

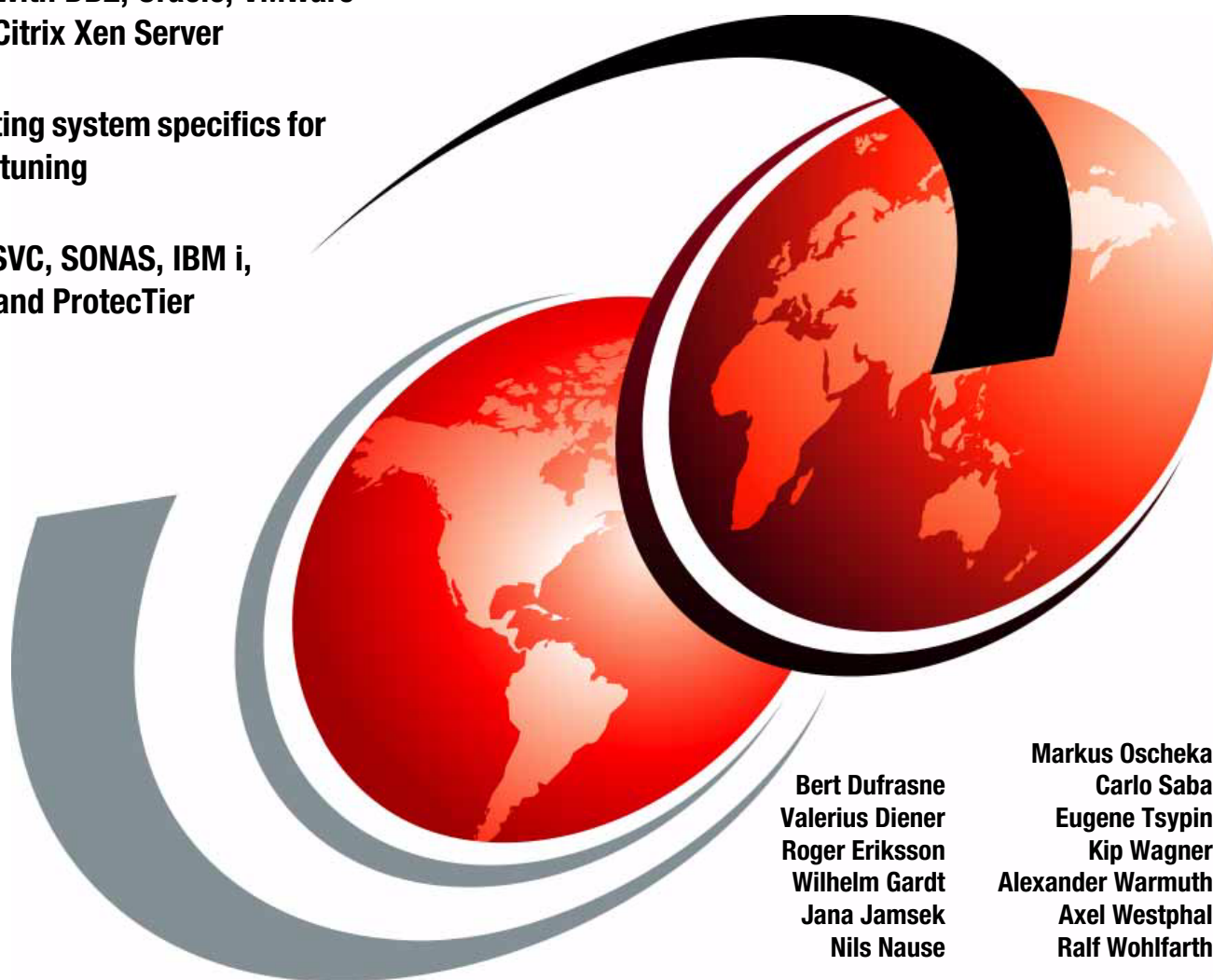
XIV Storage System

Host Attachment and Interoperability

Integrate with DB2, Oracle, VMware
ESX, and Citrix Xen Server

Get operating system specifics for
Host Side tuning

Use with SVC, SONAS, IBM i,
N Series, and ProtecTier



Bert Dufrasne
Valerius Diener
Roger Eriksson
Wilhelm Gardt
Jana Jamsek
Nils Nause

Markus Oscheka
Carlo Saba
Eugene Tsylin
Kip Wagner
Alexander Warmuth
Axel Westphal
Ralf Wohlfarth

Redbooks



International Technical Support Organization

XIV Storage System: Host Attachment and Interoperability

April 2011

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

First Edition (April 2011)

This edition applies to Version 10.2.2 of the IBM XIV Storage System Software and Version 2.5 of the IBM XIV Storage System Hardware.

© Copyright International Business Machines Corporation 2011. 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!	xiv
Comments welcome	xv
Stay connected to IBM Redbooks	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	6
1.1.3 Host Attachment Kits	7
1.1.4 FC versus iSCSI access	7
1.2 Fibre Channel (FC) connectivity	8
1.2.1 Preparation steps	8
1.2.2 FC configurations	10
1.2.3 Zoning	12
1.2.4 Identification of FC ports (initiator/target)	14
1.2.5 Boot from SAN on x86/x64 based architecture	16
1.3 iSCSI connectivity	21
1.3.1 Preparation steps	22
1.3.2 iSCSI configurations	22
1.3.3 Network configuration	24
1.3.4 IBM XIV Storage System iSCSI setup	24
1.3.5 Identifying iSCSI ports	27
1.3.6 iSCSI and CHAP authentication	28
1.3.7 iSCSI boot from XIV LUN	29
1.4 Logical configuration for host connectivity	29
1.4.1 Host configuration preparation	30
1.4.2 Assigning LUNs to a host using the GUI	32
1.4.3 Assigning LUNs to a host using the XCLI	36
1.5 Performance considerations	38
1.5.1 HBA queue depth	38
1.5.2 Volume queue depth	40
1.5.3 Application threads and number of volumes	40
1.6 Troubleshooting	41
 Chapter 2. Windows Server 2008 host connectivity	43
2.1 Attaching a Microsoft Windows 2008 host to XIV	44
2.1.1 Windows host FC configuration	45
2.1.2 Windows host iSCSI configuration	51
2.1.3 Management volume LUN 0	59
2.1.4 Host Attachment Kit utilities	59
2.1.5 Installation for Windows 2003	61
2.2 Attaching a Microsoft Windows 2003 Cluster to XIV	61
2.2.1 Prerequisites	61
2.2.2 Installing Cluster Services	62

2.3 Boot from SAN	65
Chapter 3. Linux host connectivity	67
3.1 IBM XIV Storage System and Linux support overview	68
3.1.1 Issues that distinguish Linux from other operating systems	68
3.1.2 Reference material	68
3.1.3 Recent storage related improvements to Linux.	70
3.2 Basic host attachment.	72
3.2.1 Platform specific remarks	72
3.2.2 Configure for Fibre Channel attachment.	74
3.2.3 Determine the WWPN of the installed HBAs	78
3.2.4 Attach XIV volumes to an Intel x86 host using the Host Attachment Kit.	78
3.2.5 Check attached volumes.	81
3.2.6 Set up Device Mapper Multipathing	86
3.2.7 Special considerations for XIV attachment	92
3.3 Non-disruptive SCSI reconfiguration.	94
3.3.1 Add and remove XIV volumes dynamically.	94
3.3.2 Add and remove XIV volumes in zLinux	95
3.3.3 Add new XIV host ports to zLinux	96
3.3.4 Resize XIV volumes dynamically	97
3.3.5 Use snapshots and remote replication targets	99
3.4 Troubleshooting and monitoring	102
3.5 Boot Linux from XIV volumes	104
3.5.1 The Linux boot process.	104
3.5.2 Configure the QLogic BIOS to boot from an XIV volume	105
3.5.3 OS loader considerations for other platforms	106
3.5.4 Install SLES11 SP1 on an XIV volume	107
Chapter 4. AIX host connectivity	111
4.1 Attaching XIV to AIX hosts	112
4.1.1 AIX host FC configuration	113
4.1.2 AIX host iSCSI configuration.	120
4.1.3 Management volume LUN 0	124
4.1.4 Host Attachment Kit utilities	124
4.2 SAN boot in AIX	126
4.2.1 Creating a SAN boot disk by mirroring	127
4.2.2 Installation on external storage from bootable AIX CD-ROM	128
4.2.3 AIX SAN installation with NIM.	130
Chapter 5. HP-UX host connectivity	133
5.1 Attaching XIV to a HP-UX host	134
5.2 HP-UX multi-pathing solutions	136
5.3 VERITAS Volume Manager on HP-UX	138
5.4 HP-UX SAN boot.	142
5.4.1 HP-UX Installation on external storage.	142
Chapter 6. Solaris host connectivity	147
6.1 Attaching a Solaris host to XIV	148
6.2 Solaris host FC configuration	148
6.2.1 Obtain WWPN for XIV volume mapping	148
6.2.2 Installing the Host Attachment Kit.	148
6.2.3 Configuring the host	149
6.3 Solaris host iSCSI configuration	151
6.4 Solaris Host Attachment Kit utilities for FC and iSCSI	152

6.5 Partitions and filesystems	153
6.5.1 Creating partitions and filesystems with UFS	153
Chapter 7. Symantec Storage Foundation	157
7.1 Introduction	158
7.2 Prerequisites	158
7.2.1 Checking ASL availability	158
7.2.2 Installing the XIV Host Attachment Kit	160
7.2.3 Placing XIV LUNs under VxVM control	163
7.2.4 Configure multipathing with DMP	167
7.3 Working with snapshots	168
Chapter 8. VIOS clients connectivity	171
8.1 Introduction to IBM PowerVM	172
8.1.1 IBM PowerVM overview	172
8.1.2 Virtual I/O Server	173
8.1.3 Node Port ID Virtualization	174
8.2 Planning for VIOS and IBM i	175
8.2.1 Best practices	178
8.3 Connecting an PowerVM IBM i client to XIV	179
8.3.1 Creating the Virtual I/O Server and IBM i partitions	179
8.3.2 Installing the Virtual I/O Server	182
8.3.3 IBM i multipath capability with two Virtual I/O Servers	182
8.3.4 Virtual SCSI adapters in multipath with two Virtual I/O Servers	183
8.4 Mapping XIV volumes in the Virtual I/O Server	184
8.5 Match XIV volume to IBM i disk unit	186
Chapter 9. VMware ESX host connectivity	191
9.1 Introduction	192
9.2 ESX 3.5 Fibre Channel configuration	195
9.2.1 Installing HBA drivers	196
9.2.2 Scanning for new LUNs	196
9.2.3 Assigning paths from an ESX 3.5 host to XIV	198
9.3 ESX 4.x Fibre Channel configuration	202
9.3.1 Installing HBA drivers	203
9.3.2 Identifying ESX host port WWN	203
9.3.3 Scanning for new LUNs	204
9.3.4 Attaching an ESX 4.x host to XIV	205
9.3.5 Configuring ESX 4 host for multipathing with XIV	207
9.3.6 Performance tuning tips for ESX 4 hosts with XIV	213
9.3.7 Managing ESX 4 with IBM XIV Management Console for VMWare vCenter	216
9.4 XIV Storage Replication Agent for VMware SRM	217
Chapter 10. Citrix XenServer connectivity	219
10.1 Introduction	220
10.2 Attaching a XenServer host to XIV	222
10.2.1 Prerequisites	223
10.2.2 Multi-path support and configuration	223
10.2.3 Attachment tasks	225
Chapter 11. SVC specific considerations	229
11.1 Attaching SVC to XIV	230
11.2 Supported versions of SVC	230

Chapter 12. IBM SONAS Gateway connectivity	239
12.1 IBM SONAS Gateway	240
12.2 Preparing an XIV for attachment to a SONAS Gateway	241
12.2.1 Supported versions and prerequisites	241
12.2.2 Direct attached connection to XIV	242
12.2.3 SAN connection to XIV	243
12.3 Configuring an XIV for IBM SONAS Gateway	245
12.3.1 Sample configuration	246
12.4 IBM Technician can now install SONAS Gateway	249
Chapter 13. N series Gateway connectivity	251
13.1 Overview	252
13.2 Attaching N series Gateway to XIV	252
13.2.1 Supported versions	253
13.3 Cabling	253
13.3.1 Cabling example for single N series Gateway with XIV	253
13.3.2 Cabling example for N series Gateway cluster with XIV	254
13.4 Zoning	255
13.4.1 Zoning example for single N series Gateway attachment to XIV	255
13.4.2 Zoning example for clustered N series Gateway attachment to XIV	256
13.5 Configuring the XIV for N series Gateway	256
13.5.1 Create a Storage Pool in XIV	256
13.5.2 Create the root volume in XIV	257
13.5.3 N series Gateway Host create in XIV	258
13.5.4 Add the WWPN to the host in XIV	258
13.5.5 Mapping the root volume to the host in XIV gui	261
13.6 Installing Data Ontap	262
13.6.1 Assigning the root volume to N series Gateway	262
13.6.2 Installing Data Ontap	263
13.6.3 Data Ontap update	264
13.6.4 Adding data LUNs to N series Gateway	264
Chapter 14. ProtecTIER Deduplication Gateway connectivity	265
14.1 Overview	266
14.2 Preparing an XIV for ProtecTIER Deduplication Gateway	267
14.2.1 Supported versions and prerequisites	267
14.2.2 Fibre Channel switch cabling	268
14.2.3 Zoning configuration	268
14.2.4 Configuring XIV Storage System for ProtecTIER Deduplication Gateway	269
14.3 Ready for ProtecTIER software install	274
Chapter 15. XIV in database application environments	275
15.1 XIV volume layout for database applications	276
15.2 Database Snapshot backup considerations	279
Chapter 16. Snapshot Backup/Restore Solutions with XIV and Tivoli Storage Flashcopy Manager	281
16.1 IBM Tivoli FlashCopy Manager Overview	282
16.1.1 Features of IBM Tivoli Storage FlashCopy Manager	282
16.2 FlashCopy Manager 2.2 for Unix	283
16.2.1 FlashCopy Manager prerequisites	284
16.3 Installing and configuring FlashCopy Manager for SAP/DB2	285
16.3.1 FlashCopy Manager disk-only backup	286
16.3.2 SAP Cloning	289

16.4 Tivoli Storage FlashCopy Manager for Windows	291
16.5 Windows Server 2008 Volume Shadow Copy Service	292
16.5.1 VSS architecture and components	293
16.5.2 Microsoft Volume Shadow Copy Service function	295
16.6 XIV VSS provider	295
16.6.1 XIV VSS Provider installation	296
16.6.2 XIV VSS Provider configuration	297
16.7 Installing Tivoli Storage FlashCopy Manager for Microsoft Exchange	299
16.8 Backup scenario for Microsoft Exchange Server	303
Appendix A. Quick guide for VMware SRM	311
Introduction	312
Pre-requisites	313
Install and configure the database environment	314
Installing vCenter server	328
Installing and configuring vCenter client	332
Installing SRM server	337
Installing vCenter Site Recovery Manager plug-in	341
Installing XIV Storage Replication Adapter for VMware SRM	342
Configure the IBM XIV System Storage for VMware SRM	343
Configure SRM Server	343
Related publications	355
IBM Redbooks	355
Other publications	355
Online resources	356
How to get Redbooks	356
Help from IBM	356
Index	357

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 give 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.

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 5L™	IBM®	System i®
AIX®	Micro-Partitioning™	System p®
BladeCenter®	Power Architecture®	System Storage®
DB2®	Power Systems™	System x®
DS4000®	POWER5™	System z®
DS6000™	POWER6®	Tivoli®
DS8000®	PowerVM™	TotalStorage®
FICON®	POWER®	XIV®
FlashCopy®	ProtecTIER®	z/VM®
GPFS™	Redbooks®	z10™
HyperFactor®	Redpaper™	
i5/OS®	Redbooks (logo)  ®	

The following terms are trademarks of other companies:

Snapshot, and the NetApp logo are trademarks or registered trademarks of NetApp, Inc. in the U.S. and other countries.

Novell, SUSE, the Novell logo, and the N logo are registered trademarks of Novell, Inc. in the United States and other countries.

Oracle, JD Edwards, PeopleSoft, Siebel, and TopLink are registered trademarks of Oracle Corporation and/or its affiliates.

QLogic, SANblade, and the QLogic logo are registered trademarks of QLogic Corporation. SANblade is a registered trademark in the United States.

ABAP, SAP, and SAP logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries.

Java, and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

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

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.

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

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

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

Preface

This IBM® Redbooks® publication provides information for attaching the XIV® Storage System to various host operating system platforms, or in combination with databases and other storage oriented application software.

The book also presents and discusses solutions for combining the XIV Storage System with other storage platforms, host servers, or gateways.

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

The information presented here is not meant as a replacement or substitute for the Host Attachment kit publications available at:

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

The book is meant as a complement and to provide the readers with usage recommendations and practical illustrations.

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 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 Masters degree in Electrical Engineering from the Polytechnic Faculty of Mons (Belgium).



Valerius Diener holds a Bachelors in Computer Science from the University of Applied Sciences in Wiesbaden, Germany. He joined IBM as student trainee after completing his bachelor thesis and related work about virtualization solutions with XIV Storage System and the Citrix Xen server.



Roger Eriksson is a STG Lab Services consultant, based in Stockholm, Sweden who works for the European Storage Competence Center in Mainz, Germany. He is a Senior Accredited IBM Product Service Professional. Roger has over 20 years experience working on IBM servers and storage, including Enterprise and Midrange disk, NAS, SAN, System x®, System p®, and Bladecenters. He has done consulting, proof of concepts and education mainly with XIV product line since December 2008, working with both clients and various IBM teams worldwide. He holds a Technical College Graduation in Mechanical Engineering.



Wilhelm Gardt holds a degree in Computer Sciences from the University of Kaiserslautern, Germany. He worked as a software developer and subsequently as an IT specialist designing and implementing heterogeneous IT environments (SAP®, Oracle®, AIX®, HP-UX, SAN etc.). In 2001 he joined the IBM TotalStorage® Interoperability Centre (now Systems Lab Europe) in Mainz where he performed customer briefings and proof of concepts on IBM storage products. Since September 2004 he is a member of the Technical Pre-Sales Support team for IBM Storage (Advanced Technical Support).



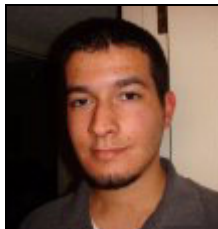
Jana Jamsek is an IT Specialist for IBM Slovenia. She works in Storage Advanced Technical Support for Europe as a specialist for IBM Storage Systems and the IBM i (i5/OS®) operating system. Jana has eight years of experience working with the IBM System i® platform and its predecessor models, and eight years of experience working with storage. She has a Masters degree in Computer Science and a degree in Mathematics from the University of Ljubljana in Slovenia.



Nils Nause is a Storage Support Specialist for IBM XIV Storage Systems and is located at IBM Mainz, Germany. Nils joined IBM in the summer of 2005, and was responsible for Proof of Concepts (PoCs) and delivering briefings for several IBM products. In July 2008 he started working for the XIV post sales support, with a special focus on Oracle Solaris attachment, and overall security aspects of the XIV Storage System. He holds a degree in Computer Science from the University of Applied Science in Wernigerode, Germany.



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 and TotalStorage solutions in environments such as AIX, Linux®, Windows®, VMware ESX, and Solaris. He has worked at IBM for nine years. He has performed many Proof of Concepts with Copy Services on DS6000/DS8000/XIV, and Performance-Benchmarks with DS4000/DS6000/DS8000/XIV. He has written several IBM Redbooks and acted as the co-project lead for Redbooks including DS6000/DS8000® Architecture and Implementation, DS6000/DS8000 Copy Services, and IBM XIV Storage System: Concepts, Architecture and Usage. He holds a degree in Electrical Engineering from the Technical University in Darmstadt.



Carlo Saba is a Test Engineer for XIV in Tucson, AZ. He has been working with the product since shortly after its introduction and is a Certified XIV Administrator. Carlo graduated from the University of Arizona in 2007 with a BSBA in MIS and minor in Spanish.



Eugene Tsypin is an IT Specialist who currently works for IBM STG Storage Systems Sales in Russia. Eugene has over 15 years of experience in the IT field, ranging from systems administration to enterprise storage architecture. He is working as Field Technical Sales Support for storage systems. His areas of expertise include performance analysis and disaster recovery solutions in enterprises utilizing the unique

capabilities and features of the IBM XIV Storage System and other IBM storage, server, and software products.



William (Kip) Wagner is an Advisory Product Engineer for XIV in Tucson, Arizona. He has more than 24 years of experience in field support and systems engineering, and is a Certified XIV Engineer and Administrator. Kip was a member of the initial IBM XIV product launch team who helped design and implement a worldwide support structure specifically for XIV. He also helped develop training material and service documentation used in the support organization. He is currently the team leader for XIV product field engineering supporting customers in North and South America. He also works with a team of engineers from around the world to provide field experience feedback into the development process to help improve product quality, reliability, and serviceability.



Alexander Warmuth is a Senior IT Specialist at IBM European Storage Competence Center. Working in technical sales support, he designs and promotes new and complex storage solutions, drives the introduction of new products, and provides advice to customers, business partners, and sales. His main areas of expertise are high end storage solutions, business resiliency, Linux, and storage. He joined IBM in 1993 and has worked in technical sales support since 2001. Alexander holds a diploma in Electrical Engineering from the University of Erlangen, Germany



Axel Westphal is working as an IT Specialist for Workshops and Proof of Concepts at the IBM European Storage Competence Center (ESCC) in Mainz, Germany. He joined IBM in 1996, working for Global Services as a System Engineer. His areas of expertise include setup and demonstration of IBM System Storage products and solutions in various environments. Since 2004 he has been responsible for storage solutions and Proof of Concepts conducted at the ESSC with DS8000, SAN Volume Controller, and XIV. He has been a contributing author to several DS6000™ and DS8000-related IBM Redbooks publications.



Ralf Wohlfarth is an IT Specialist in the IBM European Storage Competence Center in Mainz, working in technical sales support with a focus on the IBM XIV Storage System. In 1998 he joined IBM and has been working in last level product support for IBM System Storage and Software since 2004. He had the lead for post sales education during a product launch of an IBM Storage Subsystem and resolved complex customer situations. During an assignment in the US he acted as liaison into development, and has been driving product improvements into hardware and software development. Ralf holds a Masters degree in Electrical Engineering, with main subject telecommunication from the University of Kaiserslautern, Germany.

Special thanks to:

John Bynum
Worldwide Technical Support Management
IBM US, San Jose

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

Eyal Abraham
Dave Adams
Avi Aharon
Shimon Ben-David
Alice Bird
John Cherbini
Moshe Dahan
Dave Denny
Rami Elron
Orli Gan
Michael Hayut
Richard Heffel
Iddo Jacobi
Chip Jarvis
Moriel Lechtman
Carlos Lizarralde
Rosemary McCutchen
Dave Monshaw
Basil Moshous
Aviad Offer
Omri Palmon
Joe Roa
Hank Sautter
Izhar Sharon
Brian Sherman
Bill Wiegand
Juan Yanes
IBM

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>



Host connectivity

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

The term *host* in this chapter refers to a server running a supported operating system such as AIX or Windows. SVC as a host has special considerations because it acts as both a host and a storage device. SVC is covered in more detail in Chapter 11, “SVC specific considerations” on page 229.

This chapter does not include attachments from a secondary XIV used for Remote Mirroring, nor does it include data migration from an older storage system. Those topics are covered in *IBM XIV Storage System: Copy Services and Migration*, SG24-7759.

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

For the latest information, refer to the hosts attachment kit publications at:

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

1.1 Overview

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

- Fibre Channel adapters for support with the Fibre Channel Protocol (FCP)
- iSCSI software initiator for support with the iSCSI protocol

This choice gives you the flexibility to start with the less expensive iSCSI implementation, using an already available Ethernet network infrastructure, unless your workload has the need for a dedicated network for iSCSI (Note that iSCSI attachment is not supported with all platforms). Most companies have existing Ethernet connections between their locations and can use that infrastructure to implement a less expensive backup or disaster recovery setup. You can take a snapshot of a critical server and send the snapshot through iSCSI to a remote data center server for backup. In this case, you can simply use the existing network resources without the need for expensive FC switches.

As soon as workload and performance requirements justify it, you can progressively convert to the more expensive Fibre Channel infrastructure. From a technical standpoint and after HBAs and cabling are in place, the migration is easy. It only requires the XIV storage administrator to add the HBA definitions to the existing host configuration to make the logical unit numbers (LUNs) visible over FC paths.

The XIV Storage System has up to six Interface Modules, depending on the rack configuration. Figure 1-1 summarizes the number of active interface modules and the FC and iSCSI ports for different rack configurations.

Rack configuration No. of modules	6	9	10	11	12	13	14	15
Module 9 state	NA	Disabled	Disabled	Enabled	Enabled	Enabled	Enabled	Enabled
Module 8 state	NA	Enabled	Enabled	Enabled	Enabled	Enabled	Enabled	Enabled
Module 7 state	NA	Enabled	Enabled	Enabled	Enabled	Enabled	Enabled	Enabled
Module 6 state	Disabled	Disabled	Disabled	Disabled	Disabled	Enabled	Enabled	Enabled
Module 5 state	Enabled	Enabled	Enabled	Enabled	Enabled	Enabled	Enabled	Enabled
Module 4 state	Enabled	Enabled	Enabled	Enabled	Enabled	Enabled	Enabled	Enabled
Number of dedicated Data modules	4	5	6	6	7	7	8	9
Number of Active Interface modules	2	4	4	5	5	6	6	6
FC ports	8	16	16	20	20	24	24	24
iSCSI ports	0	4	4	6	6	6	6	6

Figure 1-1 XIV Rack Configuration

Each active Interface Module (Modules 4-9, if enabled) has four Fibre Channel ports, and up to three Interface Modules (Modules 7-9, if enabled) also have two iSCSI ports each. These

ports are used to attach hosts (and remote XIVs or older storage systems) to the XIV through the internal patch panel.

The patch panel simplifies cabling because the Interface Modules are pre-cabled to the patch panel so that all customer SAN and network connections are made in one central place at the back of the rack. This 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 currently not supported.

Figure 1-2 gives an example of how to connect a host through either a Storage Area Network (SAN) or an Ethernet network to a fully populated XIV Storage System. For picture clarity, the patch panel is not shown here.

Important: Host traffic can be served through any of the Interface Modules. However, I/Os are not automatically balanced by the system. It is the storage administrator's responsibility to ensure that host connections avoid single points of failure and that the host workload is adequately balanced across the connections and Interface Modules. This should be reviewed periodically or 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, and this simplifies storage management. Balancing traffic and zoning (for adequate performance and redundancy) is more critical, although not more complex, than with traditional storage systems.

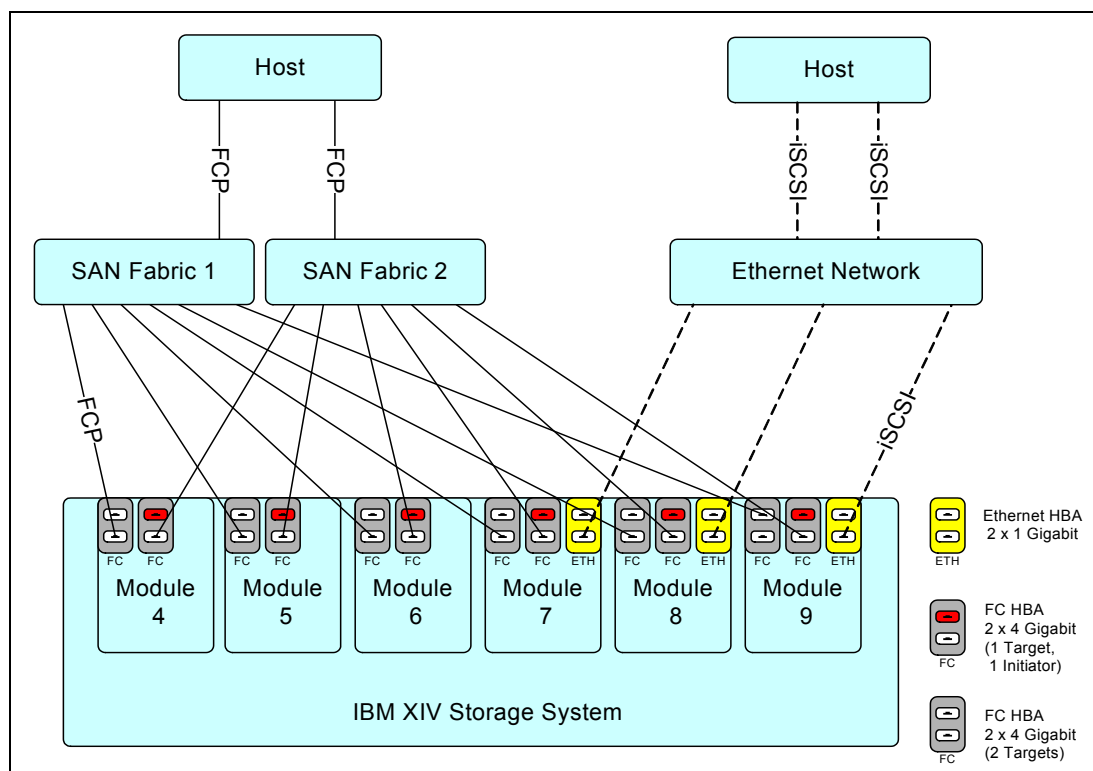


Figure 1-2 Host connectivity overview (without patch panel)

1.1.1 Module, patch panel, and host connectivity

This section presents a simplified view of the host connectivity. It is intended to explain the relationship between individual system components and how they affect host connectivity. Refer to chapter 3 of *IBM XIV Storage System: Architecture, Implementation, and Usage*, SG24-7659 for more details and an explanation of the individual components.

When connecting hosts to the XIV, there is no “one size fits all” solution that can be applied because every environment is different. However, we provide the following guidelines to ensure that there are no single points of failure and 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.
- ▶ XIV patch panel FC ports 2 and 4 (or ports 3 and 4 depending on your environment) should be used for mirroring to another XIV Storage System and for data migration from an older storage system.

Note: Most illustrations in this book show ports 1 and 3 allocated for host connectivity, and ports 2 and 4 are reserved for additional host connectivity, or remote mirror and data migration connectivity. This is generally the choice for customers who want more resiliency (ports 1 and 3 are on separate adapters), or availability (in case of adapter firmware upgrade, one connection remains available through the other adapter), and also if the workload needs more performance (each adapter has its own PCI bus). For certain environments it might be necessary to use ports 1 and 2 for host connectivity and to reserve ports 3 and 4 for mirroring. If you will 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.

If mirroring or data migration will not be used, then the remaining ports can be used for additional host connections (port 4 must first be changed from its default initiator role to target). However, additional ports provide “fan out” capability and not additional throughput.

Note: Using the remaining 12 ports will provide the ability to manage devices on additional ports, but will not necessarily provide additional storage system bandwidth.

- ▶ iSCSI hosts connect to iSCSI port 1 on Interface Modules (not possible with 6 module system).
- ▶ Hosts should have a connection path to multiple separate Interface Modules to avoid a single point of failure.
- ▶ When using SVC as a host on a fully populated XIV, all 12 available FC host ports on the XIV patch panel (ports 1 and 2 on Modules 4-9) should be used for SVC and nothing else. All other hosts access the XIV through the SVC.

Figure 1-3 illustrates an overview of FC and iSCSI connectivity for a full rack configuration.

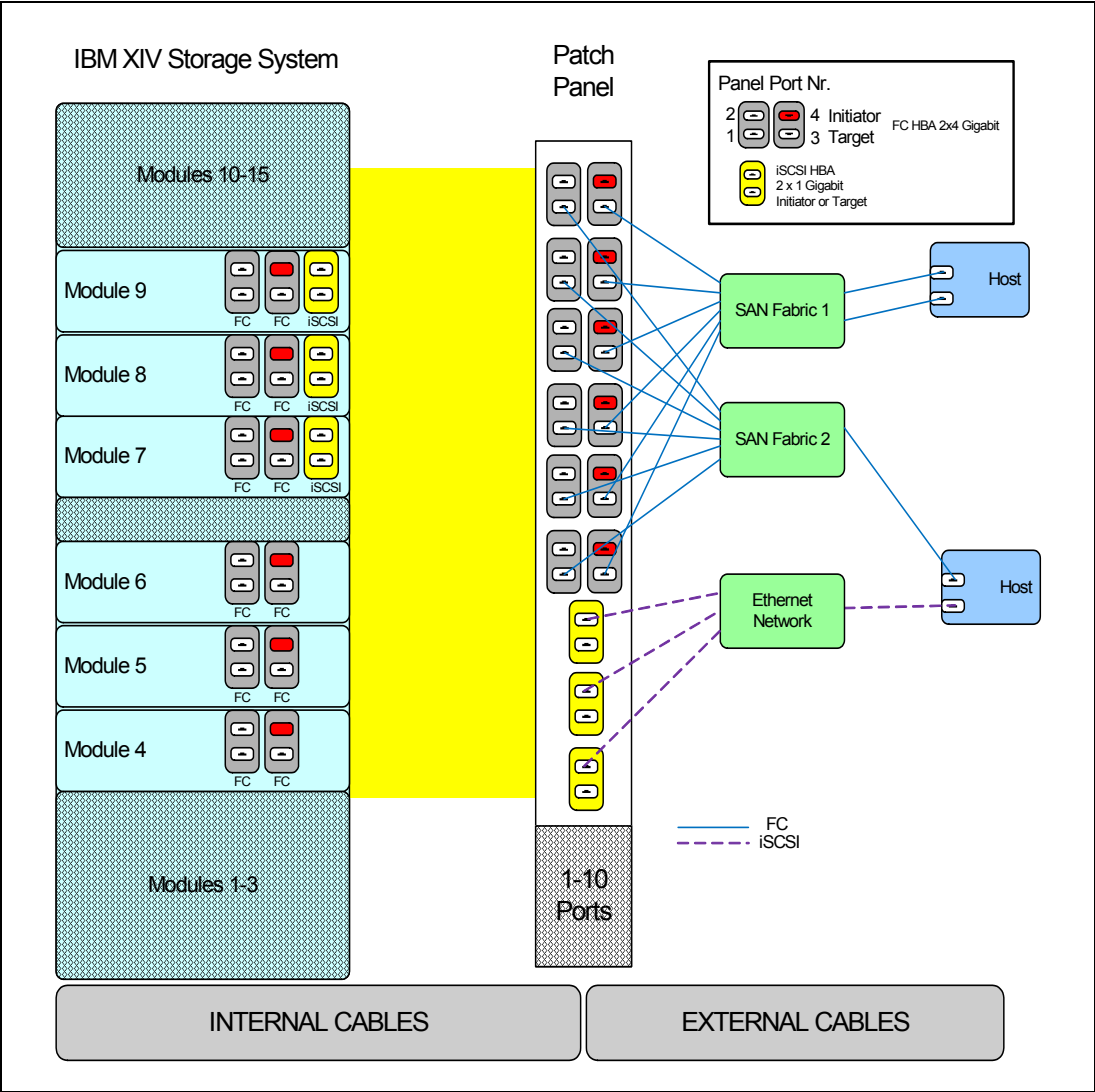


Figure 1-3 Host connectivity end-to-end view: Internal cables compared to external cables

Figure 1-4 provides an XIV patch panel to FC and a patch panel to iSCSI adapter mappings. It also shows the World Wide Port Names (WWPNs) and iSCSI Qualified Names (IQNs) associated with the ports.

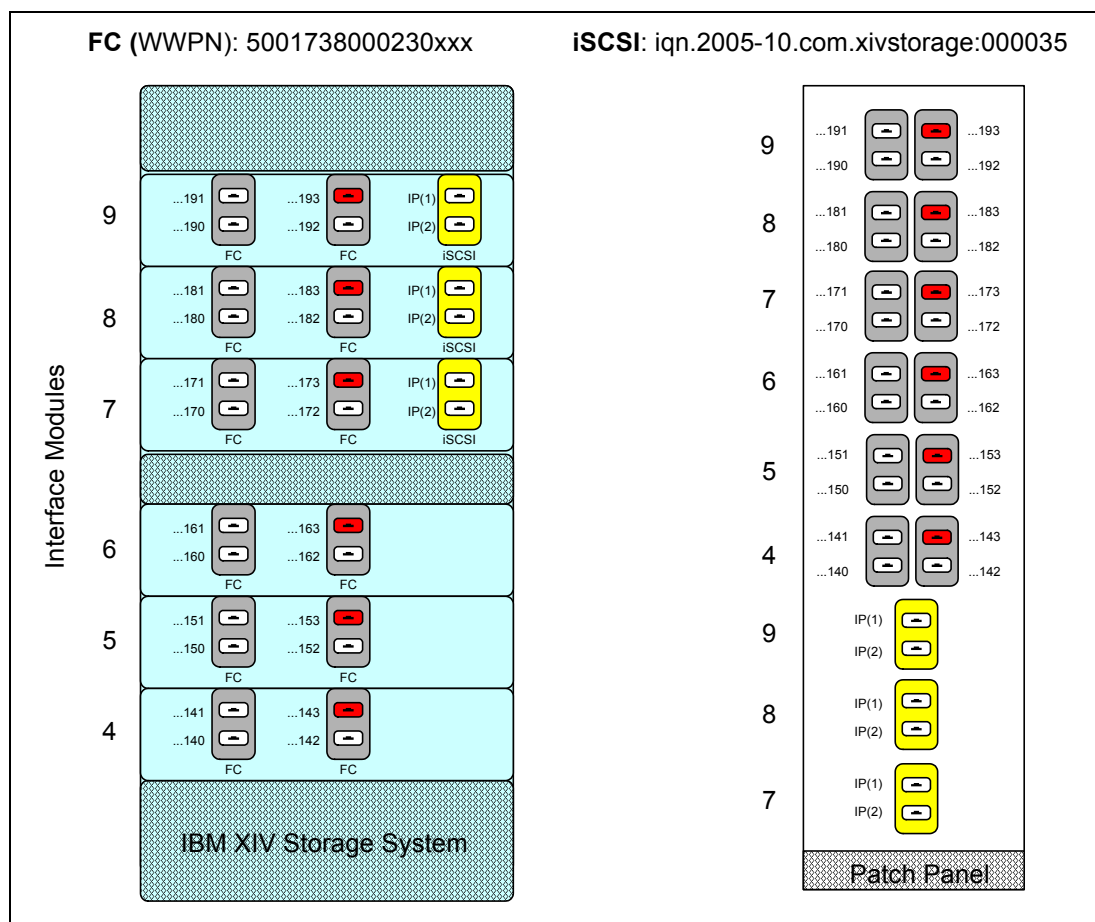


Figure 1-4 Patch panel to FC and iSCSI port mappings

A more detailed view of host connectivity and configuration options is provided in 1.2, “Fibre Channel (FC) connectivity” on page 8 and in 1.3, “iSCSI connectivity” on page 21.

1.1.2 Host operating system support

The XIV Storage System supports many operating systems, and the list is constantly growing. Here is a list of the supported operating systems at the time of writing:

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

To get the current list when you implement your XIV, refer to the IBM System Storage Interoperation Center (SSIC) at the following website:

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

1.1.3 Host Attachment Kits

Starting with version 10.1.x of the XIV system software, IBM also provides updates to all of the Host Attachment Kits (version 1.1 or later). It is *mandatory* to install the Host Attachment Kit to be able to get support from IBM, even if it might not be technically necessary for certain operating systems. Host Attachment Kits (HAKs) are built on a Python framework with the intention of providing a consistent look and feel across various OS platforms. The HAKs have the following features:

- ▶ Backwards compatibility with versions 10.0.x of the XIV system software
- ▶ Validates patch and driver versions
- ▶ Sets up multipathing
- ▶ Adjusts system tunable parameters (if required) for performance
- ▶ Provides an installation wizard
- ▶ Includes management utilities
- ▶ Includes support and troubleshooting utilities

Host Attachment Kits can be downloaded from the following URL:

http://www.ibm.com/support/search.wss?q=s5g1*&tc=STJTAG+HW3E0&rs=1319&dc=D400&dtm

1.1.4 FC versus iSCSI access

Host can attach to XIV over a FC or iSCSI topology. The version of XIV system software at the time of writing (10.2.2) supports iSCSI using the software initiator only, except for AIX where an iSCSI HBA is also supported.

The choice of connection protocol (iSCSI or FCP) should be made based on application requirements. When considering IP storage based connectivity, also look at the performance and availability of the existing corporate infrastructure.

Take the following considerations into account:

- ▶ FC hosts in a production environment should always be connected to a minimum of two separate SAN switches in independent fabrics to provide redundancy.
- ▶ For test and development, there can be single points of failure to reduce costs. However, you will have to determine if this practice is acceptable for your environment.
- ▶ When 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 and therefore should not be mixed with other IP traffic.

A host can connect through FC and iSCSI simultaneously. However, you cannot access the same LUN with both protocols. Figure 1-5 illustrates the simultaneous access to XIV LUNs from one host using both protocols.

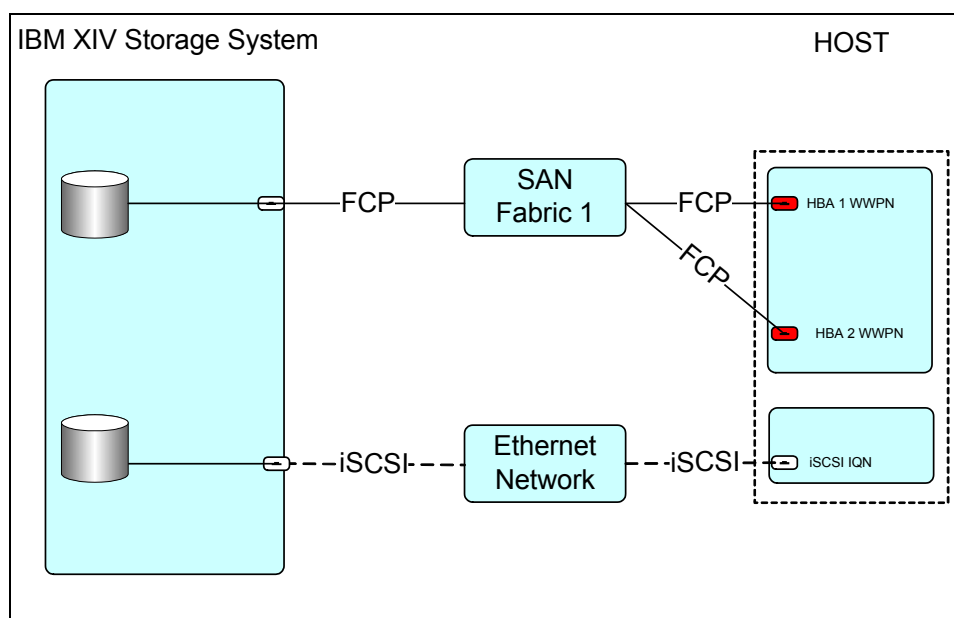


Figure 1-5 Host connectivity FCP and iSCSI simultaneously using separate host objects

1.2 Fibre Channel (FC) connectivity

This section focuses on FC connectivity that applies to the XIV Storage System in general. For operating system-specific information, refer to 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, there are a number of procedures that you must complete. The following is a list of general procedures that pertain to all hosts. However, you also need to 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 recommended or required firmware and device driver levels is available at the IBM System Storage Interoperability Center (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, use any supported driver and firmware by the HBA vendors (the latest versions are always preferred). For HBAs in Sun systems, use Sun-branded HBAs and Sun-ready HBAs only.

You should 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 on the host server to manage the total number of LUNs that you want to attach.
3. Check the optimum number of paths that should be defined. This will help in determining the zoning requirements.
4. Download, if needed, and install the latest supported HBA firmware and driver.

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, for example, when troubleshooting an HBA driver. Be aware that IBM cannot be held responsible for the content of these sites.

QLogic Corporation

The Qlogic website can be found at the following URL:

<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, which can be found at the following URL:

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

Emulex Corporation

The Emulex home page is at the following URL:

<http://www.emulex.com>

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

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

Oracle

Oracle ships its own HBAs. They are Emulex and QLogic based. However it is true that the “native” HBAs (from Emulex and QLogic) can be used to attach servers running Oracle Solaris to disk systems. In fact such “native” HBAs can even be used to run StorEdge Traffic Manager software. For more information refer to:

- ▶ 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. Refer to this information for general guidance about connecting their systems to

SAN-attached storage. However, be aware that in certain cases you will not be able to find information to help you with third-party vendors. You should always 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 the vendors' websites.

1.2.2 FC configurations

Several configurations using Fibre Channel are technically possible, and 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. There should be no single points of failure. Hosts should have as many HBAs as needed to support the operating system, application and overall performance requirements.

For test and development environments, a non-redundant configuration is often the only practical option due to cost or other constraints. Also, this will typically include one or more single points of failure.

Next, we review three typical FC configurations that are supported and offer redundancy.

Redundant configurations

The fully redundant configuration is illustrated in Figure 1-6.

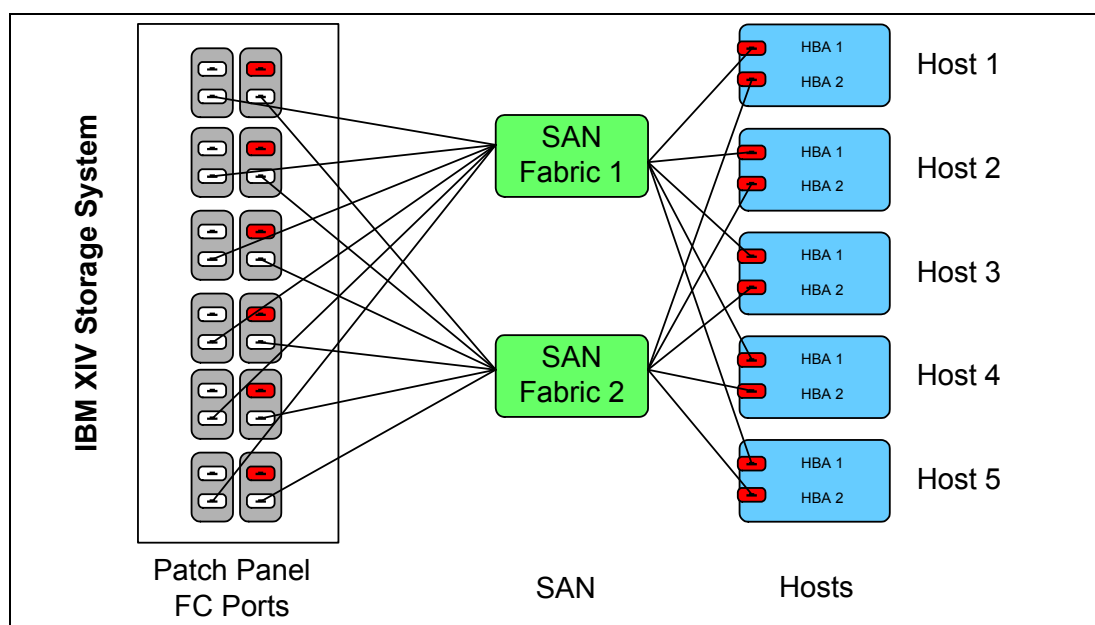


Figure 1-6 FC fully redundant configuration

In this configuration:

- ▶ 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 LUN has 12 paths.

