of access patterns.
- ▶ Performance hotspots do not exist.

Therefore, you do not need to develop a performance-optimized volume layout for database application environments with XIV. However, it is worth considering some configuration aspects during setup.

12.1.1 Common guidelines

The most unique aspect of XIV is its inherent ability to use all resources within its storage subsystem regardless of the layout of the data. However, to achieve maximum performance and availability, there are a few guidelines:

- ▶ For data, use a few large XIV volumes (typically 2 - 4 volumes). Make each XIV volume between 500 GB and 2 TB in size, depending on the database size. Using a few large XIV volumes takes better advantage of the aggressive caching technology of XIV and simplifies storage management.
- ▶ When you create the XIV volumes for the database application, make sure to plan for the extra capacity required. Keep in mind that XIV shows volume sizes in base 10 (1 KB = 1000 B). Operating systems sometimes show them in base 2 (1 KB = 1024 B). In addition, the file system also claims some storage capacity.
- ▶ Place your data and logs on separate volumes. With this configuration, you can recover to a certain point-in-time instead just going back to the last consistent snapshot image after database corruption occurs. In addition, certain backup management and automation tools such as IBM Tivoli® FlashCopy® Manager require separate volumes for data and logs.
- ▶ If more than one XIV volume is used, implement an XIV consistency group with XIV snapshots. This configuration is needed if the volumes are in the same storage pool.
- ▶ XIV offers thin provisioning storage pools. If the volume manager of the operating system fully supports thin provisioned volumes, consider creating larger volumes than needed for the database size.

12.1.2 Oracle database

Oracle database server without the ASM option does not stripe table space data across the corresponding files or storage volumes. Thus the common guidelines still apply.

Asynchronous I/O is preferable for an Oracle database on an IBM XIV Storage System. The Oracle database server automatically detects if asynchronous I/O is available on an operating system. Nevertheless, ensure that asynchronous I/O is configured. Asynchronous I/O is explicitly enabled by setting the Oracle database initialization parameter `DISK_ASYNCH_IO` to `TRUE`.

For more information about Oracle asynchronous I/O, see *Oracle Database High Availability Best Practices 11g Release 1* and *Oracle Database Reference 11g Release 1*, available at:

http://www.oracle.com/pls/db111/portal.all_books

12.1.3 Oracle ASM

Oracle Automatic Storage Management (ASM) is an alternative storage management solution to conventional volume managers, file systems, and raw devices.

The main components of Oracle ASM are disk groups. Each group includes several disks (or volumes of a disk storage system) that ASM controls as a single unit. ASM refers to the disks/volumes as ASM disks. ASM stores the database files in the ASM disk groups. These files include data files, online and offline redo logs, control files, data file copies, and Recovery Manager (RMAN) backups. However, Oracle binary and ASCII files, such as trace files, cannot be stored in ASM disk groups. ASM stripes the content of files that are stored in a disk group across all disks in the disk group to balance I/O workloads.

When you configure Oracle database using ASM on XIV, follow these guidelines to achieve better performance. These guidelines also create a configuration that is easy to manage and use.

- ▶ Use one or two XIV volumes to create an ASM disk group
- ▶ Set 8M or 16M Allocation Unit (stripe) size

With Oracle ASM, asynchronous I/O is used by default.

12.1.4 IBM DB2

DB2 offers two types of table spaces for databases: *System managed space (SMS)* and *database managed space (DMS)*. SMS table spaces are managed by the operating system. The operating system stores the database data in file system directories that are assigned when a table space is created. The file system manages the allocation and management of media storage. DMS table spaces are managed by the database manager. The DMS table space definition includes a list of files (or devices) into which the database data are stored. The files, directories, or devices where data are stored are also called *containers*.

To achieve optimum database performance and availability, take advantage of the following unique capabilities of XIV and DB2. This list focuses on the physical aspects of XIV volumes and how these volumes are mapped to the host.

- ▶ When you create a database, consider using DB2 automatic storage technology as a simple and effective way to provision storage for a database. If you use more than one XIV volume, automatic storage distributes the data evenly among the volumes. Avoid using other striping methods such as the logical volume manager of the operating system. DB2

automatic storage is used by default when you create a database by using the CREATE DATABASE command.