A redundant configuration accessing all interface modules, but using only a minimum number of six paths per LUN on the host, is depicted in Figure 1-7 on page 11.

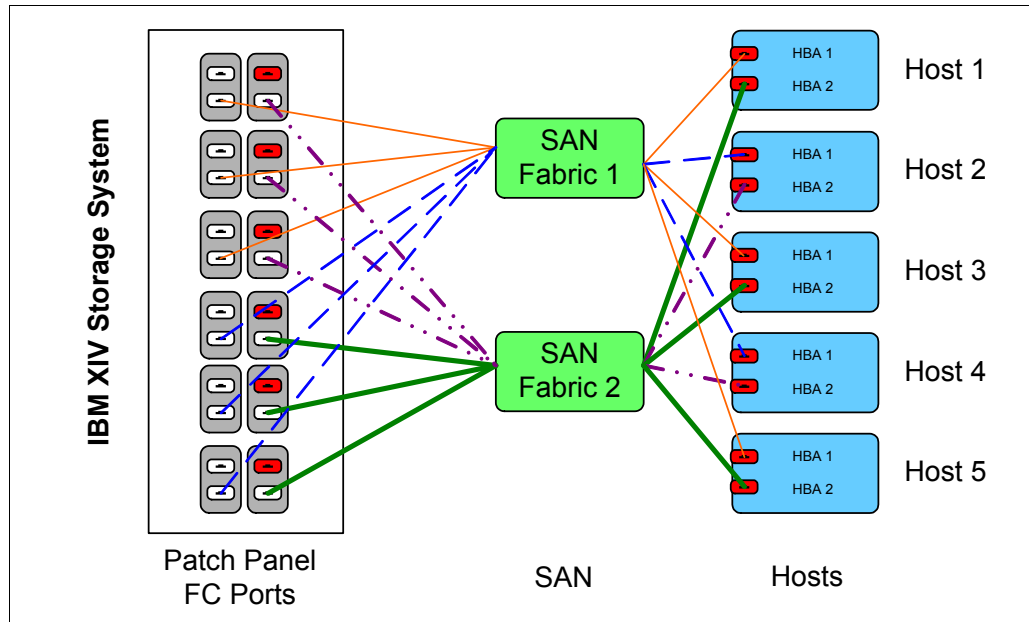


Figure 1-7 FC redundant configuration

In this configuration:

- ▶ 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 next host is using the three other paths per fabric.
- ▶ If a fabric fails still all interface modules will be used.
- ▶ Each LUN has six paths.

A simple redundant configuration is illustrated in Figure 1-8 on page 12.

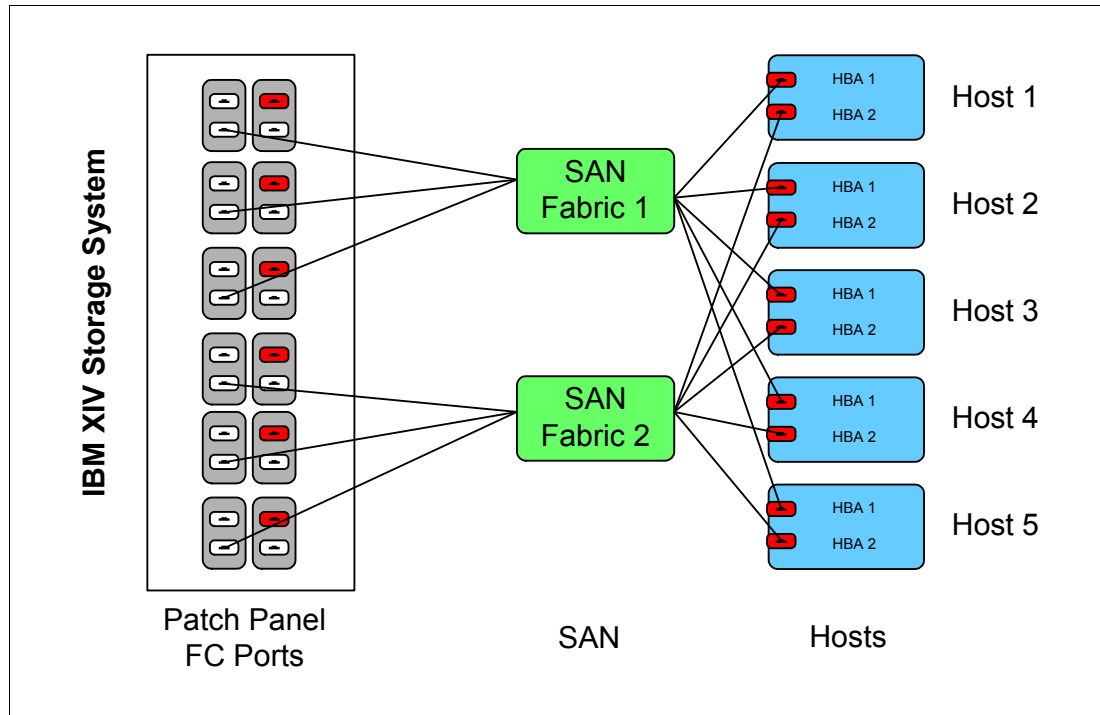


Figure 1-8 FC simple redundant configuration

In this configuration:

- ▶ 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 LUN has six paths.

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.

1.2.3 Zoning

Zoning is mandatory when 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 given 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 World Wide Port Names (WWPNs) of the HBA. Each method has its merits, and you will have to determine which is right for your environment.

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

- ▶ An error from an HBA that affects the zone or zone traffic will be isolated.

- ▶ Disk and tape traffic must be in separate zones because they have different characteristics. If they are in the same zone this can cause performance problems or have other adverse affects.
- ▶ Any change in the SAN fabric, such as a change caused by a server restarting or a new product being added to the SAN, triggers a *Registered State Change Notification* (RSCN). An RSCN requires that any device that can “see” the affected or new device to acknowledge the change, interrupting its own traffic flow.

Zoning guidelines

There are many factors that affect zoning including host type, number of HBAs, HBA driver, operating system, and applications. Therefore, it is not possible to provide a solution to cover every situation. The following list gives guidelines that can help you to avoid reliability or performance problems. However, you should also review documentation regarding your hardware and software configuration for any specific factors that need to be considered.

- ▶ Each zone (excluding those for SVC) should have one initiator HBA (the host) and multiple target HBAs (the XIV Storage System).
- ▶ Each host (excluding SVC) should have two paths per HBA unless there are other factors dictating otherwise.
- ▶ Each host should connect to ports from at least two Interface Modules.
- ▶ Do not mix disk and tape traffic on the same HBA or in the same zone.

For more in-depth information about SAN zoning, refer to section 4.7 of *Introduction to Storage Area Networks*, SG24-5470.

You can download this publication from:

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

An example of soft zoning using the “single initiator - multiple targets” method is illustrated in Figure 1-9.

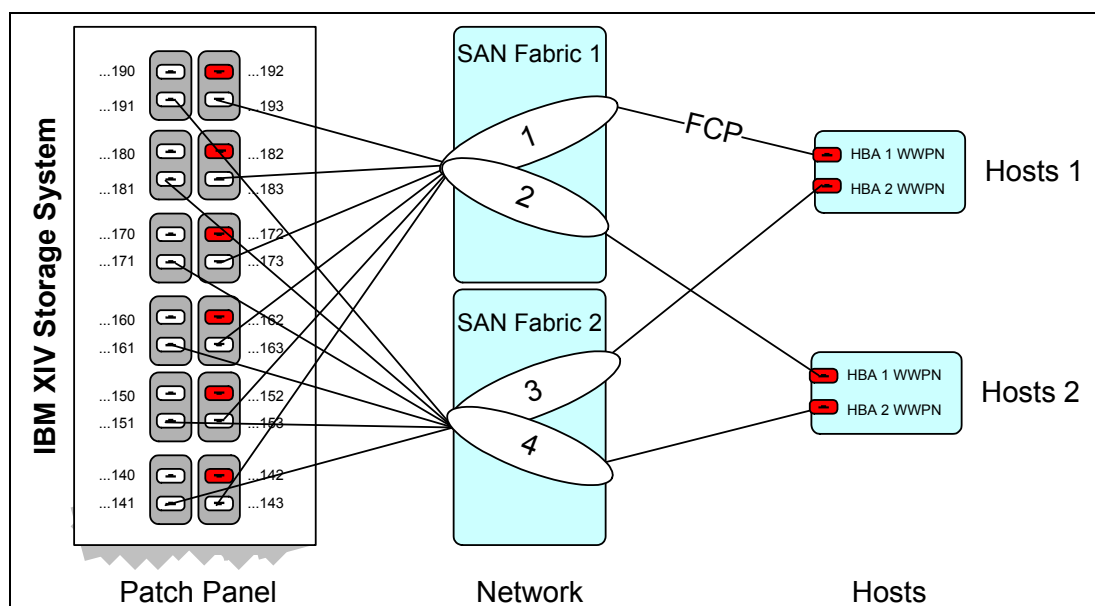


Figure 1-9 FC SAN zoning: single initiator - multiple target

Note: Use a single initiator and multiple target zoning scheme. Do not share a host HBA for disk and tape access.

Zoning considerations should include also spreading the IO workload evenly between the interfaces. For example, for a host equipped with two single port HBA, you should connect one HBA port to one port on modules 4, 5, and 6 to one port and the second HBA port to one port on modules 7, 8, and 9.

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), it is important to statically balance the workload between the paths, and monitor that the IO workload on the interfaces is balanced using the XIV statistics view in the GUI (or XIVTop).

1.2.4 Identification of FC ports (initiator/target)

Identification of a port is required for setting up the zoning, to aid with any modifications that might be required, and to assist with problem diagnosis. The unique name that identifies an FC port is called the World Wide Port Name (WWPN).

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

Example 1-1 XCLI: How to get WWPN of IBM XIV Storage System

```
>> 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:5:2	OK	yes	5001738000230151	0075001F	Target
1:FC_Port:5:3	OK	yes	5001738000230152	00021D00	Target
1:FC_Port:5:4	OK	yes	5001738000230153	00FFFFFF	Target
1:FC_Port:6:1	OK	yes	5001738000230160	00070A00	Target
1:FC_Port:6:2	OK	yes	5001738000230161	006D0713	Target
1:FC_Port:6:3	OK	yes	5001738000230162	00FFFFFF	Target
1:FC_Port:6:4	OK	yes	5001738000230163	00FFFFFF	Initiator
1:FC_Port:7:1	OK	yes	5001738000230170	00760000	Target
1:FC_Port:7:2	OK	yes	5001738000230171	00681813	Target
1:FC_Port:7:3	OK	yes	5001738000230172	00021F00	Target
1:FC_Port:7:4	OK	yes	5001738000230173	00021E00	Initiator
1:FC_Port:8:1	OK	yes	5001738000230180	00060219	Target
1:FC_Port:8:2	OK	yes	5001738000230181	00021C00	Target
1:FC_Port:8:3	OK	yes	5001738000230182	002D0027	Target
1:FC_Port:8:4	OK	yes	5001738000230183	002D0026	Initiator
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

Note that the `fc_port_list` command might not always print out the port list in the same order. When you issue the command, the rows might be ordered differently, but all the ports will be listed.

To get the same information from the XIV GUI, select the main view of an XIV Storage System, use the arrow at the bottom (circled in red) to reveal the patch panel, and move the mouse cursor over a particular port to reveal the port details including the WWPN (refer to Figure 1-10).

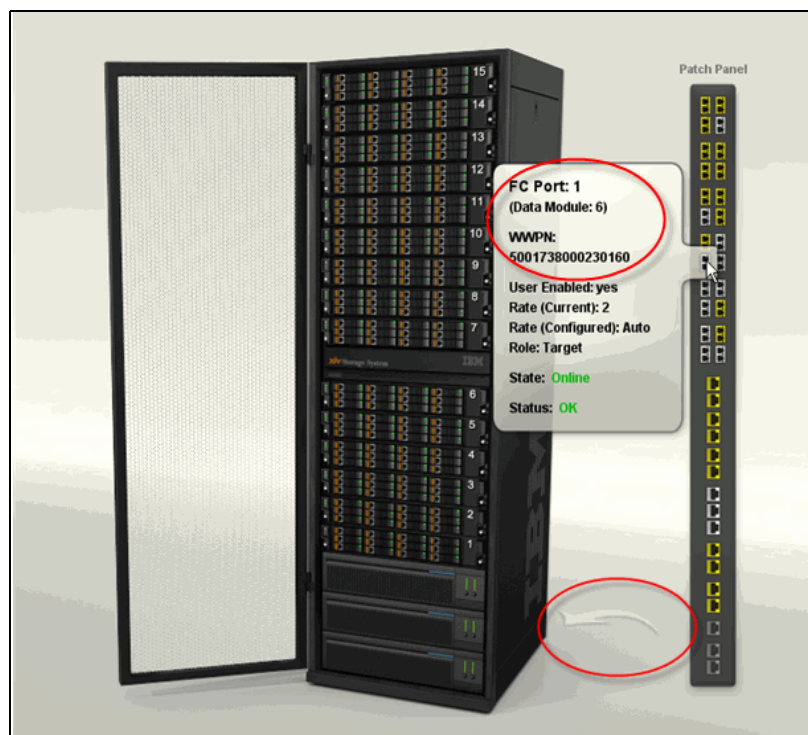


Figure 1-10 GUI: How to get WWPNs of IBM XIV Storage System

Note: 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 1-10, the WWPN is 5001738000230160, which means that the WWPN is from module 6, port 1. The WWPNs for the port are numbered from 0 to 3, whereas the physical the ports are numbered from 1 to 4.

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

Example 1-2 WWPN illustration

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

5	NAA (Network Address Authority)
001738	IEEE Company ID
NNNNN	IBM XIV Serial Number in hex
RR	Rack ID (01-ff, 0 for WWNN)
M	Module ID (1-f, 0 for WWNN)
P	Port ID (0-7, 0 for WWNN)

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. It means that the operating systems and configuration of SAN based computers can be centrally stored and managed. This can provide advantages with regards to deploying servers, backup, and disaster recovery procedures.

To boot from SAN, you need to go into the HBA configuration mode, set the HBA BIOS to be *Enabled*, select at least one XIV target port, and select a LUN to boot from. In practice you will typically configure 2-4 XIV ports as targets and you might have to enable the BIOS on two HBAs, but this will depend on the HBA, driver and operating system. Consult the documentation that comes with your HBA and operating systems.

SAN boot for AIX is separately addressed in Chapter 4, “AIX host connectivity” on page 111

SAN boot for HP-UX is described in Chapter 5, “HP-UX host connectivity” on page 133.

Boot from SAN procedures

The procedures for setting up your server and HBA to boot from SAN will vary. they are mostly 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 will differ but the logical process will be 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. See Figure 1-11.

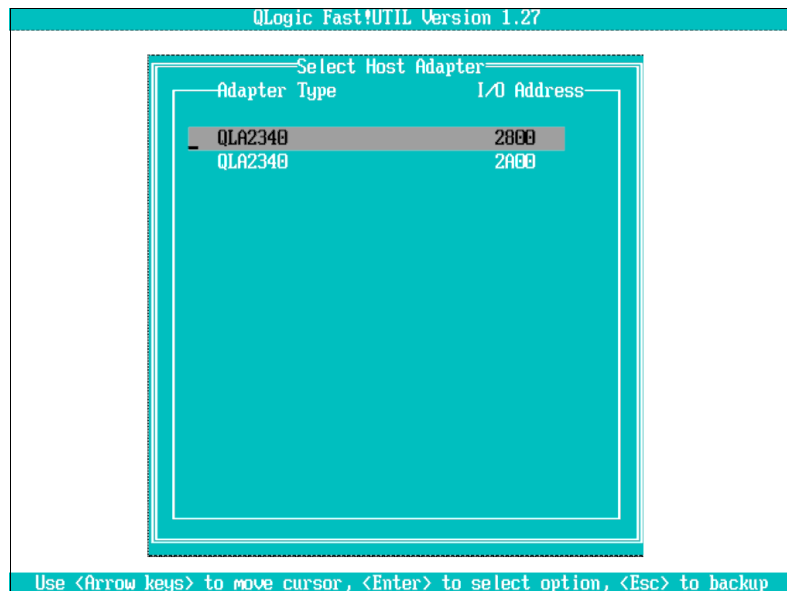


Figure 1-11 Select Host Adapter

2. You normally see one or more ports. Select a port and press Enter. This takes you to a panel as shown in Figure 1-12. Note that if you will only be enabling the BIOS on one port, then make sure to select the correct port.

Select **Configuration Settings**.

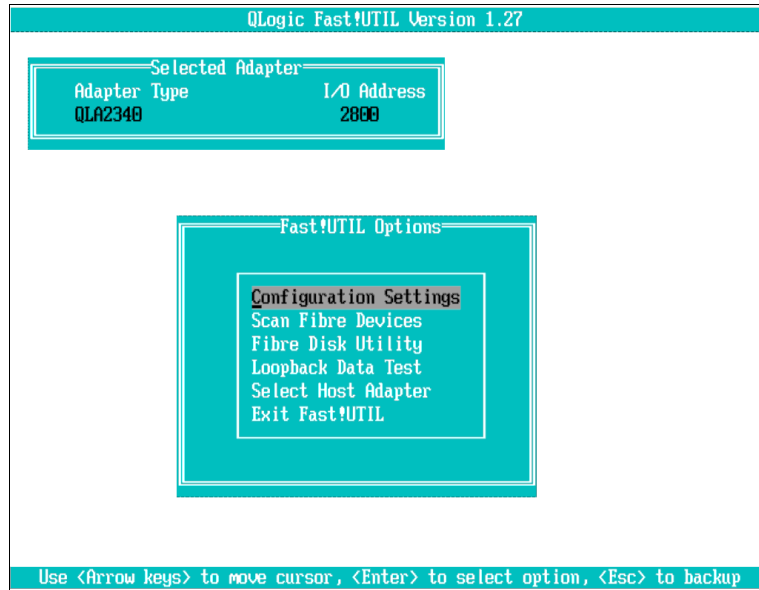


Figure 1-12 Fast!UTIL Options

3. In the panel shown in Figure 1-13, select **Adapter Settings**.

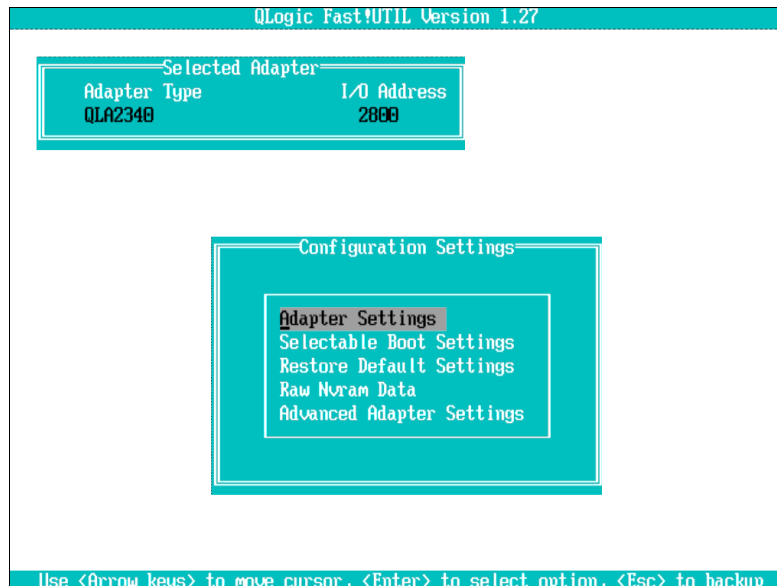


Figure 1-13 Configuration Settings

4. The **Adapter Settings** menu is displayed as shown in Figure 1-14.

Figure 1-14 Adapter Settings

```

QLogic Fast!UTIL Version 1.27

Selected Adapter:
Adapter Type      I/O Address
QLA2340           2800

Selectable Boot Settings

Selectable Boot:      Enabled
(Primary) Boot Port Name,Lun: 0000000000000000, 0
      Boot Port Name,Lun: 0000000000000000, 0
      Boot Port Name,Lun: 0000000000000000, 0
      Boot Port Name,Lun: 0000000000000000, 0

Press "C" to clear a Boot Port Name entry
  
```

Figure 1-15 Selectable Boot Settings

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

Select **Boot Port Name, Lun:** and then press Enter to get the **Select Fibre Channel Device** menu, shown in Figure 1-16.

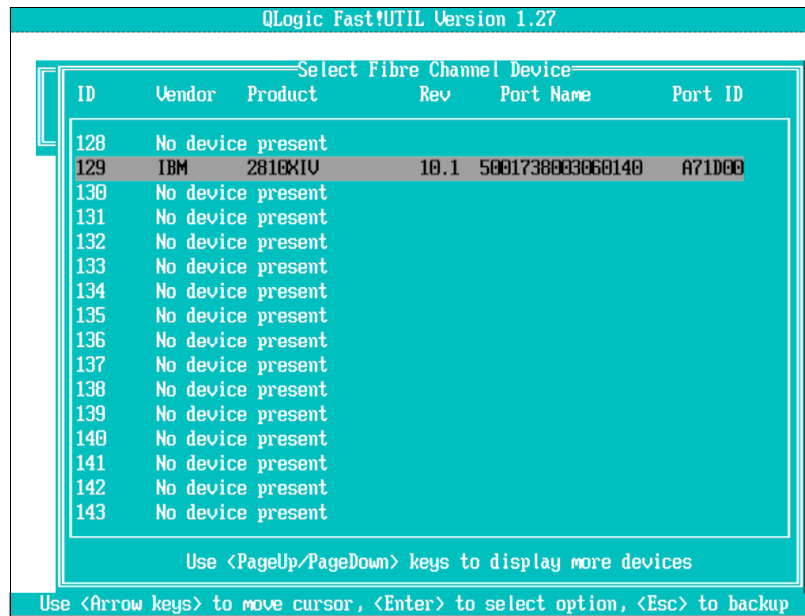


Figure 1-16 Select Fibre Channel Device

- Select the **IBM 2810XIV** device, and press Enter to display the **Select LUN** menu, seen in Figure 1-17.

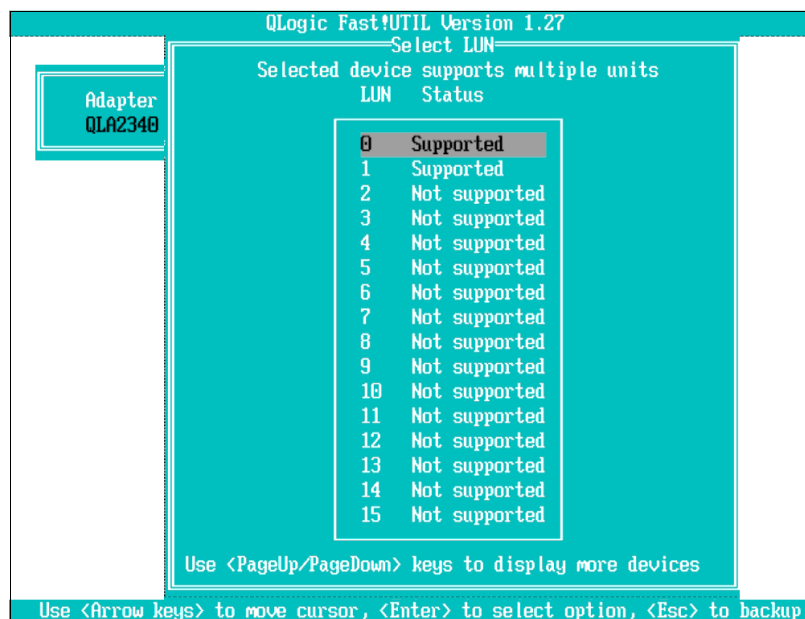


Figure 1-17 Select LUN

9. Select the boot LUN (in our case it is **LUN 0**). You are taken back to the **Selectable Boot Setting** menu and boot port with the boot LUN displayed as illustrated in Figure 1-18.

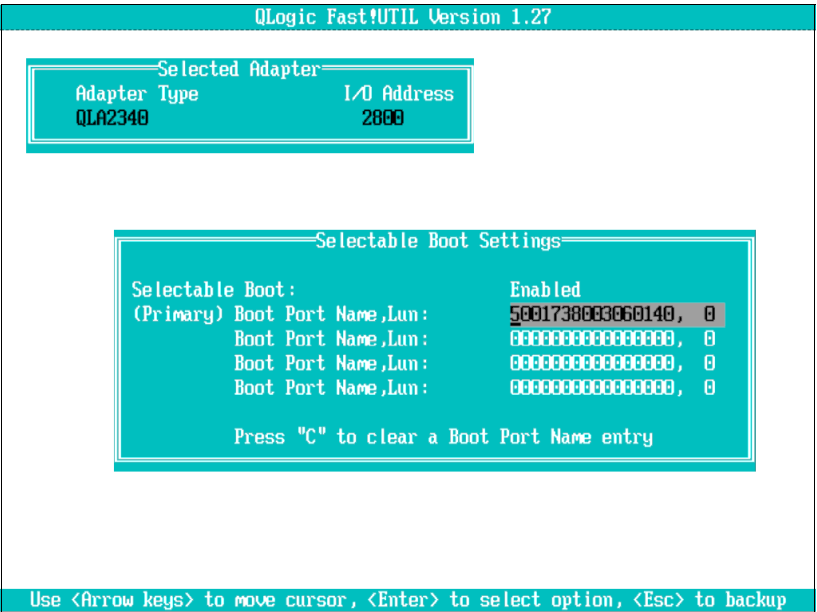


Figure 1-18 Boot Port selected

- 10.Repeat the steps 8-10 to add additional controllers. Note that any additional controllers must be zoned so that they point to the same boot LUN.
- 11.When 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 1-19.

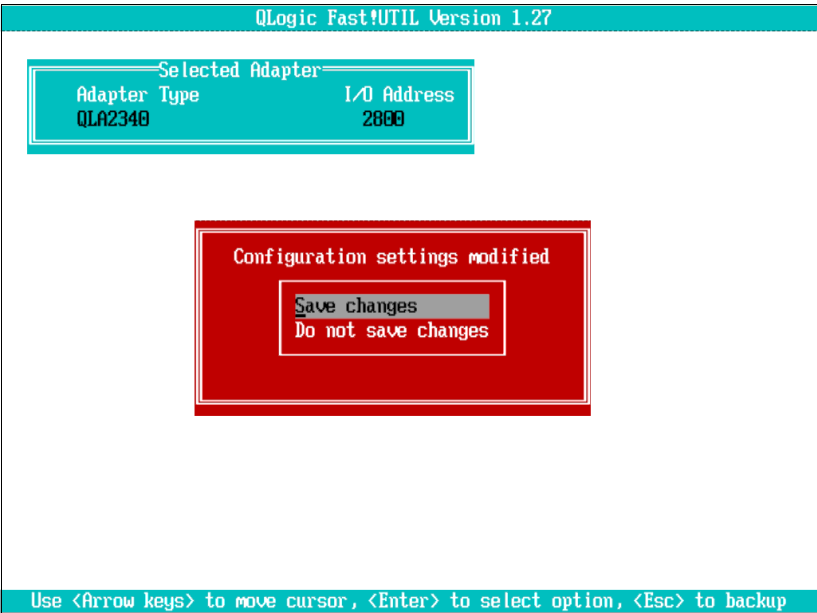


Figure 1-19 Save changes

- 12.Select **Save changes**. This takes you back to the Fast!UTIL option panel. From there, select **Exit Fast!UTIL**.

13. The **Exit Fast!UTIL** menu is displayed as shown in Figure 1-20. Select **Reboot System** to reboot and boot from the newly configured SAN drive.

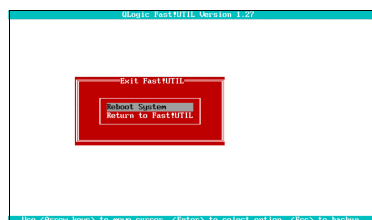


Figure 1-20 Exit Fast!UTIL

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

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, refer to the relevant section in the corresponding chapter of this book.

Currently, iSCSI hosts are only supported using the software iSCSI initiator, except for AIX. Information about iSCSI software initiator support is available at the IBM System Storage Interoperability Center (SSIC) website at the following URL:

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

Table 1-1 shows supported operating systems.

Table 1-1 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	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, there are a number of procedures that you must complete. The following list describes general procedures that pertain to all hosts. However, you need to also review any procedures that pertain to your specific hardware and operating system.

1. Connecting host to the XIV over iSCSI is done using a standard Ethernet port on the host server. The port you choose should be dedicated to iSCSI storage traffic only. This port must also be a minimum of 1 Gbps capable. This port will require an IP address, subnet mask, and gateway.

You should also review any documentation that comes with your operating system regarding iSCSI 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 on the host server to manage the total number of LUNs that you want to attach.
3. Check the optimum number of paths that should be defined. This will help in determining the number of physical connections that need to be made.
4. Install the latest supported adapter firmware and driver. If this is not the one that came 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 one (1500 bytes). Anything larger is known as a *jumbo frame*. The largest possible MTU should be specified, it is advisable to use up to 4500 bytes (which is the default value on XIV), if supported by the switches and routers.
6. Any device using iSCSI requires an iSCSI Qualified Name (IQN), which in our case is the XIV Storage System, 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.

Our XIV Storage System's name was `iqn.2005-10.com.xivstorage:000035`.

1.3.2 iSCSI configurations

Several configurations are technically possible, and 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. Ports cannot be bonded.

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

By default, there are six predefined iSCSI target ports on a fully populated XIV Storage System to serve hosts through iSCSI.

Redundant configurations

A redundant configuration is illustrated in Figure 1-21.

In this configuration:

- ▶ 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 of each of Interface Modules 7-9.

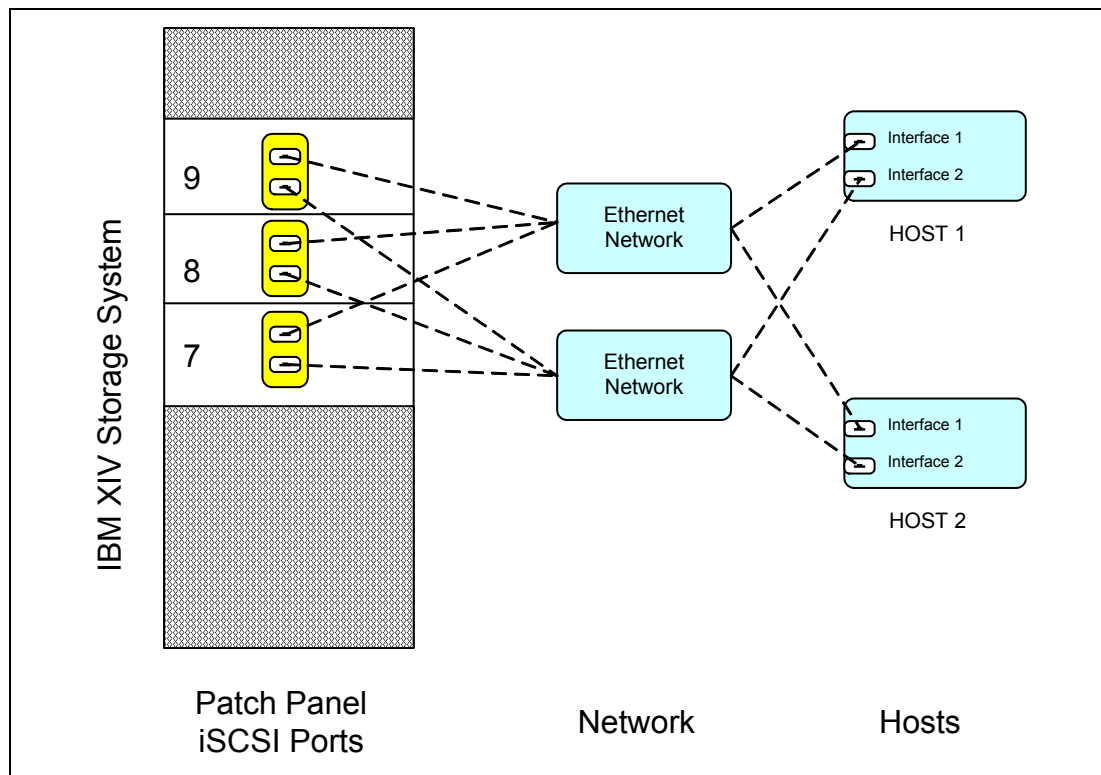


Figure 1-21 iSCSI configurations: redundant solution

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 it 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 it 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 it is typically one or two other modules through the second Ethernet interface.

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

Non-redundant configurations

Non-redundant configurations should only be used where the risks of a single point of failure are acceptable. This is typically the case for test and development environments.

Figure 1-22 illustrates a non-redundant configuration.

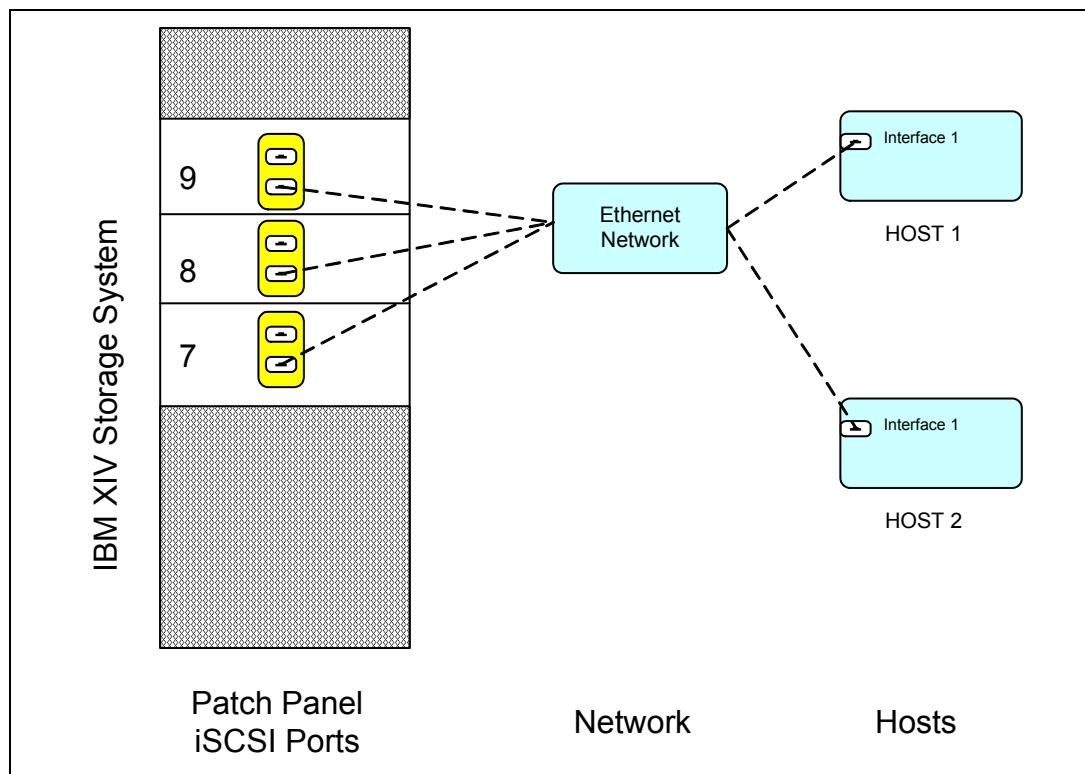


Figure 1-22 iSCSI configurations: Single switch

1.3.3 Network configuration

Disk access is susceptible to network latency. Latency can cause time-outs, delayed writes, and data loss. In order to realize the best performance from iSCSI, all iSCSI IP traffic should flow on a dedicated network. Physical switches or VLANs should be used to provide a dedicated network. This network should be a minimum of 1 Gbps and hosts should have interfaces dedicated to iSCSI only. For such configurations, additional host Ethernet ports might need to be purchased.

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 when compared to an FC connection setup.

Getting the XIV iSCSI Qualified Name (IQN)

Every XIV Storage System has a unique iSCSI Qualified Name (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 a change is required, you must engage IBM support.

The IQN is visible as part of the XIV Storage System. From the XIV GUI, from the opening GUI panel (with all the systems) right-click a system and select **Properties**. The System Properties dialog box is displayed. Select the Parameters tab, as shown in Figure 1-23.

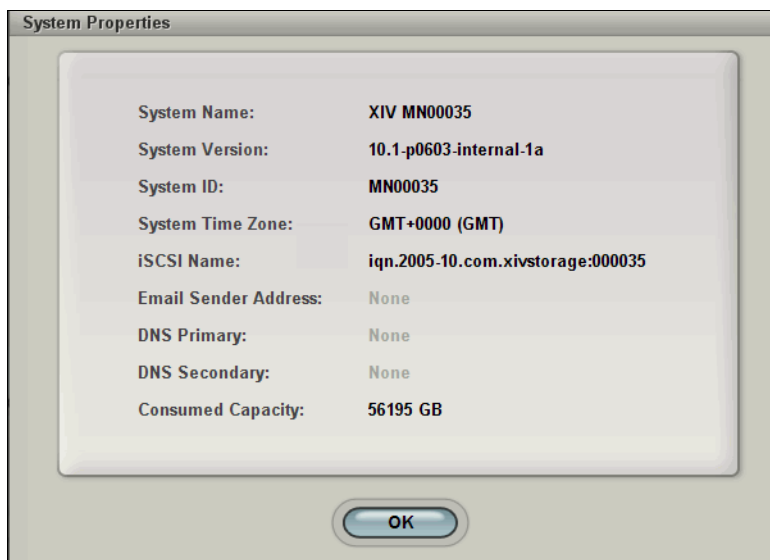


Figure 1-23 iSCSI: Use XIV GUI to get iSCSI name (IQN)

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

Example 1-3 iSCSI: use XCLI to get iSCSI name (IQN)

```
>> config_get
Name                               Value
dns_primary                        9.64.163.21
dns_secondary                      9.64.162.21
system_name                        XIV LAB 3 1300203
snmp_location                      Unknown
snmp_contact                      Unknown
snmp_community                    XIV
snmp_trap_community               XIV
system_id                         203
machine_type                      2810
machine_model                     A14
machine_serial_number             1300203
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:000203
timezone                          -7200
ntp_server                        9.155.70.61
ups_control                       yes
support_center_port_type          Management
```

iSCSI XIV port configuration using the GUI

To set up the iSCSI port using the GUI:

1. Log on to the XIV GUI, select the XIV Storage System to be configured, and move your mouse over the **Hosts and Clusters** icon. Select **iSCSI Connectivity** as shown in Figure 1-24.



Figure 1-24 GUI: iSCSI Connectivity menu option

2. The **iSCSI Connectivity** window opens. Click the Define icon at the top of the window (see Figure 1-25) to open the Define IP interface dialog.



Figure 1-25 GUI: iSCSI Define interface icon

3. Enter the following information (see Figure 1-26):
 - Name: This is a name 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 impacted if the MTU is set incorrectly.
 - Module: Select the module to configure.
 - Port number: Select the port to configure.

Figure 1-26 Define IP Interface: iSCSI setup window

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

iSCSI XIV port configuration using the XCLI

Open an XCLI session tool, and use the *ipinterface_create* command as shown in Example 1-4.

Example 1-4 XCLI: iSCSI setup

```
>> ipinterface_create ipinterface=itso_m7_p1 address=9.11.237.155
netmask=255.255.254.0 gateway=9.11.236.1 module=1:module:7 ports=1
Command executed successfully.
```

1.3.5 Identifying iSCSI ports

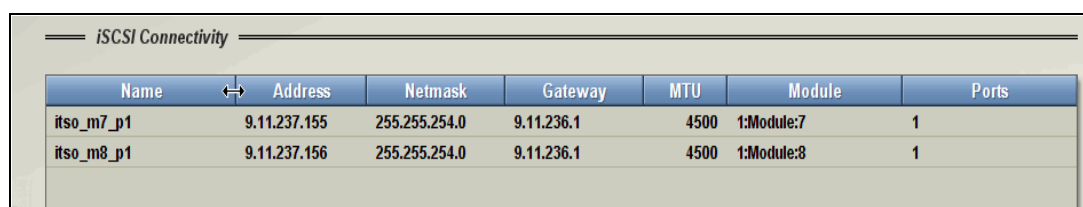
iSCSI ports can be easily identified and configured in the XIV Storage System. Use either the GUI or an XCLI command to display current settings.

Viewing iSCSI configuration using the GUI

Log on to the XIV GUI, select the XIV Storage System to be configured, and move the mouse over the **Hosts and Clusters** icon. Select **iSCSI connectivity** (see Figure 1-24 on page 26).

The iSCSI connectivity panel is displayed as shown in Figure 1-27. Right-click the port and select **Edit** from the context menu to make changes.

Note that in our example, only two of the six iSCSI ports are configured. Non-configured ports are not displayed in the GUI.



Name	Address	Netmask	Gateway	MTU	Module	Ports
itso_m7_p1	9.11.237.155	255.255.254.0	9.11.236.1	4500	1:Module:7	1
itso_m8_p1	9.11.237.156	255.255.254.0	9.11.236.1	4500	1:Module:8	1

Figure 1-27 iSCSI connectivity

View iSCSI configuration using the XCLI

The *ipinterface_list* command illustrated in Example 1-5 can be used to display configured network ports only.

Example 1-5 XCLI to list iSCSI ports with *ipinterface_list* command

```
>> ipinterface_list
Name      Type      IP Address  Network Mask  Default Gateway  MTU  Module  Ports
-----
itso_m8_p1 iSCSI     9.11.237.156 255.255.254.0 9.11.236.1      4500 1:Module:8 1
itso_m7_p1 iSCSI     9.11.237.155 255.255.254.0 9.11.236.1      4500 1:Module:7 1
management Management 9.11.237.109 255.255.254.0 9.11.236.1      1500 1:Module:4
management Management 9.11.237.107 255.255.254.0 9.11.236.1      1500 1:Module:5
management Management 9.11.237.108 255.255.254.0 9.11.236.1      1500 1:Module:6
VPN        VPN       0.0.0.0      255.0.0.0     0.0.0.0         1500 1:Module:4
VPN        VPN       0.0.0.0      255.0.0.0     0.0.0.0         1500 1:Module:6
```

Note that when you type this command, the rows might be displayed in a different order.

To see a complete list of IP interfaces, use the command *ipinterface_list_ports*. Example 1-6 on page 28 shows an example of the results of running this command.

Example 1-6 XCLI to list iSCSI ports with ipinterface_list_ports command

```
>>ipinterface_list_ports
```

Index	Role	IP Interface	Connected Component	Link Up?	Negotiated Speed (MB/s)	Full Duplex? Duplex	Module
1	Management			yes	1000	yes	1:Module:4
1	Component		1:UPS:1	yes	100	no	1:Module:4
1	Laptop			no	0	no	1:Module:4
1	VPN			no	0	no	1:Module:4
1	Management			yes	1000	yes	1:Module:5
1	Component		1:UPS:2	yes	100	no	1:Module:5
1	Laptop			no	0	no	1:Module:5
1	Remote_Support_Module			yes	1000	yes	1:Module:5
1	Management			yes	1000	yes	1:Module:6
1	Component		1:UPS:3	yes	100	no	1:Module:6
1	VPN			no	0	no	1:Module:6
1	Remote_Support_Module			yes	1000	yes	1:Module:6
1	iSCSI			unknown	N/A	unknown	1:Module:9
2	iSCSI			unknown	N/A	unknown	1:Module:9
1	iSCSI	itso_m8_p1		yes	1000	yes	1:Module:8
2	iSCSI			unknown	N/A	unknown	1:Module:8
1	iSCSI	itso_m7_p1		yes	1000	yes	1:Module:7
2	iSCSI			unknown	N/A	unknown	1:Module:7

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 in to the system.

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

Note: 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 only perform an iSCSI login to the IBM XIV Storage System 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 aren't unique, an error message is displayed. However, the command does not fail.
- ▶ The secret must be between 96 bits and 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 entered is treated as a 6 bit equivalent length.
 - Hex requires that 0x is used as a prefix for the entry. Each subsequent character entered is treated as a 4 bit equivalent length.
 - String requires that a prefix is not used (that is, it cannot be prefixed with 0b or 0x). Each character entered is treated as a 8 bit equivalent length.
- ▶ If the `iscsi_chap_secret` parameter does not conform to the required secret length (96 to 128 bits), the command fails.

- If you change the `iscsi_chap_name` or `iscsi_chap_secret` parameters, a warning message is displayed that says the changes will apply the next time the host is connected.

Configuring CHAP

Currently, you can only use the XCLI to configure CHAP. 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, it is not supported to boot through iSCSI, even if an iSCSI HBA is used.

1.4 Logical configuration for host connectivity

This section shows the tasks 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, refer to the appropriate section in 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, additional and operating system-specific configuration tasks are required. The tasks are described in subsequent chapters of this book according to the operating system of the host that is being configured.

1.4.1 Host configuration preparation

We use the environment shown in Figure 1-28 to illustrate the configuration tasks. In our example, we have two hosts: one host using FC connectivity and the other host using iSCSI. The diagram also shows the unique names of components that are used in the configuration steps.

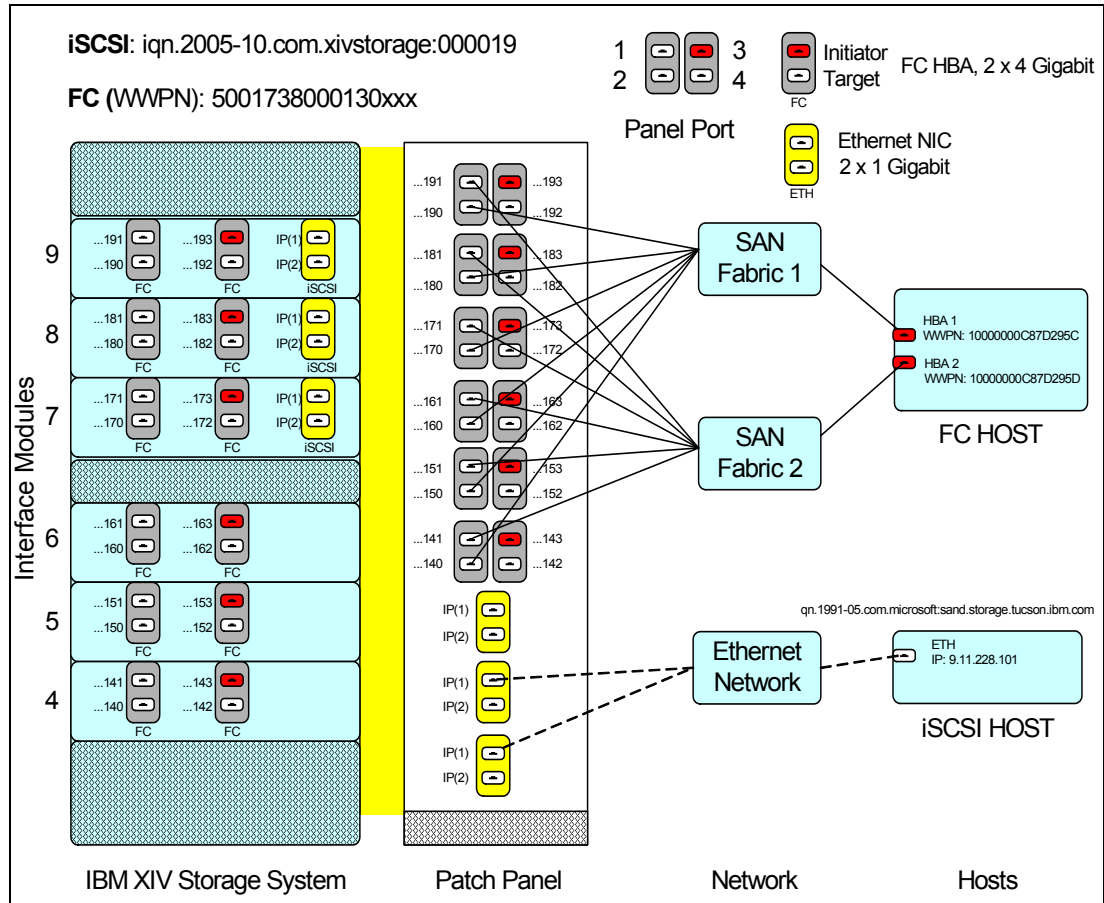


Figure 1-28 Example: Host connectivity overview of base setup

The following assumptions are made for the scenario shown in Figure 1-28:

- ▶ 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.

Hardware information

We recommend writing down the component names and IDs because this saves time during the implementation. An example is illustrated in Table 1-2 for our particular scenario.

Table 1-2 Example: 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: 10000000C87D295C HBA2 WWPN: 10000000C87D295D	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
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

Note: The OS Type is *default* for all hosts except HP-UX and zVM.

FC host specific tasks

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

For procedures showing how to configure zoning, refer to your FC switch manual. The following is an example of what the zoning details might look like for a typical server HBA zone. (Note that if using SVC as a host, there will be additional requirements that are not discussed here.)

Fabric 1 HBA 1 zone

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

```
zone: prime_sand_1
    prime_4_1; prime_5_3; prime_6_1;
    prime_7_3; sand_1
```

Fabric 2 HBA 2 zone

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

```
zone: prime_sand_2
    prime_4_1; prime_5_3; prime_6_1;
```

```
prime_7_3; sand_2
```

In the foregoing examples, 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 to allow the hosts and the XIV to communicate over IP.

1.4.2 Assigning LUNs to a host using the GUI

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

Defining a host

To define a host, follow these steps:

1. In the XIV Storage System main GUI window, move the mouse cursor over the **Hosts and Clusters** icon and select **Hosts and Clusters** (see Figure 1-29).



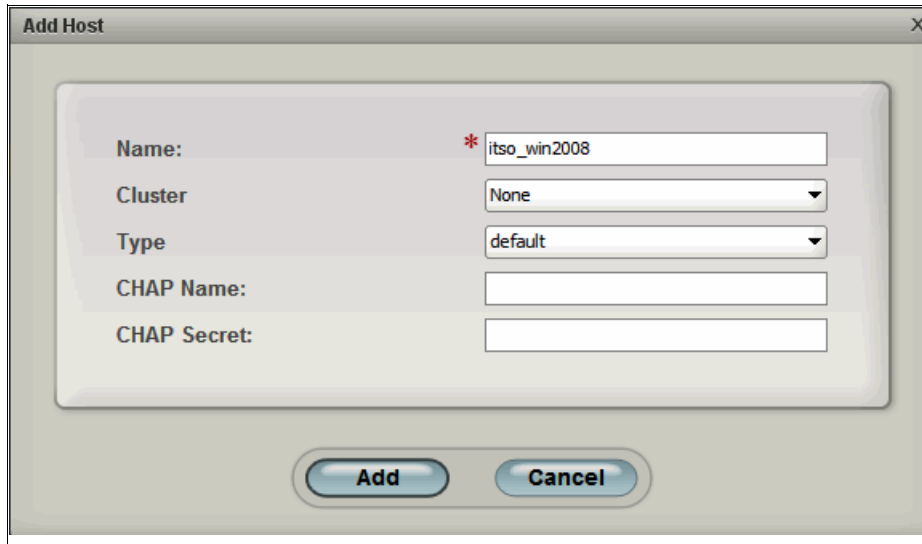
Figure 1-29 Hosts and Clusters menu

2. The Hosts window is displayed showing a list of hosts (if any) that are already defined. To add a new host or cluster, click either the **Add Host** or **Add Cluster** in the menu bar (see Figure 1-30). In our example, we select **Add Host**. The difference between the two is that **Add Host** is for a single host that will be assigned a LUN or multiple LUNs, whereas **Add Cluster** is for a group of hosts that will share a LUN or multiple LUNs.



Figure 1-30 Add new host

3. The **Add Host** dialog is displayed as shown in Figure 1-31 on page 33. Enter a name for the host. If a cluster definition was created in the previous step, it will be available in the cluster drop-down list box. To add a server to a cluster, select a cluster name. Because we do not create a cluster in our example, we select **None**. Type is **default**.



The 'Add Host' dialog box contains the following fields and controls:

- Name:** A text input field containing 'itso_win2008' with a red asterisk indicating it is required.
- Cluster:** A dropdown menu currently set to 'None'.
- Type:** A dropdown menu currently set to 'default'.
- CHAP Name:** An empty text input field.
- CHAP Secret:** An empty text input field.
- Buttons:** 'Add' and 'Cancel' buttons at the bottom.

Figure 1-31 Add host details

- Repeat steps 4 and 5 to create additional hosts. In our scenario we add another host called itso_win2008_iscsi
- Host access to LUNs is granted depending on the host adapter ID. For an FC connection, the host adapter ID is the FC HBA WWPNN, whereas for an iSCSI connection, the host adapter ID is the host IQN. To add a WWPNN or IQN to a host definition, right-click the host and select **Add Port** from the context menu (see Figure 1-32).

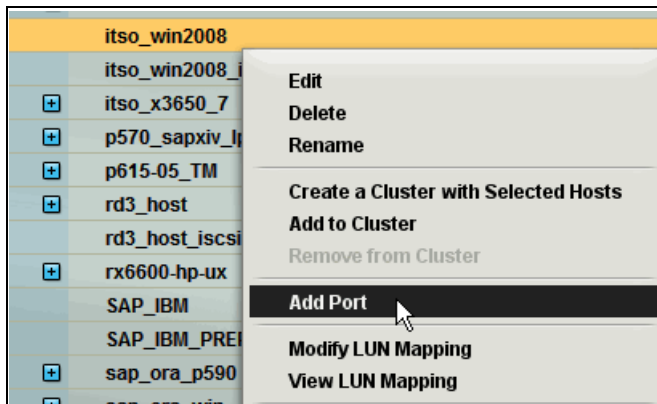


Figure 1-32 GUI example: Add port to host definition

- The Add Port window is displayed as shown in Figure 1-33 on page 34. Select port type FC or iSCSI.

In this example, an FC host is defined. Add the WWPNN for HBA1 as listed in Table 1-2 on page 31. If the host is correctly connected and has done a port login to the SAN switch at least once, the WWPNN is shown in the drop-down list box. Otherwise, you must manually enter the WWPNN. Adding ports from the drop-down list is preferable because it is less prone to error. However, if hosts have not yet been connected to the SAN or zoned, then manually adding the WWPNNs is the only option.

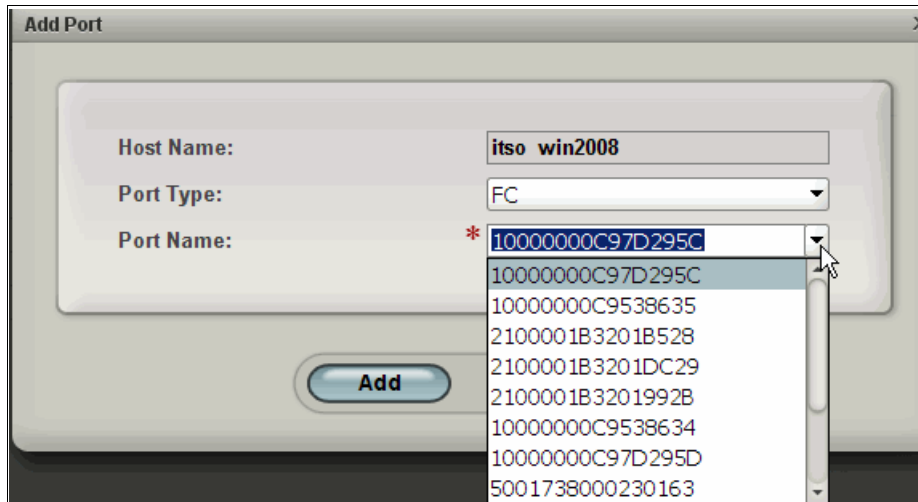


Figure 1-33 GUI example: Add FC port WWPN

Repeat steps 5 and 6 to add the second HBA WWPN; ports can be added in any order.

7. To add an iSCSI host, in the **Add Port** dialog, specify the port type as iSCSI and enter the IQN of the HBA as the iSCSI Name. Refer to Figure 1-34.



Figure 1-34 GUI example: Add iSCSI port

8. The host will appear with its ports in the Hosts dialog box as shown in Figure 1-35.

[-] itso_win2008	defa
10000000C97D295C	FC
10000000C97D295D	FC
[-] itso_win2008_iscsi	defa
iqn.1991-05.com.microsoft:sand.storage.tucson.ibm.com	iSC

Figure 1-35 List of hosts and ports

In this example, the hosts *itso_win2008* and *itso_win2008_iscsi* are in fact the same physical host. However, they have been 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 using the following steps:

1. While still in the **Hosts and Clusters** configuration pane, right-click the host to which the volume is to be mapped and select **Modify LUN Mappings** from the context menu (see Figure 1-36).

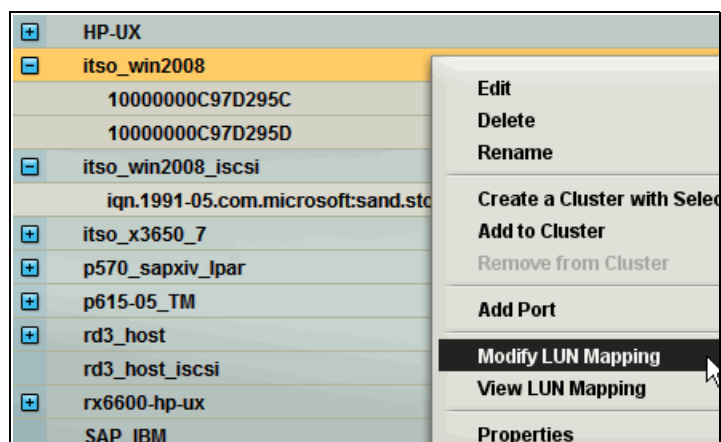


Figure 1-36 Map LUN to host

2. The Volume to LUN Mapping window opens as shown in Figure 1-37.
 - ▶ Select an available volume from the left pane.
 - ▶ The GUI will suggest a LUN ID to which to map the volume. However, this can be changed to meet your requirements.
 - ▶ Click **Map** and the volume is assigned immediately.

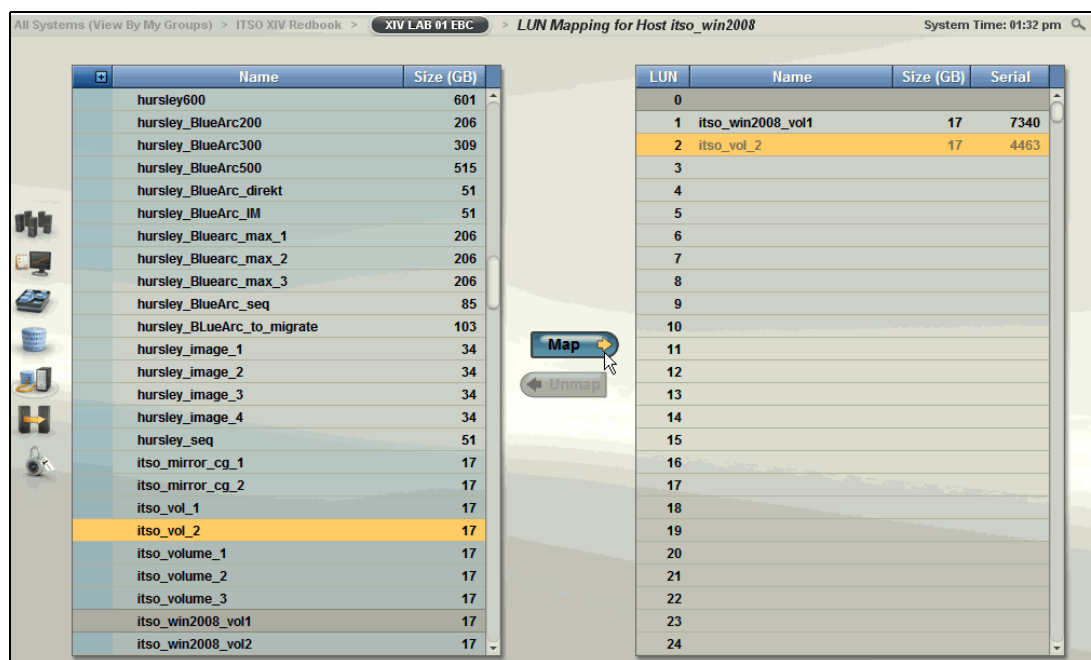


Figure 1-37 Map FC volume to FC host

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

- To complete this example, 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. See Figure 1-38.

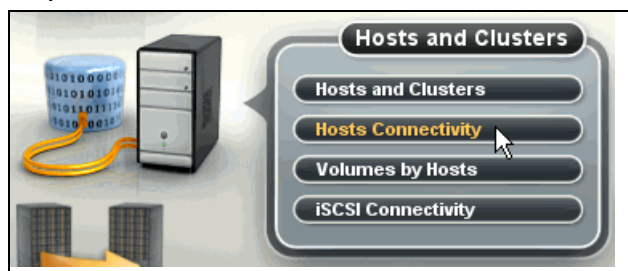


Figure 1-38 Hosts Connectivity

- The host connectivity window is displayed. In our example, the ExampleFHost was expected to have dual path connectivity to every module. However, only two modules (5 and 6) show as connected (refer to Figure 1-39), and the iSCSI host has no connection to module 9.

	Name	1:4	1:5	1:6	1:7	1:8	1:9
+	Bladecenter4_ebc_1						
+	Bladecenter4_ebc_2						
+	Bladecenter4_ebc_3						
+	Bladecenter4_ebc_4						
+	Bladecenter4v9						
-	ExampleFHost						
	210000E08B8FA29E			✓ ₂			
	210100E08BAFA29E		✓ ₂				
+	ExampleiSCSIhost						
	iqn.2000-04.com.qlogic:...				✓ ₁	✓ ₁	

Figure 1-39 GUI example: Host connectivity matrix

- The setup of the new FC or iSCSI hosts on the XIV Storage System is complete.
At this stage there might be operating-system-dependent steps that need to be performed. These are described in the operating system chapters.

1.4.3 Assigning LUNs to a host using the XCLI

There are a number of steps 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

Follow these steps to use the XCLI to prepare for a new host:

- Create a host definition for your FC and iSCSI hosts, using the **host_define** command. Refer to Example 1-7.

Example 1-7 XCLI example: Create host definition

```
>>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 1-8, the WWPN of the FC host for HBA1 and HBA2 is added with the **host_add_port** command and by specifying an **fcaddress**.

Example 1-8 Create FC port and add to host definition

```
>> host_add_port host=itso_win2008 fcaddress=10000000c97d295c
Command executed successfully.

>> host_add_port host=itso_win2008 fcaddress=10000000c97d295d
Command executed successfully.
```

In Example 1-9, the IQN of the iSCSI host is added. Note this is the same **host_add_port** command, but with the **iscsi_name** parameter.

Example 1-9 Create iSCSI port and add 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, follow these steps:

1. The final configuration step is to map LUNs to the host definition. Note that for a cluster, the volumes are mapped to the cluster host definition. There is no difference for FC or iSCSI mapping to a host. Both commands are shown in Example 1-10.

Example 1-10 XCLI example: Map volumes to hosts

```
>> 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. To complete the example, power up the server and check the host connectivity status from the XIV Storage System point of view. Example 1-11 shows the output for both hosts.

Example 1-11 XCLI example: Check host connectivity

```
>> host_connectivity_list host=itso_win2008
```

Host	Host Port	Module	Local FC port	Type
itso_win2008	10000000C97D295C	1:Module:6	1:FC_Port:6:1	FC
itso_win2008	10000000C97D295C	1:Module:4	1:FC_Port:4:1	FC
itso_win2008	10000000C97D295D	1:Module:5	1:FC_Port:5:3	FC
itso_win2008	10000000C97D295D	1:Module:7	1:FC_Port:7:3	FC

```
>> 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 1-11 on page 37, there are two paths per host FC HBA and two paths for the single Ethernet port that was configured.

3. The setup of the new FC or iSCSI hosts on the XIV Storage System is now complete.

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

1.5 Performance considerations

There are several key points when configuring the host for optimal performance. Because the XIV Storage System is distributing the data across all the disks, an additional layer of volume management at the host, such as Logical Volume Manager (LVM), might hinder performance for workloads. Multiple levels of striping can also create an imbalance across a specific resource. Therefore, it is usually better to disable host striping of data for XIV Storage System volumes and allow the XIV Storage System to manage the data unless the application requires striping on LVM level.

Based on your host workload, you might need to modify the maximum transfer size that the host generates to the disk to obtain the peak performance. For applications with large transfer sizes, if a smaller maximum host transfer size is selected, the transfers are broken up, causing multiple round-trips between the host and the XIV Storage System. By making the host transfer size as large or larger than the application transfer size, fewer round-trips occur, and the system experiences improved performance. If the transfer is smaller than the maximum host transfer size, the host only transfers the amount of data that it has to send.

Due to the distributed data features of the XIV Storage System, high performance is achieved by parallelism. Specifically, the system maintains a high level of performance as the number of parallel transactions occur to the volumes. Ideally, the host workload should be tailored to use multiple threads. If that is not possible, spread the work across multiple volumes if the application works on thread per volume base.

1.5.1 HBA queue depth

The XIV Storage architecture was designed to perform under real-world customer production workloads: lots of I/O requests at the same time. Queue depth is an important host HBA setting because it essentially controls how much data is allowed to be “in flight” onto the SAN from the HBA. A queue depth of 1 requires that each I/O request be completed before another is started. A queue depth greater than one indicates that multiple host I/O requests could be waiting for responses from the storage system. Therefore, the higher the host HBA queue depth, the more parallel I/O goes to the XIV Storage System.

The XIV Storage architecture eliminates the storage concept of a large central cache. Instead, each component 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 is where the host queue depth becomes an important factor in maximizing XIV Storage I/O performance.

Configure large host HBA queue depths - start with a queue depth of 64 per HBA - to ensure that you exploit the parallelism of the XIV architecture.

Figure 1-40 shows a queue depth comparison for a database I/O workload (70 percent reads, 30 percent writes, 8k block size, DBO = Database Open).

The performance numbers in this example are valid for this special test at an IBM lab only. The numbers do not describe the general capabilities of IBM XIV Storage System as you might observe them in your environment.

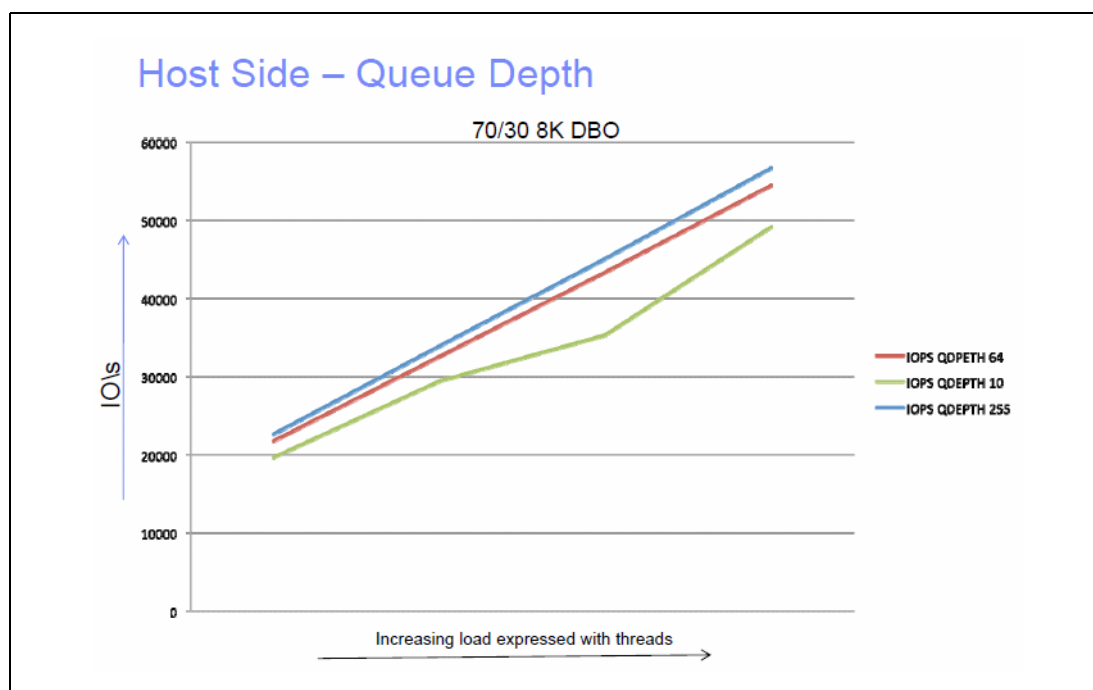


Figure 1-40 Host side queue depth comparison

Note: Although higher queue depth in general yields better performance with XIV, you must consider the limitations per port on the XIV side. For example, each HBA port on the XIV Interface Module is designed and set to sustain up to 1400 concurrent I/Os (except for port 3 when port 4 is defined as initiator, in which case port 3 is set to sustain up to 1000 concurrent I/Os). With a queue depth of 64 per host port suggested above, one XIV port is limited to 21 concurrent host ports given that each host will fill up the entire 64 queue depth for each request.

Other HBAs and operating systems will have their own procedures for configuring queue depth. Refer to your documentation for more information.

Figure 1-41 on page 40 shows an example of the Emulex HBAnyware utility used on a Windows server to change queue depth.

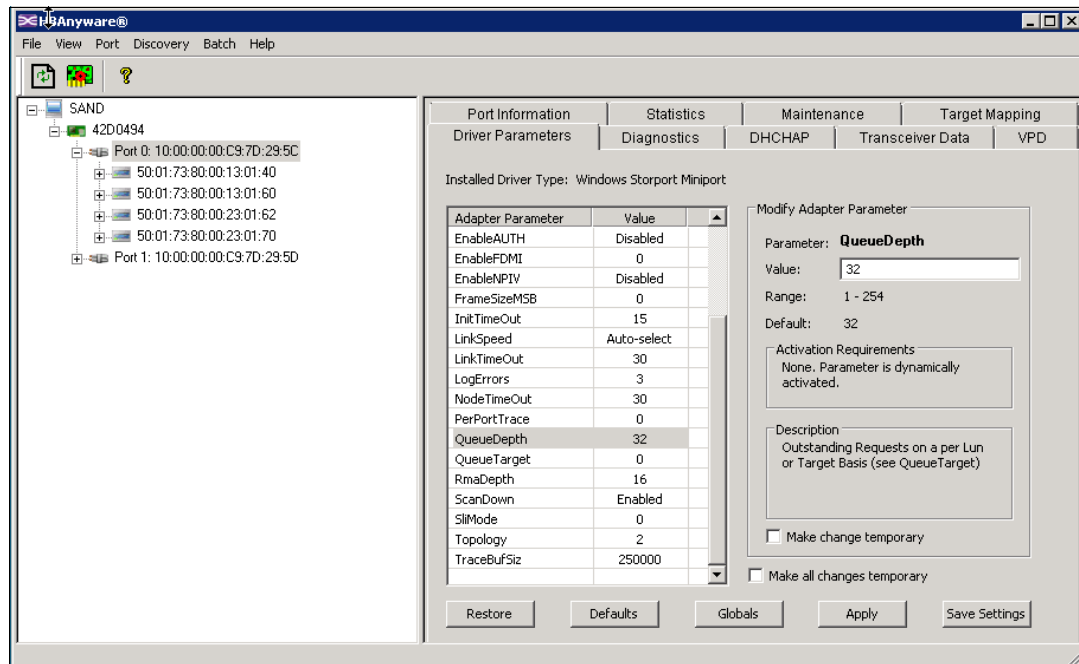


Figure 1-41 Emulex queue depth

1.5.2 Volume queue depth

The disk queue depth is an important OS setting that controls how much data is allowed to be “in flight” for a certain XIV volume to the HBA. The disk queue depth depends on the number of XIV volumes attached to this host from one XIV system and the HBA queue depth. For example, if you have a host with just one XIV volume attached and two HBAs with the recommended HBA queue depth of 64, you would need to configure a disk queue depth of 128 for this XIV volume to be able to fully utilize the queue of the HBAs.

1.5.3 Application threads and number of volumes

The overall design of the XIV grid architecture excels with applications that employ threads to handle the parallel execution of I/Os. It is critical to engage a large number of threads per process when running applications on XIV.

Generally there is no need to create a large number of small LUNs except when the application needs to use multiple LUNs to allocate or create multiple threads to handle the I/O. However, more LUNs might be needed to utilize queues on the host HBA side and the host Operating System side. However, if the application is sophisticated enough to define multiple threads independent of the number of LUNs, or the number of LUNs has no effect on application threads, there is no compelling reason to have multiple LUNs.

1.6 Troubleshooting

Troubleshooting connectivity problems can be difficult. However, the XIV Storage System does have built-in tools to assist with this. Table 1-3 contains a list of some of the built-in tools. For further information, refer to the XCLI manual, which can be downloaded from the XIV Information Center at:

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

Table 1-3 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



Windows Server 2008 host connectivity

This chapter explains specific considerations for attaching to XIV a Microsoft Windows Server 2008 host and attaching a Microsoft Windows 2003 Cluster.

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

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

and the Host Attachment publications at:

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

2.1 Attaching a Microsoft Windows 2008 host to XIV

This section discusses specific instructions for Fibre Channel (FC) and Internet Small Computer System Interface (iSCSI) connections. All the information here relates only to Windows Server 2008 (and not other versions of Windows) unless otherwise specified.

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

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

Also, refer to the XIV Storage System *Host System Attachment Guide for Windows - Installation Guide*, which is available at:

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

Prerequisites

To successfully attach a Windows host to XIV and access storage, a number of prerequisites need to be met. The following is a generic list, but your environment might have additional requirements:

- ▶ Complete the cabling.
- ▶ Complete the zoning.
- ▶ Install Service Pack 1 or later.
- ▶ Install any other updates if required.
- ▶ Install hot fix KB958912.
- ▶ Install hot fix KB932755 if required.
- ▶ Refer to KB957316 if booting from SAN.
- ▶ 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 SP1 and later (x86, x64)
- ▶ Windows Server 2003 SP1 and later (x86, x64)
- ▶ Windows 2000 Server SP4 (x86) available through RPQ

Supported FC HBAs

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

<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). For HBAs in Sun systems, use Sun-branded HBAs and Sun-ready HBAs only.

Multi-path support

Microsoft provides a multi-path framework and development kit called the Microsoft Multi-path I/O (MPIO). The driver development kit allows storage vendors to create Device Specific Modules (DSMs) for MPIO and 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 it to Windows as a single LUN. The Windows MPIO drivers enables a true active/active path policy, allowing 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>

Boot from SAN support

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

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

2.1.1 Windows host FC configuration

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

Installing HBA drivers

Windows 2008 includes drivers for many HBAs, however, it is likely that they will not be the latest version for your HBA. You should install the latest available driver that is supported. HBAs drivers are available from the IBM, Emulex, and QLogic websites. They will come with instructions that should be followed to complete the installation.

With Windows operating systems, the queue depth settings are specified as part of the host adapter's configuration through the BIOS settings or using a specific software 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 trying to evenly spread the I/O load across all available ports, taking 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 an built-in feature of Windows 2008. Follow these steps to install it:

1. Using *Server Manager*, select **Features Summary**, then right-click and select **Add Features**. In the **Select Feature** page, select **Multi-Path I/O**. See Figure 2-1 on page 46.
2. Follow the instructions on the panel to complete the installation. This might require a reboot.

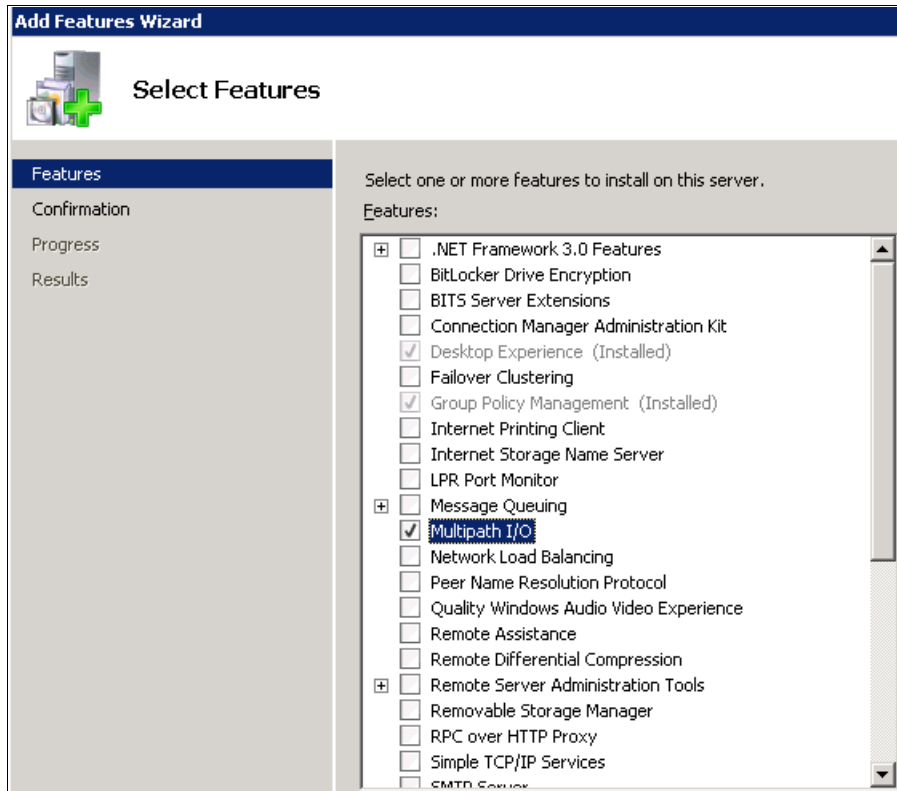


Figure 2-1 Selecting the Multipath I/O feature

3. To check that the driver has been installed correctly, load **Device Manager** and verify that it now includes **Microsoft Multi-Path Bus Driver** as illustrated in Figure 2-2.

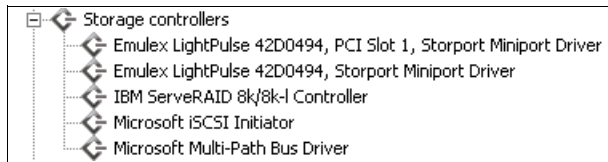


Figure 2-2 Microsoft Multi-Path Bus Driver

Windows Host Attachment Kit installation

The Windows 2008 Host Attachment Kit must be installed to gain access to XIV storage. Note that there are specific versions of the Host Attachment Kit for specific versions of Windows, and this is further sub-divided into 32-bit and 64-bit versions. The Host Attachment Kit can be downloaded from the following website:

http://www.ibm.com/support/search.wss?q=ssg1*&tc=STJTAG+HW3E0&rs=1319&dc=D400&dtm

The following instructions are based on the installation performed at the time of writing. You should also refer to the instructions in the *Windows Host Attachment Guide* because these instructions are subject to change over time. The instructions included here show the GUI installation. For command line instructions, refer to the *Windows Host Attachment Guide*.

Before installing the Host Attachment Kit, any other multipathing software that was previously installed must be removed. Failure to do so can lead to unpredictable behavior or even loss of data.

Install the XIV HAK (it is a mandatory prerequisite for support) using these steps:

1. Run the **XIV_host_attach-1.5.2-windows-x64.exe** file. When the setup file is run, it first determines if the python engine (*xpyv*) is required. If required, it will be automatically installed when you click Install as shown in Figure 2-3. Proceed with the installation following the installation wizard instructions.

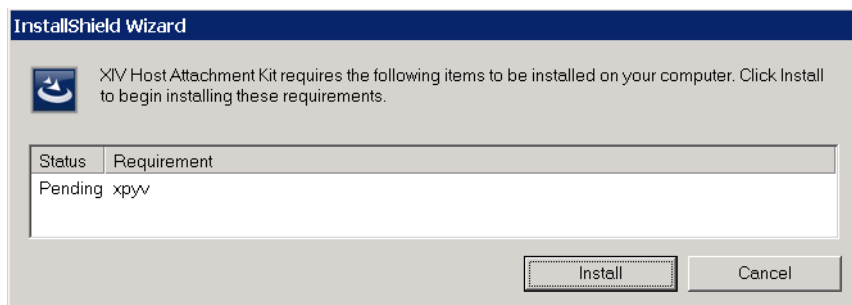


Figure 2-3 Determining requirements on Xpyv installation

2. After xpyv is installed, the XIV HAK installation wizard is launched. Follow the installation wizard instructions and select the complete installation option.

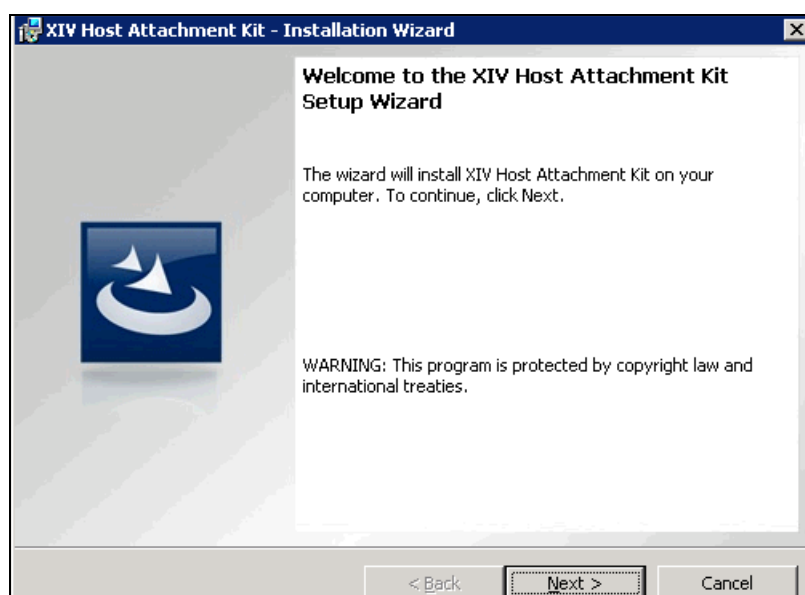


Figure 2-4 Welcome to XIV HAK installation wizard

3. Next, you need to run the XIV Host Attachment Wizard as shown in Figure 2-5. Click Finish to proceed.

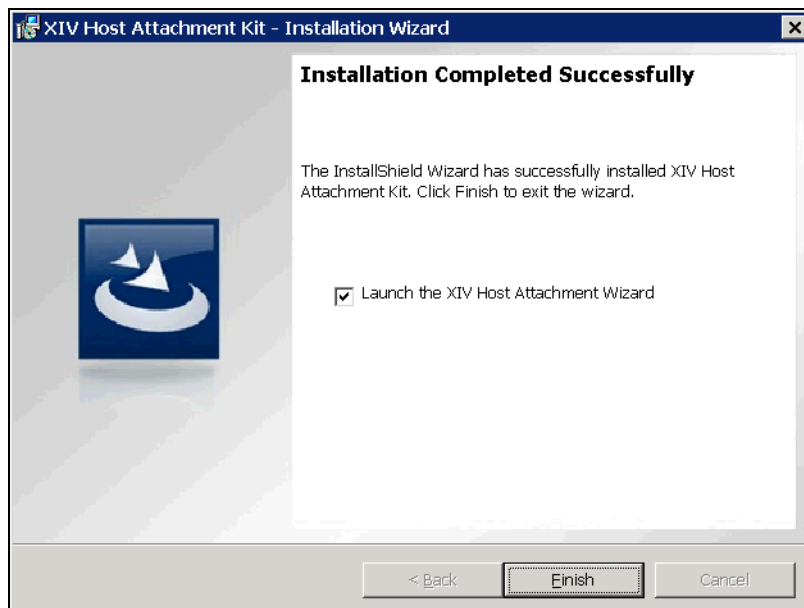


Figure 2-5 XIV HAK installation wizard Installation Completed

4. Answer questions from the XIV Host Attachment wizard as indicated Example 2-1. At the end, you need to reboot your host.

Example 2-1 First run of the XIV Host Attachment wizard

```
Welcome to the XIV Host Attachment wizard, version 1.5.2.
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 : f
```

```
-----
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...
-----
```

```
A reboot is required in order to continue.
Please reboot the machine and restart the wizard
```

```
Press [ENTER] to exit.
```

5. After you have rebooted, run the XIV Host Attachment Wizard again (from the **Start** button on your desktop, select **All programs** then select **XIV**, then click **XIV Host Attachment Wizard**. Answer to the questions prompted by the wizard as indicated in Example 2-2.

Example 2-2 Attaching host over FC to XIV using the XIV Host Attachment Wizard

Welcome to the XIV Host Attachment wizard, version 1.5.2.
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 : f

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:
21:00:00:e0:8b:87:9e:35: [QLogic QLA2340 Fibre Channel Adapter]: QLA2340
21:00:00:e0:8b:12:a3:a2: [QLogic QLA2340 Fibre Channel Adapter]: QLA2340
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)
6000105	10.2	Yes	All	FC	sand
1300203	10.2	No	None	FC,iSCSI	--

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: sand]:
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.

At this point your Windows host should have all the required software to successfully attach to the XIV Storage System.

Scanning for new LUNs

Before you can scan for new LUNs, your host needs to be created, configured, and have LUNs assigned. See Chapter 1, “Host connectivity” on page 1 for information about how to do this. The following instructions assume that these operations have been completed.

1. Click **Server Manager** → **Device Manager** → **Action** → **Scan for hardware changes**. In the Device Manager tree under **Disk Drives**, your XIV LUNs will appear as shown in Figure 2-6.

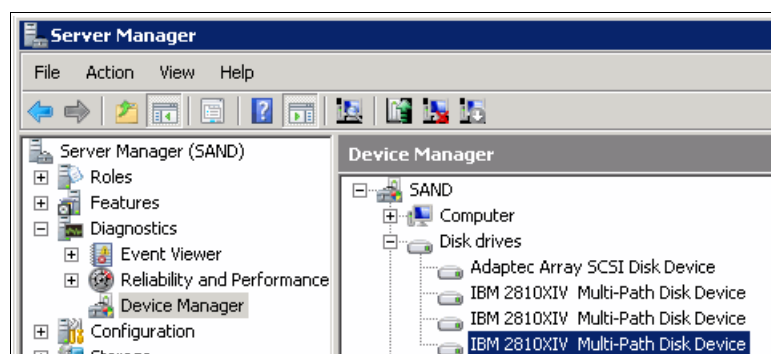


Figure 2-6 Multi-Path disk devices in Device Manager

The number of objects named **IBM 2810XIV SCSI Disk Device** will depend on the number of LUNs mapped to the host.

2. Right-click one of the **IBM 2810XIV SCSI Device** objects and select **Properties**. Click the **MPIO** tab to set the load balancing as shown in Figure 2-7.

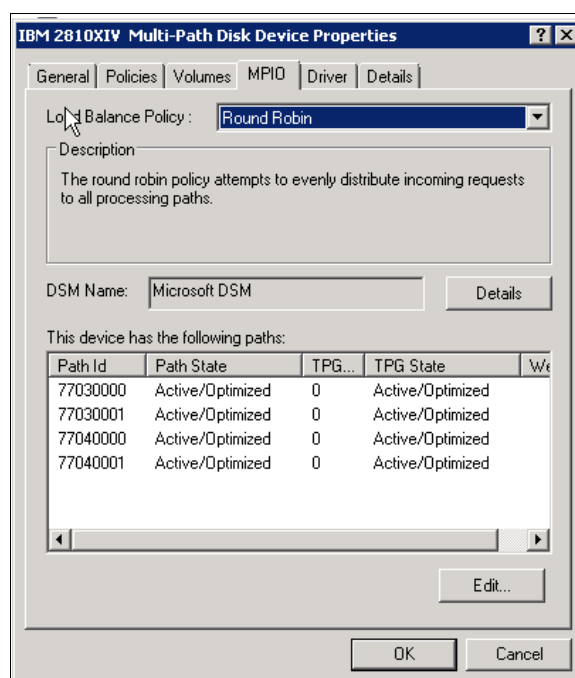


Figure 2-7 MPIO load balancing

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

The possible options are:

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

3. The mapped LUNs on the host can be seen under **Disk Management** as illustrated in Figure 2-8.

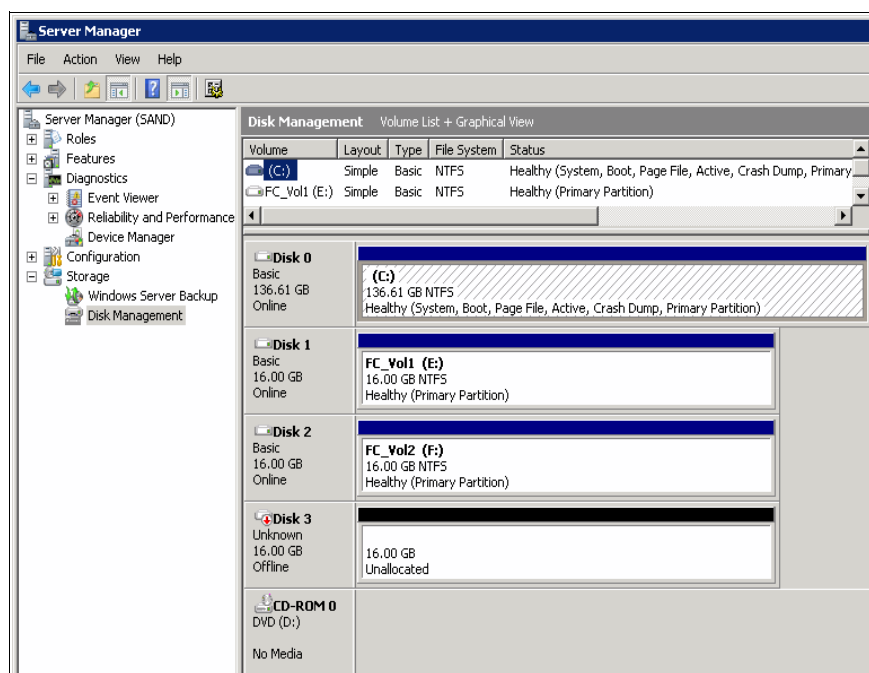


Figure 2-8 Mapped LUNs appear in Disk Management

2.1.2 Windows host iSCSI configuration

To establish the physical iSCSI connection to the XIV Storage System, refer to 1.3, “iSCSI connectivity” on page 21.

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

Supported iSCSI HBAs

For Windows, XIV does not support hardware iSCSI HBAs. The only adapters supported are standard Ethernet interface adapters using 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 45.

To install the Windows Host Attachment Kit, use the procedure explained under “Windows Host Attachment Kit installation” on page 46 until you reach the step where you need to reboot.

After rebooting, start the Host Attachment Kit installation wizard again, and follow the procedure given in Example 2-3.

Example 2-3 Running XIV Host Attachment Wizard on attaching to XIV over iSCSI

```
C:\Users\Administrator.SAND>xiv_attach
```

```
-----  
Welcome to the XIV Host Attachment wizard, version 1.5.2.  
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): 9.11.237.155  
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): 9.11.237.156  
Is this host defined in the XIV system to use CHAP? [default: 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)  
6000105   10.2   Yes          All            FC        sand  
1300203   10.2   No           None           FC,iSCSI  --  
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: sand ]:  
Please enter a username for system 1300203 : [default: admin ]: itso  
Please enter the password of user itso for system 1300203:
```

```
Press [ENTER] to proceed.
```

You can now proceed with mapping XIV volumes to the defined Windows host, then configuring the Microsoft iSCSI software initiator.

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 shown in Figure 2-9.

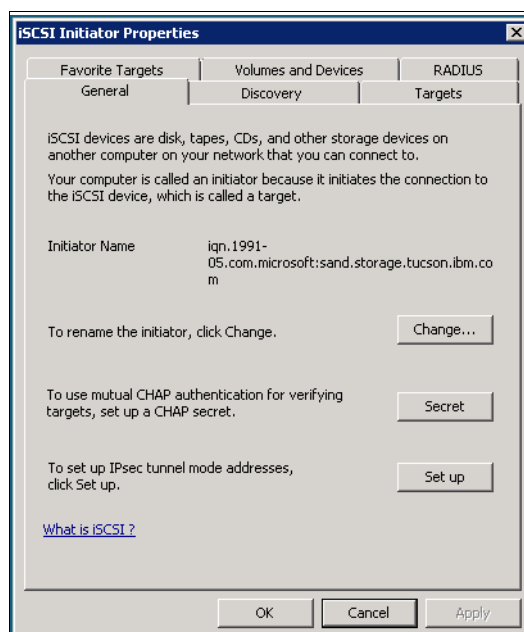


Figure 2-9 iSCSI Initiator Properties

2. Note the server's iSCSI Qualified Name (IQN) from the General tab (in our example, *iqn.1991-05.com.microsoft:sand.storage.tucson.ibm.com*). Copy this IQN to your clipboard and use it to define this host on the XIV Storage System.
3. Select the **Discovery** tab and click **Add Portal** button in the **Target Portals** pane. Use one of your XIV Storage System's iSCSI IP addresses. Figure 2-10 shows the results.