- ▶ If more than one XIV volume is used for data, place the volumes in a single XIV consistency group. In a partitioned database environment, create one consistency group per partition. Pooling all data volumes together per partition facilitates XIV creating a consistent snapshot of all volumes within the group. Do not place your database transaction logs in the same consistency group as the data they describe.
- ▶ For log files, use only one XIV volume and match its size to the space required by the database configuration guidelines. Although the ratio of log storage capacity is heavily dependent on workload, a good general rule is 15% to 25% of total allocated storage to the database.
- ▶ In a partitioned DB2 database environment, use separate XIV volumes per partition to allow independent backup and recovery of each partition.

12.1.5 DB2 parallelism options for Linux, UNIX, and Windows

When there are multiple containers for a table space, the database manager can use parallel I/O. *Parallel I/O* is the process of writing to, or reading from, two or more I/O devices simultaneously. It can result in significant improvements in throughput. DB2 offers two types of query parallelism:

- ▶ *Interquery parallelism* is the ability of the database to accept queries from multiple applications at the same time. Each query runs independently of the others, but DB2 runs all of them at the same time. DB2 database products have always supported this type of parallelism.
- ▶ *Intraquery parallelism* is the simultaneous processing of parts of a single query, by using either intrapartition parallelism, interpartition parallelism, or both.

Prefetching is important to the performance of intrapartition parallelism. DB2 retrieves one or more data or index pages from disk in the expectation that they are required by an application.

The DB2_PARALLEL_IO registry variable influences parallel I/O in a table space. With parallel I/O off, the parallelism of a table space is equal to the number of containers. With parallel I/O on, the parallelism of a table space is equal to the number of containers multiplied by the value that is given in the DB2_PARALLEL_IO registry variable.

In IBM lab tests, the best performance was achieved with XIV Storage System by setting this variable to 32 or 64 per table space. Example 12-1 shows how to configure DB2 parallel I/O for all table spaces with the **db2set** command on AIX.

Example 12-1 Enabling DB2 parallel I/O

```
# su - db2xiv
$ db2set DB2_PARALLEL_IO=*:64
```

For more information about DB2 parallelism options, see *DB2 for Linux, UNIX, and Windows Information Center* available at:

<http://publib.boulder.ibm.com/infocenter/db21uw/v9r7>

12.1.6 Microsoft SQL Server

Organizations that use Microsoft SQL Server 2008 R2 in a business-critical database environment require high performance and availability. Enterprise-class storage software such as the IBM XIV Storage System more than satisfy this requirement. To achieve optimal performance and growth results, follow the experience-based guidelines in this section.

Database storage and server configuration

For optimum SQL 2008 R2 performance and availability, take advantage of the unique capabilities of XIV and SQL.

Both SQL and XIV perform best when host bus adapter (HBA) queue depths are high. SQL Server applications are generally I/O-intensive, with many concurrent outstanding I/O requests. As a result, the HBA queue depth must be high enough for optimal performance. By increasing the HBA queue depth, greater amounts of parallel I/O get distributed as a result of the XIV grid architecture. To maximize the benefits of the XIV parallel architecture, use a queue depth of 128 for all HBAs attaching to SQL servers.

Remember: Although a higher queue depth in general yields better performance with the XIV, consider the XIV Fibre Channel (FC) port limitations. The FC ports of the XIV can handle a maximum queue depth of 1400.

XIV volume design

XIV distributes data for each volume across all disks regardless of the quantity or size of the volumes. Its grid architecture handles most of the performance tuning and self-healing without intervention. However, to achieve maximum performance and availability, consider the following guidelines:

- ▶ For data files, use a few large volumes (typically 2 - 4 volumes). Make each volume between 500G - 2 TB, depending on the database size. XIV is optimized to use all drive spindles for each volume. Using small numbers of large volumes takes better advantage of the aggressive caching technology of XIV and simplifies storage management. The grid architecture of XIV is different from the traditional model of small database RAID arrays and a large central cache.
- ▶ For log files, use only a single XIV volume.
- ▶ In a partitioned database environment, use separate XIV volumes per partition to enable independent backup and recovery of each partition.
- ▶ Place database and log files on separate volumes. If database corruption occurs, placing them on separate volumes allows point-in-time recovery rather than recovery of the last consistent snapshot image. In addition, some backup management and automation tools, such as Tivoli FlashCopy Manager, require separate volumes for data and logs.
- ▶ Create XIV consistency groups to take simultaneous snapshots of multiple volumes that are concurrently used by SQL. Keep in mind that volumes can belong only to a single consistency group. There are a few different SQL CG snapshot concepts that are based on backup/recovery preferences. For full database recoveries, place data and log volumes in the same consistency group. For point-in-time recovery, create two separate consistency groups: One for logs and one for data. Creating two groups ensures that the logs do not get overwritten, thus allowing point-in-time transaction log restores. You can also use XIV consistency group snapshots with your preferred transaction log backup practices.