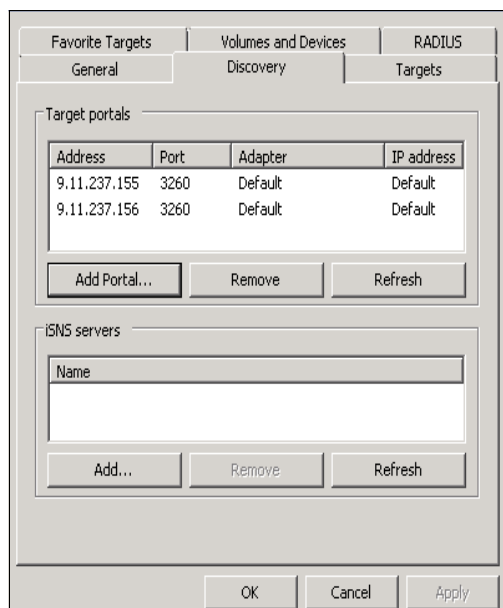


Figure 2-10 iSCSI targets portals defined

Repeat this step for additional target portals

4. To view IP addresses for the iSCSI ports in the XIV GUI, move the mouse cursor over the **Hosts and Clusters** menu icons in the main XIV window and select **iSCSI Connectivity** as shown in Figure 2-11 and Figure 2-12.



Figure 2-11 iSCSI Connectivity

Name	Address	Netmask	Gateway	MTU	Module	Ports
itso_m7_p1	9.11.237.155	255.255.254.0	9.11.236.1	4500	1:Module:7	1
itso_m8_p1	9.11.237.156	255.255.254.0	9.11.236.1	4500	1:Module:8	1

Figure 2-12 iSCSI connectivity

Alternatively, you can issue the Extended Command Line Interface (XCLI) command as shown in Example 2-4.

Example 2-4 List iSCSI interfaces

```
>> ipinterface_list
```

Name	Type	IP Address	Network Mask	Default Gateway	MTU	Module	Ports
itso_m8_p1	iSCSI	9.11.237.156	255.255.254.0	9.11.236.1	4500	1:Module:8	1
management	Management	9.11.237.109	255.255.254.0	9.11.236.1	1500	1:Module:4	
VPN	VPN	0.0.0.0	255.0.0.0	0.0.0.0	1500	1:Module:4	
management	Management	9.11.237.107	255.255.254.0	9.11.236.1	1500	1:Module:5	
management	Management	9.11.237.108	255.255.254.0	9.11.236.1	1500	1:Module:6	
VPN	VPN	0.0.0.0	255.0.0.0	0.0.0.0	1500	1:Module:6	
itso_m7_p1	iSCSI	9.11.237.155	255.255.254.0	9.11.236.1	4500	1:Module:7	1

You can see that the iSCSI addresses used in our test environment are 9.11.237.155 and 9.11.237.156.

5. The XIV Storage System will be discovered by the initiator and displayed in the **Targets** tab as shown in Figure 2-13. At this stage the Target will show as **Inactive**.

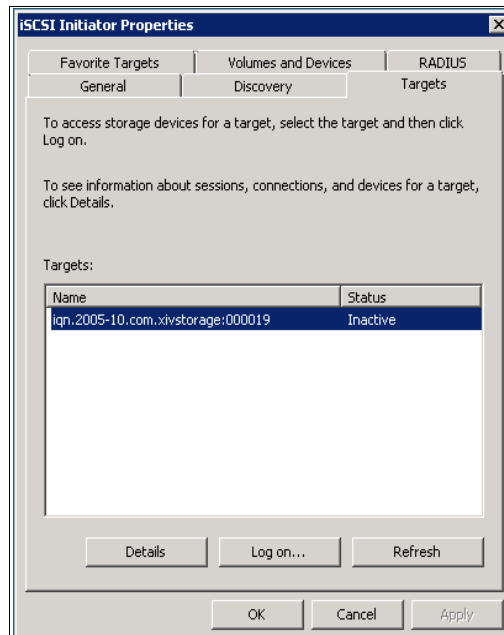


Figure 2-13 A discovered XIV Storage with Inactive status

6. To activate the connection, click **Log On**. In the **Log On to Target** pop-up window, select the **Enable multi-path** and **Automatically restore this connection when the system boots** check boxes as shown in Figure 2-14.

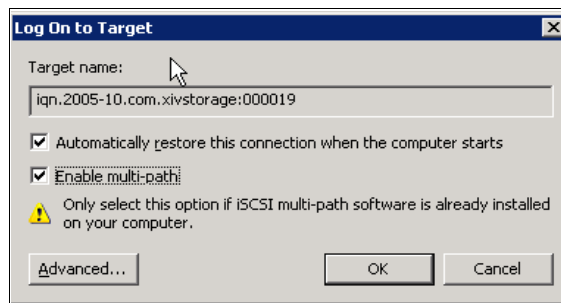


Figure 2-14 Log On to Target

- Click **Advanced**. The **Advanced Settings** window is displayed. Select the **Microsoft iSCSI Initiator** from the **Local adapter** drop-down. In the **Source IP** drop-down, click the first host IP address to be connected. In the **Target Portal**, select the first available IP address of the XIV Storage System as shown in Figure 2-15. Click **OK**, and you will be returned to the parent window. Click **OK** again.

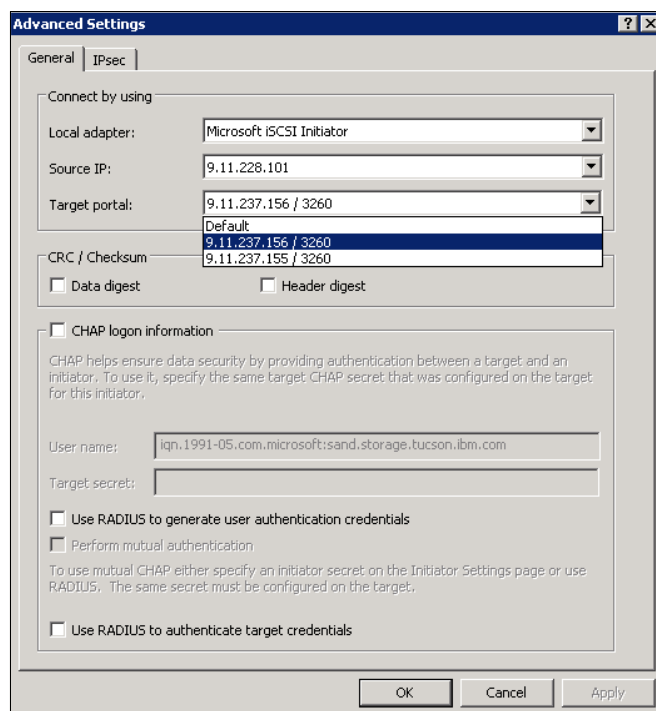


Figure 2-15 Advanced Settings in the Log On to Target panel

- The iSCSI Target connection status now shows as **Connected** as shown in Figure 2-16.

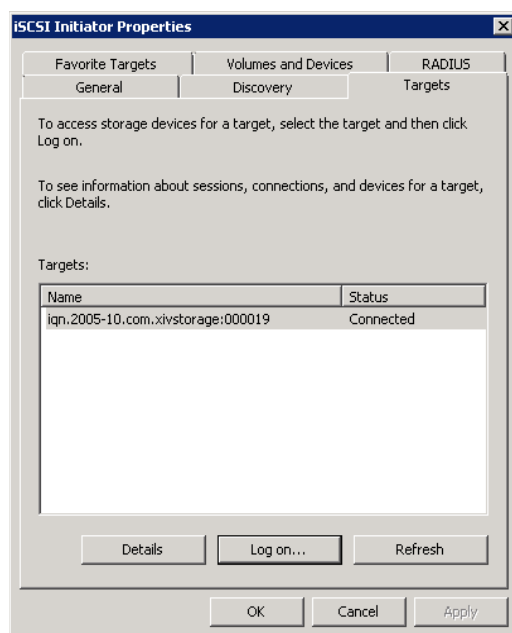


Figure 2-16 A discovered XIV Storage with Connected status

9. The redundant paths are not yet configured. To do so, repeat this process for all IP addresses on your host and all Targets Portals (XIV iSCSI targets). This establishes connection sessions to all of the desired XIV iSCSI interfaces from all of your desired source IP addresses.

After the iSCSI sessions are created to each target portal, you can see details of the sessions. Click the **Targets** tab, highlight the target, and click **Details** to verify the sessions of the connection as seen in Figure 2-17.

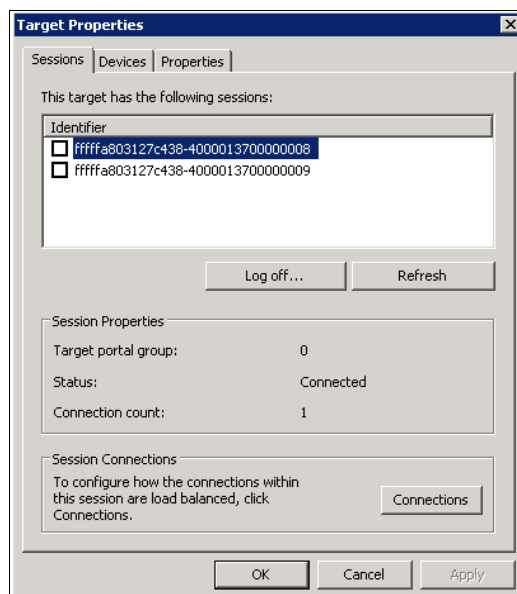


Figure 2-17 Target connection details

Depending on your environment, numerous sessions might appear.

10. To see further details or change the load balancing policy, click the **Connections** button as shown in Figure 2-18.

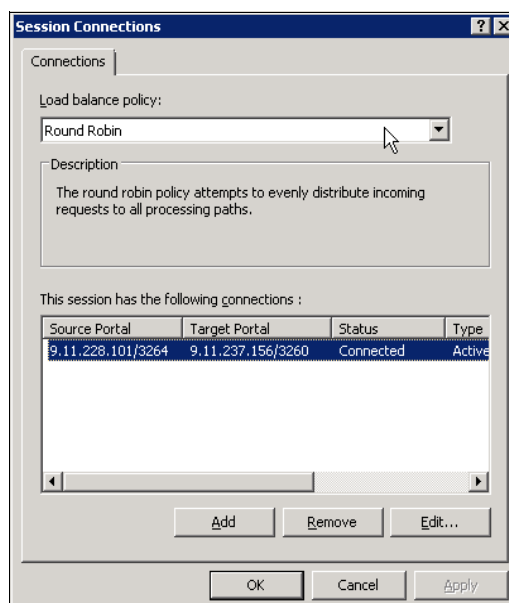


Figure 2-18 Connected sessions

The default load balancing policy should be **Round Robin**. Change this only if you are confident that another option is better suited to your environment.

The possible options are as follows:

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

11. At this stage, if you have already mapped volumes to the host system, you will see them under the **Devices** tab. If no volumes are mapped to this host yet, you can assign them now.

Another way to verify your assigned disks is to open the Windows Device Manager as shown in Figure 2-19.

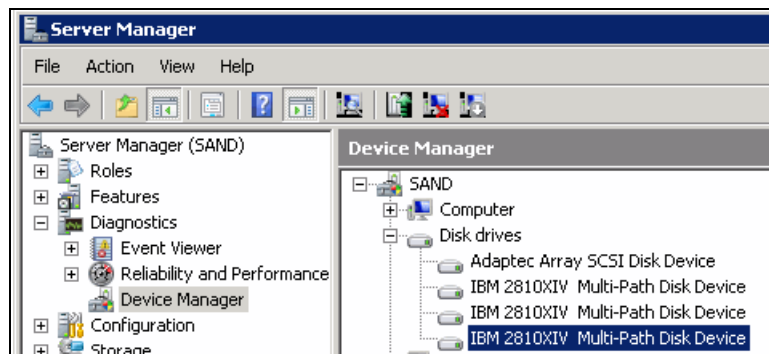


Figure 2-19 Windows Device Manager with XIV disks connected through iSCSI

12. The mapped LUNs on the host can be seen in **Disk Management** as illustrated in Figure 2-20.

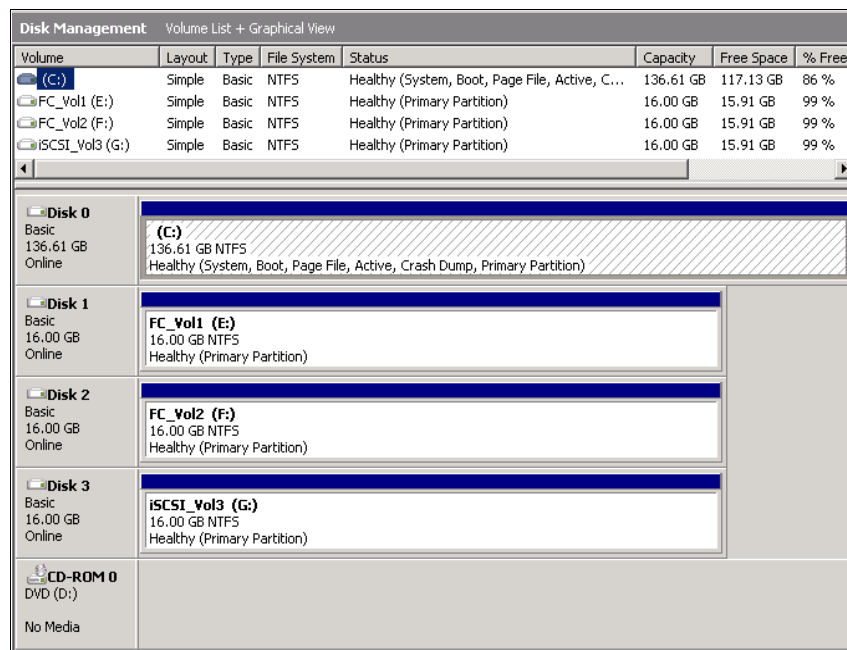


Figure 2-20 Mapped LUNs appear in Disk Management

2.1.3 Management volume LUN 0

In Device Manager additional devices under **XIV SCSI Array** appear as shown in Figure 2-21.

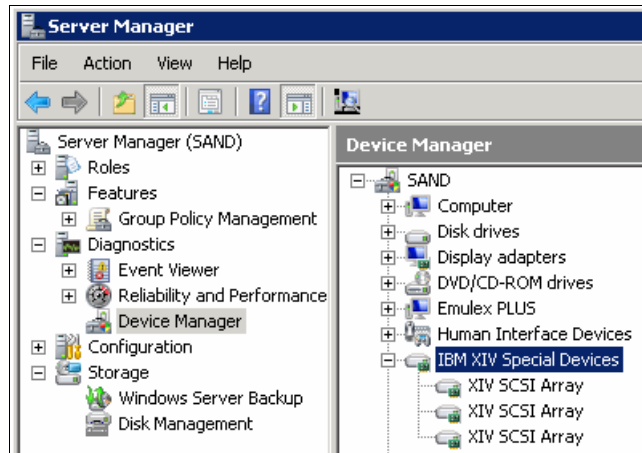


Figure 2-21 XIV special LUNs

The net result of this is that the mapping slot on the XIV for LUN 0 will be reserved and disabled, but the slot can be enabled and used as normal with no ill effects. From a Windows point of view, these devices can be ignored.

2.1.4 Host Attachment Kit utilities

The Host Attachment Kit (HAK) includes the following utilities.

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 2-5.

Example 2-5 xiv_devlist

```
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 requires Administrator privileges. The 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. To run, go to a command prompt and enter **xiv_diag** as shown in Example 2-6.

Example 2-6 xiv_diag

```
C:\Users\Administrator.SAND>xiv_diag
Please type in a path to place the xiv_diag file in [default: C:\Windows\Temp]:
Creating archive xiv_diag-results_2010-10-27_18-49-32
INFO: Gathering System Information (1/2)... DONE
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 (1/2)... DONE
INFO: Gathering Windows Setup API (2/2)... DONE
INFO: Gathering Hardware Registry Subtree... DONE
INFO: Gathering xiv_devlist... 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 Multipath I/O Agent Data... 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: Closing xiv_diag archive file DONE
Deleting temporary directory... DONE
INFO: Gathering is now complete.
INFO: You can now send C:\Windows\Temp\xiv_diag-results_2010-10-27_18-49-32.tar.gz to
IBM-XIV for review
INFO: Exiting.
```

wfetch

This is a simple CLI utility for downloading files from HTTP, HTTPS, and FTP sites. It runs on most UNIX®, Linux, and Windows operating systems.

2.1.5 Installation for Windows 2003

The installation for Windows 2003 follows a set of procedures similar to that of Windows 2008 with the exception that Windows 2003 does not have native MPIO support.

MPIO support for Windows 2003 is installed as part of the Host Attachment Kit.

Review the prerequisites and requirements outlined in the XIV Host Attachment Kit.

2.2 Attaching a Microsoft Windows 2003 Cluster to XIV

This section discusses the attachment of Microsoft Windows 2003 cluster nodes to the XIV Storage System.

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

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

Also, refer to the XIV Storage System *Host System Attachment Guide for Windows - Installation Guide*, which is available at:

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

This section only focuses on the implementation of a two node Windows 2003 Cluster using FC connectivity and assumes that all of the following prerequisites have been completed.

2.2.1 Prerequisites

To successfully attach a Windows cluster node to XIV and access storage, a number of prerequisites need to be met. Here is a generic list: your environment might have additional requirements:

- ▶ Complete the cabling.
- ▶ Configure the zoning.
- ▶ Install Windows Service Pack 2 or later.
- ▶ Install any other updates if required.
- ▶ Install hot fix KB932755 if required.
- ▶ Install the Host Attachment Kit.
- ▶ Ensure that all nodes are part of the same domain.
- ▶ Create volumes to be assigned 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
- ▶ Windows Server 2003 SP2

Supported configurations of Windows Cluster Server

Windows Cluster Server is supported in the following configurations:

- ▶ 2 node: All versions

- ▶ 4 node: Windows 2003 x64
- ▶ 4 node: Windows 2008 x86

If other configurations are required, you will need a Request for Price Quote (RPQ). This is a process by which IBM will test a specific customer configuration to determine if it can be certified and supported. Contact your IBM representative for more information.

Supported FC HBAs

Supported FC HBAs are available from IBM, Emulex, and QLogic. Further details on driver versions are available from SSIC at the following URL:

<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). For HBAs in Sun systems, use Sun-branded HBAs and Sun-ready HBAs only.

Multi-path support

Microsoft provides a multi-path framework and development kit called the Microsoft Multi-path I/O (MPIO). The driver development kit allows storage vendors to create Device Specific Modules (DSM) for MPIO and 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 it to Windows as a single LUN. The Windows MPIO drivers enable a true active/active path policy allowing I/O over multiple paths simultaneously.

MPIO support for Windows 2003 is installed as part of the Windows Host Attachment Kit.

Further information about Microsoft MPIO support is available at the following URL:

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

2.2.2 Installing Cluster Services

In our scenario described next, we install a two node Windows 2003 Cluster. Our procedures assume that you are familiar with Windows 2003 Cluster and focus on specific requirements for attaching to XIV. For further details about installing a Windows 2003 Cluster, refer to the following URL:

<http://www.microsoft.com/downloads/details.aspx?familyid=96F76ED7-9634-4300-9159-89638F4B4EF7&displaylang=en>

To install the cluster, follow these steps:

1. Set up a cluster specific configuration. This includes:
 - Public network connectivity
 - Private (Heartbeat) network connectivity
 - Cluster Service account
2. Before continuing, ensure that at all times, only one node can access the shared disks until the cluster service has been installed on the first node. To do this, turn off all nodes except the first one (Node 1) that will be installed.
3. On the XIV system, click **Hosts and Clusters** → **Hosts and Clusters**. Create a cluster and put both nodes into the cluster as depicted in Figure 2-22 on page 63.

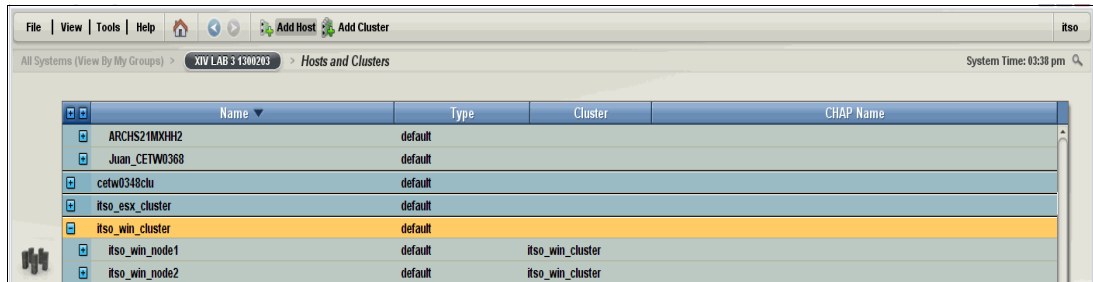


Figure 2-22 XIV cluster with Node 1

You can see that an XIV cluster named *itso_win_cluster* has been created and both nodes have been put in. Node2 must be turned off.

4. Map the quorum and data LUNs to the cluster as shown in Figure 2-23.

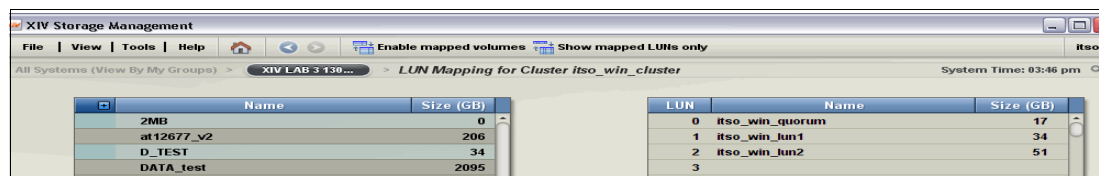


Figure 2-23 Mapped LUNs

You can see here that three LUNs have been mapped to the XIV cluster (but not to the individual nodes).

5. On Node1, scan for new disks, then initialize, partition, and format them with NTFS.

Microsoft has some best practices for drive letter usage and drive naming. For more information, refer to the following document:

<http://support.microsoft.com/?id=318534>

For our scenario, we use the following values:

- Quorum drive letter = Q
- Quorum drive name = DriveQ
- Data drive 1 letter = R
- Data drive 1 name = DriveR
- Data drive 2 letter = S
- Data drive 2 name = DriveS

The following requirements are for shared cluster disks:

- These disks must be basic disks.
- For 64-bit versions of Windows 2003, they must be MBR disks.

Refer to Figure 2-24 on page 64 for what this would look like on Node1.

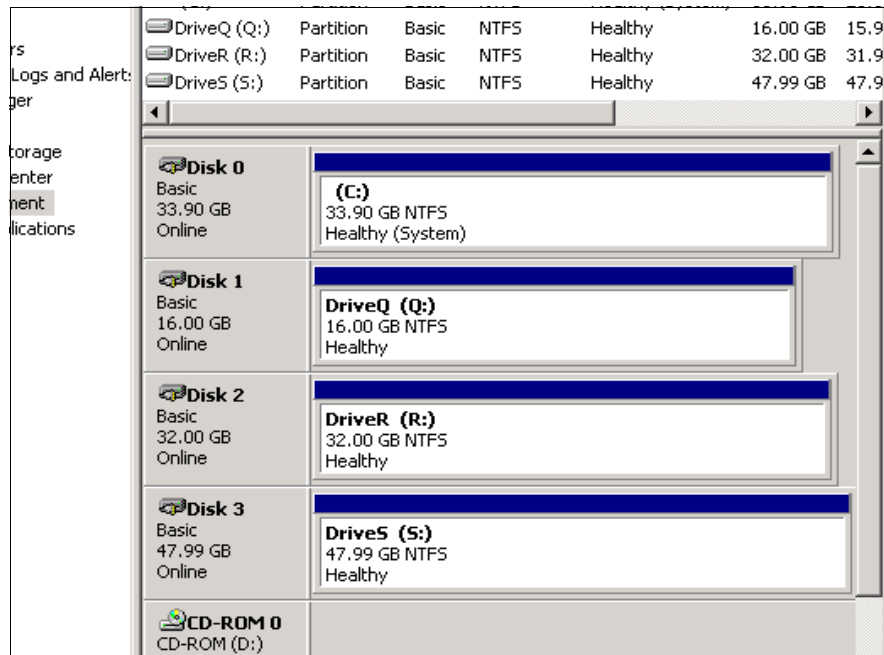


Figure 2-24 Initialized, partitioned, and formatted disks

6. 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 Node1 off.
7. Turn on Node2 and scan for new disks. All the disks should appear (in our case, three disks). They will already be initialized and partitioned. However, they might need formatting again. You will still have to set drive letters and drive names, and these must be identical to those set in step 4.
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, then turn Node2 off.
9. Turn Node1 back on, launch Cluster Administrator, and create a new cluster. Refer to documentation from Microsoft if necessary for help with this task.
10. After the cluster service is installed on Node1, turn on Node2. Launch **Cluster Administrator** on Node2 and install Node2 into the cluster.
11. Change the boot delay time on the nodes so that Node2 boots one minute after Node1. If you have more nodes, then continue this pattern so that, for instance, Node3 boots one minute after Node2, and so on. The reason for this is that if all the nodes boot at once and try to attach to the quorum resource, the cluster service might fail to initialize.
12. At this stage the configuration is complete with regard to the cluster attaching to the XIV system. However, there might be some post-installation tasks to complete. Refer to the Microsoft documentation for more information. Figure 2-25 on page 65 shows resources split between the two nodes.

Name	State	Owner	Group	Resource Type
Disk S:	Online	ITSO_WIN_NODE2	Group 1	Physical Disk
Cluster IP Address	Online	ITSO_WIN_NODE1	Cluster Group	IP Address
Cluster Name	Online	ITSO_WIN_NODE1	Cluster Group	Network Name
Disk Q:	Online	ITSO_WIN_NODE1	Cluster Group	Physical Disk
Disk R:	Online	ITSO_WIN_NODE2	Group 0	Physical Disk

Figure 2-25 Cluster resources shared between nodes

2.3 Boot from SAN

Booting from SAN creates a number of possibilities that are not available when booting from local disks. It means that the operating systems and configuration of SAN based computers can be centrally stored and managed. This can provide advantages with regard to deploying servers, backup, and disaster recovery procedures.

SAN boot is described in 1.2.5, “Boot from SAN on x86/x64 based architecture” on page 16.



Linux host connectivity

In this chapter we discuss the specifics for attaching IBM XIV Storage System to host systems running Linux

It is not our intent to repeat everything contained in the sources of information listed in section 3.1.2, “Reference material” on page 68. However, to avoid you having to look for and read through several documents just to be able to configure a Linux server for XIV attachment, we will provide a comprehensive guide without going into too much detail. When we show examples, we will usually use the Linux console commands because they are more generic than the GUIs provided by various vendors.

In this guide we cover all hardware architectures that are supported for XIV attachment:

- ▶ Intel® x86 and x86_64, both Fibre Channel and iSCSI, using the *XIV Host Attachment Kit (HAK)*
- ▶ IBM Power Systems™
- ▶ IBM System z®

Although older Linux versions are supported to work with the IBM XIV Storage System, we limit the scope here to the most recent enterprise level distributions:

- ▶ Novell SUSE Linux Enterprise Server 11, Service Pack 1 (SLES11 SP1)
- ▶ Redhat Enterprise Linux 5, Update 5 (RH-EL 5U5).

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

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

and the Host Attachment publications at:

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

3.1 IBM XIV Storage System and Linux support overview

Linux is an open source, UNIX-like kernel. The term *Linux* is often used to mean the whole operating system of GNU/Linux. In this chapter we also use it with that meaning. The Linux kernel, along with the tools and software needed to run an operating system, are maintained by a loosely organized community of thousands of (mostly) volunteer programmers.

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.
- ▶ Depending on the target group, the distributions differ largely in the kind of support that is available.
- ▶ Linux is available for almost all computer architectures.
- ▶ Linux is rapidly evolving.

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

IBM only supports 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, 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 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 System Storage Interoperation Center (SSIC) at the following address:

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

There are exceptions to this strategy when the market demand justifies the test and support effort.

Tip: XIV also supports certain versions of CentOS, which is technically identical to RH-EL.

3.1.2 Reference material

There is a wealth of information available that helps you set up your Linux server to attach it to an XIV storage system.

Host System Attachment Guide for Linux

The *Host System Attachment Guide for Linux*, GA32-0647 provides instructions to prepare an Intel IA-32-based machine for XIV attachment, including:

- ▶ Fibre Channel and iSCSI connection
- ▶ Configuration of the XIV system
- ▶ Install and use the XIV Host Attachment Kit (HAK)
 - Discover XIV volumes

- Set up multipathing
- ▶ Prepare a system that boots from the XIV

The guide doesn't cover other hardware platforms such as IBM Power Systems or IBM System z. You can find this information in the XIV Information Center:

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

Online Storage Reconfiguration Guide

This is part of the documentation provided by Redhat for Redhat Enterprise Linux 5. Although written specifically for Redhat 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 the following URL:

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

DM Multipath Configuration and Administration

This is also part of the Redhat Enterprise Linux 5 documentation. It contains a lot of useful information for anyone who works with Device Mapper Multipathing (DM-MP), and is not only valid for Redhat Enterprise Linux:

- ▶ How Device Mapper Multipathing works
- ▶ How to setup and configure DM-MP within Redhat Enterprise Linux 5
- ▶ Troubleshooting DM-MP

This publication is available at the following URL:

http://docs.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux/5/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 interested in storage related subjects. The most interesting topics for our purposes are:

- ▶ Setup and configure multipath I/O
- ▶ Setup a system to boot from multipath devices
- ▶ Combine multipathing with Logical Volume Manager and Linux Software RAID

This publication is available at the following URL:

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 that contains information about Linux on Power Architecture®, including:

- ▶ A discussion forum

- An announcement section
- Technical articles

It can be found at the following URL:

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

Fibre Channel Protocol for Linux and z/VM on IBM System z

This IBM Redbook is a comprehensive guide to storage attachment using Fibre Channel to z/VM® and Linux on z/VM. It describes:

- The 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

This publication is available at the following URL:

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

Other sources of information

The Linux distributors' documentation pages are good starting points when it comes to installation, configuration, and administration of Linux servers, especially when you have to deal with server-platform-specific issues:

- Documentation for Novell SUSE Linux Enterprise Server is available at:
<http://www.novell.com/documentation/suse.html>
- Documentation for Redhat Enterprise Linux is available at:
<http://www.redhat.com/docs/manuals/enterprise/>

IBM System z dedicates its own web page to storage attachment using FCP at the following URL:

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

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

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

3.1.3 Recent storage related improvements to Linux

In this section we provide a summary of storage-related improvements that have been introduced to Linux in recent years. Details about usage and configuration are available in the subsequent sections.

Past issues

Below is a partial list of storage related issues that could be seen in older Linux versions and which have been addressed in recent versions:

- Limited number of devices that could be attached

- ▶ Gaps in LUN sequence leading to incomplete device discovery
- ▶ Limited dynamic attachment of devices
- ▶ Non-persistent device naming could lead to re-ordering
- ▶ 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 tried to anticipate all devices that would ever be used for a system and created the nodes for them. This often led to a confusing number of existing nodes on the one hand and missing ones on the other.

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

With these improvements, we now can be sure that the required device nodes exist immediately after a device is detected, and there only device nodes that are actually 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. In the case of a disk drive, or SAN attached volume, this name contains a string that uniquely identifies the volume. This ensures that every time this volume is attached to the system, it gets the same name.

Multipathing

Linux now has its own built-in multipathing solution. It is based on the *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 the issues we had when only proprietary multipathing solutions existed:

- ▶ Proprietary multipathing solutions were only supported for certain kernel versions. Therefore, systems could follow the distribution's update schedule.
- ▶ They often were binary only and would not be supported by the Linux vendors because they could not debug them.
- ▶ A mix of different vendors storage systems on the same server, or even different types of the same vendor, usually was not possible because the multipathing solutions could not coexist.

Today, DM-MP is the only multipathing solution fully supported by both Redhat and Novell for their enterprise Linux distributions. It is available on all hardware platforms and supports all block devices that can have more than one path. IBM has adopted a strategy to support 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 disk from Linux. SAN attached volumes are usually not detected automatically, because adding a

volume to an XIV host object does not create a hotplug trigger event such as inserting a USB storage device does. SAN attached volumes are discovered during user-initiated device scans and then automatically integrated into the system, including multipathing.

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

Dynamic LUN resizing

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

3.2 Basic host attachment

In this section we explain the steps you must take to make XIV volumes available to your Linux host. We start with remarks about the ways to attach storage for the different hardware architectures. Then we describe the 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 platform for Linux, the Intel x86 (32 or 64 bit) architecture, only allows direct mapping of XIV volumes to hosts through Fibre Channel fabrics and HBAs, or IP networks. The other platforms, IBM System z and IBM Power Systems, provide additional mapping methods to allow better exploitation of their much more advanced virtualization capabilities.

IBM Power Systems

Linux, running in an LPAR on an IBM Power system, can get storage from an XIV either directly, through an exclusively assigned Fibre Channel HBA, or through a Virtual IO Server (VIOS) running on the system.

We do not discuss direct attachment here because it basically works the same way as with the other platforms. VIOS attachment requires specific considerations that we discuss below. We do not go into details about the way VIOS works and how it is configured. You can refer to Chapter 8, “VIOS clients connectivity” on page 171 if you need such information. There also are IBM Redbooks available that cover this:

- ▶ *PowerVM Virtualization on IBM System p: Introduction and Configuration* (SG24-7940)
<http://www.redbooks.ibm.com/abstracts/sg247940.html>
- ▶ *IBM PowerVM Virtualization Managing and Monitoring* (SG247590)
<http://www.redbooks.ibm.com/abstracts/sg247590.html>

Virtual vscsi disks through VIOS

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

Example 3-1 Virtual SCSI disks

```
p6-570-lpar13:~ # ls SCSI
[0:0:1:0]    disk    AIX      VDASD          0001  /dev/sda
```

[0:0:2:0]	disk	AIX	VDASD	0001	/dev/sdb
-----------	------	-----	-------	------	----------

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

Virtual Fibre Channel adapters through NPIV

IBM PowerVM™, the hypervisor of the IBM Power machine, can use the N-Port ID Virtualization (NPIV) capabilities of modern SANs and Fibre Channel HBAs to provide virtual HBAs for the LPARs. The mapping of these to the LPARs is done by the VIOS.

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

LoP distributions come with a Kernel module for the virtual HBA, which is called `ibmvfc`. After it is loaded, it 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 appear as though they were connected through a physical adapter as shown in Example 3-2.

Example 3-2 Volumes mapped through NPIV virtual HBAs

p6-570-lpar13:~ # ls SCSI					
[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 will show up more than once (once per path) and have to be managed by a DM-MP.

System z

For Linux running on an IBM System z server (zLinux), there are even more storage attachment choices and therefore potential confusion. Here is a short overview:

zLinux running natively in a System z LPAR

When you run zLinux directly on a System z LPAR there are two ways to attach disk storage:

The FICON® channel adapter cards in a System z machine can operate in *Fibre Channel Protocol* (FCP) mode. FCP is the protocol that 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 make use of this FCP capability and connect directly to FB storage devices. zLinux provides the Kernel module `zfcp` to operate the FICON adapter in FCP mode. Note that a channel card can only run either in FCP or FICON mode. Also, in FCP mode it must be dedicated to a single LPAR and cannot be shared.

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

zLinux running in a virtual machine under z/VM

Running a number of virtual Linux instances in a z/VM environment is much more common. z/VM provides granular and flexible assignment of resources to the virtual machines (VMs). It also allows you 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 running in FCP mode to a VM. A Linux instance running in this VM can operate the card using the `zfcp` driver and access the attached XIV FB volumes.

To maximize the utilization of the FCP adapters, 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 running in the VMs must make sure that each VM only uses the volumes that it is supposed to.

- ▶ FCP attachment of SCSI devices through NPIV

To overcome the issue described above, *N-Port ID Virtualization* (NPIV) was introduced for System z, z/VM, and zLinux. It allows to create 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 *World Wide Port Names* (WWPNs). To the XIV they look exactly like physical HBAs. You can create Host Connections for them and map volumes. This allows you to assign XIV volumes directly to the Linux virtual machine. No other instance can access these, even if it uses the same physical adapter card.

Tip: zLinux can also use *Count-Key-Data* devices. This is the traditional mainframe method to access disks. However, the XIV storage system doesn't support the CKD protocol, so we will not discuss it further.

3.2.2 Configure for Fibre Channel attachment

In the following sections we describe how Linux is configured to access XIV volumes. A *Host Attachment Kit (HAK)* is available for the Intel x86 platform to ease the configuration. Thus many of the manual steps we describe in the following sections are only necessary for the other supported platforms. However, the description might be helpful even if you only run Intel servers because it provides insight in the Linux storage stack. It is also useful information if you have to 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 platform. There is a unified driver for all types of QLogic FC HBAs. The name of the kernel module is `qla2xxx`. 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 Systems platform. There also is a unified driver that works with all Emulex FC HBAs. The kernel module name is `lpfc`. A supported version is also included in the enterprise Linux distributions (both for Intel x86 and Power Systems).
- ▶ Brocade: *Converged Network Adapters* (CNA) that operate as FC and Ethernet adapters, and which are relatively new to the market. They are supported on the Intel x86 platform for FC attachment to the XIV. The kernel module version 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. Note that there may be support issues with the Linux distributor because of the modifications done to the Kernel. The FC kernel module for the CNAs is called `bfa`. The driver can be downloaded from this URL:

<http://www.brocade.com/sites/dotcom/services-support/drivers-downloads/CNA/Linux.page>

- IBM FICON Express: the HBAs for the System z platform. They can either operate in FICON (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 also be removed again, as long as they are not in use as shown in Example 3-3.

Example 3-3 Load and unload a Linux Fibre Channel HBA Kernel module

```
x3650lab9:~ # modprobe qla2xxx
x3650lab9:~ # modprobe -r qla2xxx
```

Upon loading, 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`**, as shown in Example 3-4.

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

```
x3650lab9:~ #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, ....
```

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

Example 3-5 Detailed information about a specific kernel module

```
x3650lab9:~ # 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 `zfcp` driver for zLinux automatically scans and registers the attached volumes 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. See “zLinux running in a virtual machine under z/VM” on page 74. and “Add XIV volumes to a zLinux system” on page 84.

Using the FC HBA driver at installation time

You can use XIV volumes already attached to a Linux system at installation time. This 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 module(s), scan for volumes, and offer them in the installation options.

When you have an unsupported driver version included with your Linux distribution, you either have to replace it immediately after installation or (if it doesn't work at all) 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 (see “Loading the Linux Fibre Channel drivers” on page 74).

Important:

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

See 3.5, “Boot Linux from XIV volumes” on page 104 for details.

Make the FC driver available early in the boot process

If the SAN attached XIV volumes are needed early in the Linux boot process, for example if all or part of the system is located on these volumes, it is necessary to include the HBA driver into the *Initial RAM Filesystem* (initRAMFS) image. The initRAMFS is a way that allows the Linux boot process to provide certain system resources before the real system disk is set up.

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

Note: The *initRAMFS* was introduced a few years ago and replaced the *Initial RAM Disk* (initrd). Today, people often still refer to `initrd`, although they mean initRAMFS.

SUSE Linux Enterprise Server

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

Example 3-6 Tell 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 adding the HBA driver module name to the configuration file, you re-build the initRAMFS with the **mkinitrd** command. It will create and install the image file with standard settings and to standard locations as illustrated in Example 3-7.

Example 3-7 Create the initRAMFS

```
x3650lab9:~ # 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, you can use parameters for **mkinitrd** (see the man page for **mkinitrd** on your Linux system).

Redhat Enterprise Linux

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

Example 3-8 Tell RH-EL to include a kernel module in the initRAMFS

```
[root@x3650lab9 ~]# 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 adding the HBA driver module to the configuration file, you re-build the initRAMFS with the **mkinitrd** command. The Redhat version of **mkinitrd** requires parameters, the name and location of the image file to create, and the Kernel version it is built for as illustrated in Example 3-9.

Example 3-9 Create the initRAMFS

```
[root@x3650lab9 ~]# mkinitrd /boot/initrd-2.6.18-194.el5.img 2.6.18-194.el5
```

If the image file with the specified name already exists, you need the **-f** option to force **mkinitrd** to overwrite the existing one. The command will show more detailed output with the **-v** option.

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

Example 3-10 Determine the Kernel version

```
[root@x3650lab9 ~]# uname -r
2.6.18-194.el5
```

3.2.3 Determine the WWPN of the installed HBAs

To create a host port on the XIV that allows you to map volumes to a certain HBA, you need the HBA's *World Wide Port Name* (WWPN). The WWPN, along with a lot more information about the HBA 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 and then determine their WWPNs.

Example 3-11 Find the WWPNs of the FC HBAs

```
[root@x3650lab9 ~]# ls /sys/class/fc_host/
host1 host2
# cat /sys/class/fc_host/host1/port_name
0x10000000c93f2d32
# cat /sys/class/fc_host/host2/port_name
0x10000000c93d64f5
```

Note: The *sysfs* contains a lot more information. It is also used to modify the hardware configuration. We will see it more often in the next sections.

Map volumes to a Linux host, as described in 1.4, “Logical configuration for host connectivity” on page 29.

Tip: For Intel based host systems, the XIV Host Attachment Kit can create the XIV host and host port objects for you automatically from within the Linux operating system. See 3.2.4, “Attach XIV volumes to an Intel x86 host using the Host Attachment Kit” on page 78.

3.2.4 Attach 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 (HAK).

Install the HAK

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

Attention: Although it is possible to configure Linux on Intel x86 servers manually for XIV attachment, use the HAK. The HAK is required for in case you need support from IBM because it provides data collection and troubleshooting tools.

Download the latest HAK for Linux from the XIV support site:

[http://www.ibm.com/support/entry/portal/Troubleshooting/Hardware/System_Storage/Disk_systems/Enterprise_Storage_Servers/XIV_Storage_System_\(2810,_2812\)/](http://www.ibm.com/support/entry/portal/Troubleshooting/Hardware/System_Storage/Disk_systems/Enterprise_Storage_Servers/XIV_Storage_System_(2810,_2812)/)

To install the Host Attachment Kit, additional 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 of the Host Attachment Kit package will terminate, and you will be notified of the missing package. The required packages are listed in Figure 3-1.

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 3-1 Required Linux packages

To install the HAK, copy the downloaded package to your Linux server, open a terminal session, and change to the directory where the package is located. Unpack and install HAK according using the commands in Example 3-12.

Example 3-12 Install the HAK package

```
# tar -zxvf XIV_host_attach-1.5.2-sles11-x86.tar.gz
# cd XIV_host_attach-1.5.2-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, will be different depending on your HAK version, Linux distribution, and hardware platform. The installation script will prompt you for this information. After running the script you can review the installation log file `install.log` residing in the same directory.

The HAK provides the utilities you need to configure the Linux host for XIV attachment. They are located in the `/opt/xiv/host_attach` directory.

Note: You must be logged in as root or have root privileges to use the Host Attachment Kit.

The main executables and scripts reside in the directory `/opt/xiv/host_attach/bin`. The install script includes this directory in the command search path of the user root. Thus the commands can be executed from every working directory.

Configure the host for Fibre Channel using the HAK

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, provided that you have a user ID and password for an XIV storage administrator account. Example 3-13 on page 80 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 1.5.2.
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. Refer to 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.
#.
```

Configure the host for iSCSI using the HAK

You use the same command, **xiv_attach**, to configure the host for iSCSI attachment of XIV volumes. See Example 3-14 for an illustration. 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 1.5.2.
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 Check attached volumes

The HAK provides tools to verify mapped XIV volumes. Without the HAK you can use Linux native methods to do so. We will describe both methods.

Example 3-15 illustrates the HAK method for an iSCSI attached volume. The **xiv_devlist** command lists all XIV devices attached to a host.

Example 3-15 Verifying mapped XIV LUNs using the HAK tool (iSCSI example)

```

[/]# 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
...

```

Note: The **xiv_attach** command already enables and configures multipathing. Therefore the **xiv_devlist** command only shows multipath devices.

Without HAK, or if you want to see the individual devices representing each of the paths to an XIV volume, you can use the `lsccsi` command to check whether there are any XIV volumes 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 connected through four paths. Linux creates a SCSI disk device for each of the paths.

Example 3-16 List attached SCSI devices

[root@x3650lab9 ~]# lsccsi					
[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

Tip: The RH-EL installer does not install `lsccsi` 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 `lsccsi` output is the internal Linux SCSI address. It is, for historical reasons, made up like a traditional parallel SCSI address. It consists of four fields:

1. HBA ID: each HBA in the system, be it parallel SCSI, FC, or even a SCSI emulator, gets a host adapter instance when it is initiated.
2. Channel ID: this is always zero. It was formerly used as an identifier for the channel in multiplexed parallel SCSI HBAs.
3. Target ID: for parallel SCSI, this is the real target ID (the one you set using a jumper on the disk drive). For Fibre Channel, it represents a remote port that is connected to the HBA. With it, we can distinguish between multiple paths, and between multiple storage systems.
4. 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 3-2 illustrates how the SCSI addresses are generated.