Using XIV grid architecture with SQL I/O parallelism

The overall design of the XIV grid architecture excels with applications that employ multiple threads to handle the parallel execution of I/O from a single server. Multiple threads from multiple servers perform even better.

In a SQL environment, there are several ways to achieve parallelism to take advantage of the XIV grid architecture:

- ▶ Inter-query parallelism: A single database can accept queries from multiple applications simultaneously. Each query runs independently and simultaneously.
- ▶ Intra-query parallelism: Simultaneous processing of parts of a single query that uses inter-partition parallelism, intra-partition parallelism, or both.
 - Intra-partition parallelism: A single query is broken into multiple parts.
 - Inter-partition parallelism: A single query is broken into multiple parts across multiple partitions of a partitioned database on a single server or multiple servers. The query runs in parallel.
- ▶ Depending on the server hardware and database solution, the maximum degree of parallelism (MAXDOP) can be configured. For more information about configuring the MAXDOP option in SQL, see the Microsoft Knowledge Base topic at:
<http://support.microsoft.com/kb/2806535/en-us>
- ▶ SQL backups and restores are I/O intensive. SQL Server uses both I/O parallelism and intra-partition parallelism when it runs backup and restore operations. Backups use I/O parallelism by reading from multiple database files in parallel, and asynchronously writing to multiple backup media in parallel.
- ▶ For batch jobs that are single threaded with time limitations, follow these guidelines:
 - Use large database buffers to take advantage of prefetching. Allocate as much server memory as possible.
 - Run multiple batch jobs concurrently. Even though each batch job takes approximately the same amount of time, the overall time frame for combined tasks is less.
 - If possible, break down large batch jobs into smaller jobs. Schedule the jobs to run simultaneously from the same host or multiple hosts. For example, instead of backing up data files sequentially, back up multiple data files concurrently.
 - When you run queries, DDL, DML, data loading, backup, recovery, and replication, follow guidelines that enhance parallel execution of those tasks.

Microsoft SQL Server automatically tunes many of the server configuration options, so little, if any, tuning by a system administrator is required. When performance tuning SQL, thoroughly test the databases. Many factors can influence the outcome, including custom stored procedures and applications. Generally, avoid using too many performance modifications to keep the storage configuration simple and streamlined.

For more information, see *IBM XIV Storage Tips for Microsoft SQL Server 2008 R2* at:

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101758>

12.2 Guidelines for SAP

Many organizations rely on integrated SAP solutions to run almost every aspect of their business operations rapidly and efficiently. Typically, the SAP applications at the heart of the business are mission-critical, and without them, enterprise operations are severely affected. Therefore, ensure that underlying IT infrastructure can provide the necessary performance, productivity, and reliability to support the SAP landscape. Issues such as system complexity and management play a significant role in forming a suitable infrastructure strategy that meets the business need.

In general, SAP stores all data in one of the following external databases:

- ▶ DB2 for Linux, UNIX, Windows from IBM
- ▶ DB2 for IBM i from IBM
- ▶ MaxDB from SAP
- ▶ MS SQL Server from Microsoft
- ▶ Oracle from Oracle
- ▶ Sybase from SAP

Normally the transaction load of a non-SAP database differs greatly from the load behavior of an SAP application server. Often non-SAP databases have many random write with 4k blocks to support small and fast transactions with a low latency. This tendency is in contrast to an SAP online transaction processing (OLTP) System, which has mostly a 70/30% read/write workload. SAP Business Warehouse systems are online analytical processing (OLAP) systems, which have even more sequential shares in the I/O behavior of the database.

12.2.1 Number of volumes

The XIV Storage System distributes data across all disk drives for each allocated volume. For better performance with XIV, allocate fewer volumes of a larger size. Using the minimum number of volumes that are needed for data separation and keeping the volume sizes larger allows XIV to better use cache algorithms. Cache algorithms include pre-fetch and least recently used (LRU). The only exception is multi-pathing solutions that cannot use the I/O paths in round-robin mode, or allow round-robin with low queue depths only. In this last case, allocate the same number of volumes as the number of available paths to use all XIV interfaces. This configuration is called *static load balancing*.