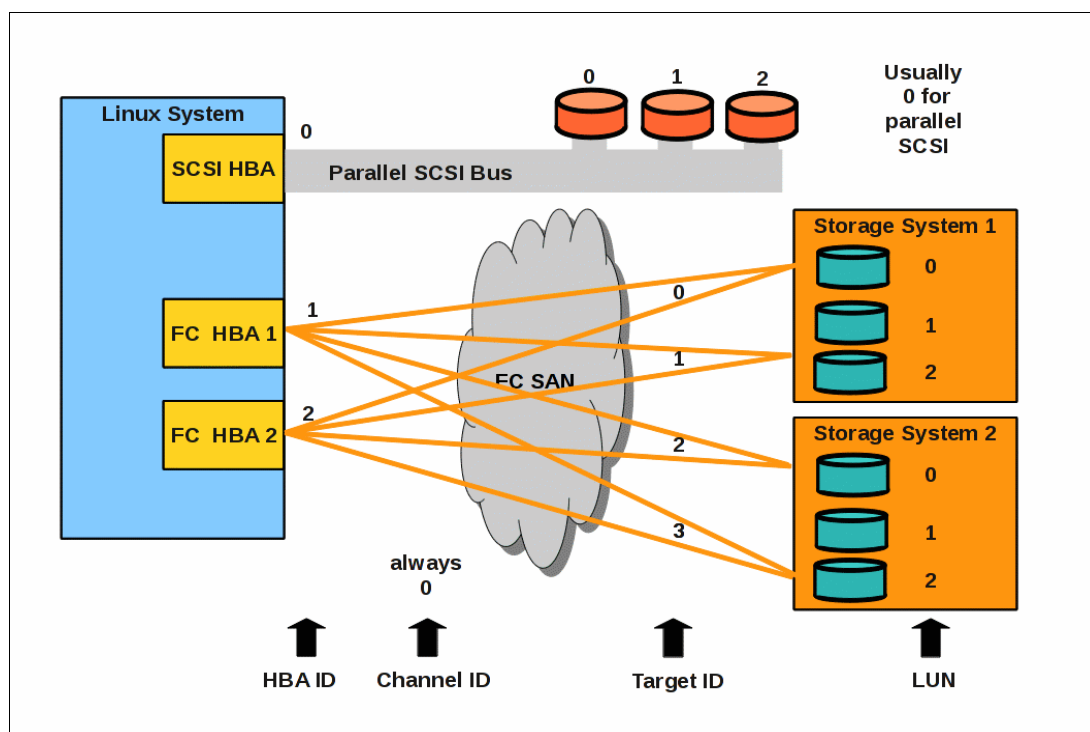


Figure 3-2 Composition of Linux internal SCSI Addresses

Identify a particular XIV Device

The udev subsystem creates device nodes for all attached devices. In the case of disk drives, it not only sets up the traditional `/dev/sdx` nodes, but also some other representatives. The most interesting 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 that is composed of parts of the *World Wide Node Name* (WWNN) of the XIV system and the XIV volume serial number in hexadecimal notation as shown in Example 3-17.

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

```
x3650lab9:~ # ls -l /dev/disk/by-id/ | cut -c 44-
...
scsi-20017380000cb051f -> ../../sde
scsi-20017380000cb0520 -> ../../sdf
scsi-20017380000cb2d57 -> ../../sdb
scsi-20017380000cb3af9 -> ../../sda
scsi-20017380000cb3af9-part1 -> ../../sda1
scsi-20017380000cb3af9-part2 -> ../../sda2
...
```

Note: The WWNN of the XIV system we use for the examples in this chapter is `0x5001738000cb0000`. It has 3 zeroes between the vendor ID and the system ID, whereas the representation in `/dev/disk/by-id` has 4 zeroes

Note: The XIV volume with the serial number **0x3af9** is partitioned (actually, it is the system disk). It contains two partitions. Partitions show up in Linux as individual block devices.

Note that 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.

In `/dev/disk/by-path` there are nodes for all paths to all XIV volumes. Here you can see the physical connection to the volumes: starting with the PCI identifier of the HBAs, through the remote port, represented by the XIV WWPN, to the LUN of the volumes as illustrated in Example 3-18.

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

```
x3650lab9:~ # 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
```

Add XIV volumes to a zLinux system

Attention: Due to hardware restraints, we had to use SLES10 SP3 for the examples shown here. The procedures, commands, and configuration files of other distributions can differ.

Only in recent Linux distributions for System z does the `zfcp` driver automatically scan for connected volumes. Here we show how you 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.

Our zLinux has two FC HBAs assigned through z/VM. In z/VM we can determine the device numbers of these adapters as can be seen 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 zLinux tool to list the FC HBAs is `lszfcp`. It shows the enabled adapters only. Adapters that are not listed correctly can be enabled using the `chccwdev` command as illustrated in Example 3-20.

Example 3-20 List and enable zLinux FCP adapters

```
lnxvm01:~ # lszfcp
0.0.0501 host0

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

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

For SLES 10, the volume configuration files reside 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

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

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

The configuration files contain a remote (XIV) port - 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

```
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 - 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 we provide here is the LUN that we find in the XIV LUN map, as shown in Figure 3-3, padded with zeroes, such that it reaches a length of eight bytes.

The screenshot shows a window titled "LUN Mapping for ITSO_zLinux". Inside, there is a tab labeled "Mapped Volumes". Below the tab is a table with two columns: "LUN" and "Volume". The table contains four rows of data:

LUN	Volume
1	ITSO_zLinux_1
2	ITSO_zLinux_2
3	ITSO_zLinux_3
4	ITSO_zLinux_4

Figure 3-3 XIV LUN map

RH-EL uses the file `/etc/zfc.conf` to configure SAN attached volumes. It contains the same kind of information in a different format, which we show in Example 3-23. The three bottom lines in the example are comments that explain the format. They do not have to be actually present in the file.

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

```
lnxvm01:~ # cat /etc/zfc.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 Set up Device Mapper Multipathing

To gain redundancy and optimize performance, you usually 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. To utilize the path redundancy and increased I/O bandwidth, and at the same time maintain data integrity, you need an additional layer in the Linux storage stack to recombine the multiple disk instances into one device.

Today Linux has its own native multipathing solution. It is based on the *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 is able to manage path failover and failback and load balancing for various storage architectures. Figure 3-4 illustrates how DM-MP is integrated into the Linux storage stack.

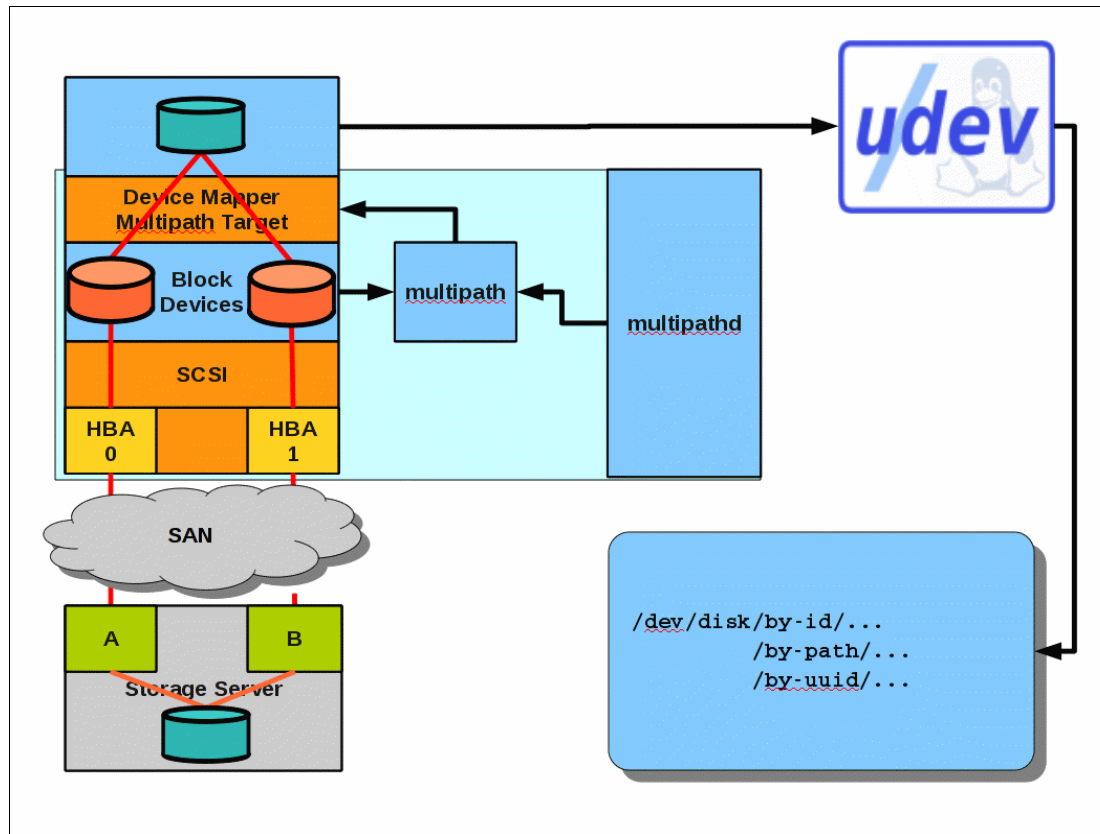


Figure 3-4 Device Mapper Multipathing in the Linux storage stack

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 representing the paths.
- ▶ The `multipath` tool scans the device (path) configuration and builds the instructions for the Device Mapper. These include the composition of the multipath devices, failover and failback patterns, and load balancing behavior. Currently there is work in progress to move the functionality of the tool to the multipath background daemon. Therefore it will disappear in the future.
- ▶ The multipath background daemon `multipathd` constantly monitors the state of the multipath devices and the paths. In case of events 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 tell `udev` what device nodes to create, so that multipath devices can be accessed and are persistent.

Configure 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

We do not go into much detail for `/etc/multipath.conf` here. The publications in 3.1.2, “Reference material” on page 68 contain all the information. In 3.2.7, “Special considerations for XIV attachment” on page 92 we will go through the settings that are recommended specifically for XIV attachment.

One option, however, that shows up several times in the next sections needs some explanation here. 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 Specify 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 will not be used again for a new one. If it is re-attached, it will get 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 below.

Enable multipathing for SLES 11

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

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

Example 3-25 Start 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
```

In order to have DM-MP start automatically at each system start, you must add these start scripts to the SLES 11 system start process. Refer to Example 3-26.

Example 3-26 Configure automatic start of DM-MP in SLES 11

```
x3650lab9:~ # insserv boot.multipath
x3650lab9:~ # insserv multipathd
```

Enable 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 and they will be ignored the next time DM-MP scans for devices. Refer to Example 3-27.

Example 3-27 Disable 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 Start DM-MP in RH-EL 5

```
[root@x3650lab9 ~]# /etc/init.d/multipathd start
Starting multipathd daemon: [ OK ]
```

In order to have DM-MP start automatically at each system start, you must add the following start script in Example 3-29 to the RH-EL 5 system start process.

Example 3-29 Configure automatic start of DM-MP in RH-EL 5

```
[root@x3650lab9 ~]# chkconfig --add multipathd
[root@x3650lab9 ~]# chkconfig --levels 35 multipathd on
[root@x3650lab9 ~]# chkconfig --list multipathd
multipathd      0:off  1:off  2:off  3:on   4:off  5:on   6:off
```

Check and change 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, a variety of options are available. Use the `help` command to get a list. We show some of more important ones in the following examples and in 3.3, “Non-disruptive SCSI reconfiguration” on page 94.

The `show topology` command illustrated in Example 3-30 prints out a detailed view of the current DM-MP configuration, including the state of all available paths.

Example 3-30 Show multipath topology

```
x3650lab9:~ # multipathd -k"show top"
20017380000cb0520 dm-4 IBM,2810XIV
[size=16G][features=0][hwhandler=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][hwhandler=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 located in separate path groups. Thus, there is no load balancing between the paths. DM-MP must be configured with a XIV-specific multipath.conf file to enable load balancing (see 3.2.7, “Special considerations for XIV attachment” on page 108 and “Multipathing” on page 87). The HAK does this 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 scanning the paths and configuration files. Use it to add new multipath devices after adding new XIV volumes. See section 3.3.1, “Add and remove XIV volumes dynamically” on page 94 for more information.

Example 3-31 Reconfigure DM-MP

```

multipathd> reconfigure
ok

```

Attention: The `multipathd -k` command prompt of SLES11 SP1 supports the `quit` and `exit` commands to terminate. That of RH-EL 5U5 is a little older and must still be terminated using the Ctrl + d key combination.

Tip: You can also issue commands in a “one-shot-mode” by enclosing them in double quotes and typing them directly, without space, behind the `multipath -k`.

An example would be `multipathd -k"show paths"`

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 will be removed from DM-MP and all further development and improvements go into `multipathd`.

Access DM-MP devices in SLES 11

The device nodes 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
...
```

Attention: The Device Mapper itself creates its default device nodes in the /dev directory. They are called /dev/dm-0, /dev/dm-1, and so forth. These nodes are not persistent. They can change with configuration changes and should not be used for device access.

SLES 11 creates an additional set of device nodes for multipath devices. It overlays the former single path device nodes in /dev/disk/by-id. This means that any device mapping (for example mounting a file system) you did with one of these nodes works exactly the same as before starting DM-MP. It will just use 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-20017380000cb051f -> ../../dm-5
scsi-20017380000cb0520 -> ../../dm-4
scsi-20017380000cb2d57 -> ../../dm-0
scsi-20017380000cb3af9 -> ../../dm-1
...
```

If you set the `user_friendly_names` option in `/etc/multipath.conf`, SLES 11 will create 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 volumes' unique IDs in their names.

Access DM-MP devices in RH-EL 5

RH-EL sets the `user_friendly_names` option in its default `/etc/multipath.conf` file. The devices 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
```

There also is a second set of device nodes containing the unique IDs of the volumes in their name, regardless of whether user friendly names are specified or not. You find them in the directory `/dev/mpath`. Refer to Example 3-35.

Example 3-35 RH-EL 5 device nodes in /dev/mpath

```
[root@x36501ab9 ~]# ls -l /dev/mpath/ | cut -c 39-

20017380000cb051f -> ../../dm-5
20017380000cb0520 -> ../../dm-4
20017380000cb2d57 -> ../../dm-0
```

```
20017380000cb3af9 -> ../../dm-1
```

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 using the **fdisk** command or any other partitioning tool. To make new partitions on DM-MP devices available you can use the **partprobe** command. It triggers udev to set up new block device nodes for the partitions, as illustrated in Example 3-36.

Example 3-36 Use the partprobe command to register newly created partitions

```
x36501ab9:~ # fdisk /dev/mapper/20017380000cb051f
...
<all steps to create a partition and write the new partition table>
...
x36501ab9:~ # ls -l /dev/mapper/ | cut -c 48-

20017380000cb051f
20017380000cb0520
20017380000cb2d57
20017380000cb3af9
...
x36501ab9:~ # partprobe
x36501ab9:~ # ls -l /dev/mapper/ | cut -c 48-

20017380000cb051f
20017380000cb051f-part1
20017380000cb0520
20017380000cb2d57
20017380000cb3af9
...
```

Example 3-36 was created with SLES 11. The method works as well for RH-EL 5 but the partition names might be different.

Note: 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 has special considerations that specifically apply to XIV.

Configure multipathing

You have to create an XIV-specific `multipath.conf` file to optimize the DM-MP operation for XIV. Here we provide the contents of this file as it is created by the HAK. The settings that are relevant for XIV are shown in Example 3-37 on page 93.

Example 3-37 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 32
        path_checker tur
        failback 15
        no_path_retry 5
        polling_interval 3
    }
}
```

We discussed the `user_friendly_names` parameter already in 3.2.6, “Set up Device Mapper Multipathing” on page 86. You can add it to file or leave it out, as you like. The values for `failback`, `no_path_retry`, `path_checker`, and `polling_interval` control the behavior of DM-MP in case of path failures. Normally they should not be changed. If your situation requires a modification of these parameters, refer to the publications in 3.1.2, “Reference material” on page 68. The `rr_min_io` setting specifies the number of IO requests that are sent to one path before switching to the next one. The value of 32 shows good load balancing results in most cases. However, you can adjust it as necessary.

System z specific multipathing settings

Testing of zLinux with multipathing has shown that the `dev_loss_tmo` parameter should be set to 90 seconds, and the `fast_io_fail_tmo` parameter to 5 seconds. Modify the `/etc/multipath.conf` file and add the settings shown in Example 3-38.

Example 3-38 System z specific multipathing settings

```
...
defaults {
    ...
    dev_loss_tmo      90
    fast_io_fail_tmo  5
    ...
}
...
```

You can make the changes effective by using the **reconfigure** command in the interactive **multipathd -k** prompt.

Disable QLogic failover

The QLogic HBA kernel modules have limited built in multipathing capabilities. Because multipathing is managed by DM-MP, you must make sure that the Qlogic failover support is disabled. Use the **modinfo q1a2xxx** command shown in Example 3-39 to check.

Example 3-39 Check for enabled QLogic failover

```
x3650lab9:~ # modinfo q1a2xxx | grep version
version:      8.03.01.04.05-05-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 3-40.

Example 3-40 Disable QLogic failover

```
x36501ab9:~ # cat /etc/modprobe.conf
...
options qla2xxx ql2xfailover=0
...
```

After modifying this file you must run the **depmod -a** command to refresh the Kernel driver dependencies. Then reload the `qla2xxx` module to make the change effective. If you have included the `qla2xxx` module in the `InitRAMFS`, you must create a new one.

3.3 Non-disruptive SCSI reconfiguration

This section reviews actions that can be taken on the attached host in a nondisruptive manner.

3.3.1 Add and remove 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 3-41, you use a command interface provided by **sysfs**.

Example 3-41 Scan 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 you discover which SCSI instances your FC HBAs have, then you 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 given value.

Note: If you have the KAK 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 it is necessary to force the FC HBA perform a port login to recognize newly added devices. This can be done with the following command that 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 you must follow a certain sequence to avoid system hangs due to incomplete I/O requests:

1. Stop all applications that use the device and make sure 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.

The last step is illustrated in Example 3-42.

Example 3-42 Remove both paths to a disk device

```
x36501ab9:~ # echo 1 > /sys/class/scsi_disk/0\0:0\0:3/device/delete
x36501ab9:~ # echo 1 > /sys/class/scsi_disk/1\0:0\0:3/device/delete
```

The device paths (or disk devices) are represented by their Linux SCSI address (see “Linux SCSI addressing explained” on page 82). Run the **multipathd -k"show topology"** command after removing 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 detaching the device on the storage system level.

Tip: 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 Add and remove XIV volumes in zLinux

The mechanisms to scan and attach new volumes, as shown in 3.3.1, “Add and remove XIV volumes dynamically” on page 94, do not work the same in zLinux. There are commands available that discover and show the devices connected to the FC HBAs, but they do not do the logical attachment to the operating system automatically. In SLES10 SP3 we use the **zfcp_san_disc** command for discovery.

Example 3-43 shows how to discover and list the connected volumes, in this case one remote port or path, with the **zfcp_san_disc** command. You must run this command for all available remote ports.

Example 3-43 List LUNs connected through a specific remote port

```
lnxvm01:~ # zfcp_san_disc -L -p 0x5001738000cb0191 -b 0.0.0501
0x0001000000000000
0x0002000000000000
0x0003000000000000
0x0004000000000000
```

Tip: In more recent distributions **zfcp_san_disc** is not available anymore. Remote ports are automatically discovered. The attached volumes can be listed using the **ls1uns** script.

After discovering the connected volumes, do the logical attachment 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 3-44 shows such a **sysfs** structure for a specific XIV port.

Example 3-44 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
```

As shown in Example 3-45, add LUN 0x0003000000000000 to both available paths using the **unit_add** metafile.

Example 3-45 Add 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
```

Attention: You must perform discovery using **zfcplib** whenever new devices, remote ports, or volumes are attached. Otherwise the system will not recognize them even if you do the logical configuration.

New disk devices that you attached this way automatically get device nodes and are added to DM-MP.

If you want to remove a volume from zLinux you must follow the same sequence as for the other platforms to avoid system hangs due to incomplete I/O requests:

1. Stop all applications that use the device and make sure 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, using a method similar to the 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, you must use the methods shown in “Add XIV volumes to a zLinux system” on page 84 to create the configuration files to be used at the next system start.

3.3.3 Add new XIV host ports to zLinux

If you connect new XIV ports or a new XIV system to the zLinux system, you must logically attach the new remote ports. Example 3-46 on page 97 discovers and shows the XIV ports that are connected to our HBAs.

Example 3-46 Show connected remote ports

```
lnxvm01:~ # zfcplib -W -b 0.0.0501
0x5001738000cb0191
0x5001738000cb0170
lnxvm01:~ # zfcplib -W -b 0.0.0601
0x5001738000cb0160
0x5001738000cb0181
```

In the next step we attach the new XIV ports logically to the HBAs. As Example 3-47 shows, there is already a remote port attached to HBA 0.0.0501. It is the one path we already have available to access the XIV volume. We add the second connected XIV port to the HBA.

Example 3-47 List attached remote ports, attach 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
```

In Example 3-48, we add the second new port to the other HBA in the same way.

Example 3-48 Attach 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 Resize XIV volumes dynamically

At the time of writing this publication, only SLES11 SP1 is capable of utilizing the additional capacity of dynamically enlarged XIV volumes. Reducing the size is not supported. Here we briefly describe the sequence.

First, we create an ext3 file system on one of the XIV multipath devices and mount it. The **df** command in Example 3-49 shows the available capacity.

Example 3-49 Check the size and available space on a mounted file system

```
x3650lab9:~ # df -h /mnt/itso_0520/
Filesystem      Size  Used Avail Use% Mounted on
/dev/mapper/2001738000cb0520
                  16G  173M   15G   2% /mnt/itso_0520
```

Now we 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 we initiate a rescan of each SCSI disk device (path) through **sysfs** as shown in Example 3-50.

Example 3-50 Rescan all disk devices (paths) of a XIV volume

```
x3650lab9:~ # echo 1 > /sys/class/scsi_disk/0\:0\:0\:4/device/rescan
```

```
x3650lab9:~ # echo 1 > /sys/class/scsi_disk/1\:0\:0\:4/device/rescan
```

The message log shown in Example 3-51 indicates the change in capacity.

Example 3-51 Linux message log indicating the capacity change of a SCSI device

```
x3650lab9:~ # tail /var/log/messages
...
Oct 13 16:52:25 lnvm01 kernel: [ 9927.105262] sd 0:0:0:4: [sdh] 100663296
512-byte logical blocks: (51.54 GB/48 GiB)
Oct 13 16:52:25 lnvm01 kernel: [ 9927.105902] sdh: detected capacity change from
17179869184 to 51539607552
...
```

In the next step shown in Example 3-52, we indicate the device change to DM-MP using the **resize_map** command of **multipathd**. Afterwards we can see the updated capacity in the output of **show topology**.

Example 3-52 Resize a multipath device

```
x3650lab9:~ # multipathd -k"resize map 20017380000cb0520"
ok
x3650lab9:~ # multipathd -k"show top map 20017380000cb0520"
20017380000cb0520 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]
```

Finally we resize the file system and check the new capacity as shown in Example 3-53.

Example 3-53 Resize file system and check capacity

```
x3650lab9:~ # resize2fs /dev/mapper/20017380000cb0520
resize2fs 1.41.9 (22-Aug-2009)
Filesystem at /dev/mapper/20017380000cb0520 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/20017380000cb0520 to 12582912 (4k)
blocks.
The filesystem on /dev/mapper/20017380000cb0520 is now 12582912 blocks long.

x3650lab9:~ # df -h /mnt/itso_0520/
Filesystem              Size  Used Avail Use% Mounted on
/dev/mapper/20017380000cb0520
                        48G  181M   46G   1% /mnt/itso_0520
```

Restriction: At the time of writing this publication there are several restrictions to the dynamic volume increase process:

- ▶ 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 still popular ext2 file system cannot.

3.3.5 Use snapshots and remote replication targets

The XIV snapshot and remote replication solutions create byte-wise 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, like file system identifier or LVM signature, however, is identical to that of the source. This can lead to confusion and data integrity problems if you plan to use the target on the same Linux system as the source.

In this section we describe some ways to avoid integrity issues. We also highlight some potential traps that could lead to problems.

File system directly residing on a XIV volume

The copy of a file systems that was created directly on a SCSI disk device (single path) or a DM-MP device, without additional virtualization layer, such as RAID or LVM, can be used on the same host as the source without modification. If you follow the sequence outlined below carefully and avoid the highlighted traps, this will work without problems. We describe the procedure using the example of 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 3-54 using a device node that is bound to the volume's unique identifier and not to any metadata that is stored on the device itself.

Example 3-54 Mount the source volume

```
x3650lab9:~ # mount /dev/mapper/20017380000cb0520 /mnt/itso_0520/
x3650lab9:~ # mount
...
/dev/mapper/20017380000cb0520 on /mnt/itso_0520 type ext3 (rw)
```

2. Make sure the data on the source volume is consistent, for example 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 our 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 (see 3.3.1, “Add and remove XIV volumes dynamically” on page 94 for details). DM-MP will automatically integrate the snapshot target. Refer to Example 3-55.

Example 3-55 Check DM-MP topology for target volume

```
x3650lab9:~ # multipathd -k"show top"
20017380000cb0520 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]
...
20017380000cb1f93 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. As shown in Example 3-56, mount the target volume to a separate mount point using a device node that is created from the unique identifier of the volume.

Example 3-56 Mount the target volume

```
x36501ab9:~ # mount /dev/mapper/20017380000cb1f93 /mnt/itso_fc/
x36501ab9:~ # mount
...
/dev/mapper/20017380000cb0520 on /mnt/itso_0520 type ext3 (rw)
/dev/mapper/20017380000cb1f93 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.

Attention: udev also creates device nodes that relate to the file system 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 residing 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, this metadata has to be modified and made unique to ensure data integrity. Otherwise LVM could mix volumes from the source and the target sets.

A script called `vgimportclone.sh` is publicly available that automates the modification of the metadata and thus supports the import of LVM volume groups that reside on a set of replicated disk devices. It can be downloaded from the following URL:

<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 this URL:

<http://www.cl.cam.ac.uk/cgi-bin/manpage?8+vgimportclone>

Tip: The `vgimportclone` script is part of the standard LVM tools for RH-EL 5. The SLES 11 distribution does not contain the script by default.

Perform the following steps in the correct order to ensure consistent data on the target volumes and avoid mixing up the source and target. We describe the sequence with the help of a small example. We have a volume group containing a logical volume that is striped over two XIV volumes. We use snapshots to create a point in time copy of both volumes. Then we make both the original logical volume and the cloned one 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 using the LVM logical volume device, as shown in Example 3-57.

Example 3-57 Mount 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 the data on the source volume is consistent, for example 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 (see 3.3.1, “Add and remove XIV volumes dynamically” on page 94 for details). DM-MP will automatically integrate the snapshot targets. Refer to Example 3-58.

Example 3-58 Check DM-MP topology for target volume

```
x3650lab9:~ # multipathd -k"show topology"
...
20017380000cb1fe4 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]
20017380000cb1fe5 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]
```

Note: To avoid data integrity issues, it is important that no LVM configuration commands are issued until step 5 is complete.

5. As illustrated in Example 3-59, run the `vgimportclone.sh` script against the target volumes, providing a new volume group name.

Example 3-59 Adjust the target volumes' LVM metadata

```
x3650lab9:~ # ./vgimportclone.sh -n vg_itso_snap /dev/mapper/20017380000cb1fe4
/dev/mapper/20017380000cb1fe5
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 3-60.

Example 3-60 Activate volume group on target device and mount 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

In this section, we discuss topics related to troubleshooting and monitoring.

Linux Host Attachment Kit utilities

The Host Attachment Kit (HAK) now includes the following utilities:

► **xiv_devlist**

xiv_devlist is the command allowing validation of the attachment configuration. This command generates a list of multipathed devices available to the operating system. In Example 3-61 you can see the options of the **xiv_devlist** command.

Example 3-61 Options of 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)
  -o FIELDS, --options=FIELDS
                           Fields to display; Comma-separated, no spaces. Use -l
                           to see the list of fields
  -H, --hex                 Display XIV volume and machine IDs in hexadecimal base
  -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
```

► **xiv_diag**

The 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. To run, go to a command prompt and enter **xiv_diag**. See Example 3-62.

Example 3-62 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.

```

Alternative ways to check SCSI devices

The Linux kernel maintains a list of all attached SCSI devices in the `/proc` pseudo filesystem as illustrated in Example 3-63. `/proc/scsi/scsi` 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 3-63 An alternative 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 shown in Example 3-64 can be used to list all block devices, including their partition information and capacity, but without SCSI address, vendor and model information as shown in Example 3-64.

Example 3-64 Output of fdisk -l

```

x3650lab9:~ # 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

...

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, but is not necessarily installed by default. The **iostat** command reads data provided by the kernel in `/proc/stats` and prints it in human readable format. See the man page of **iostat** for more details.

Generic SCSI tools

For Linux there is a set of tools that allow low-level access to SCSI devices. They are called the *sg_tools*. They communicate with SCSI devices through the generic SCSI layer, which 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`, `/dev/sdb`, and any other device node that represents a SCSI device directly.

Useful *sg_tools* include the following:

- ▶ **sg_inq /dev/sgx** prints SCSI Inquiry data, such as the volume serial number.
- ▶ **sg_scan** prints the scsi host, channel, target, 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. It also allows you to manipulate the mode pages.

3.5 Boot Linux from XIV volumes

In this section we describe how you can configure a system to load the Linux kernel and operating system from a SAN-attached XIV volume. To do so, we use an example that is based on SLES11 SP1 on an x86 server with QLogic FC HBAs. We note and describe where other distributions and hardware platforms have deviations from our example. We do not explain here how to configure the HBA BIOS to boot from SAN attached XIV volume (see 1.2.5, “Boot from SAN on x86/x64 based architecture” on page 16).

3.5.1 The Linux boot process

In order to understand the configuration steps required to boot a Linux system from SAN-attached XIV volumes, you need a basic understanding of the Linux boot process. Therefore we briefly summarize the 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 (for example the BIOS of x86 servers). When a system is turned on, it first performs 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) that uses those basic I/O routines to read a specific location on the defined system disk and starts executing the code it contains. This code either is part of the boot loader of the operating system or it branches to the location where the boot loader resides.

If we want to boot from a SAN attached disk, we must make sure that the OS loader can access this 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)*.

Note: For zLinux under z/VM, the OS loader is not part of the firmware but rather the z/VM program `ipl`.

2. The boot loader

The boot loader's purpose is to start the operating system kernel. To do this, it must know the physical location of the kernel image on the system disk, read it in, unpack it if it is compressed, and start it. All of this is still done using the basic I/O routines 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 Boot Loader*) for x86 systems
- `zipl` for System z
- `yaboot` for Power Systems

3. The kernel and the InitRAMFS

After the kernel is unpacked and running, it takes control over the system hardware. It starts and sets up memory management, interrupt handling, and the built in hardware drivers for the hardware that is common on all systems (MMU, clock, and so on).

It reads and unpacks the InitRAMFS image, again using the same basic I/O routines. The InitRAMFS contains additional 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 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: networking, services, user interfaces, etc. At this point the hardware is already completely abstracted. Therefore `init()` is not platform dependent, nor are there any SAN-boot specifics.

A detailed description of the Linux boot process for x86 based systems can be found on *IBM Developerworks* at:

<http://www.ibm.com/developerworks/linux/library/l-linuxboot/>

3.5.2 Configure the QLogic BIOS to boot from an XIV volume

The first step is to configure the HBA to load a BIOS extension that provides the basic input-output capabilities for a SAN attached disk. Refer to 1.2.5, “Boot from SAN on x86/x64 based architecture” on page 16 for details.

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 detailed instructions you can refer to 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 x86 the specific way to start loading an operating system. In this section we briefly describe how this is done on the other platforms we support and what you have to consider.

IBM Power Systems

When you install Linux on a IBM Power System server or LPAR, the Linux installer sets the boot device in the firmware to the drive which you're installing on. There are no special precautions to take 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 IPLed from traditional CKD disk devices or from Fibre-Channel-attached Fixed-Block (SCSI) devices. To IPL from SCSI disks, the SCSI IPL feature (FC 9904) must be installed and activated on the System z server. SCSI IPL is generally available on recent System z machines (z10™ and later)

Attention: 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. zLinux running natively in a System z LPAR

After installing zLinux you have to provide the device from which the LPAR runs the *Initial Program Load (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. zLinux running under z/VM

Within z/VM we start an operating system with the **IPL** command. With the command we provide the z/VM device address of the device where the Linux boot loader and Kernel is installed.

When booting from SCSI disk, we do not have a z/VM device address for the disk itself (see 3.2.1, "Platform specific remarks" on page 72, section "System z" on page 73. We must provide the information on 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 3-65 on page 107.

Example 3-65 Set and query 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 we provide 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 Install 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 additional considerations you have to take are:

- ▶ Identify the right XIV volume(s) to install on.
- ▶ Enable multipathing during installation.

Note: After the SLES11 installation program (YAST) is running, the installation is mostly hardware platform independent. It works the same when running on an x86, IBM Power System or System z server.

1. Start the installation process, for example by booting from an installation DVD, and follow the installation configuration windows as usual, until you come to the “Installation Settings” window as shown in Figure 3-5.

Note: The zLinux installer does not automatically list the available disks for installation. You will see a **Configure Disks** panel before you get to the **Installation Settings**, where you can discover and attach the disks that are needed to install the system using a graphical user interface. At least one disk device is required to perform the installation.

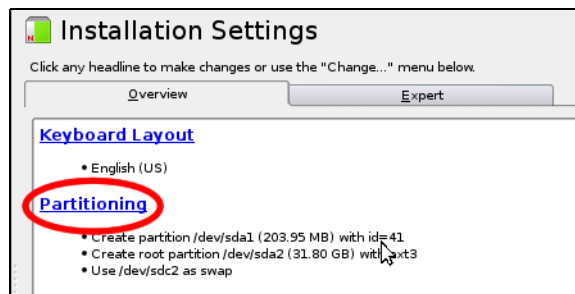


Figure 3-5 SLES11 SP1 Installation Settings

2. Click **Partitioning** to perform the configuration steps required to define the XIV volume as the system disk. This takes you to the “Preparing Hard Disk: Step 1” window, as shown in Figure 3-6 on page 108. Here you make sure that the **Custom Partitioning (for experts)** button is selected and click **Next**. It does not matter which disk device is selected in the **Hard Disk** field.

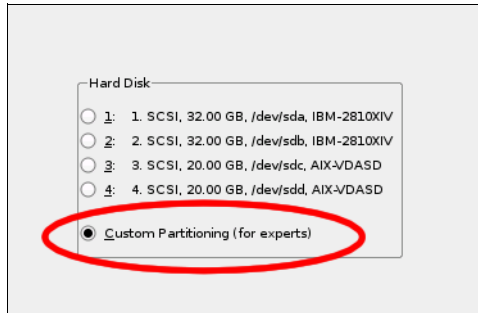


Figure 3-6 "Preparing Hard Disk: Step 1" window

3. Enable multipathing in the Expert Partitioner window. After selecting **Hard disks** in the navigation section on the left side, the tool offers the **Configure** button in the bottom right corner of the main panel. Click it and select **Configure Multipath**. The procedure is illustrated in Figure 3-7.

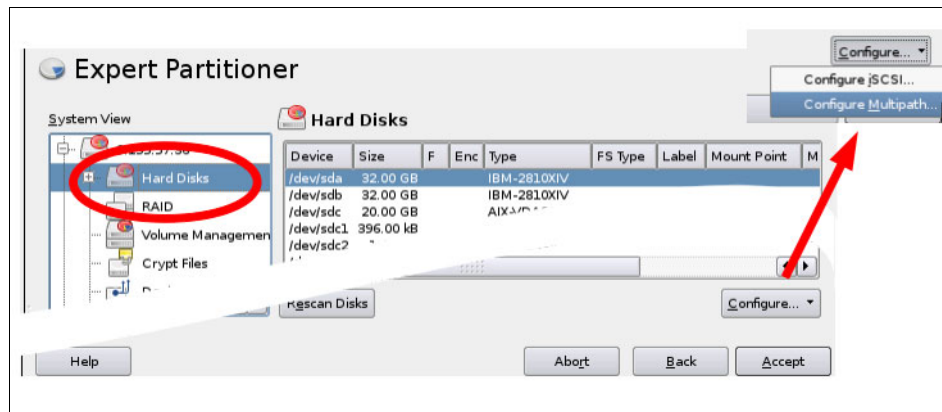


Figure 3-7 Enable multipathing in partitioner

4. The tool asks for confirmation and then rescans the disk devices. When finished, it presents an updated list of hard disks that also shows the multipath devices it has found, as you can see in Figure 3-8.

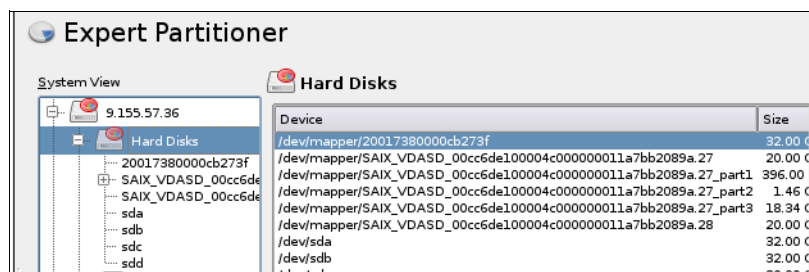


Figure 3-8 Select multipath device for installation

Select the multipath device (XIV volume) you want to install to and click the **Accept** button. The next window you see is the partitioner. From here on you create and configure the required partitions for your system the same way you would do on a local disk.

5. You can also use the automatic partitioning capabilities of YAST after the multipath devices have been detected. Just click the **Back** button until you see the initial partitioning window again. It now shows the multipath devices instead of the disks, as illustrated in Figure 3-9 on page 109.

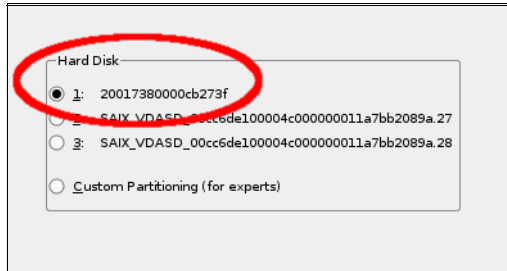


Figure 3-9 “Preparing Hard Disk: Step 1” window with multipath devices

6. Select the multipath device you want to install on, click **Next**, and select the partitioning scheme you want.

Important: All supported platforms can boot Linux from multipath devices. In some cases, however, the tools that install the boot loader only can write to simple disk devices. In these cases you must install the boot loader with multipathing deactivated. SLES10 and SLES11 allow this by adding the parameter `multipath=off` to the boot command in the boot loader. The boot loader for IBM Power Systems and System z must be re-installed 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. Please see the Linux distribution specific documentation in 3.1.2, “Reference material” on page 68 for more details.

The installer doesn’t implement any device specific settings, such as creating the `/etc/multipath.conf` file. You must do this manually after the installation as explained in 3.2.7, “Special considerations for XIV attachment” on page 92. Because DM-MP is already started during the processing of the `InitRAMFS`, you also have to build a new `InitRAMFS` image after changing the DM-MP configuration (see “Make the FC driver available early in the boot process” on page 76).

Tip: 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 also support multipathing already. Enable 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



AIX host connectivity

This chapter explains specific considerations and describes the host attachment-related tasks for the AIX operating system platform

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

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

as well as the Host Attachment publications at:

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

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. The Fibre Channel connectivity is discussed first, 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 the XIV Infocenter at:

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

The XIV Storage System supports different versions of the AIX operating system, either via Fibre Channel (FC) or iSCSI connectivity.

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

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

General notes for all AIX releases:

- ▶ XIV Host Attachment Kit 1.5.2 for AIX supports all AIX releases (except for AIX 5.2 and earlier)
- ▶ Dynamic LUN expansion with LVM requires XIV firmware version 10.2 or later

Prerequisites

If the current AIX operating system level installed on your system is not a level that is compatible with XIV, you must upgrade prior to attaching the XIV storage. To determine the maintenance package or technology level currently installed on your system, use the `oslevel` command as shown in Example 4-1.

Example 4-1 AIX: Determine current AIX version and maintenance level

```
# oslevel -s  
6100-05-01-1016
```

In our example, the system is running AIX 6.1.0.0 technology level 5 (61TL5). Use this information in conjunction with the SSIC to ensure that the attachment will be an IBM-supported configuration.

In the event that AIX maintenance items are needed, consult the IBM Fix Central website to 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.

In addition to proper cabling, if using FC switched connections, you must ensure that you have a correct zoning (using the WWPN numbers of the AIX host).

4.1.1 AIX host FC configuration

Attaching the XIV Storage System to an AIX host using Fibre Channel involves the following activities from the host side:

- ▶ Identify the Fibre Channel host bus adapters (HBAs) and determine their WWPN values.
- ▶ Install the XIV-specific AIX Host Attachment Kit.
- ▶ Configure multipathing.

Identifying FC adapters and attributes

In order to allocate XIV volumes to an AIX host, the first step is to 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 AIX: Listing FC adapters

```
# lsdev -Cc adapter | grep fcs
fcs0    Available 01-08 FC Adapter
fcs1    Available 02-08 FC Adapter
```

This example shows that, in our example, we have two FC ports.

Another useful command that is shown in Example 4-3 returns not just the ports, but also where the Fibre Channel adapters reside in the system (in which PCI slot). This command can be used to physically identify in what slot a specific adapter is placed.

Example 4-3 AIX: Locating FC adapters

```
# lsslot -c pci | grep fcs
U787B.001.DNW28B7-P1-C3  PCI-X capable, 64 bit, 133MHz slot  fcs0
U787B.001.DNW28B7-P1-C4  PCI-X capable, 64 bit, 133MHz slot  fcs1
# lsdev -Cc adapter | grep fcs
fcs0    Available 01-08 FC Adapter
fcs1    Available 02-08 FC Adapter
```

To obtain the Worldwide Port Name (WWPN) of each of the POWER system FC adapters, you can use the **lscfg** commands as shown in Example 4-4.

Example 4-4 AIX: Finding Fibre Channel adapter WWN

```
# lscfg -v1 fcs0
fcs0          U787B.001.DNW28B7-P1-C3-T1  FC Adapter
  Part Number.....80P4543
  EC Level.....A
  Serial Number.....1D5450889E
  Manufacturer.....001D
  Customer Card ID Number.....280B
  FRU Number.....      80P4544
  Device Specific.(ZM).....3
  Network Address.....10000000C94F9DF1
  ROS Level and ID.....02881955
  Device Specific.(Z0).....1001206D
  Device Specific.(Z1).....00000000
  Device Specific.(Z2).....00000000
  Device Specific.(Z3).....03000909
  Device Specific.(Z4).....FF801413
  Device Specific.(Z5).....02881955
```

```

Device Specific.(Z6).....06831955
Device Specific.(Z7).....07831955
Device Specific.(Z8).....20000000C94F9DF1
Device Specific.(Z9).....TS1.91A5
Device Specific.(ZA).....T1D1.91A5
Device Specific.(ZB).....T2D1.91A5
Device Specific.(ZC).....00000000
Hardware Location Code.....U787B.001.DNW28B7-P1-C3-T1

```

You can also print the WWPN of an HBA directly by issuing this command:

```
lscfg -vl <fcs#> | grep Network
```

where *<fcs#>* stands for an instance of a FC HBA to query.

At this point, you can define the AIX host system on the XIV Storage System and assign the WWPN to the host. If the FC connection was correctly done, the zoning enabled, and the fibre channel adapters are in an available state on the host, these ports will be selectable from the drop-down list as shown in Figure 4-1. After creating the AIX host, map the XIV volumes to the host.

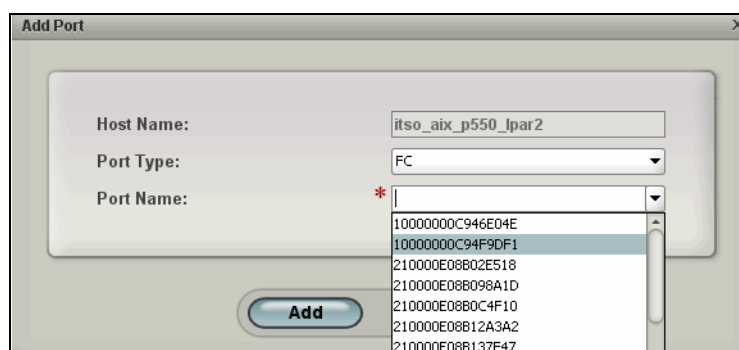


Figure 4-1 Selecting ports from the drop-down list in the XIV GUI

Tip: If the WWPNs are not displayed in the drop-down list box, run the **cfgmgr** command on the AIX host to activate the HBAs. If you still do not see the WWPNs, remove the **fcsX** with the command **rmdev -Rdl fcsX**, then run **cfgmgr** again.

With older AIX releases, you may see a warning when you run the **cfgmgr** or **xiv_fc-admin -R** command as shown in Example 4-5. This warning can be ignored. You can also find a fix at the following URL:

<http://www-01.ibm.com/support/docview.wss?uid=isgl1Z75967>

Example 4-5 *cfgmgr* warning message

```

# cfgmgr
cfgmgr: 0514-621 WARNING: The following device packages are required for
device support but are not currently installed.
devices.fcp.array

```

Installing the XIV Host Attachment Kit for AIX

For AIX to correctly recognize the disks mapped from the XIV Storage System as *MPIO 2810 XIV Disk*, the *XIV Host Attachment Kit for AIX* (XIV HAK) is required on the AIX system. This package will also enable multipathing. At the time of writing this material, the XIV HAK 1.5.2 was used. The files can be downloaded from:

http://www.ibm.com/support/search.wss?q=s5g1*&tc=STJTAG+HW3E0&rs=1319&dc=D400&dtm

Important: Although AIX now natively supports XIV via ODM changes that have been back-ported to several older AIX releases, it is still important to install the XIV HAK for support and for access to the latest XIV utilities like `xiv_diag`. The output of these `xiv` utilities is mandatory for IBM support when opening an XIV-related service call on an AIX platform.

To install the HAK, follow these steps:

1. Download or copy the downloaded HAK to your AIX system.
2. From the AIX prompt, change to the directory where your XIV package is located and execute the `gunzip -c XIV_host_attachment-1.5-*.tar.gz | tar xvf -` command to extract the file.
3. Switch to the newly created directory and run the install script as shown in Example 4-6.

Example 4-6 AIX XIV HAK installation

```
# ./install.sh
Welcome to the XIV Host Attachment Kit installer.

NOTE: This installation defaults to round robin multipathing,
if you would like to work in fail-over mode, please set the environment
variables before running this installation.

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.
```

When the installation has completed, listing the disks should display the correct number of disks seen from the XIV storage. They are labeled as XIV disks, as illustrated in Example 4-7.

Example 4-7 AIX: XIV labeled FC disks

```
# lsdev -Cc disk
hdisk0 Available          Virtual SCSI Disk Drive
hdisk1 Available 01-08-02 MPIO 2810 XIV Disk
hdisk2 Available 01-08-02 MPIO 2810 XIV Disk
```

The Host Attachment Kit 1.5.2 provides an interactive command line utility to configure and connect the host to the XIV storage system. The command `xiv_attach` starts a wizard, that does attach the host to the XIV. Example 4-8 on page 116 shows part of the `xiv_attach` command output.

Example 4-8 xiv_attach command output

```
# xiv_attach
-----
Welcome to the XIV Host Attachment wizard, version 1.5.2.
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 : f
-----
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:
10:00:00:00:C9:4F:9D:F1: fcs0: [IBM]: N/A
10:00:00:00:C9:4F:9D:6A: fcs1: [IBM]: N/A
Press [ENTER] to proceed.
Would you like to rescan for new storage devices now? [default: yes ]:
.
```

AIX Multi-path I/O (MPIO)

AIX 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 basic information about MPIO and the management of MPIO devices, refer to the online guide *AIX 5L System Management Concepts: Operating System and Devices* from the AIX documentation website at:

<http://publib.boulder.ibm.com/infocenter/pseries/v5r3/index.jsp>

Configuring XIV devices as MPIO or non-MPIO devices

Configuring XIV devices as MPIO provides the optimum solution. In some cases, you could be using a third party multipathing solution for managing other storage devices and want to manage the XIV 2810 device with the same solution. This usually requires the XIV devices to be configured as non-MPIO devices.

AIX provides a command to migrate a device between MPIO and non-MPIO. The **manage_disk_drivers** command can be used to change how the XIV device is configured (MPIO or non-MPIO). The command causes all XIV disks to be converted. It is not possible to convert one XIV disk to MPIO and another XIV disk to non-MPIO.

- ▶ To migrate XIV 2810 devices from MPIO to non-MPIO, run the following command:
manage_disk_drivers -o AIX_non_MPIO -d 2810XIV
- ▶ To migrate XIV 2810 devices from non-MPIO to MPIO, run the following command:
manage_disk_drivers -o AIX_AAPCM -d 2810XIV

After running either of these commands, the system will need to be rebooted in order 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 from round_robin to fail_over mode or from fail_over to round_robin mode. The default disk behavior mode is round_robin, with a queue depth setting of forty. To check the disk behavior algorithm and queue depth settings, refer to Example 4-9.

Example 4-9 AIX: Viewing disk behavior and queue depth

# lsattr -El hdisk1 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 may need to be adjusted. Generally, in such an environment have a queue_depth between 64-256 and max_tranfer=0x100000.

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 and make sure 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, but 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. Obviously, you do not want to submit more IOs per XIV port than the 1400 maximum it can handle. The limit for the number of queued IOs for a HBA on AIX systems is 2048 (this is controlled by the **num_cmd_elems** attribute for the HBA).

Typical values are 40 to 64 as the queue depth per LUN, and 512-2048 per HBA in AIX.

Note: If you are booting your AIX system from a SAN attached LUN, make sure to have an mksysb image or at least a crash consistent snapshot of the boot LUN before changing HBA settings.

- To check the queue depth, periodically run **iostat -D 5**. If **avgwqsz** (average wait queue size) or **sqfull** are consistently greater zero, increase the queue depth (max 256).

See Table 4-1 and Table 4-2 for minimum level of service packs and the HAK version to determine the exact specification based on the AIX version installed on the host system.

Table 4-1 AIX 5.3 minimum level service packs and HAK Versions

AIX Release	APAR	Bundled in	HAK Version
AIX 5.3 TL 7*	IZ28969	SP 6	1.5.2
AIX 5.3 TL 8*	IZ28970	SP 4 - SP 8	1.5.2
AIX 5.3 TL 9*	IZ28047	SP 0 - SP 5	1.5.2
AIX 5.3 TL 10	IZ28061	SP 0 - SP 2	1.5.2

Table 4-2 AIX 6.1 minimum level service packs and HAK Versions

AIX Release	APAR	Bundled in	HAK Version
AIX 61 TL0*	IZ28002	SP 6	1.5.2
AIX 6.1 TL1*	IZ28004	SP 2	1.5.2
AIX 6.1 TL2*	IZ28079	SP 0	1.5.2
AIX 6.1 TL3	IZ30365	SP 0 - SP 2	1.5.2
AIX 6.1 TL 4	IZ59789	SP0	1.5.2

For all the AIX releases that are marked with * the queue depth is limited to 1 in round robin mode. Queue depth is limited to 256 when using MPIO with the fail_over mode.

As noted earlier, the default disk behavior algorithm is round_robin with a queue depth of 40. If the appropriate AIX levels and APAR list have been met, then the queue depth restriction is lifted and the settings can be adjusted. To adjust the disk behavior algorithm and queue depth setting, see Example 4-10.

Example 4-10 AIX: Change disk behavior algorithm and queue depth command

```
# chdev -a algorithm=round_robin -a queue_depth=40 -l <hdisk#>
```

In the command above, <hdisk#> stands for a particular instance of an hdisk.

If you want the fail_over disk behavior algorithm, after making the changes in Example 4-10, load balance the I/O across the FC adapters and paths by setting 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

There are commands to change priority attributes for paths that can specify a preference for the path 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/Os unless there is a path failure, in which case the other path will be used. After a path failure and recovery, if you have IY79741 installed, I/O will be redirected down the path with the highest priority. Otherwise, if you want the I/O to go down the primary path, use **chpath** to disable the secondary path, and then re-enable it. If the priority attribute is the same for all paths, the first path listed with **lspath -H1 <hdisk>** will be the primary path. Set the primary path to be used by setting its priority value to 1, and 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, there will be 255 I/Os sent down pathB.

To change the path priority of an MPIO device, use the **chpath** command. An example of this is shown as part of a procedure in Example 4-13 on page 119.

Initially, use the **lspath** command to display the operational status for the paths to the devices as shown in Example 4-11 on page 119.

Example 4-11 AIX: The lspath command shows the paths for hdisk2

```
# lspath -l hdisk2 -F status:name:parent:path_id:connection
Enabled:hdisk2:fscsi0:0:5001738000130140,20000000000000
Enabled:hdisk2:fscsi0:1:5001738000130150,20000000000000
Enabled:hdisk2:fscsi0:2:5001738000130160,20000000000000
Enabled:hdisk2:fscsi0:3:5001738000130170,20000000000000
Enabled:hdisk2:fscsi0:4:5001738000130180,20000000000000
Enabled:hdisk2:fscsi0:5:5001738000130190,20000000000000
Enabled:hdisk2:fscsi1:6:5001738000130142,20000000000000
Enabled:hdisk2:fscsi1:7:5001738000130152,20000000000000
Enabled:hdisk2:fscsi1:8:5001738000130162,20000000000000
Enabled:hdisk2:fscsi1:9:5001738000130172,20000000000000
Enabled:hdisk2:fscsi1:10:5001738000130182,20000000000000
Enabled:hdisk2:fscsi1:11:5001738000130192,20000000000000
```

The **lspath** command can also be used to read the attributes of a given path to an MPIO capable device as shown in Example 4-12.

The *<connection>* info is either “<SCSI ID>, <LUN ID>” for SCSI, (for example, “5,0”) or “<WWN>, <LUN ID>” for FC devices.

Example 4-12 AIX: The lspath command reads attributes of the 0 path for hdisk2

```
# lspath -AHE -l hdisk2 -p fscsi0 -w "5001738000130140,20000000000000"
attribute value          description  user_settable

scsi_id    0x133e00        SCSI ID      False
node_name  0x5001738000690000    FC Node Name False
priority   2                Priority      True
```

As just noted, the **chpath** command is used to perform change operations on a specific path. It can either change the operational status or tunable attributes associated with a path. It cannot perform both types of operations in a single invocation.

Example 4-13 illustrates the use of the **chpath** command with an XIV Storage System, which sets the primary path to fscsi0 using the first path listed (there are two paths from the switch to the storage for this adapter). Then for the next disk, we set the priorities to 4, 1, 2, and 3 respectively. If we are in fail-over mode and assuming the I/Os are relatively balanced across the hdisks, this setting will balance the I/Os evenly across the paths.

Example 4-13 AIX: The chpath command

```
# chpath -l hdisk2 -p fscsi0 -w 5001738000130160,20000000000000 -a priority=2 path Changed

# chpath -l hdisk2 -p fscsi1 -w 5001738000130140,20000000000000 -a priority=3 path Changed

# chpath -l hdisk2 -p fscsi1 -w 5001738000130160,20000000000000 -a priority=4 path Changed
```

The **rmpath** command unconfigures or undefines, or both, one or more paths to a target device. It is not possible to unconfigure (undefine) the last path to a target device using the **rmpath** command. The only way to unconfigure (undefine) the last path to a target device is to unconfigure the device itself (for example, use the **rmdev** command).

4.1.2 AIX host iSCSI configuration

At the time of writing, AIX 5.3 and AIX 6.1 operating systems are supported for iSCSI connectivity with XIV (for iSCSI hardware and software initiator). For iSCSI no Host Attachment Kit is required.

To make sure that your system is equipped with the required filesets, run the **lslpp** command as shown in Example 4-14. We used the AIX Version 6.1 operating system with Technology Level 05 in our examples.

Example 4-14 Verifying installed iSCSI filesets in AIX

```
# lslpp -la "*.iscsi*"
```

Fileset	Level	State	Description

Path: /usr/lib/objrepos			
devices.common.IBM.iscsi.rte			
	6.1.4.0	COMMITTED	Common iSCSI Files
	6.1.5.0	COMMITTED	Common iSCSI Files
devices.iscsi.disk.rte	6.1.4.0	COMMITTED	iSCSI Disk Software
	6.1.5.0	COMMITTED	iSCSI Disk Software
devices.iscsi.tape.rte	6.1.0.0	COMMITTED	iSCSI Tape Software
devices.iscsi_sw.rte	6.1.4.0	COMMITTED	iSCSI Software Device Driver
	6.1.5.1	COMMITTED	iSCSI Software Device Driver
Path: /etc/objrepos			
devices.common.IBM.iscsi.rte			
	6.1.4.0	COMMITTED	Common iSCSI Files
	6.1.5.0	COMMITTED	Common iSCSI Files
devices.iscsi_sw.rte	6.1.4.0	COMMITTED	iSCSI Software Device Driver

Current limitations when using iSCSI

The code available at the time of preparing this book had the following limitations when using the iSCSI software initiator in AIX:

- ▶ iSCSI is supported via a single path. No MPIO support is provided.
- ▶ The **xiv_iscsi_admin** 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 using 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.

Note: Volume Groups are activated during a different boot phase than the iSCSI software driver. For this reason, it is not possible to activate iSCSI Volume Groups during the boot process

- ▶ Do not span Volume Groups across non-iSCSI devices.

I/O failures

To avoid I/O failures, consider these recommendations:

- ▶ If connectivity to iSCSI target devices is lost, I/O failures occur. To prevent I/O failures and file system corruption, stop all I/O activity and unmount iSCSI-backed file systems before doing anything that will cause long term loss of connectivity to the active iSCSI targets.
- ▶ If a loss of connectivity to iSCSI targets occurs while applications are attempting I/O activities with iSCSI devices, I/O errors will eventually occur. It might not be possible to unmount iSCSI-backed file systems because the underlying iSCSI device stays busy.
- ▶ File system maintenance must be performed if I/O failures occur due to 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

The software initiator is configured using the System Management Interface Tool (SMIT) as shown in this procedure:

1. Select **Devices**.
2. Select **iSCSI**.
3. Select **iSCSI Protocol Device**.
4. Select **Change / Show Characteristics of an iSCSI Protocol Device**.
5. After selecting the desired device, verify that the iSCSI Initiator Name value. The Initiator Name value is used by the iSCSI Target during login.

Note: A default initiator name is assigned when the software is installed. This initiator name can be changed by the user to match local network naming conventions.

You can also issue the **lsattr** command to verify the `initiator_name` parameter as shown in Example 4-15.

Example 4-15 Check initiator name

```
# lsattr -El iscsi0 | grep initiator_name
initiator_name iqn.com.ibm.de.mainz.p550-tic-1v5.hostid.099b426e iSCSI Initiator
Name
```

6. 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 driver during configuration.

After the software initiator is configured, define iSCSI targets that will be accessed by the iSCSI software initiator. To specify those targets:

1. Determine your iSCSI IP addresses in the XIV Storage System. To get that information, select **iSCSI Connectivity** from the **Host and LUNs** menu as shown in Figure 4-2 on page 122.



Figure 4-2 iSCSI Connectivity

2. The iSCSI connectivity panel in Figure 4-3 shows all the available iSCSI ports. Use an MTU size of 4500.

Name	Address	Netmask	Gateway	MTU	Module	Ports
M7_P1	9.155.90.183	255.255.255.0	9.155.90.1	4500	1:Modul...	1
M7_P2_FW_1	9.155.50.63	255.255.255.0	9.155.50.1	4500	1:Modul...	2
M9_P1	9.155.90.186	255.255.255.0	9.155.90.1	4500	1:Modul...	1
M9_P2_FW_3	9.155.50.61	255.255.255.0	9.155.50.1	4500	1:Modul...	2

Figure 4-3 XIV iSCSI ports

If you are using XCLI, issue the `ipinterface_list` command, as shown in Example 4-16 in the XCLI. Use 4500 as the MTU size.

Example 4-16 List iSCSI interfaces

XIV LAB 3 1300203>>ipinterface_list

Name	Type	IP Address	Network Mask	Default Gateway	MTU	Module	Ports
M9_P1	iSCSI	9.155.90.186	255.255.255.0	9.155.90.1	4500	1:Module:9	1

3. Find the iSCSI name (IQN) of the XIV Storage System. To get this information, navigate to the basic system view in the XIV GUI, right-click the XIV Storage box itself, and select **Properties** → **Parameters**. The System Properties window appears as shown in Figure 4-4.

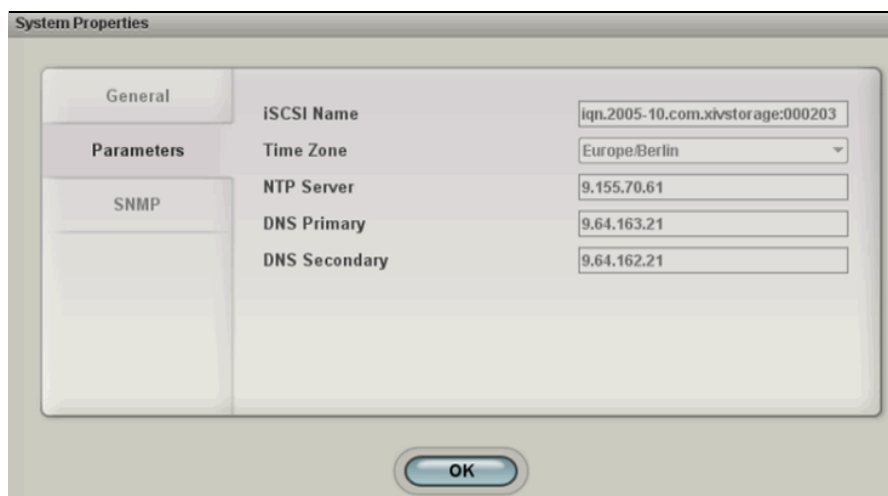


Figure 4-4 Verifying iSCSI name in XIV Storage System

If you are using XCLI, issue the `config_get` command. Refer to Example 4-17 on page 123.

Example 4-17 The config_get command in XCL

```
XIV LAB 3 1300203>>config_get
Name                               Value
dns_primary                        9.64.163.21
dns_secondary                      9.64.162.21
system_name                        XIV LAB 3 1300203
snmp_location                      Unknown
snmp_contact                      Unknown
snmp_community                    XIV
snmp_trap_community               XIV
system_id                         203
machine_type                      2810
machine_model                     A14
machine_serial_number             1300203
email_sender_address
email_reply_to_address
email_subject_format              {severity}: {description}
iscsi_name                        iqn.2005-10.com.xivstorage:000203
ntp_server                        9.155.70.61
support_center_port_type          Management
```

4. Go back to the AIX system and edit the `/etc/iscsi/targets` file to include the iSCSI targets needed during device configuration:

Note: The iSCSI targets file defines the name and location of the iSCSI targets that the iSCSI software initiator will attempt to access. This file is read any time that the iSCSI software initiator driver 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 gigabit Ethernet network that is separate from other network traffic.

Include your specific connection information in the targets file as shown in Example 4-18. Insert a HostName PortNumber and iSCSIName similar to what is shown in this example.

Example 4-18 Inserting connection information into /etc/iscsi/targets file in AIX operating system

```
9.155.90.186 3260 iqn.2005-10.com.xivstorage:000203
```

5. After editing the `/etc/iscsi/targets` file, enter the following command at the AIX prompt:

cfgmgr -l iscsi0

This command will reconfigure the software initiator driver, and causes the driver to attempt to communicate with the targets listed in the `/etc/iscsi/targets` file, and to define a new hdisk for each LUN found on the targets.

Note: 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 TCP Large Send, TCP send and receive flow control, and Jumbo Frame features of the AIX Gigabit Ethernet Adapter and the iSCSI Target interface.

Tune network options and interface parameters for maximum iSCSI I/O throughput on the AIX system:

1. Enable the RFC 1323 network option.
2. Set up the `tcp_sendspace`, `tcp_recvspace`, `sb_max`, and `mtu_size` network options and network interface options to appropriate values:
 - The iSCSI software initiator's maximum transfer size is 256 KB. Assuming that the system maximums for `tcp_sendspace` and `tcp_recvspace` are set to 262144 bytes, an `ifconfig` command used to configure a gigabit Ethernet interface might look like:

```
ifconfig en2 10.1.2.216 mtu 4500 tcp_sendspace 262144 tcp_recvspace 262144
```
3. Set the `sb_max` network option to at least 524288, and preferably 1048576.
4. Set the `mtu_size` to 4500, which is the default and maximum size. 9k frames are not supported.
5. For certain iSCSI targets, the TCP Nagle algorithm must be disabled for best performance. Use the `no` command to set the `tcp_nagle_limit` parameter to 0, which will disable the Nagle algorithm.

4.1.3 Management volume LUN 0

According to the SCSI standard, XIV Storage System maps itself in every map to LUN 0. This LUN serves as the “well known LUN” for that map, and the host can issue SCSI commands to that LUN that are not related to any specific volume. This device appears as a normal `hdisk` in the AIX operating system, and because it is not recognized by Windows by default, it appears with an unknown device's question mark next to it.

Exchange management of LUN 0 to a real volume

You might want to eliminate this management LUN on your system, or need to assign the LUN 0 number to a specific volume. In that case, map your volume to the first place in the mapping view and it will replace the management LUN to your volume and assign the zero value to it.

4.1.4 Host Attachment Kit utilities

The Host Attachment Kit includes a couple of useful utilities briefly described below.

► `xiv_devlist`

The `xiv_devlist` utility lists XIV and non-XIV volumes that are mapped to the AIX host. Example 4-19 on page 125 shows the output of this command. Two XIV disks 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. It's a utility you do not want to miss anymore.

Example 4-19 xiv_devlist output

```
# xiv_devlist
XIV Devices
-----
Device      Size    Paths Vol Name   Vol Id   XIV Id   XIV Host
-----
/dev/hdisk1 34.4GB 12/12  itso_aix_2 7343     6000105 itso_aix_p550_lpar2
-----
/dev/hdisk2 34.4GB 12/12  itso_aix_1 7342     6000105 itso_aix_p550_lpar2
-----

Non-XIV Devices
-----
Device      Size    Paths
-----
/dev/hdisk0 32.2GB 2/2
-----
```

The following options are available for the **xiv_devlist** command:

- **-t xml** to provide XML output format
- **--hex** to display volume ID and System ID in hexadecimal format
- **-o all** to add all available fields to the table
- **--xiv-only** to list only XIV volumes
- **-d** to write debugging information to a file

► **xiv_diag**

The **xiv_diag** utility gathers diagnostic data from the AIX operating system and saves it in a zip file. This file can be send to IBM support for analysis. See Example 4-20 for sample output.

Example 4-20 xiv_diag output

```
# xiv_diag
Please type in a path to place the xiv_diag file in [default: /tmp]:
Creating archive xiv_diag-results_2010-9-27_17-0-31
INFO: Gathering xiv_devlist logs... DONE
INFO: Gathering xiv_attach logs... DONE
INFO: Gathering snap: output... DONE
INFO: Gathering /tmp/ibmsupt.xiv directory... 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_2010-9-27_17-0-31.tar.gz to IBM-XIV
for review.
INFO: Exiting.
```

► **xiv_fc_admin** and **xiv_iscsi_admin**

Both of these utilities are used to perform administrative attachment and querying fibre channel and iSCSI attachment related information.

For more details about these utilities, please refer to the XIV Host Attachment Guide for AIX:

<http://www-01.ibm.com/support/docview.wss?uid=ssg1S4000802>

4.2 SAN boot in AIX

This section contains a step-by-step illustration of SAN boot implementation for the IBM POWER System (formerly System p) in an AIX v5.3 environment. Similar steps can be followed for an AIX v6.1 environment.

When using AIX SAN boot in conjunction with XIV, the default MPIO is used. During the boot sequence, AIX uses the information from the bootlist to find valid paths to a LUN/hdisk that contains a valid boot logical volume (hd5). However, a maximum of only five paths can be defined in the bootlist, while the recommended XIV multipathing setup results in more than five paths to a hdisk. In fact a fully redundant configuration establishes 12 paths (see Figure 1-6 on page 10).

For example, consider that we have 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, you would issue the command:

```
/ > bootlist -m normal hdisk0 hdisk1
```

Using the bootlist command to display the list of boot devices gives the output shown in Example 4-21.

Example 4-21 Displaying the bootlist

```
/ > bootlist -m normal -o
hdisk0 blv=hd5
hdisk0 blv=hd5
hdisk0 blv=hd5
hdisk0 blv=hd5
hdisk0 blv=hd5
```

Example 4-21 shows that hdisk1 is not present in the bootlist, and so the system could not boot from hdisk1 if we were to lose the paths to hdisk0.

There is a workaround in AIX 6.1 TL06 and AIX 7.1 to control the bootlist using the **pathid** parameter as in the following command:

```
bootlist -m normal hdisk0 pathid=0 hdisk0 pathid=1 hdisk1 pathid=0 hdisk1 pathid=1
```

There are various possible implementations of SAN boot with AIX:

- ▶ To implement SAN boot on a system with an already installed AIX operating system, you can mirror of the rootvg volume to the SAN disk.
- ▶ To implement SAN boot for a new system, start the AIX installation from a bootable AIX CD install package or use the Network Installation Manager (NIM).

The method known as *mirroring* is simpler to implement than the more complete and more sophisticated method using the Network Installation Manager.

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. If it is not already available, you must locate an available system where you can install AIX on an internal SCSI disk.

To create a boot disk on the XIV system, perform the following steps:

1. Select a logical drive that is the same size or larger than the size of rootvg that currently resides on the internal SCSI disk. Ensure that your AIX system can see the new disk. You can verify this with the **lspv** command. Verify the size with **bootinfo** and use **lsdev** to make sure that you are using an XIV (external) disk
2. Add the new disk to the rootvg volume group by clicking **smitty vg** → **Set Characteristics of a Volume Group** → **Add a Physical Volume from a Volume Group** (see Figure 4-5).

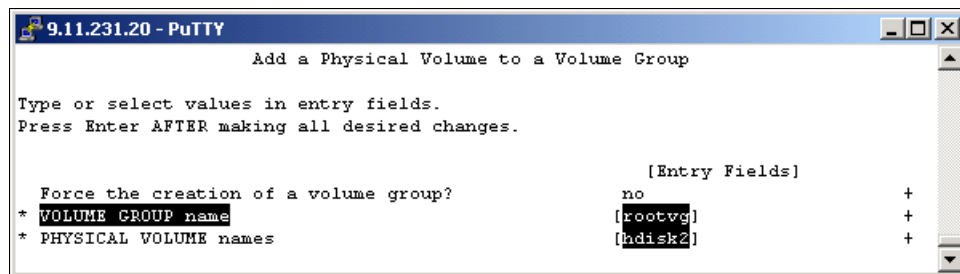


Figure 4-5 Add the disk to the rootvg

3. Create the mirror of rootvg. If the rootvg is already mirrored you can create a third copy on the new disk by clicking **smitty vg** → **Mirror a Volume Group**, then selecting the rootvg and the new hdisk.

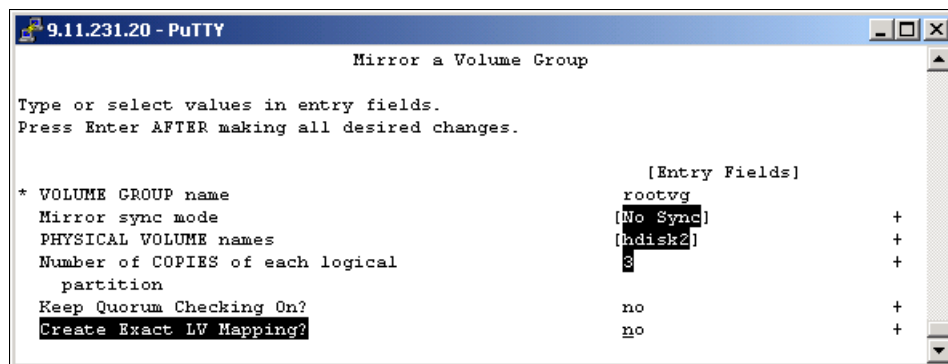


Figure 4-6 Create a rootvg mirror

4. Verify that all partitions are mirrored (Figure 4-7 on page 128) with **lsvg -l rootvg**, recreate the boot logical drive, and change the normal boot list with the following commands:

```
bosboot -ad hdiskx
```

```
bootlist -m normal hdiskx
```

```

9.11.231.20 - PuTTY
fastll10> lsvg -l rootvg
rootvg:
LV NAME          TYPE      LPs  PPs  PVs  LV STATE  MOUNT POINT
hd5              boot      1    3    3    closed/stale  N/A
hd6              paging    32   96    3    open/stale   N/A
hd8              jfslog    1    3    3    open/stale   N/A
hd4              jfs       2    6    3    open/stale   /
hd2              jfs      255   765    3    open/stale   /usr
hd9var           jfs       2    6    3    open/stale   /var
hd3              jfs       2    6    3    open/stale   /tmp
hd1              jfs       1    3    3    open/stale   /home
hd10opt          jfs       7   21    3    open/stale   /opt
fastll10> bosboot -ad hdisk2

0301-177 A previous bosdebug command has changed characteristics of this
        boot image. Use bosdebug -L to display what these changes are.

bosboot: Boot image is 24027 512 byte blocks.
fastll10> bootlist -m normal hdisk2
fastll10> _

```

Figure 4-7 Verify that all partitions are mirrored

5. Remove the original mirror copy by clicking **smitty vg-> Unmirror a Volume Group**. Choose the rootvg volume group, then the disks that you want to remove from mirror and run the command.
6. Remove the disk from the volume group rootvg by clicking **smitty vg-> Set Characteristics of a Volume Group-> Remove a Physical Volume from a Volume Group**, select rootvg for the volume group name ROOTVG and the internal SCSI disk you want to remove, and run the command.
7. Execute the following commands again (see step 4):

```

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.

4.2.2 Installation on external storage from bootable AIX CD-ROM

To install AIX on XIV System disks, make 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, the host is properly connected to the SAN, the zoning configuration is updated, and at least one LUN is mapped to the host.

Note: If the system cannot see the SAN fabric at login, you can configure the HBAs at the server open firmware prompt.

Because by nature a SAN allows access to a large number of 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 machine being installed can only discover the disks to be installed to. After the installation has completed, you can then reopen the zoning so the machine 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 as follows:

- If possible, assign Physical Volume Identifiers (PVIDs) to all disks from an already installed AIX system that can access the disks. This can be done using the command:

```
chdev -a pv=yes -l hdiskX
```

Where *X* is the appropriate disk number. Create a table mapping PVIDs to physical disks. The PVIDs will be visible from the install menus by selecting option **77 display more disk info** (AIX 5.3 install) when selecting a disk to install to. Or you could use the PVIDs to do an unprompted Network Installation Management (NIM) install.

- Another way to ensure the selection of the correct disk is to use Object Data Manager (ODM) commands. Boot from the AIX installation CD-ROM and from the main install menu, click **Start Maintenance Mode for System Recovery → Access Advanced Maintenance Functions → Enter the Limited Function Maintenance Shell**. At the prompt, issue the command:

```
odmget -q "attribute=lun_id AND value=0xNN..N" CuAt
```

or

```
odmget -q "attribute=lun_id" CuAt (list every stanza with lun_id attribute)
```

Where *0xNN..N* is the *lun_id* that you are looking for. This command prints out the ODM stanzas for the hdisks that have this *lun_id*. 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* by adding a byte of zeroes to the front and 6 bytes of zeroes to the end. For example, *lun_id* 2 becomes 0x0002000000000000. Note that usually the *lun_id* will be displayed without the leading zeroes. Care must be taken when installing because the installation procedure will allow installation to *lun_ids* outside of this range.

Installation procedure

To install on external storage, follow these steps:

1. Insert an AIX CD that has a bootable image into the CD-ROM drive.
2. Select **CD-ROM** as the install device to make the system boot from the CD. The way to change the bootlist varies model by model. In most System p models, this can be done by using the System Management Services (SMS) menu. Refer to the user's guide for your model.
3. Let the system boot from the AIX CD image after you have left the SMS menu.
4. After a few minutes the console should display a window that directs you to press the specified key on the device to be used as the system console.
5. A window is displayed that 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 that have been set for this machine in order to select a Fibre Channel-attached disk as a target disk. Type **2** and press Enter.
7. At the Installation and Settings window you should type **1** to change the system settings and select the **New and Complete Overwrite** option.
8. You are presented with the Change (the destination) Disk window. Here you can select the Fibre Channel disks that are mapped to your system. To make sure and get more information, type **77** to display the detailed information window. The system shows the PVID. Type **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 deselects the device. Be sure to select an XIV disk.

9. After you have selected the Fibre Channel-attached disks, the Installation and Settings window is displayed with the selected disks. Verify the installation settings. If everything looks okay, type 0 and press Enter, and the installation process begins.

Important: Be sure that you have made the correct selection for root volume group because the existing data in the destination root volume group will be destroyed during BOS installation.

10. When the system reboots, a window message displays the address of the device from which the system is reading the boot image.

4.2.3 AIX SAN installation with NIM

Network Installation Manager (NIM) is a client server infrastructure and service that allows remote install of the operating system, manages software updates, and can be configured to install and update third-party applications. Although both the NIM server and client file sets are part of the operating system, a separate NIM server has to be configured to keep the configuration data and the installable product file sets.

We assume that the NIM environment is deployed and all of the necessary configurations on the NIM master are already done:

- ▶ The NIM server is properly configured as the NIM master and the basic NIM resources have been defined.
- ▶ The Fibre Channel Adapters are already installed on the machine 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 machine (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, refer to the *AIX 5L Version 5.3: Installing AIX* reference, SC23-4887-02.

Installation procedure

Prior the installation, you should modify the `bosinst.data` file, where the installation control is stored.

Insert your appropriate values at the following stanza:

`SAN_DISKID`

This 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 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//0x20000000000000`

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:

1. Enter the command:

```
# smit nim_bosinst
```

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 is capable of performing a non prompted BOS installation.
5. Select the **RESOLV_CONF to use for network configuration** option, and select a `resolv_conf` resource.
6. Select 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 va09s
```




HP-UX host connectivity

This chapter explains specific considerations for attaching the XIV system to a HP-UX host.

For the latest information, refer to the hosts attachment kit publications at:

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

HP-UX manuals are available at the HP Business Support Centre:

<http://www.hp.com/go/hpux-core-docs>

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

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

5.1 Attaching XIV to a HP-UX host

At the time of writing this book, then XIV Storage System software release 10.2 supports Fibre Channel attachment to HP Integrity and PA-RISC servers running HP-UX 11iv2 (11.23) and HP-UX 11iv3 (11.31). For details and up-to-date information about supported environments, refer to the IBM System Storage Interoperation Center (SSIC) at:

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

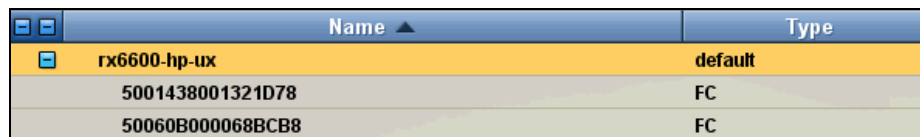
The HP-UX host attachment process with XIV is described in detail in the Host Attachment Guide for HPUX which is available at the IBM Support Portal:

[http://www.ibm.com/support/entry/portal/Overview/Hardware/System_Storage/Disk_systems/Enterprise_Storage_Servers/XIV_Storage_System_\(2810,_2812\)](http://www.ibm.com/support/entry/portal/Overview/Hardware/System_Storage/Disk_systems/Enterprise_Storage_Servers/XIV_Storage_System_(2810,_2812))

The attachment process includes getting the world-wide identifiers (WWN) of the host Fibre Channel adapters, SAN zoning, definition of volumes and host objects on the XIV storage system, mapping the volumes to the host, and installation of the XIV Host Attachment Kit which can also be downloaded from the above URL.

This section focusses 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 and Figure 5-2 show the host object and the volumes that were defined for the HP-UX server used for the examples in this book.



Name ▲	Type
rx6600-hp-ux	default
5001438001321D78	FC
50060B000068BCB8	FC

Figure 5-1 XIV host object for the HP-UX server



LUN	Volume
1	rx6600-hpux_1
2	rx6600-hpux_2
3	rx6600-hpux_3
4	rx6600-hpux_4
5	rx6600_hpux1131_bootdisk

Figure 5-2 XIV volumes mapped to the HP-UX server

The HP-UX utility `ioscan` displays the host's Fibre Channel adapters and `fcmsutil` displays details of these adapters including the WWN. See Example 5-1 on page 135.

Example 5-1 HP Fibre Channel adapter properties

```
# ioscscan -fnk|grep fcd
fc          0  0/3/1/0      fcd          CLAIMED      INTERFACE      HP AB379-60101
4Gb Dual Port PCI/PCI-X Fibre Channel Adapter (FC Port 1)
/dev/fcd0

fc          2  0/7/1/0      fcd          CLAIMED      INTERFACE      HP AB379-60101
4Gb Dual Port PCI/PCI-X Fibre Channel Adapter (FC Port 1)
/dev/fcd1

# fcmsutil /dev/fcd0

Vendor ID is = 0x1077
Device ID is = 0x2422
PCI Sub-system Vendor ID is = 0x103C
PCI Sub-system ID is = 0x12D7
PCI Mode = PCI-X 266 MHz
ISP Code version = 4.2.2
ISP Chip version = 3
Topology = PTTOPT_FABRIC
Link Speed = 4Gb
Local N_Port_id is = 0x133900
Previous N_Port_id is = None
N_Port Node World Wide Name = 0x5001438001321d79
N_Port Port World Wide Name = 0x5001438001321d78
Switch Port World Wide Name = 0x203900051e031124
Switch Node World Wide Name = 0x100000051e031124
N_Port Symbolic Port Name = rx6600-1_fcd0
N_Port Symbolic Node Name = rx6600-1_HP-UX_B.11.31
Driver state = ONLINE
Hardware Path is = 0/3/1/0
Maximum Frame Size = 2048
Driver-Firmware Dump Available = NO
Driver-Firmware Dump Timestamp = N/A
Driver Version = @(#) fcd B.11.31.0809.%319 Jul  7 2008
```

The XIV Host Attachment Kit includes scripts to facilitate HP-UX attachment to XIV. For example the `xiv_attach` script identifies the host's Fibre Channel adapters that are connected to XIV storage systems and the name of the host object defined on the XIV storage system for this host (if already created) and supports rescanning for new storage devices. See Example 5-2.

Example 5-2 xiv_attach script output

```
# /usr/bin/xiv_attach
-----
Welcome to the XIV Host Attachment wizard, version 1.5.
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:
5001438001321d78: /dev/fcd0: []:
50060b000068bcb8: /dev/fcd1: []:
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    Ver    Host Defined  Ports Defined  Protocol  Host Name(s)
6000105   10.2   Yes          All            FC        rx6600-hp-ux
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 ]: no

```

```

Press [ENTER] to proceed.

```

```

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

```

5.2 HP-UX multi-pathing solutions

Up to HP-UX 11iv2, *pvlinks* was HP's multipathing solution on HP-UX and was built into the Logical Volume Manager (LVM). Devices have been addressed by a specific path of hardware components such as adapters and controllers. Multiple I/O paths from the server to the device resulted in the creation of multiple device files in HP-UX. This addressing method is now called legacy addressing.

HP introduced HP Native Multi-Pathing with HP-UX 11iv3. The earlier *pvlinks* multi-pathing is still available, but Native Multi-Pathing should be used because HP Native-Multipathing provides I/O load balancing across the available I/O paths while *pvlinks* 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 leverages the so-called Agile View Device Addressing which addresses a device by its world-wide identifier (WWID) as an object. Thus 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. Consequently with this addressing method only one device file is created for the device.

Example 5-3 shows the HP-UX view 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     0  0/0/2/1.0x0.0x10  UsbScsiAdaptor  CLAIMED    LUN_PATH  USB SCSI Stack
Adaptor
disk     4  64000/0xfa00/0x0  esdisk        CLAIMED    DEVICE     HP
DH072ABAA6
                        /dev/disk/disk4  /dev/rdisk/disk4
disk     5  64000/0xfa00/0x1  esdisk        CLAIMED    DEVICE     HP
DH072ABAA6

```

```

/dev/disk/disk5      /dev/disk/disk5_p2  /dev/disk/disk5
/dev/rdisk/disk5_p2 /dev/disk/disk5_p1  /dev/disk/disk5_p3  /dev/rdisk/disk5_p1
/dev/rdisk/disk5_p3
disk 1299 64000/0xfa00/0x64 esdisk CLAIMED DEVICE IBM 2810XIV
      /dev/disk/disk1299 /dev/rdisk/disk1299
disk 1300 64000/0xfa00/0x65 esdisk CLAIMED DEVICE IBM 2810XIV
      /dev/disk/disk1300 /dev/rdisk/disk1300
disk 1301 64000/0xfa00/0x66 esdisk CLAIMED DEVICE IBM 2810XIV
      /dev/disk/disk1301 /dev/rdisk/disk1301
disk 1302 64000/0xfa00/0x67 esdisk CLAIMED DEVICE IBM 2810XIV
      /dev/disk/disk1302 /dev/rdisk/disk1302

# ioscan -m dsf /dev/disk/disk1299
Persistent DSF          Legacy DSF(s)
=====
/dev/disk/disk1299      /dev/dsk/c153t0d1
                        /dev/dsk/c155t0d1

```

If device special files are missing on the HP-UX server, there are two options to create them. The first one is a reboot of the host, which is disruptive. The alternative is to run the command **insf -eC disk**, which will reinstall the special device files for all devices of the class disk.

Finally, 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 to create a volume group in a Logical Volume Manager (LVM) environment. The rest is standard HP-UX system administration, not XIV-specific and therefore is not discussed in this book.

HP Native Multi-Pathing is automatically used to 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/disk1299
Physical volume "/dev/rdisk/disk1299" has been successfully created.
# pvcreate /dev/rdisk/disk1300
Physical volume "/dev/rdisk/disk1300" has been successfully created.

# vgcreate vg02 /dev/disk/disk1299 /dev/disk/disk1300
Increased the number of physical extents per physical volume to 4095.
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, there are two volume managers to choose from:

- ▶ The HP Logical Volume Manager (LVM).
- ▶ The VERITAS Volume Manager (VxVM). In this context however it is important to recall that any I/O is handled in pass-through mode and therefore executed by Native Multipathing, **not** by DMP.

According to the *HP-UX System Administrator's Guide: Overview*, B3921-90011, Edition 5 Sept. 2010, both volume managers can coexist on an HP-UX server. See:

<http://bizsupport1.austin.hp.com/bc/docs/support/SupportManual/c02281492/c02281492.pdf>

You can use both simultaneously (on separate physical disks), but usually you will choose one or the other and use it exclusively.

The configuration of XIV volumes on HP-UX with LVM has been described earlier in this chapter. 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

```
# vxdisk list
DEVICE      TYPE          DISK          GROUP          STATUS
c2t0d0      auto:none     -             -              online invalid
c2t1d0      auto:none     -             -              online invalid
c10t0d1     auto:none     -             -              online invalid
c10t6d0     auto:none     -             -              online invalid
c10t6d1     auto:none     -             -              online invalid
c10t6d2     auto:none     -             -              online invalid

# 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
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,?] c10t6d0 c10t6d1

Here are the disks selected. Output format: [Device_Name]

c10t6d0 c10t6d1

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

There is no active disk group named 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)

A new disk group will be created named dg01 and the selected disks will be added to the disk group with default disk names.

c10t6d0 c10t6d1

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 the desired format [cdsdisk,hpdisk,q,?] (default: cdsdisk) hpdisk

Enter desired private region length
[<privlen>,q,?] (default: 1024)

Initializing device c10t6d0.

Initializing device c10t6d1.

VxVM NOTICE V-5-2-120

Creating a new disk group named dg01 containing the disk device c10t6d0 with the name dg0101.

VxVM NOTICE V-5-2-88

Adding disk device c10t6d1 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
c2t0d0	auto:none	-	-	online invalid
c2t1d0	auto:none	-	-	online invalid
c10t0d1	auto:none	-	-	online invalid
c10t6d0	auto:hpdisk	dg0101	dg01	online
c10t6d1	auto:hpdisk	dg0102	dg01	online
c10t6d2	auto:none	-	-	online invalid

The graphical equivalent for the **vxdiskadm** utility is the VERITAS Enterprise Administrator (VEA). Figure 5-3 on page 141 shows the presentation of disks by this graphical user interface.

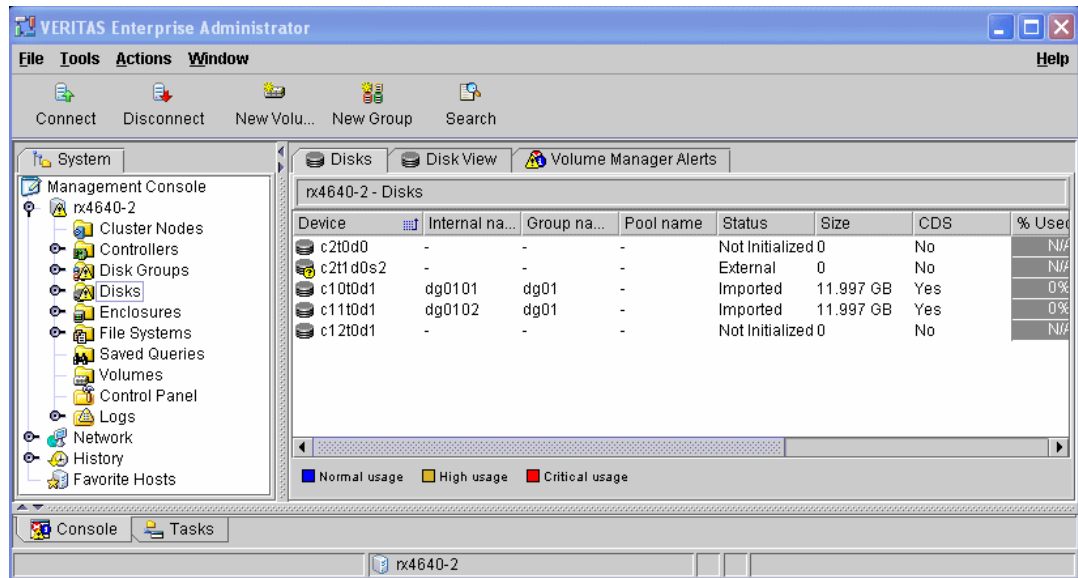


Figure 5-3 Disk presentation by VERITAS Enterprise Administrator

Also in this example, after having creating the diskgroups and the VxVM disks, you will need to create file systems and mount them.

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 certain storage system this service is provided by an Array Support Library (ASL) that can be downloaded from the Symantec websites. 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. See Example 5-6.

Example 5-6 VxVM command to list Array Support Libraries

```
# vxddladm listsupport
LIBNAME          VID
=====
...
libvxxiv.sl      XIV, IBM

# vxddladm listsupport libname=libvxxiv.sl
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 enable easier identification of the attached disk storage devices by serially numbering the attached storage systems of the same type and the volumes of a single storage system being assigned to this host.

Example 5-7 shows that four volumes of one XIV system are assigned to that HP-UX host. VxVM controls the devices XIV1_3 and XIV1_4 and the disk group name is dg02. HP's Logical Volume Manager (LVM) controls the remaining XIV devices.

Example 5-7 VxVM disk list

# vxdisk list				
DEVICE	TYPE	DISK	GROUP	STATUS
Disk_0s2	auto:LVM	-	-	LVM
Disk_1	auto:none	-	-	online invalid
XIV1_0	auto:LVM	-	-	LVM
XIV1_1s2	auto:LVM	-	-	LVM
XIV1_2	auto:LVM	-	-	LVM
XIV1_3	auto:cdsdisk	dg0201	dg02	online
XIV1_4	auto:cdsdisk	dg0202	dg02	online

An ASL overview is available at the following URL:

<http://www.symantec.com/business/support/index?page=content&id=TECH21351>

ASL packages for XIV and HP-UX 11iv3 are available for download from this web page:

<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 running HP-UX 11iv3 (11.31). Boot management is provided by the Extensible Firmware Interface (EFI). Earlier systems ran another boot manager and thus the SAN boot process may differ.

There are various possible implementations of SAN boot with HP-UX:

- ▶ To implement SAN boot for a new system, you can start the HP-UX installation from a bootable HP-UX CD or DVD install package or use a network-based installation, for example Ignite-UX.
- ▶ To implement SAN boot on a system with an already installed HP-UX operating system, is possible by mirroring of the system disk(s) volume to the SAN disk.

5.4.1 HP-UX Installation on external storage

To install HP-UX on XIV system volumes make sure that you have an appropriate SAN configuration: The host is properly connected to the SAN, the zoning configuration is updated, and at least one LUN is mapped to the host.

Because by nature a SAN allows access to a large number of devices, identifying the volume to install to can be difficult. We recommend the following method to facilitate the discovery of the *lun_id* to HP-UX device file correlation:

1. If possible, zone the switch and change the LUN mapping on the XIV storage system such that the machine being installed can only discover the disks to be installed to. After the installation has completed, you can then reopen the zoning so the machine can discover all necessary devices.
2. If possible, temporarily attach the volumes to an already installed HP-UX system. Note down the hardware paths of the volumes to later compare them to the other system's hardware paths during installation. Example 5-8 shows the output of the **ioscan** command that creates a hardware path list.
3. In any case note down the volumes' 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 shown in Example 5-8. This disk's hardware path name includes the string 0x5. See Figure 5-2 on page 134.