12.2.2 Separation of database logs and data files

Create separate XIV volumes for database logs and data files. IBM DB2 databases also provide a separate volume or set of volumes for the local database directory. This guideline is not specific to XIV. It is valid for all storage systems, and helps to preserve data consistency with Snapshot or FlashCopy. Separate volume groups for data and logs are a prerequisite for creating data consistency. Create the following LUNs before you create a snapshot of the database environment:

- ▶ At least one for database log
- ▶ One for the SAP binary files (/usr/sap)
- ▶ One for the archive
- ▶ One for each expected terabyte of data
- ▶ One for SAPDATA, with space to grow up to 2-terabyte LUN size (including the temp space)

12.2.3 Storage pools

XIV storage pools are a logical entity. You can resize a pool or move volumes between pools with no impact on an application. XIV currently allows a maximum number of 256 storage pools. You can consolidate multiple SAP systems on one XIV Storage System, and have applications share XIV. In these cases, create separate XIV storage pools for the different SAP environments and applications. This increases clarity and eases storage management. Typically, such a system includes three different SAP storage pools for development, quality assurance, and production systems.

12.3 Database Snapshot backup considerations

The *Tivoli Storage FlashCopy Manager* software product creates consistent database snapshots backups. It offloads the data from the snapshot backups to an external backup/restore system like Tivoli Storage Manager.

Even without a specialized product, you can create a consistent snapshot of a database. To ensure consistency, the snapshot must include the database, file systems, and storage.

This section gives hints and tips to create consistent storage-based snapshots in database environments. For more information about storage-based snapshot backup of a DB2 database, see the “Snapshot” chapter of *IBM XIV Storage System: Copy Services and Migration*, SG24-7759.

12.3.1 Snapshot backup processing for Oracle and DB2 databases

If a snapshot of a database is created, particular attention must be paid to the consistency of the copy. The easiest, and most unusual, way to provide consistency is to stop the database before you create the snapshot pairs. If a database cannot be stopped for the snapshot, pre- and post-processing actions must be run to create a consistent copy.

An XIV Consistency Group comprises multiple volumes. Therefore, take the snapshot of all the volumes at the same time. It is ideal for applications that span multiple volumes that have their transaction logs on one set of volumes and their database on another.

When you create a backup of the database, synchronize the data so that it is consistent at the database level as well. If the data is inconsistent, a database restore is not possible because the log and the data are different. In this situation, part of the data might be lost.

If consistency groups and snapshots are used to back up the database, database consistency can be established without shutting down the application. To do so, complete these steps:

1. Suspend the database I/O. With Oracle, an I/O suspend is not required if the backup mode is enabled. Oracle handles the resulting inconsistencies during database recovery.
2. If the database is in file systems, write all modified file system data back to the storage system. This process flushes the file systems buffers before creating the snapshots for a *file system sync*.
3. Optionally, run file system freeze/thaw operations (if supported by the file system) before or after the snapshots. If file system freezes are omitted, file system checks are required before mounting the file systems allocated to the snapshots copies.
4. Use snapshot-specific consistency groups.

Transaction log files handling has the following considerations:

- ▶ For an offline backup of the database, create snapshots of the XIV volumes on which the data files and transaction logs are stored. A snapshot restore thus brings back a restartable database.
- ▶ For an online backup of the database, consider creating snapshots of the XIV volumes with data files only. If an existing snapshot of the XIV volume with the database transactions logs is restored, the most current logs files are overwritten. It might not be possible to recover the database to the most current point-in-time by using the forward recovery process of the database.

12.3.2 Snapshot restore

An XIV snapshot is run on the XIV volume level. Thus, a snapshot restore typically restores complete databases. Certain databases support online restores at a filegroup (Microsoft SQL Server) or table space (Oracle, DB2) level. Partial restores of single table spaces or databases files are therefore possible with these databases. However, combining partial restores with storage-based snapshots requires exact mapping of table spaces or database files with storage volumes. The creation and maintenance of such an IT infrastructure requires immense effort and is therefore impractical.

A full database restore always requires shutting down the database. If file systems or a volume manager are used on the operating system level, the file systems must be unmounted and the volume groups deactivated as well.