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
        0/3/1/0.0x5001738000cb0170.0x0    0/3/1/0.19.6.255.0.0.0
        0/3/1/0.0x5001738000cb0182.0x0    0/3/1/0.19.1.0.0.0.0
        0/3/1/0.0x5001738000cb0192.0x0    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
        0/7/1/0.0x5001738000cb0192.0x0    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
        0/7/1/0.0x5001738000cb0192.0x0    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
        0/3/1/0.0x5001738000690160.0x0    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
        0/7/1/0.0x5001738000690190.0x0    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
        0/3/1/0.0x5001738000690160.0x2000000000000
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

```

Installation procedure

The examples and screen shots in this chapter refer to a HP-UX installation on HP's Itanium-based Integrity systems. On older HP PA-RISC systems the processes to boot the server and select disk(s) to install HP-UX to is 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*, BA927-90045, Edition 8 Sept. 2010, available at the following URL:

<http://bizsupport2.austin.hp.com/bc/docs/support/SupportManual/c02281370/c02281370.pdf>

Follow these steps to install HP-UX 11iv3 on an XIV volume from DVD - on a HP Integrity system:

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. Select **Boot from DVD** and continue. See Figure 5-4

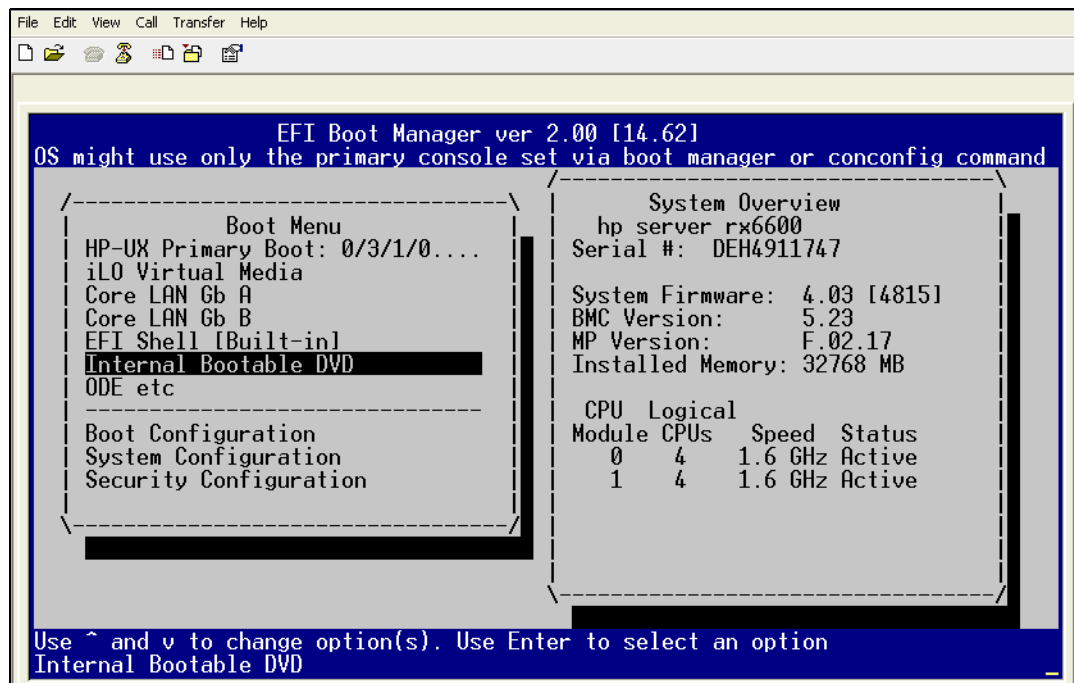


Figure 5-4 Boot device selection with EFI Boot Manager

3. The server boots from the installation media. Wait for the "HP-UX installation and recovery process" panel and select **Install HP-UX**. See Figure 5-5 on page 145.

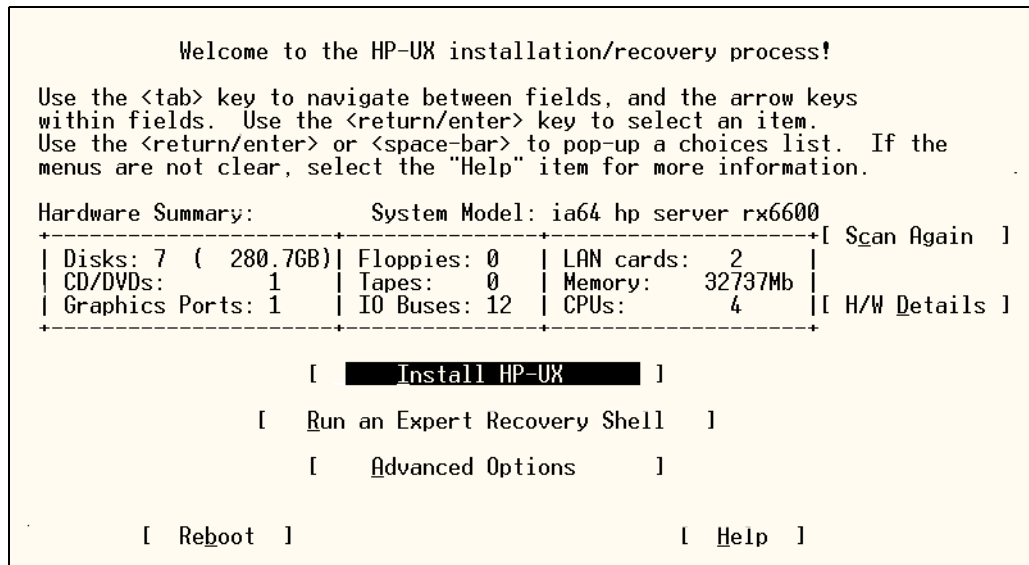


Figure 5-5 HP-UX installation panel: start OS installation

- In a subsequent step 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. See Figure 5-6.

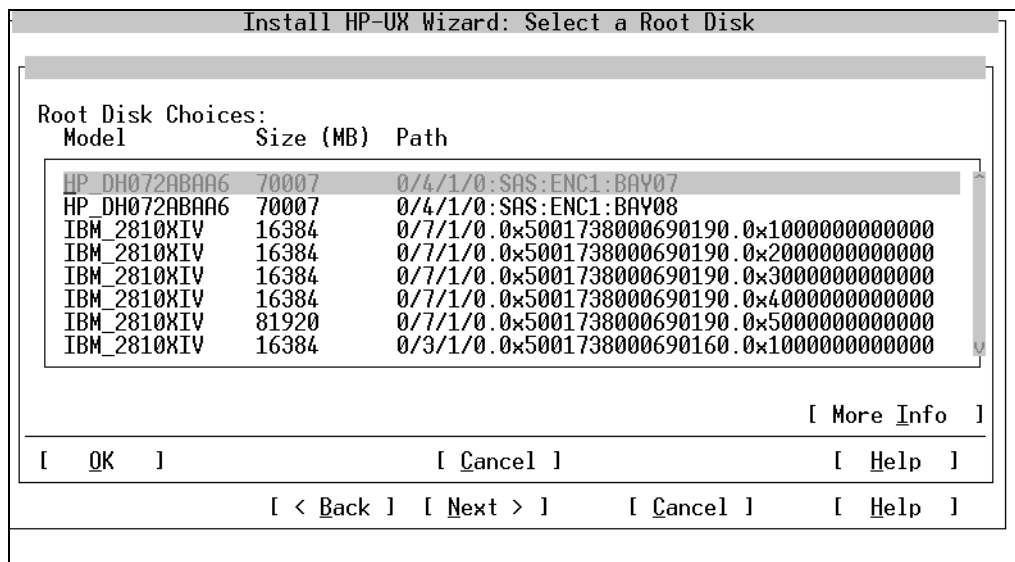


Figure 5-6 HP-UX installation panel: select a root disk

- The remaining steps of a HP-UX installation on a SAN disk do not differ from installation on an internal disk.

Creating a SAN boot disk by mirroring

The section "Mirroring the Boot Disk" of the HP manual "HP-UX System Administrator's Guide: Logical Volume Management HP-UX 11i Version 3 (B3921-90014, Edition 5, Sept. 2010)" includes a detailed description of the boot disk mirroring process. The manual is available at the following URL:

<http://bizsupport1.austin.hp.com/bc/docs/support/SupportManual/c02281490/c02281490.pdf>

The storage-specific part is the identification of the XIV volume to install to on HP-UX. Refer to 5.4.1, “HP-UX Installation on external storage” on page 142 for hints on this identification process.



Solaris host connectivity

This chapter explains specific considerations for attaching the XIV system to a Solaris host.

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

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

and the Host Attachment publications at:

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

6.1 Attaching a Solaris host to XIV

Before starting with the configuration, the network setup needs to be done first. After the connection is established in the SAN for the FC connectivity. For the iSCSI connection the iSCSI ports needs to be configured first on the system. A description can be found in 1.3, “iSCSI connectivity” on page 21.

Note: You can also have a coexistence of Fibre Channel and iSCSI connections to attach various hosts.

Do not use both Fibre Channel and iSCSI connections for the same LUN at the same host.

6.2 Solaris host FC configuration

This section describes attaching a Solaris host to XIV over Fibre Channel and provides detailed descriptions and installation instructions for the various software components required. Check in advance on the IBM SSIC to make sure that your HBA and the firmware is supported.

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

Our environment to prepare the examples that we present in the remainder of this section consists of a SUN Sparc v480 running with Solaris 10 U8.

6.2.1 Obtain WWPN for XIV volume mapping

To map the volumes to the Solaris host, you must know the World Wide Port Names (WWPNs) of the HBAs. WWPNs can be found in the `fcinfo`. Refer to Example 6-1 for details.

Example 6-1 WWPNs of the HBAs

```
# fcinfo hba-port | grep HBA
HBA Port WWN: 210000e08b137f47
HBA Port WWN: 210000e08b0c4f10
```

6.2.2 Installing the Host Attachment Kit

To install the HAK, open a terminal session and go to the directory where the package was downloaded. Execute the command shown in Example 6-2 to extract the archive.

Example 6-2 Extracting the HAK

```
# gunzip -c XIV_host_attach-<version>-<os>-<arch>.tar.gz | tar xvf -
```

Change to the newly created directory and invoke the Host Attachment Kit installer, as you can see in Example 6-3.

Example 6-3 Starting the installation

```
# cd XIV_host_attach-<version>-<os>-<arch>
# /bin/sh ./install.sh
```

Follow the prompts. After running the installation script, review the installation log file `install.log` residing in the same directory.

6.2.3 Configuring the host

Use the utilities provided in the Host Attachment Kit to configure the Solaris host. Host Attach Kit packages are installed in `/opt/xiv/host_attach` directory.

Note: You must be logged in as root or have root privileges to use the Host Attachment Kit.

The main executable files will be installed in the folder seen 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, you need to set up your SAN zoning first so that the XIV is visible for the Host. To start the configuration, run the `xiv_attach` command. This is mandatory for support. Referring to Example 6-5 for an example of the host configuration with the command.

Note: After running the `xiv_attach` command for the first time, the server will need to be rebooted.

Example 6-5 Sample results of the xiv_attach command

`# xiv_attach`

Welcome to the XIV Host Attachment wizard, version 1.5.2.
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 : f

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...

A reboot is required in order to continue.
Please reboot the machine and restart the wizard

Press [ENTER] to exit.

After the system reboot, you have to start `xiv_attach` again to finish the system to XIV configuration for the Solaris host, as seen in Example 6-6 on page 150.

Example 6-6 Fibre Channel host attachment configuration after reboot

xiv_attach

Welcome to the XIV Host Attachment wizard, version 1.5.2.
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 : f

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:
210000e08b0c4f10: /dev/cfg/c2: [QLogic Corp.]: QLA2340
210000e08b137f47: /dev/cfg/c3: [QLogic Corp.]: QLA2340
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)
6000105	10.2	No	None	FC	--
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: sun-v480R-tic-1]: sun-s1e-1
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 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.

Note: A rescan of for new XIV luns can be done with **xiv_fc_admin -R**.

The command **/opt/xiv/host_attach/bin/xiv_devlist** or just **xiv_devlist**, which can be executed from each working directory, will show you the mapped volumes and the number of paths to the IBM XIV Storage System, as shown in Example 6-7 on page 151.

Example 6-7 Showing mapped volumes and available pathes

```
# xiv_devlist -x
XIV Devices
```

Device	Size	Paths	Vol Name	Vol Id	XIV Id	XIV Host
/dev/dsk/c4t0017380000CB1174d0	17.2GB	4/4	itso_1	4468	1300203	sun-sle-1
/dev/dsk/c4t0017380000CB1175d0	17.2GB	4/4	itso_2	4469	1300203	sun-sle-1

6.3 Solaris host iSCSI configuration

This chapter will introduce how to connect an iSCSI volume to the server.

Our example environment consists of a SUN Sparc v480 running with Solaris 10 U8.

Execute the command **xiv_attach** first, as shown in Example 6-8.

Example 6-8 xiv_attach for iscsi

```
# xiv_attach
```

```
Welcome to the XIV Host Attachment wizard, version 1.5.2.
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...
The wizard needs to configure the host for the XIV system.
Do you want to proceed? [default: yes ]:
Please wait while the host is being configured...
The host is now being 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 ]:
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)
6000105   10.2   No              None              FC          --
1300203   10.2   No              None              FC          --
This host is defined on all iSCSI-attached XIV storage arrays

Press [ENTER] to proceed.
```

```
The XIV host attachment wizard successfully configured this host
Press [ENTER] to exit.
```

If you do not know the iSCSI qualified name (IQN) of your server, you can discover it with the **xiv_iscsi_admin -P** command as seen in the Example 6-9.

Example 6-9 IQN

```
# xiv_iscsi_admin -P
iqn.1986-03.com.sun:01:0003ba4dbd8a.4c84dec9
```

Define a iSCSI host and map a volume on the XIV system.

After mapping the volumes to the server, rescan the iscsi using the **xiv_iscsi_admin -R** command. You will 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	Paths	Vol Name	Vol Id	XIV Id	XIV Host
/dev/dsk/c4t001738000CB1174d0	17.2GB	2/2	itso_1	4468	1300203	sun-iscsi

6.4 Solaris Host Attachment Kit utilities for FC and iSCSI

The Host Attachment Kit (HAK) now includes the following utilities:

► **xiv_devlist**

xiv_devlist is the command allowing validation of the attachment configuration. This command generates a list of multipathed devices available to the operating system. In Example 6-11 you can see 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)
  -o FIELDS, --options=FIELDS
                           Fields to display; Comma-separated, no spaces. Use -l
                           to see the list of fields
  -H, --hex                 Display XIV volume and machine IDs in hexadecimal base
  -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
```

► **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. To run, go to a command prompt and enter **xiv_diag**. Example results are shown in Example 6-12.

Example 6-12 xiv_diag command

```
[/]  
[/]# xiv_diag  
Please type in a path to place the xiv_diag file in [default: /tmp]:  
Creating archive xiv_diag-results_2010-9-30_10-33-21  
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 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_2010-9-30_10-33-21.tar.gz to IBM-XIV  
for review.  
INFO:  
Exiting.
```

6.5 Partitions and filesystems

This section illustrates the creation and use of partition and filesystems from XIV-provided storage.

6.5.1 Creating partitions and filesystems with UFS

This section describes how to create a partition and filesystem with UFS, on mapped XIV volumes. We used Solaris 10 on a Sparc to illustrate the process shown in Example 6-13.

Example 6-13 Mapped XIV volumes

```
# xiv_devlist -x  
XIV Devices  
-----  
Device                               Size   Paths Vol Name  Vol Id  XIV Id  XIV Host  
-----  
/dev/dsk/c4t0017380000CB1174d0 17.2GB 4/4   itso_1   4468   1300203 sun-sle-1  
-----  
/dev/dsk/c4t0017380000CB1175d0 17.2GB 4/4   itso_2   4469   1300203 sun-sle-1  
-----
```

Example 6-14 Solaris format tool

```
# format
Searching for disks...done

c4t0017380000CB1175d0: configured with capacity of 15.98GB

AVAILABLE DISK SELECTIONS:
  0. c1t0d0 <SUN72G cyl 14087 alt 2 hd 24 sec 424>
    /pci@9,600000/SUNW,qlc@2/fp@0,0/ssd@w500000e0102e9dd1,0
  1. c1t1d0 <SUN72G cyl 14087 alt 2 hd 24 sec 424>
    /pci@9,600000/SUNW,qlc@2/fp@0,0/ssd@w500000e010183a51,0
  2. c4t0017380000CB1174d0 <IBM-2810XIV-10.2 cyl 2046 alt 2 hd 128 sec 128>
    /scsi_vhci/ssd@g0017380000cb1174
  3. c4t0017380000CB1175d0 <IBM-2810XIV-10.2 cyl 2046 alt 2 hd 128 sec 128>
    /scsi_vhci/ssd@g0017380000cb1175
Specify disk (enter its number): 3
selecting c4t0017380000CB1175d0
[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 command **partition** in the format tool to change the partition table. You can see the newly defined table 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: 2046 + 2 (reserved cylinders)

Part      Tag      Flag      Cylinders      Size      Blocks
  0       root      wm         0              0      (0/0/0)        0
  1       swap      wu         0              0      (0/0/0)        0
  2    backup      wu      0 - 2045     15.98GB  (2046/0/0) 33521664
  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      0 - 2045     15.98GB  (2046/0/0) 33521664
  7 unassigned      wm         0              0      (0/0/0)        0

partition> label
Ready to label disk, continue? yes

partition> quit
...
format> quit

```

Verify the new table as shown in Example 6-3.

Example 6-16 Verifying the table

```

# prtvtoc /dev/rdisk/c4t0017380000CB1175d0s2
* /dev/rdisk/c4t0017380000CB1175d0s2 partition map
*
* Dimensions:
*   512 bytes/sector
*   128 sectors/track
*   128 tracks/cylinder
* 16384 sectors/cylinder
*   2048 cylinders
*   2046 accessible cylinders
*
* Flags:
*   1: unmountable
*  10: read-only
*
*
* Partition  Tag  Flags      First      Sector      Last
*           Tag  Flags      Sector      Count      Sector  Mount Directory
*     2       5   01          0 33521664 33521663
*     5       0   00          0 33521664 33521663

```

You can now create a new filesystem on the partition/volume like in Example 6-17 on page 156.

Example 6-17 Making a new filesystem

```
# newfs /dev/rdisk/c4t0017380000CB1175d0s2
newfs: construct a new file system /dev/rdisk/c4t0017380000CB1175d0s2: (y/n)? y
/dev/rdisk/c4t0017380000CB1175d0s2:      33521664 sectors in 5456 cylinders of 48
tracks, 128 sectors
      16368.0MB in 341 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:
  32540064, 32638496, 32736928, 32835360, 32933792, 33032224, 33130656,
  33229088, 33327520, 33425952
```

You can optionally check the filesystem as seen in Example 6-18.

Example 6-18 Checking the filesystem

```
# fsck /dev/rdisk/c4t0017380000CB1175d0s2
** /dev/rdisk/c4t0017380000CB1175d0s2
** 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, 16507097 free (9 frags, 2063386 blocks, 0.0% fragmentation)
```

After mounting the volume as shown in Example 6-19, you can start using the volume with a ufs filesystem.

Example 6-19 Mount the volume to Solaris

```
# mount /dev/dsk/c4t0017380000CB1175d0s2 /XIV_vol/
bash-3.00# df -h
Filesystem              size  used  avail capacity  Mounted on
/dev/dsk/c1t1d0s0        16G   4.2G   12G     27%      /
/devices                 0K     0K     0K      0%    /devices
...
...
swap                    3.4G   184K   3.4G      1%    /tmp
swap                    3.4G    32K   3.4G      1%    /var/run
/dev/dsk/c1t1d0s7        51G   500M   50G      1%    /export/home
/dev/dsk/c4t0017380000CB1175d0s2
                        16G    16M   16G      1%    /XIV_vol
```



Symantec Storage Foundation

This chapter explains specific considerations for host connectivity and describes the host attachment-related tasks for the OS platforms that use Symantec Storage Foundation instead of their built-in functionality.

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

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

and the Host Attachment publications at:

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

7.1 Introduction

The Symantec Storage Foundation, formerly known as the Veritas Volume Manager (VxVM) and Veritas Dynamic Multipathing (DMP), is available for the OS platform as a unified method of volume management at the OS level.

At the time of writing, XIV supports the use of VxVM and DMP with several operating systems including HP-UX, AIX, Redhat Enterprise Linux, SUSE Linux, Linux on Power, and Solaris.

Depending on the OS version and hardware platform, only specific versions and releases of Veritas Volume Manager are supported when connecting to XIV. In general, IBM supports VxVM versions 4.1, 5.0, and 5.1.

For most of the OS and VxVM versions mentioned above, IBM supports space reclamation on thin provisioned volumes.

Refer to the System Storage Interoperability Center for the latest and detailed information about the Operating Systems and VxVM versions supported:

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

In addition, you can also find information about attaching the IBM XIV Storage System to hosts with VxVM and DMP at the Symantec website:

<https://vos.symantec.com/asl>

7.2 Prerequisites

In addition to common prerequisites, such as cabling, SAN zoning defined, volumes created and mapped to the host, the following must also be completed to successfully attach XIV to host systems 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 DMP multipathing with IBM XIV.

Make sure you have installed all the patches and updates available for your Symantec Storage Foundation installation. For instructions, refer to your Symantec Storage Foundation documentation.

7.2.1 Checking ASL availability

To illustrate here attachment to XIV and configuration for hosts using VxVM with DMP as logical volume manager, we used Solaris version 10 on SPARC. The steps are similar for most Unix and Linux hosts.

To check for the presence of ASL on your host system, log on to the host as root and execute the command shown in Example 7-1 on page 159.

Example 7-1 Check the availability ASL for IBM XIV Storage System

```
# vxddladm listversion
LIB_NAME                ASL_VERSION            Min. VXVM version
=====
```

If the command output does not show that the required ASL is already installed, you will need to locate the installation package. The installation package for the ASL is available at:

<https://vos.symantec.com/asl>

You will need to specify the vendor of your storage system, your operating system, and the version of your Symantec Storage Foundation. After you have specified that information, you will be redirected to a web page from where you can download the appropriate ASL package for your environment and installation instructions.

Proceed with the ASL installation according to the instructions. Example 7-2 illustrates the ASL installation for the Symantec Storage Foundation version 5.0 on Solaris version 10, on a SPARC server.

Example 7-2 Installing ASL for the IBM Storage System

```
# vxdctl mode
mode: enabled
# cd /export/home
# pkgadd -d .
```

The following packages are available:

```
1  VRTSibmxiv      Array Support Library for IBM xiv and XIV Nextra
                        (sparc) 1.0,REV=09.03.2008.11.56
```

Select package(s) you wish to process (or 'all' to process all packages). (default: all) [?,??,q]: 1

Processing package instance <VRTSibmxiv> from </export/home>

Array Support Library for IBM xiv and XIV Nextra(sparc) 1.0,REV=09.03.2008.11.56
Copyright L 1990-2006 Symantec Corporation. All rights reserved.

Symantec and the Symantec Logo are trademarks or registered trademarks of Symantec Corporation or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.

The Licensed Software and Documentation are deemed to be "commercial computer software" and "commercial computer software documentation" as defined in FAR Sections 12.212 and DFARS Section 227.7202.

Using </etc/vx> as the package base directory.

Processing package information.

Processing system information.

3 package pathnames are already properly installed.

Verifying disk space requirements.

Checking for conflicts with packages already installed.

Checking for setuid/setgid programs.

This package contains scripts which will be executed with super-user permission during the process of installing this package.

```

Do you want to continue with the installation of <VRTSibmxiv> [y,n,?] y

Installing Array Support Library for IBM xiv and XIV Nextra as <VRTSibmxiv>

## Installing part 1 of 1.
/etc/vx/aslkey.d/libvxxiv.key.2
/etc/vx/lib/discovery.d/libvxxiv.so.2
[ verifying class <none> ]
## Executing postinstall script.
Adding the entry in supported arrays
Loading The Library

```

```

Installation of <VRTSibmxiv> was successful.
# vxddladm listversion
LIB_NAME                      ASL_VERSION          Min. VXVM version
=====
libvxxiv.so                   vm-5.0-rev-2         5.0
# vxddladm listsupport
LIBNAME          VID          PID
=====
libvxxiv.so      XIV, IBM      NEXTRA, 2810XIV

```

At this stage, you are ready to install the required XIV Host Attachment Kit (HAK) for your platform. You can check the HAK availability for your platform at this URL:

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

Proceed with the XIV HAK installation process.

If the XIV HAK is not available for your platform, you need to define your host on the XIV system and map the LUNs to the hosts. For the details on how to define hosts and map the LUNs refer to Chapter 1, “Host connectivity” on page 1.

7.2.2 Installing the XIV Host Attachment Kit

When available for an XIV supported host platform, installing the corresponding XIV Host Attachment Kit is required for support.

To install the HAK in our Solaris/SPARC experimentation scenario, open a terminal session and go to the directory where the package was downloaded.

To extract files from the archive, execute the command shown in Example 7-3.

Example 7-3 Extracting the HAK

```
# gunzip -c XIV_host_attach-<version>-<os>-<arch>.tar.gz | tar xvf -
```

Change to the newly created directory and invoke the Host Attachment Kit installer, as seen in Example 7-4.

Example 7-4 Starting the installation

```
# cd XIV_host_attach-<version>-<os>-<arch>
# /bin/sh ./install.sh
```

Follow the prompts. After running the installation script, review the installation log file `install.log` residing in the same directory.

Configuring the host

Use the utilities provided in the Host Attachment Kit to configure the host. The Host Attachment Kit packages are installed in `/opt/xiv/host_attach` directory.

Note: You must be logged in as root or have root privileges to use the Host Attachment Kit.

Execute the `xiv_attach` utility as shown in Example 7-5. The command can also be launched from any working directory.

Example 7-5 Launch xiv_attach

```
/opt/xiv/host_attach/bin/xiv_attach
```

```
-----
Welcome to the XIV Host Attachment wizard, version 1.5.2.
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 : fc
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...
-----
```

```
A reboot is required in order to continue.
Please reboot the machine and restart the wizard
```

```
Press [ENTER] to exit.
```

At this stage, for the Solaris on SUN server as used in our example, you must reboot the host before proceeding to the next step. After the system reboot, start `xiv_attach` again to complete the host system configuration for XIV attachment as shown in Example 7-6 on page 162.

Example 7-6 Fibre Channel host attachment configuration after reboot

xiv_attach

Welcome to the XIV Host Attachment wizard, version 1.5.2.
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 : f

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:
210000e08b0c4f10: /dev/cfg/c2: [QLogic Corp.]: QLA2340
210000e08b137f47: /dev/cfg/c3: [QLogic Corp.]: QLA2340
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)
6000105	10.2	No	None	FC	--
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: sun-v480R-tic-1]: sun-s1e-1
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 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.

Now you can map your XIV volumes (LUNs) to the host system. You can use the XIV GUI for that task, as illustrated in 1.4, “Logical configuration for host connectivity” on page 29.

After the LUN mapping is completed, you need to discover the mapped LUNs on your host by executing the command **xiv_cf_admin -R**.

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. Refer to Example 7-7 on page 163.

Example 7-7 Showing mapped volumes and available paths

```
# xiv_devlist -x
XIV Devices
```

Device	Size	Paths	Vol Name	Vol Id	XIV Id	XIV Host
/dev/vx/dmp/xiv0_0	17.2GB	4/4	itso_vol_2	4341	1300203	sun-v480
/dev/vx/dmp/xiv1_0	17.2GB	2/2	itso_vol_1	4462	6000105	sun-v480

7.2.3 Placing XIV LUNs under VxVM control

To place XIV LUNs under VxVM control you need discover new devices on your hosts using the **vxdiskconfig** command, or you can use command **vxctl -f enable** and then check for new devices discovered by executing the **vxdisk list** command as illustrated in Example 7-8.

Example 7-8 Discovering and checking new disks on your host

```
# vxctl -f enable
# vxdisk -f scandisks
# vxdisk list
```

DEVICE	TYPE	DISK	GROUP	STATUS
clt0d0s2	auto:none	-	-	online invalid
clt1d0s2	auto:none	-	-	online invalid
xiv0_0	auto	-	-	no label
xiv1_0	auto	-	-	no label

After you have discovered the new disks on the host, you might need to format the disks: refer to your OS specific Symantec Storage Foundation documentation. In our example, we need to format the disks.

Run the **vxdiskadm** command as shown in Example 7-9. 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 7-9 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,?] list

DEVICE	DISK	GROUP	STATUS
clt0d0	-	-	online invalid
clt1d0	-	-	online invalid
xiv0_0	vgxiv02	vgxiv	online
xiv0_1	-	-	no label
xiv1_0	vgxiv01	vgxiv	online

Select disk devices to add: [<pattern-list>,all,list,q,?] xiv0_1

Here is the disk selected. Output format: [Device_Name]

xiv0_1

Continue operation? [y,n,q,?] (default: y)

You can choose to add this disk to an existing disk group, a new disk group, or leave the disk 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 disk available for future use, specify a disk group name of "none".

Which disk group [<group>,none,list,q,?] (default: none) vxiv

Use a default disk name for the disk? [y,n,q,?] (default: y)

Add disk as a spare disk for vxiv? [y,n,q,?] (default: n)

Exclude disk from hot-relocation use? [y,n,q,?] (default: n)

Add site tag to disk? [y,n,q,?] (default: n)

The selected disks will be added to the disk group vxiv with default disk names.

xiv0_1

Continue with operation? [y,n,q,?] (default: y)

The following disk device has a valid VTOC, but does not appear to have been initialized for the Volume Manager. If there is data on the disk that should NOT be destroyed you should encapsulate the existing disk partitions as volumes instead of adding the disk as a new disk.
Output format: [Device_Name]

xiv0_1

Encapsulate this device? [y,n,q,?] (default: y) n

xiv0_1

Instead of encapsulating, initialize? [y,n,q,?] (default: n) y

Initializing device xiv0_1.

Enter desired private region length

[<privlen>,q,?] (default: 65536)

VxVM NOTICE V-5-2-88

Adding disk device xiv0_1 to disk group vxiv with disk name vxiv03.

Add or initialize other disks? [y,n,q,?] (default: n)

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: q

Goodbye.

When you done putting XIV LUNs under VxVM control, check the results of your work using the **vxdisk list** command as shown in Example 7-10.

Example 7-10 Showing the results on putting XI V LUN's under VxVM control

```

# vxdisk list
DEVICE      TYPE          DISK          GROUP          STATUS
c1t0d0s2    auto:none      -             -              online invalid
c1t1d0s2    auto:none      -             -              online invalid
xiv0_0      auto:cdsdisk   vxxiv02       vxxiv          online
xiv0_1      auto:cdsdisk   vxxivthin01   vxxivthin      online
xiv1_0      auto:cdsdisk   vxxiv01       vxxiv          online

# vxdg list
NAME        STATE          ID
vxxiv       enabled,cds    1287499674.11.sun-v480R-tic-1
vxxivthin   enabled,cds    1287500956.17.sun-v480R-tic-1

# vxdg list vxxiv
Group:      vxxiv
dgid:       1287499674.11.sun-v480R-tic-1
import-id:  1024.10
flags:      cds
version:    150
alignment:  8192 (bytes)
ssb:        on
autotagging: on
detach-policy: global
dg-fail-policy: dgdisable
copies:     nconfig=default nlog=default
config:     seqno=0.1061 permilen=48144 free=48138 templen=3 loglen=7296
config disk xiv0_0 copy 1 len=48144 state=clean online
config disk xiv1_0 copy 1 len=48144 state=clean online
log disk xiv0_0 copy 1 len=7296

```



```
log disk xiv1_0 copy 1 len=7296
```

The XIV LUNs that were just added are now available for volume creation and data storage.

Check that you get adequate performance, and if required configure DMP multipathing settings.

7.2.4 Configure multipathing with DMP

The Symantec Storage Foundation version 5.0 used MinimumQ *iopolicy* by default for the enclosures on an Active/Active storage systems. The best practices when attaching hosts to the IBM XIV Storage System recommend setting the *iopolicy* parameter to round-robin and enabling the use of all paths.

1. Identify names of enclosures that reside on the XIV Storage System. Log on to the host as root user, execute the **vxddmpadm listenclosure all** command, and examine the results to determine which enclosure names belong to an XIV Storage System. As illustrated in Example 7-11, in our scenario we can identify the enclosure names as xiv0 and xiv1.

Example 7-11 Identifying names of enclosures which are seated on an IBM XIV Storage System

#vxddmpadm listenclosure all					
ENCLR_NAME	ENCLR_TYPE	ENCLR_SNO	STATUS	ARRAY_TYPE	LUN_COUNT
=====					
xiv0	XIV	00CB	CONNECTED	A/A	2
disk	Disk	DISKS	CONNECTED	Disk	2
xiv1	XIV	0069	CONNECTED	A/A	1

2. The next step is to change the *iopolicy* parameter for the identified enclosures by executing the command **vxddmpadm setattr enclosure <identified enclosure name> iopolicy=round-robin** for each identified enclosure.
3. Check the results of the change by executing the command **vxddmpadm getattr enclosure <identified enclosure name>** as shown in Example 7-12.

Example 7-12 Changing DMP settings on 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	partitionsize	512	512
xiv0	use_all_paths	-	-
xiv0	failover_policy	Global	Global
xiv0	recoveryoption[throttle]	Nothrottle[0]	Timebound[10]
xiv0	recoveryoption[errorretry]	Fixed-Retry[5]	Fixed-Retry[5]
xiv0	redundancy	0	0
xiv0	failovermode	explicit	explicit

4. For heavy workloads, increase the queue depth parameter up to 64 or 128. You can do this by executing the command **vxddmpadm gettune dmp_queue_depth** to get information about current settings and if required execute **vxddmpadm settune dmp_queue_depth=<new queue depth value>** to adjust the settings as shown in Example 7-13 on page 168.

Example 7-13 Changing queue depth parameter

```
# vxdmpadm gettune dmp_queue_depth
```

Tunable	Current Value	Default Value
dmp_queue_depth	32	32

```
# vxdmpadm settune dmp_queue_depth=96
```

Tunable value will be changed immediately

```
# vxdmpadm gettune dmp_queue_depth
```

Tunable	Current Value	Default Value
dmp_queue_depth	96	32

7.3 Working with snapshots

Version 5.0 of Symantec Storage Foundation introduced a new functionality 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 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 known to VxVM and is compared to the UDID that is stored in the disk's private region. If the UDID does not match, the `udid_mismatch` flag is set on the disk. This 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, which can then be used for the offline backup or processing.

After creating a snapshot for LUNs used on a host under VxVM control, you need, in XIV, to enable writing on the snapshots and map them to your host. When this process is complete, the snapshot LUNS can be imported on the host.

1. Check that the created snapshots are visible for your host by executing the commands **vxdctl enable** and **vxdisk list** as shown in Example 7-14.

Example 7-14 Identifying created snapshots on host side

```
#vxdctl enable
```

```
#vxdisk list
```

DEVICE	TYPE	DISK	GROUP	STATUS
disk_0	auto:none	-	-	online invalid
disk_1	auto:none	-	-	online invalid
xiv0_0	auto:cdsdisk	vgxiv02	vgxiv	online
xiv0_4	auto:cdsdisk	vgsnap01	vgsnap	online
xiv0_5	auto:cdsdisk	vgsnap02	vgsnap	online
xiv0_6	auto:cdsdisk	-	-	online udid_mismatch
xiv0_7	auto:cdsdisk	-	-	online udid_mismatch
xiv1_0	auto:cdsdisk	vgxiv01	vgxiv	online

2. Import the created snapshot on your host by executing the command **vxvg -n <name for new volume group> -o useclonedev=on,updateid -C import <name of original volume group>**, and then executing the **vxdisk list** command to ensure that the LUNs were imported. See Example 7-15

Example 7-15 Import snapshots on to your host

```
# vxvg -n vgsnap2 -o useclonedev=on,updateid -C import vgsnap
```

VxVM vxvg WARNING V-5-1-1328 Volume 1vol: Temporarily renumbered due to conflict

vxdisk list

DEVICE	TYPE	DISK	GROUP	STATUS
disk_0	auto:none	-	-	online invalid
disk_1	auto:none	-	-	online invalid
xiv0_0	auto:cdsdisk	vgxiv02	vgxiv	online
xiv0_4	auto:cdsdisk	vgsnap01	vgsnap	online
xiv0_5	auto:cdsdisk	vgsnap02	vgsnap	online
xiv0_6	auto:cdsdisk	vgsnap02	vgsnap2	online clone_disk
xiv0_7	auto:cdsdisk	vgsnap01	vgsnap2	online clone_disk
xiv1_0	auto:cdsdisk	vgxiv01	vgxiv	online

Now you ready to use XIV snapshots on your host.



VIOS clients connectivity

This chapter explains XIV connectivity to Virtual I/O Server (VIOS) clients, including AIX, Linux on Power and in particular, IBM i. VIOS is a component of Power VM that provides the ability for LPARs (VIOS clients) to share resources.

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

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

and the Host Attachment publications at:

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

8.1 Introduction to IBM PowerVM

Virtualization on IBM Power Systems servers can provide a rapid and cost-effective response to many business needs. Virtualization capabilities are becoming an important element in planning for IT floor space and servers. With growing commercial and environmental concerns, there is pressure to reduce the power footprint of servers. Also, with the escalating cost of powering and cooling servers, consolidation and efficient utilization of the servers is becoming critical.

Virtualization on Power Systems servers enables an efficient utilization of servers by reducing the following:

- ▶ 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 special software appliance tied to IBM Power Systems, that is the converged IBM i and IBM p server platforms. It is licensed on a Power-Systems-processor basis. PowerVM is a virtualization technology for AIX, IBM i, and Linux environments on IBM POWER processor-based systems.

PowerVM offers a secure virtualization environment with the following major 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

They provide logical partitioning technology by using either the Hardware Management Console (HMC) or the Integrated Virtualization Manager (IVM), dynamic logical partition (LPAR) operations, Micro-Partitioning™ and VIOS capabilities, and Node 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, clients can create up to three partitions on a server (two client partitions and one for the VIOS and IVM). They can use virtualized disk and optical devices, and try the shared processor pool. All virtualization features, such as

Micro-Partitioning, shared processor pool, VIOS, PowerVM LX86, shared dedicated capacity, NPIV, and virtual tape, can be managed by using the IVM.

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, such as Micro-Partitioning, Shared Processor Pool, Virtual I/O Server, PowerVM Lx86, Shared Dedicated Capacity, NPIV, and Virtual Tape, can be managed by using an Hardware Management Console or the IVM.

PowerVM Enterprise Edition

PowerVM Enterprise Edition is offered exclusively on IBM POWER6® 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 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.

The VIOS owns the physical I/O resources such as Ethernet and SCSI/FC adapters. It virtualizes those resources for its client LPARs to share them remotely by using the built-in hypervisor services. These client LPARs can be created quickly, typically owning only real memory and shares of CPUs 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 is used to make storage devices, such as the IBM XIV Storage System server, that do not support the IBM i proprietary 520-byte/sectors format that is available to IBM i clients of VIOS.

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 that are created as part of its partition profile.

The client partition with its corresponding VSCSI client adapters defined in its partition profile connects to the VIOS VSCSI server adapters by using the hypervisor. VIOS performs 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.

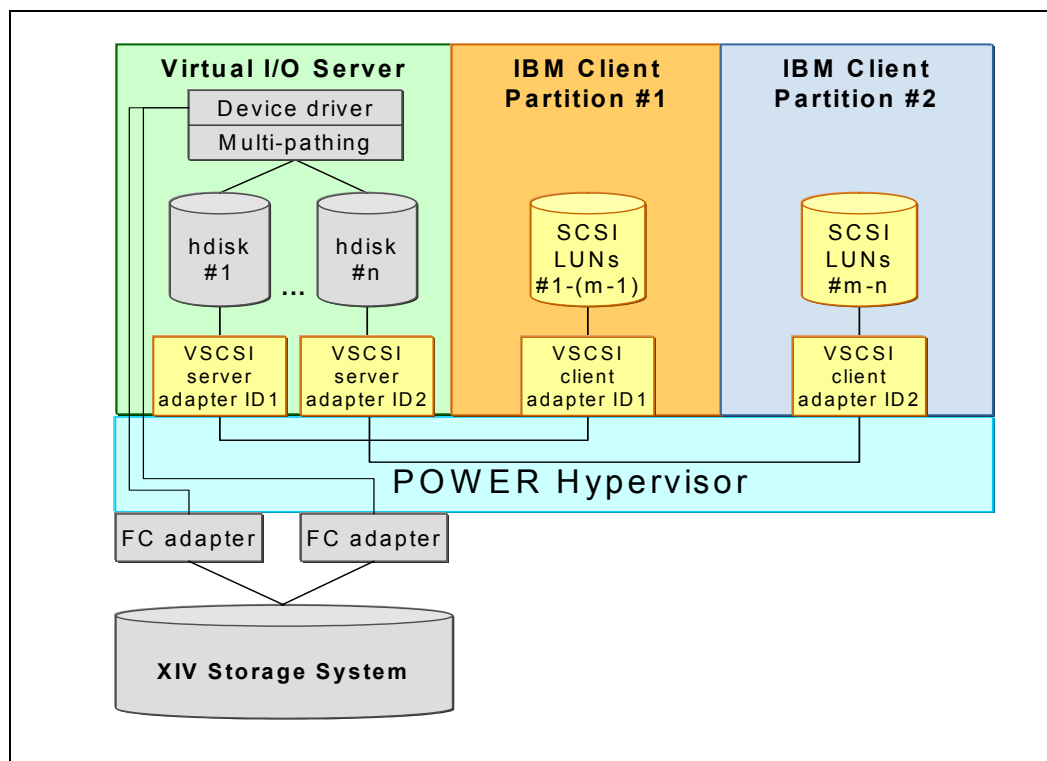


Figure 8-1 VIOS virtual SCSI support

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 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 by using NPIV. Thus no devices are shown in the VIOS connected to NPIV adapters. The discovery is left for the virtual client and all the devices 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.

Note: 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, such as the IBM System Storage DS8000. To connect other storage devices, 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.

Note: Although PowerVM and VIOS themselves are supported on both POWER5™ and POWER6 systems, IBM i, being a client of VIOS, is supported only on POWER6 systems.

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

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

and the Host Attachment publications at:

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

Requirements

When connecting the IBM XIV Storage System server to an IBM i operating system by using the Virtual I/O Server (VIOS), you must have the following requirements in place:

- ▶ The IBM i partition and the VIOS partition must reside on a POWER6 processor-based IBM Power Systems server or POWER6 processor-based IBM Power Blade servers.

POWER6 processor-based IBM Power Systems include the following models:

- 8203-E4A (IBM Power 520 Express)
- 8261-E4S (IBM Smart Cube)
- 9406-MMA
- 9407-M15 (IBM Power 520 Express)
- 9408-M25 (IBM Power 520 Express)
- 8204-E8A (IBM Power 550 Express)
- 9409-M50 (IBM Power 550 Express)
- 8234-EMA (IBM Power 560 Express)
- 9117-MMA (IBM Power 570)
- 9125-F2A (IBM Power 575)
- 9119-FHA (IBM Power 595)

The following servers are POWER6 processor-based IBM Power Blade servers:

- IBM BladeCenter JS12 Express
- IBM BladeCenter JS22 Express
- IBM BladeCenter JS23
- IBM BladeCenter JS43

- ▶ You must have one of the following PowerVM editions:
 - PowerVM Express Edition, 5765-PVX
 - PowerVM Standard Edition, 5765-PVS
 - PowerVM Enterprise Edition, 5765-PVE
- ▶ You must have Virtual I/O Server Version 2.1.1 or later.
Virtual I/O Server is delivered as part of PowerVM.
- ▶ You must have IBM i, 5761-SS1, Release 6.1 or later.

- ▶ You must have one of the following Fibre Channel (FC) adapters supported to connect the XIV system to the VIOS partition in POWER6 processor-based IBM Power Systems server:
 - 2 Gbps PCI-X 1-port Fibre Channel adapter, feature number 1957
 - 2 Gbps PCI-X 1-port Fibre Channel adapter, feature number 1977
 - 2 Gbps PCI-X 1-port Fibre Channel adapter, feature number 5716
 - 2 Gbps PCI-X Fibre Channel adapter, feature number 6239
 - 4 Gbps PCI-X 1-port Fibre Channel adapter, feature number 5758
 - 4 Gbps PCI-X 2-port Fibre Channel adapter, feature number 5759
 - 4 Gbps PCIe 1-port Fibre Channel adapter, feature number 5773
 - 4 Gbps PCIe 2-port Fibre Channel adapter, feature number 5774
 - 4 Gbps PCI-X 1-port Fibre Channel adapter, feature number 1905
 - 4 Gbps PCI-X 2-port Fibre Channel adapter, feature number 1910
 - 8 Gbps PCIe 2-port Fibre Channel adapter, feature number 5735

Note: Not all listed Fibre Channel adapters are supported in every POWER6 server listed in the first point. For more information about which FC adapter is supported with which server, see *IBM Power 520 and Power 550 (POWER6) System Builder*, SG24-7765, and the IBM Redpaper™ publication *IBM Power 570 and IBM Power 595 (POWER6) System Builder*, REDP-4439.

The following Fibre Channel host bus adapters (HBAs) are supported to connect the XIV system to a VIOS partition on IBM Power Blade servers JS12 and JS22:

- LP1105-BCv - 4 Gbps Fibre PCI-X Fibre Channel Host Bus Adapter, P/N 43W6859
- IBM SANblade QMI3472 PCIe Fibre Channel Host Bus Adapter, P/N 39Y9306
- IBM 4 Gb PCI-X Fibre Channel Host Bus Adapter, P/N 41Y8527

The following Fibre Channel HBAs are supported to connect the XIV system to a VIOS partition on IBM Power Blade servers JS23 and JS43:

- IBM SANblade QMI3472 PCIe Fibre Channel Host Bus Adapter, P/N 39Y9306
- IBM 44X1940 QLOGIC ENET & 8 Gbps Fibre Channel Expansion Card for BladeCenter
- IBM 44X1945 QMI3572 QLOGIC 8 Gbps Fibre Channel Expansion Card for BladeCenter
- IBM 46M6065 QMI2572 QLogic 4 Gbps Fibre Channel Expansion Card for BladeCenter
- IBM 46M6140 Emulex 8 Gb Fibre Channel Expansion Card for BladeCenter
- IBM SANblade QMI3472 PCIe Fibre Channel Host Bus Adapter, P/N 39Y9306

- ▶ You must have IBM XIV Storage System firmware 10.0.1b or later.

Supported SAN switches

For the list of supported SAN switches when connecting the XIV system to the IBM i operating system, see the System Storage Interoperation Center at the following address:

http://www-03.ibm.com/systems/support/storage/config/ssic/displaysssearchwithoutjs.wss?start_over=yes

Physical Fibre Channel adapters and virtual SCSI adapters

It is possible to connect up to 4,095 logical unit numbers (LUNs) per target and up to 510 targets per port on a VIOS physical FC adapter. Because you can assign up to 16 LUNs to one virtual SCSI (VSCSI) adapter, you can use the number of LUNs to determine the number of virtual adapters that you need.

Note: When the IBM i operating system and VIOS reside on an IBM Power Blade server, you can define only one VSCSI adapter in the VIOS to assign to an IBM i client. Consequently the number of LUNs to connect to the IBM i operating system is limited to 16.

Queue depth in the IBM i operating system and Virtual I/O Server

When connecting 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
SCSI command tag queuing in the IBM i operating system enables 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 given 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 a given 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 the type of connected storage, type of physical adapter, and type of multipath driver or Host Attachment kit that is used. Typically for the XIV system, the queue depth per physical disk is 32, the queue depth per 4 Gbps FC adapter is 200, and 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 by using the following command:

```
chdev -dev hdiskxx -attr queue_depth=32
```

This command ensures that the queue depth in the VIOS matches the IBM i queue depth for an XIV LUN.

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, you connect the XIV system to an IBM i client with two or more VIOS partitions, with one VSCSI adapter in the IBM i operating system assigned 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 utilize multipath by using two VIOS partitions.

See 8.3.3, “IBM i multipath capability with two Virtual I/O Servers” on page 182, for more information.

8.2.1 Best practices

In this section, we present general best practices 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 recommended approach is to maximize the utilization of all the XIV resources at all times.

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. It is *not* necessary for each host instance to connect to each interface module. However, when the host has more than one physical connection, it is beneficial to 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 such that each host adapter connects to each XIV interface module and through each SAN switch. Refer to 1.2.2, “FC configurations” on page 10 and 1.2.3, “Zoning” on page 12.

Note: Create a separate zone for each host adapter that connects to a switch, for each zone that contains the connection from the host adapter, and for all connections to the XIV system.

Queue depth

SCSI command tag queuing for LUNs on the IBM XIV Storage System server enables 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 existing 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 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/Os. The multi-threaded applications will 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. In the following sections, we explain how to set up the environment on a POWER6 system to connect the XIV system to an IBM i client with multipath through two VIOS partitions.

8.3.1 Creating the Virtual I/O Server and IBM i partitions

In this section, we describe the steps to create a VIOS partition and an IBM i partition through the POWER6 Hardware Management Console (HMC). We also explain how to create VSCSI adapters in the VIOS and the IBM i partition and how to assign them so that the IBM i partition can work as a client of the VIOS.

For more information about how to create the VIOS and IBM i client partitions in the POWER6 server, 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 logical partition (LPAR) for VIOS:

1. Insert the PowerVM activation code in the HMC if you have not already done so. Click **Tasks** → **Capacity on Demand (CoD)** → **Advanced POWER Virtualization** → **Enter Activation Code**.
2. Create the new partition. In the left pane, select **Systems Management** → **Servers**. In the right pane, select the server to use for creating the new VIOS partition. Then click **Tasks** → **Configuration** → **Create Logical Partition** → **VIO Server** (see Figure 8-2).

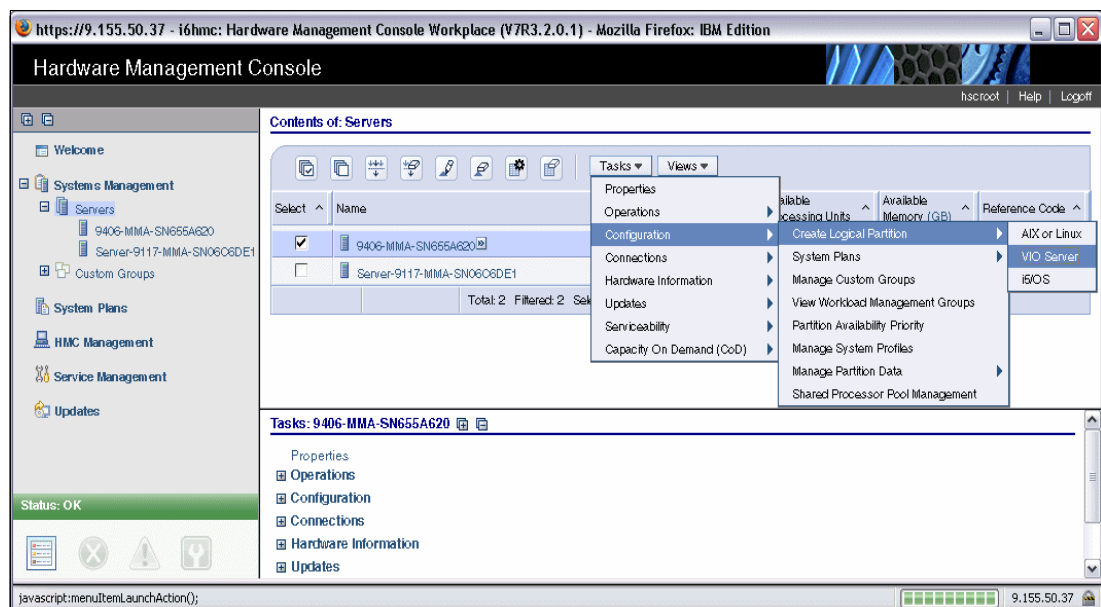


Figure 8-2 Creating the VIOS partition

3. In the Create LPAR wizard:
 - a. Type the partition ID and name.
 - b. Type the partition profile name.
 - c. Select whether the processors in the LPAR will be dedicated or shared. We recommend that you select **Dedicated**.
 - d. Specify the minimum, desired, and maximum number of processors for the partition.
 - e. Specify the minimum, desired, and maximum amount of memory in the partition.
4. In the I/O panel (see Figure 8-3), select the I/O devices to include in the new LPAR. In our example, we included the RAID controller to attach the internal SAS drive for the VIOS boot disk and DVD_RAM drive. We include the physical Fibre Channel (FC) adapters to connect to the XIV server. As shown in Figure 8-3, we add them as **Required**.

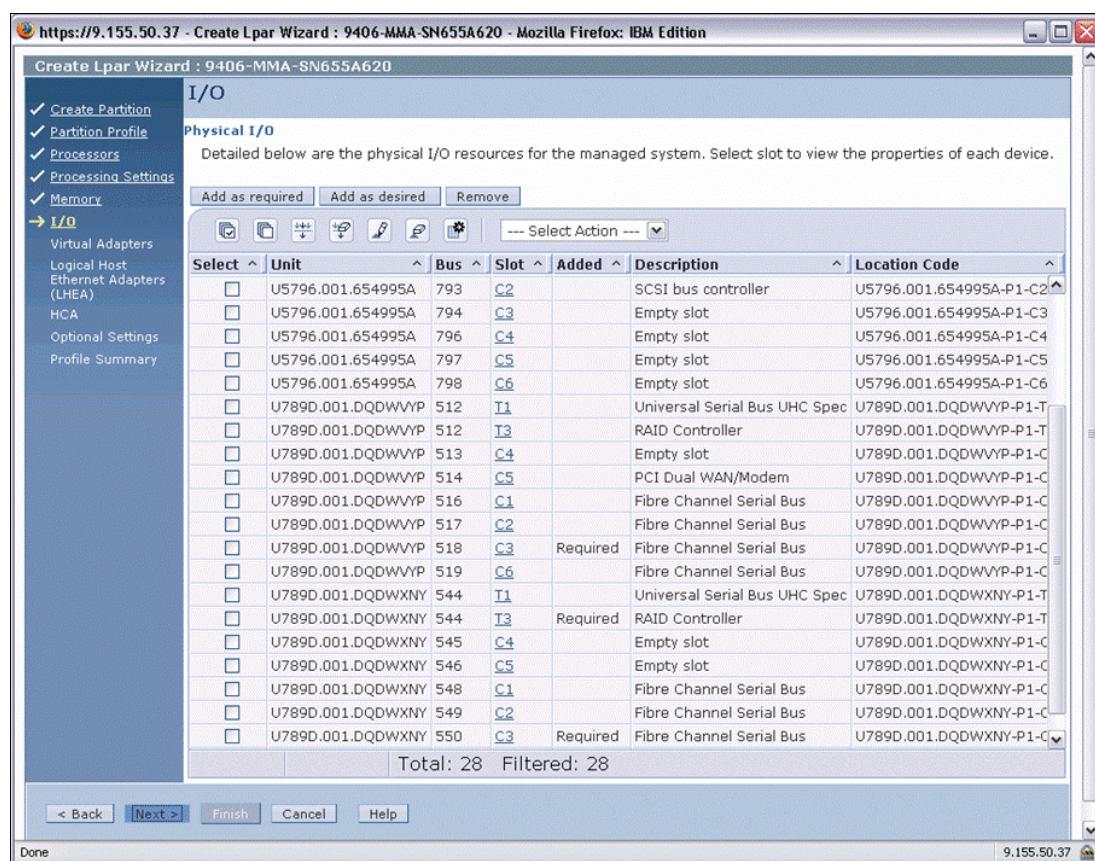


Figure 8-3 Adding the I/O devices to the VIOS partition

5. In the Virtual Adapters panel, create an Ethernet adapter by selecting **Actions** → **Create** → **Ethernet Adapter**. Mark it as **Required**.
6. Create the VSCSI adapters that will be assigned to the virtual adapters in the IBM i client:
 - a. Select **Actions** → **Create** → **SCSI Adapter**.
 - b. In the next window, either leave the **Any Client partition can connect** selected or limit the adapter to a particular client.

If DVD-RAM will be virtualized to the IBM i client, you might want to create another VSCSI adapter for DVD-RAM.

7. Configure the logical host Ethernet adapter:
 - a. Select the logical host Ethernet adapter from the list.
 - b. In the next window, click **Configure**.
 - c. Verify that the selected logical host Ethernet adapter is not selected by any other partitions, and select **Allow all VLAN IDs**.
8. In the Profile Summary panel, review the information, and click **Finish** to create the LPAR.

Creating an IBM i partition in the POWER6 server

To create an IBM i partition that will be the VIOS client, follow these steps:

1. From the HMC, select **Systems Management** → **Servers**. In the right panel, select the server in which you want to create the partition. Then select **Tasks** → **Configuration** → **Create Logical Partition** → **i5/OS**.
2. In the Create LPAR wizard:
 - a. Insert the Partition ID and name.
 - b. Insert partition Profile name.
 - c. Select whether the processors in the LPAR will be dedicated or shared. We recommend that you select **Dedicated**.
 - d. Specify the minimum, desired, and maximum number of processors for the partition.
 - e. Specify the minimum, desired, and maximum amount of memory in the partition.
 - f. In the I/O panel, if the IBM i client partition is not supposed to own any physical I/O hardware, click **Next**.
3. In the Virtual Adapters panel, select **Actions** → **Create** → **Ethernet Adapter** to create a virtual Ethernet adapter.
4. In the Create Virtual Ethernet Adapter panel, accept the suggested adapter ID and the VLAN ID. Select **This adapter is required for partition activation** and click **OK** to continue.
5. In the Virtual Adapters panel, select **Actions** → **Create** → **SCSI Adapter** to create the VSCSI client adapters on the IBM i client partition that is used for connecting to the corresponding VIOS.
6. 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, you can repeat this step to create another VSCSI client adapter to connect to the VIOS VSCSI server adapter that is used for virtualizing the DVD-RAM.
7. Configure the logical host Ethernet adapter:
 - a. Select the logical host Ethernet adapter from the drop-down list and click **Configure**.
 - b. In the next panel, ensure that no other partitions have selected the adapter and select **Allow all VLAN IDs**.
8. In the OptiConnect Settings panel, if OptiConnect is not used in IBM i, click **Next**.

9. In the Load Source Device panel, if the connected XIV system will be used to boot from a storage area network (SAN), select the virtual adapter that connects to the VIOS.

Note: The IBM i Load Source device resides on an XIV volume.

10. In the Alternate Restart Device panel, if the virtual DVD-RAM device will be used in the IBM i client, select the corresponding virtual adapter.
11. In the Console Selection panel, select the default of HMC for the console device. Click **OK**.
12. Depending on the planned configuration, click **Next** in the three panels that follow until you reach the Profile Summary panel.
13. In the Profile Summary panel, check the specified configuration and click **Finish** to create the IBM i LPAR.

8.3.2 Installing the Virtual I/O Server

For information about how to install the VIOS in a partition of the POWER6 server, see the Redbooks publication *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). In our example, we mirror it across two RAID0 arrays of hdisk0 and hdisk1 to help protect the VIOS from potential CEC internal SAS disk drive failures.

Configuring Virtual I/O Server network connectivity

To set up network connectivity in the VIOS:

1. In the HMC terminal window, logged in as padmin, enter the following command:

```
lsdev -type adapter | grep ent
```

Look for the logical host Ethernet adapter resources. In our 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 syntax and specifying 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, it is possible to establish multipath to a set of LUNs, with each path using a connection through a separate VIOS. This topology provides redundancy in case either 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 will probably establish no more than two such connections.

8.3.4 Virtual SCSI adapters in multipath with two Virtual I/O Servers

In our setup, we use two VIOS and two VSCSI adapters in the IBM i partition, where each adapter is assigned to a virtual adapter in one VIOS. We connect the same set of XIV LUNs to each VIOS through two physical FC adapters in the VIOS multipath and map them to VSCSI adapters serving IBM i partition. This way, the IBM i partition sees the LUNs through two paths, each path using one VIOS. Figure 8-5 shows our setup.

For our testing, we did not use separate switches as shown in Figure 8-5, but rather used separate blades in the same SAN Director. In a real, production environment, use separate switches as shown in Figure 8-5.

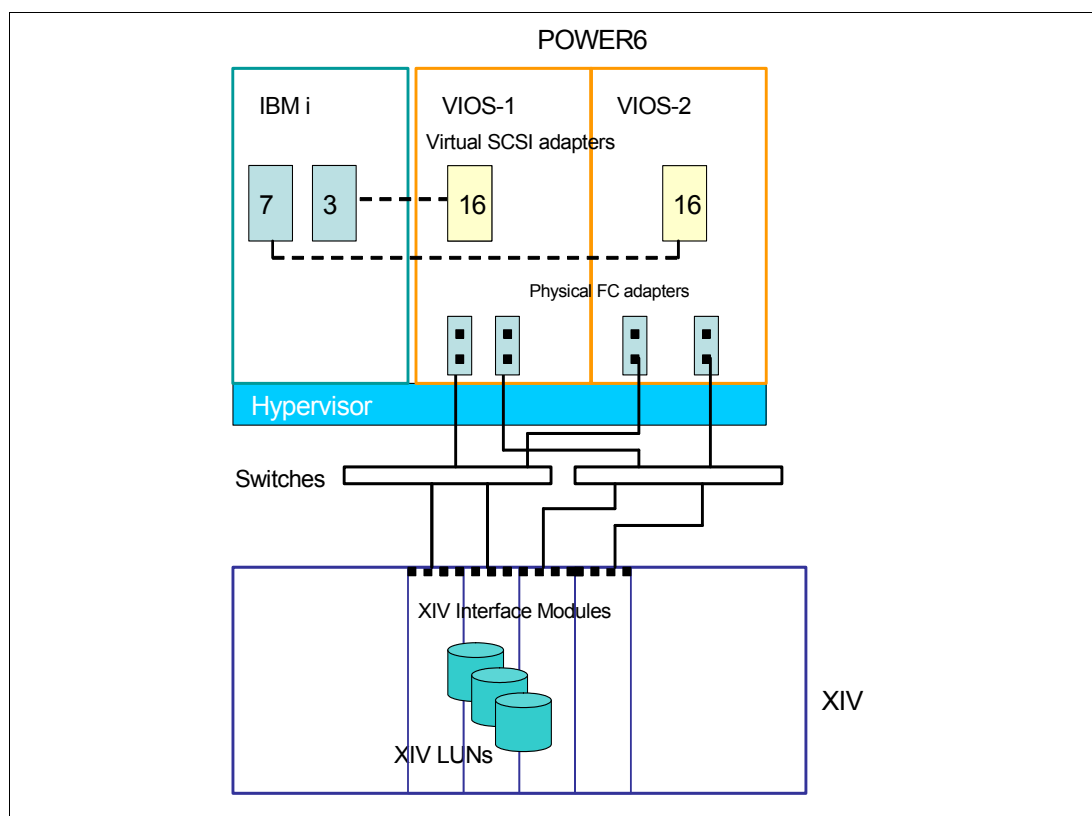


Figure 8-5 Setup for multipath with two VIOS

To connect XIV LUNs to an IBM i client partition in multipath with two VIOS:

Important: Perform steps 1 through 5 in each of the two VIOS partitions.

1. After the LUNs are created in the XIV system, use the XIV Storage Management GUI or Extended Command Line Interface (XCLI) to map the LUNs to the VIOS host as shown in 8.4, “Mapping XIV volumes in the Virtual I/O Server” on page 184.
2. Log in to VIOS as administrator. In our example, we use PUTTY to log in as described in 6.5, “Configuring VIOS virtual devices,” in *IBM i and Midrange External Storage*, SG24-7668.

Type the **cfgdev** command so that the VIOS can recognize newly attached LUNs.

3. In the VIOS, remove the SCSI reservation attribute from the LUNs (hdisks) that will be connected through two VIOS by entering the following command for each hdisk that will connect to the IBM i operating system in multipath:

```
chdev -dev hdiskX -attr reserve_policy=no_reserve
```

4. Set the attributes of Fibre Channel adapters in the VIOS to **fc_err_recov=fast_fail** and **dyntrk=yes**. When the attributes are set to these values, the error handling in FC adapter allows faster transfer to the alternate paths in case of problems with one FC path. To make multipath within one VIOS work more efficiently, specify these values by entering the following command:

```
chdev -dev fscsi0 -attr fc_err_recov=fast_fail dyntrk=yes-perm
```

5. To get more bandwidth by using multiple paths, enter the following command for each hdisk (hdiskX):

```
chdev -dev hdiskX -perm -attr algorithm=round_robin
```

6. Map the disks that correspond to the XIV LUNs to the VSCSI adapters that are assigned to the IBM i client. First, check the IDs of assigned virtual adapters. Then complete the following steps:
 - a. 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.
 - b. 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 **lsmap -all** to view the virtual adapters.
 - c. 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, perform the following tasks in each VIOS:

1. Connect to the VIOS partition. In our example, we use a PUTTY session to connect.
2. In the VIOS, enter the **cfgdev** command to discover the newly added XIV LUNs, which makes the LUNs available as disk devices (hdisks) in the VIOS. In our example, the LUNs that we added correspond to hdisk132 - hdisk140, as shown in Figure 8-6 on page 185.

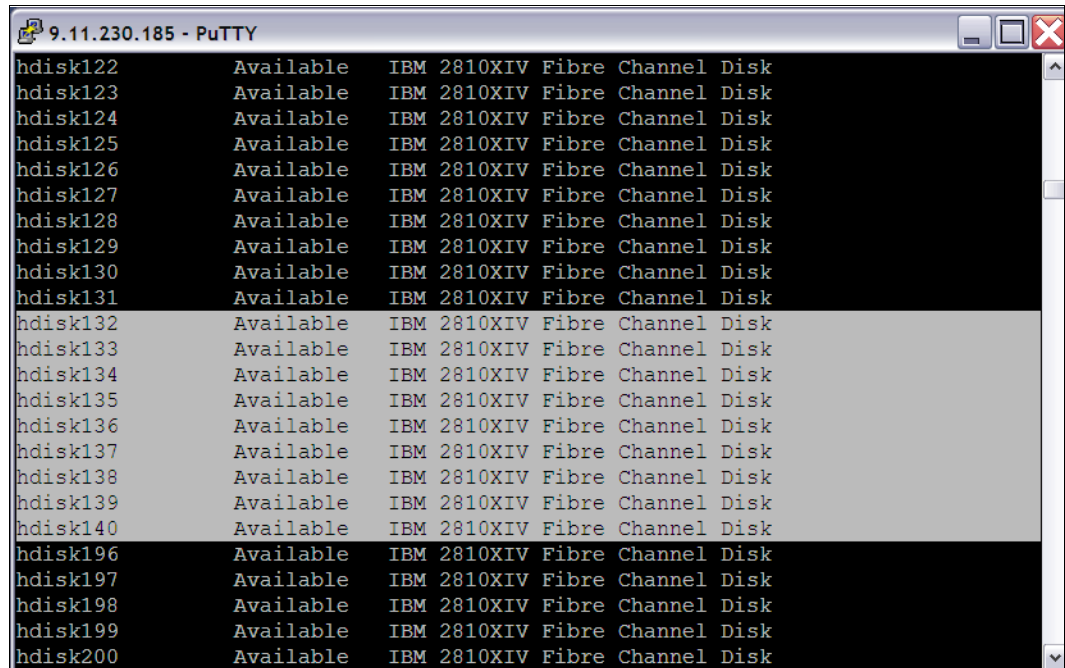


Figure 8-6 The hdisks of the added XIV volumes

As we previously explained, to realize a multipath setup for IBM i, we connected (mapped) each XIV LUN to both VIOS partitions. Before assigning 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 and whether any other devices are mapped to it.
 - a. Enter the following command to display the virtual slot of the adapter and see if any other devices are assigned to it:

```
lsmap -vadapter <name>
```

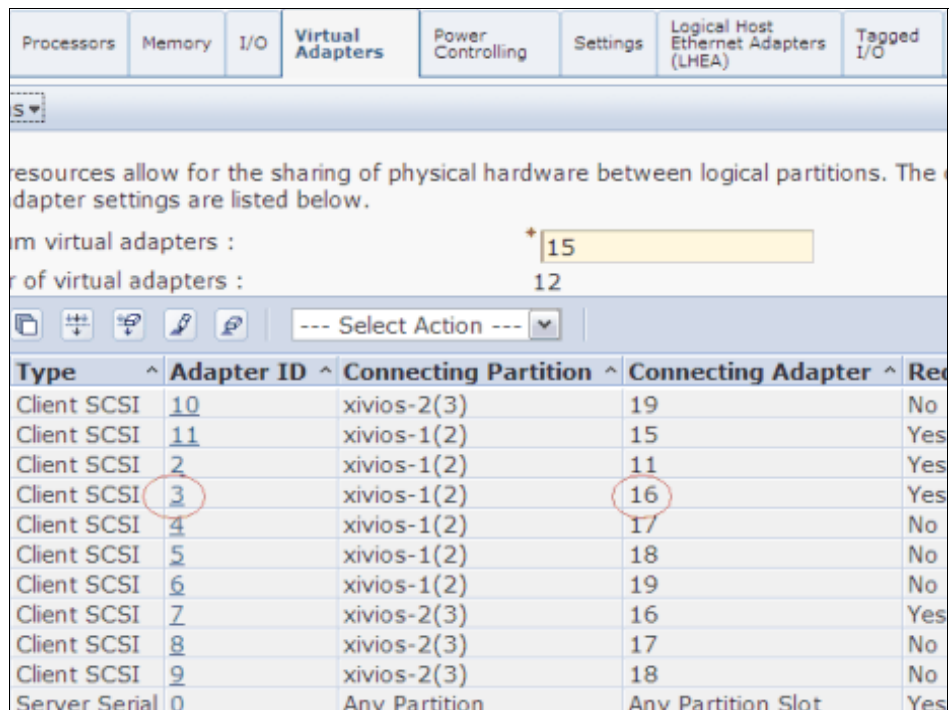
In our setup, no other devices are assigned to the adapter, and the relevant slot is C16 as seen in Figure 8-7.

\$ lsmap -vadapter vhost5		
SVSA	Physloc	C1

vhost5	U9117.MMA.10AF384-V2-C16	0x
VTD		
NO VIRTUAL TARGET DEVICE FOUND		

Figure 8-7 Virtual SCSI adapter in the VIOS

- b. From the HMC, edit the profile of the IBM i partition. Select the partition and choose **Configuration** → **Manage Profiles**. Then select the profile and click **Actions** → **Edit**.
- c. In the partition profile, click the Virtual Adapters tab and make sure that a client VSCSI adapter is assigned to the server adapter with the same ID as the virtual slot number. In our example, client adapter 3 is assigned to server adapter 16 (thus matching the virtual slot C16) as shown in Figure 8-8.



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 8-8 Assigned virtual adapters

5. Map the relevant hdisks to the VSCSI adapter by entering the following command:

```
mkvdev -vdev hdiskx -vadapter <name>
```

In our example, we map the XIV LUNs to the adapter vhost5, and we give to each LUN the virtual device name by using the **-dev** parameter as shown in Figure 8-9.

```
$ mkvdev -vdev hdisk132 -vadapter vhost5 -dev vadamaboot132
vadamaboot132 Available
```

Figure 8-9 Mapping the LUNs in the VIOS

After completing these steps for each VIOS, the LUNs are available to the IBM i client in multipath (one path through each VIOS).

8.5 Match XIV volume to IBM i disk unit

To identify which IBM i disk unit is a particular XIV volume, follow these steps:

1. In VIOS, use the command **XIV_devlist** to list the VIOS disk devices and their associated XIV volumes. This command comes as a part of XIV Host Attachment Kit for VIOS, and must be run in VIOS root-command line, therefore, use the following sequence of steps to execute it:

- **eom_setup_env** (to initiate the OEM software installation and setup environment)
- **# XIV_devlist** (to list the hdisks and corresponding XIV volumes)
- **# exit** (to return to the VIOS prompt)

The output of **XIV_devlist** command in one of the VIO servers in our setup is shown in Figure 8-10. As can be seen in the picture, hdisk5 corresponds to the XIV volumes ITS0_i_1 with serial number 4353.

XIV Devices						
Device	Size	Paths	Vol Name	Vol Id	XIV Id	XIV Host
/dev/hdisk5	154.6GB	2/2	ITS0_i_1	4353	1300203	VIOS_1
/dev/hdisk6	154.6GB	2/2	ITS0_i_CG.snap_group_00001.I TS0_i_4	4497	1300203	VIOS_1
/dev/hdisk7	154.6GB	2/2	ITS0_i_3	4355	1300203	VIOS_1
/dev/hdisk8	154.6GB	2/2	ITS0_i_CG.snap_group_00001.I TS0_i_6	4499	1300203	VIOS_1
/dev/hdisk9	154.6GB	2/2	ITS0_i_5	4357	1300203	VIOS_1
/dev/hdisk10	154.6GB	2/2	ITS0_i_CG.snap_group_00001.I TS0_i_2	4495	1300203	VIOS_1
/dev/hdisk11	154.6GB	2/2	ITS0_i_7	4359	1300203	VIOS_1
/dev/hdisk12	154.6GB	2/2	ITS0_i_8	4360	1300203	VIOS_1

Figure 8-10 VIOS devices and matching XIV volumes

2. In VIOS use the command **lsmap -vadapter vhostx** for the virtual adapter that connects your disk devices to observe which virtual SCSI device is which hdisk. This can be seen in Figure 8-11 on page 188.

```

$ lsmap -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-L2000000000000
Mirrored     false

```

Figure 8-11 Hdisk to vscsi device mapping

- For a particular virtual SCSI device observe the corresponding LUN id by using VIOS command **lsdev -dev vtscsiX -vpd**. In our example, the virtual LUN id of device vtscsi0, is 1, as can be seen in Figure 8-12.

```

$
$ lsdev -dev vtscsi0 -vpd
vtscsi0      U9117.MMA.06C6DE1-V15-C20-L1  Virtual Target Device - Disk
$

```

Figure 8-12 LUN id of a virtual SCSI device

- In IBM i command **STRSST** to start the use System Service Tools (SST). You will need the SST User id and Password to sign in the SST. After you are in SST select Option 3. Work with disk units → Option 1. Display disk configuration → Option 1. Display disk configuration status. In the Disk Configuration Status panel, use F9 to display disk unit details. In the Display Disk Unit Details panel the columns Ctl specifies which LUN id belongs to which disk unit as seen in Figure 8-13 on page 189. In our 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	Ctl	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 8-13 LUN ids of IBM i disk units



VMware ESX host connectivity

This chapter explains OS-specific considerations for host connectivity and describes the host attachment related tasks for ESX version 3.5 and ESX version 4. In addition, this chapter also includes information related to XIV Site Replication Agent.

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

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

and the Host Attachment publications at:

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

9.1 Introduction

Today's virtualization technology is transforming business. Whether the virtualization goal is to consolidate servers, centralize services, implement disaster recovery, set up remote or thin client desktops, or create clouds for optimized resource use, companies are increasingly virtualizing their environments.

Organizations often deploy server virtualization in the hope of gaining economies of scale in consolidating underutilized resources to a new platform. Equally crucial to a server virtualization scenario is the storage itself. Many who have implemented server virtualization but neglected to take storage into account find themselves facing common challenges such as uneven resource sharing, and performance and reliability degradation.

The IBM XIV Storage System, with its grid architecture, automated load balancing, and ease of management, provides best-in-class virtual enterprise storage for virtual servers. In addition, IBM XIV end-to-end support for VMware solutions, including vSphere and VCenter, provides hotspot-free server-storage performance, optimal resource use, and an on-demand storage infrastructure that enables a simplified growth, key to meeting enterprise virtualization goals.

IBM collaborates with VMware on ongoing strategic, functional, and engineering levels. IBM XIV system leverages this technology partnership to provide robust solutions and release them quickly. Along with other IBM storage platforms, the XIV system is installed at VMware's Reference Architecture Lab and other VMware engineering development labs, where it is used for early testing of new VMware product release features. Among other VMware product projects, IBM XIV took part in the development and testing of VMware ESX 4.1.

IBM XIV engineering teams have ongoing access to VMware co-development programs such as developer forums and a comprehensive set of developer resources such as toolkits, source code, and application program interfaces. This access translates to excellent virtualization value for customers.

Note: For more details on some of the topics presented here, refer to the IBM White Paper on which this introduction is based: *A Perfect Fit: IBM XIV Storage System with VMware for Optimized Storage-Server Virtualization*, available at:

http://www.xivstorage.com/materials/white_papers/a_perfect_fit_ibm_xiv_and_vmware.pdf

VMware offers a comprehensive suite of products for server virtualization.

- ▶ VMware ESX server: a production-proven virtualization layer run on physical servers that allows processor, memory, storage, and networking resources to be provisioned to multiple virtual machines.
- ▶ VMware Virtual Machine File System (VMFS): a high-performance cluster file system for virtual machines.
- ▶ VMware Virtual Symmetric Multi-Processing (SMP): enables a single virtual machine to use multiple physical processors simultaneously.
- ▶ VirtualCenter Management Server: a central point for configuring, provisioning, and managing virtualized IT infrastructure.
- ▶ VMware Virtual Machine: a representation of a physical machine by software. A virtual machine has its own set of virtual hardware (for example, RAM, CPU, network adapter, and hard disk storage) upon which an operating system and applications are loaded. The operating system sees a consistent, normalized set of hardware regardless of the actual

physical hardware components. VMware virtual machines contain advanced hardware features such as 64-bit computing and virtual symmetric multiprocessing.

- ▶ Virtual Infrastructure Client (VI Client): is an interface allowing administrators and users to connect remotely to the VirtualCenter Management Server or individual ESX installations from any Windows PC.
- ▶ VMware vCenter Server: formerly VMware VirtualCenter, centrally manages VMware vSphere environments giving IT administrators dramatically improved control over the virtual environment compared to other management platforms.
- ▶ Virtual Infrastructure Web Access: a web interface for virtual machine management and remote consoles access.
- ▶ VMware VMotion: enables the live migration of running virtual machines from one physical server to another with zero downtime, continuous service availability, and complete transaction integrity.
- ▶ VMware Site Recovery Manager (SRM): a business continuity and disaster recovery solution for VMware ESX servers.
- ▶ VMware Distributed Resource Scheduler (DRS): allocates and balances computing capacity dynamically across collections of hardware resources for virtual machines.
- ▶ VMware High Availability (HA): provides easy-to-use, cost-effective high availability for applications running in virtual machines. In the event of server failure, affected virtual machines are automatically restarted on other production servers that have spare capacity.
- ▶ VMware Consolidated Backup: provides an easy-to-use, centralized facility for agent-free backup of virtual machines that simplifies backup administration and reduces the load on ESX installations.
- ▶ VMware Infrastructure SDK: provides a standard interface for VMware and third-party solutions to access VMware Infrastructure.

IBM XIV provides end-to-end support for VMware (including vSphere, Site Recovery Manager, and soon VAAI), with ongoing support for VMware virtualization solutions as they evolve and are developed. Specifically, IBM XIV works in concert with the following VMware products:

- ▶ vSphere ESX
- ▶ vSphere Hypervisor (ESXi)
- ▶ vCenter server
- ▶ vSphere VMotion

When the VMWare ESX server virtualizes server environment, the VMware Site Recovery Manager enables administrators of virtualized environments to automatically fail over the whole environment or parts of it to a backup site.

The VMware SRM (Site Recovery Manager) provides automation for:

- ▶ Planning and testing vCenter inventory migration from one site to another
- ▶ Executing vCenter inventory migration on schedule or for emergency failover.

VMware Site Recovery Manager utilizes the mirroring capabilities of the underlying storage to create a copy of the data on a second location (e.g. a backup data center). This ensures that, at any time, two copies of the data are maintained and production can run on either of them.

IBM XIV Storage System has a Storage Replication Agent that integrates the IBM XIV Storage System with VMware Site Recovery Manager.

In addition, IBM XIV leverages VAAI to take on storage-related tasks that were previously performed by VMware. Transferring the processing burden dramatically reduces performance overhead, speeds processing and frees up VMware for more mission-critical tasks, such as adding applications.

VAAI improves I/O performance and data operations. When hardware acceleration is enabled with XIV, operations like VM provisioning, VM cloning, and VM migration complete dramatically faster, and with minimal impact on the ESX server, increasing scalability.

IBM XIV uses the following T10-compliant SCSI primitives to achieve the above levels of integration and related benefits:

- ▶ **Full Copy:** Divests copy operations from VMware ESX to the IBM XIV Storage System. This feature allows for rapid movement of data by off-loading block level copy, move, and snapshot operations to the IBM XIV Storage System. It also enables VM deployment by VM cloning and storage cloning at the block level within and across LUNs. Benefits include expedited copy operations, minimized host processing/resource allocation, reduced network traffic, and considerable boosts in system performance.
- ▶ **Block Zeroing:** Off-loads repetitive block-level write operations within virtual machine disks to IBM XIV. This feature reduces server workload and improves server performance. Space reclamation and thin provisioning allocation are more effective with this feature.

Support for VAAI Block Zeroing allows VMware to better utilize XIV architecture and gain overall dramatic performance improvements with VM provisioning and on-demand VM provisioning when VMware typically zero-fills a large amount of storage space.

Block Zeroing allows the VMware host to save bandwidth and communicate faster by minimizing the amount of actual data sent over the path to IBM XIV. Similarly, it allows IBM XIV to minimize its own internal bandwidth consumption and virtually apply the write much faster.
- ▶ **Hardware Assisted Locking:** Intelligently relegates resource reservation down to the selected block level instead of the LUN, significantly reducing SCSI reservation contentions, lowering storage resource latency, and enabling parallel storage processing, particularly in enterprise environments where LUNs are more likely to be used by multiple applications or processes at once.

To implement virtualization with VMware and XIV Storage System you need deploy at minimum one ESX server that can be used for the Virtual Machines deployment, and one vCenter server.

You can implement a high availability solution in your environment by adding and deploying an additional server (or servers) running under VMware ESX and implementing the VMware High Availability option for your ESX servers.

To further improve the availability of your virtualized environment, simplify business continuity and disaster recovery solutions. Implement ESX servers, vCenter server, and another XIV storage system at the recovery site. You should also install VMware Site Recovery Manager and use the Site Replication Agent to integrate Site Recovery Manager with your XIV storage systems at both sites. Note that the Site Recovery Manager itself can also be implemented as a virtual machine on the ESX server.

Finally, you need redundancy at the network and SAN levels for all your sites.

A full disaster recovery solution is depicted Figure 9-1.

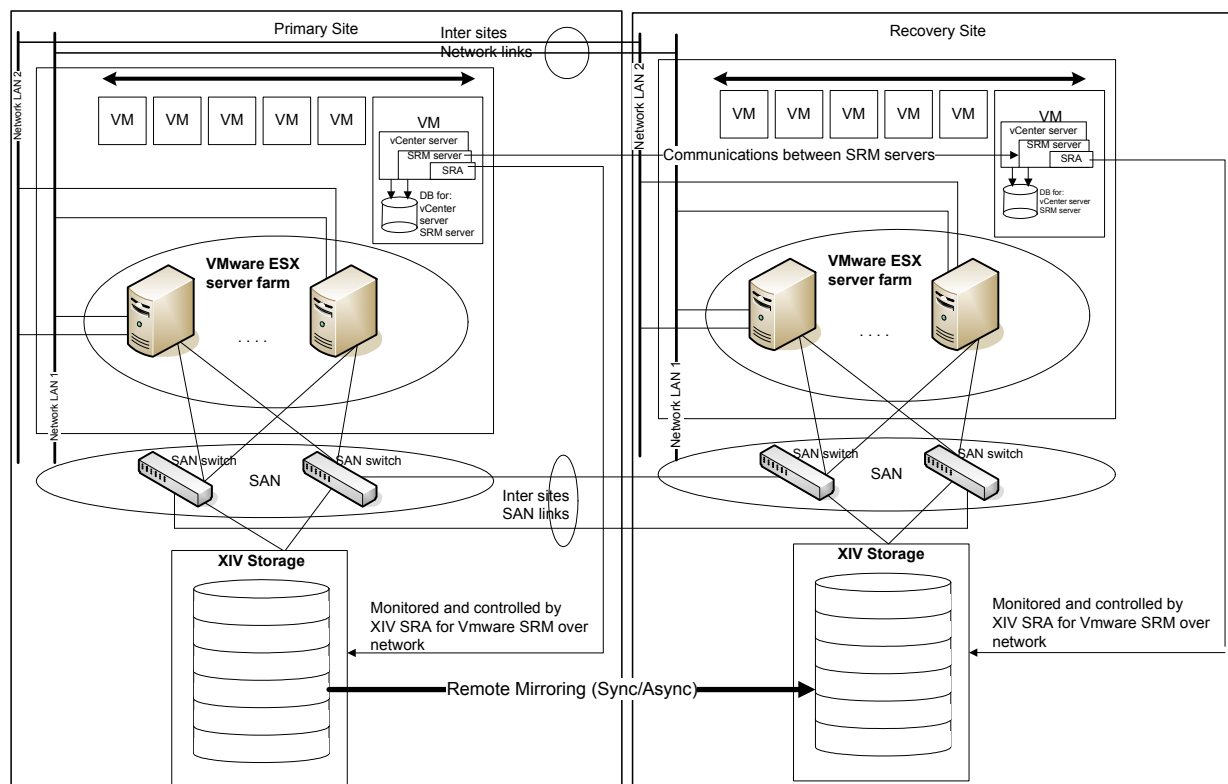


Figure 9-1 Example of the simple architecture for the virtualized environment built on the VMware and IBM XIV Storage System

The rest of this chapter is divided into three major sections. The first two sections address specifics for VMware ESX server 3.5 or 4 respectively, and the final section relates to XIV Storage Replication Agent for VMware Site Recovery Manager.

9.2 ESX 3.5 Fibre Channel configuration

This section describes attaching VMware ESX 3.5 hosts through Fibre Channel. Refer also to http://publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/Host_System_Attachment_Guide_for_VMware.pdf

Details about Fibre Channel configuration on VMware ESX server 3.5 can be found at: http://www.vmware.com/pdf/vi3_35/esx_3/r35u2/vi3_35_25_u2_san_cfg.pdf

Refer also to:

http://www.vmware.com/pdf/vi3_san_design_deploy.pdf

Follow these steps to configure the VMware host for FC attachment with multipathing:

1. Check HBAs and FC connections from your host to XIV Storage System.
2. Configure the host, volumes and host mapping in the XIV Storage System.
3. Discover the volumes created on XIV.

9.2.1 Installing HBA drivers

VMware ESX includes drivers for all the HBAs that it supports. VMware strictly controls the driver policy, and only drivers provided by VMware must be used. Any driver updates are normally included in service/update packs.

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

<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.

Refer also to:

<http://www.vmware.com/resources/compatibility/search.php>

9.2.2 Scanning for new LUNs

Before you can scan for new LUNs on ESX, your host needs to be added and configured on the XIV Storage System. See 1.4, “Logical configuration for host connectivity” on page 29 for information about how to do this.

ESX hosts that access the same shared LUNs should be grouped in a cluster (XIV cluster) and the LUNs assigned to the cluster. Refer to Figure 9-2 and Figure 9-3 for how this might typically be set up.

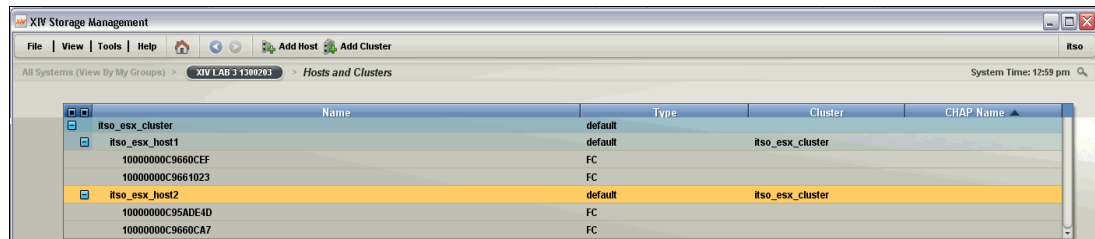


Figure 9-2 ESX host cluster setup in XIV GUI

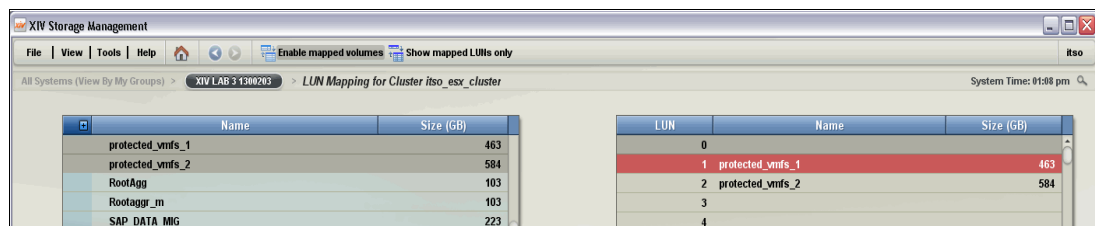


Figure 9-3 ESX LUN mapping to the cluster

To scan for and configure new LUNs, follow these instructions:

1. After completing the host definition and LUN mappings in the XIV Storage System, click the Configuration tab for your host, and select **Storage Adapters** as shown in Figure 9-4 on page 197.

Here you can see vmhba2 highlighted but a rescan will scan across all adapters. The adapter numbers might be enumerated differently on the different hosts; this is not an issue.

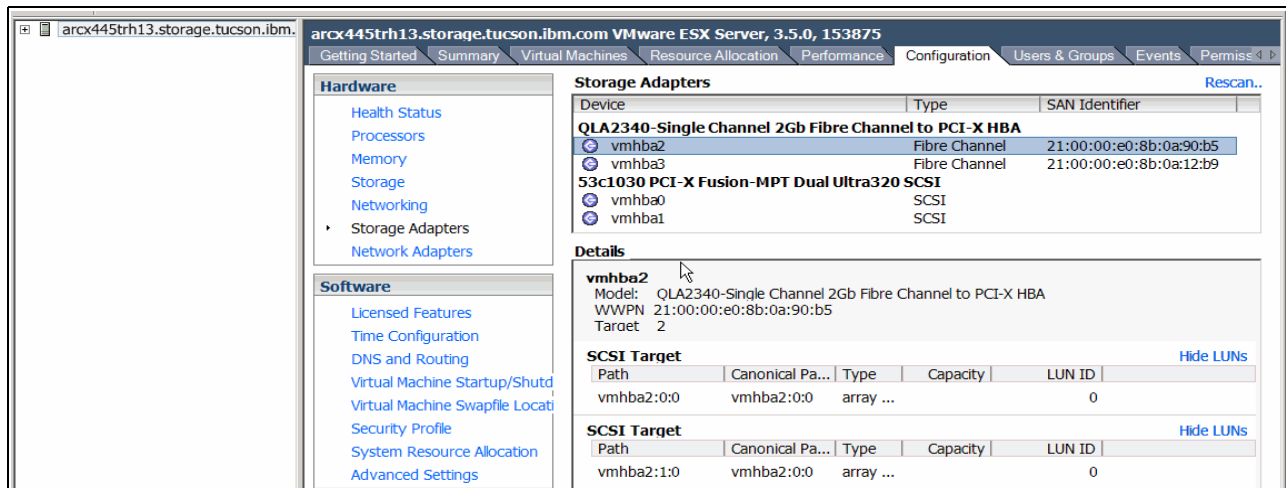


Figure 9-4 Select Storage Adapters

2. Select the **Scan for New Storage Devices** and **Scan for New VMFS Volumes** check boxes, and then click **OK** to scan for new storage devices as shown in Figure 9-5.

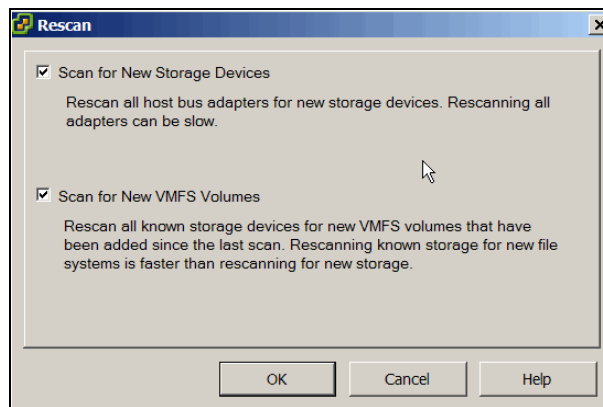


Figure 9-5 Rescan for New Storage Devices

3. The new LUNs assigned will appear in the Details pane as shown in Figure 9-6 on page 198.

Details				
vmhba2				
Model: QLA2340/2340L				
WWPN 21:00:00:e0:8b:0a:90:b5				
Target 2				
SCSI Target				
Path	Canonical Pa...	Type	Capacity	LUN ID
vmhba2:2:0	vmhba2:2:0	disk	32.00 GB	0
vmhba2:2:1	vmhba2:2:1	disk	32.00 GB	1
SCSI Target				
Path	Canonical Pa...	Type	Capacity	LUN ID
vmhba2:3:0	vmhba2:2:0	disk	32.00 GB	0
vmhba2:3:1	vmhba2:2:1	disk	32.00 GB	1

Figure 9-6 FC discovered LUNs on vmhba2

Here, you observe that controller vmhba2 can see two LUNs (LUN 0 and LUN 1) circled in green and they are visible on two targets (2 and 3) circled in red. The other controllers on the host will show the same path and LUN information.

For detailed information about how to use LUNs with virtual machines, refer to the VMware guides, available at the following URLs:

http://www.vmware.com/pdf/vi3_35/esx_3/r35u2/vi3_35_25_u2_admin_guide.pdf

http://www.vmware.com/pdf/vi3_35/