The following are the high-level tasks that are required to run a full database restore from a storage-based snapshot:

1. Stop application and shut down the database.
2. Unmount the file systems (if applicable) and deactivate the volume groups.
3. Restore the XIV snapshots.
4. Activate the volume groups and mount the file systems.
5. Recover database, either by using complete forward recovery or incomplete recovery to a certain point in time.
6. Start the database and application.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

For information about ordering these publications, see “How to get Redbooks” on page 304. Note that some of the documents referenced here might be available in softcopy only.

- ▶ *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- ▶ *IBM System Storage TS7650, TS7650G, and TS7610*, SG24-7652
- ▶ *IBM XIV Storage System Gen3 Architecture, Implementation, and Usage*, SG24-7659
- ▶ *IBM XIV Storage System: Copy Services and Migration*, SG24-7759
- ▶ *Implementing the IBM System Storage SAN Volume Controller V5.1*, SG24-6423
- ▶ *Introduction to Storage Area Networks and System Networking*, SG24-5470
- ▶ *Using the IBM XIV Storage System in OpenStack Cloud Environments*, REDP-4971
- ▶ *XIV Storage System in VMware Environments*, REDP-4965
- ▶ *Solid-State Drive Caching in the IBM XIV Storage System*, REDP-4842

Other publications

These publications are also relevant as further information sources:

- ▶ *IBM XIV Storage System Application Programming Interface*, GC27-3916
- ▶ *IBM XIV Storage System Host Attachment Guide: Host Attachment Kit for AIX*, GA32-0643
- ▶ *IBM XIV Storage System Host Attachment Guide: Host Attachment Kit for HPUX*, GA32-0645
- ▶ *IBM XIV Storage System Host Attachment Guide: Host Attachment Kit for Linux*, GA32-0647
- ▶ *IBM XIV Storage System Host Attachment Guide: Host Attachment Kit for Solaris*, GA32-0649
- ▶ *IBM XIV Storage System Host Attachment Guide: Host Attachment Kit for Windows*, GA32-0652
- ▶ *IBM XIV Storage System Planning Guide*, GC27-3913
- ▶ *IBM XIV Storage System Pre-Installation Network Planning Guide for Customer Configuration*, GC52-1328-01
- ▶ *IBM XIV Storage System: Product Overview*, GC27-3912
- ▶ *IBM XIV Remote Support Proxy Installation and User's Guide*, GA32-0795
- ▶ *IBM XIV Storage System User Manual*, GC27-3914
- ▶ *IBM XIV Storage System XCLI Utility User Manual*, GC27-3915

Online resources

These websites are also relevant as further information sources:

- ▶ IBM XIV Storage System Information Center:
<http://publib.boulder.ibm.com/infocenter/ibmxiv/r2/index.jsp>
- ▶ IBM XIV Storage System series website:
<http://www.ibm.com/systems/storage/disk/xiv/index.html>
- ▶ System Storage Interoperation Center (SSIC):
<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

How to get Redbooks

You can search for, view, or download Redbooks, Redpapers, Technotes, draft publications and Additional materials, as well as order hardcopy Redbooks publications, at this website:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services



Redbooks

IBM XIV Storage System Host Attachment and Interoperability

(0.5" spine)
0.475" x 0.873"
250 <-> 459 pages



IBM XIV Storage System

Host Attachment and Interoperability



Redbooks®

Integrate with DB2, VMware ESX, and Microsoft HyperV

Get operating system specifics for host side tuning

Use XIV with IBM i, N series, and ProtecTIER

This IBM Redpaper Redbooks publication provides information for attaching the IBM XIV Storage System to various host operating system platforms, including IBM i.

The book provides information and references for combining the XIV Storage System with other storage platforms, host servers, or gateways, including IBM N Series, and IBM ProtecTIER. It is intended for administrators and architects of enterprise storage systems.

The book also addresses using the XIV storage with databases and other storage-oriented application software that include:

- ▶ IBM DB2
- ▶ VMware ESX
- ▶ Microsoft HyperV
- ▶ SAP

The goal is to give an overview of the versatility and compatibility of the XIV Storage System with various platforms and environments.

The information that is presented here is not meant as a replacement or substitute for the Host Attachment kit publications. It is meant as a complement and to provide readers with usage guidance and practical illustrations.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks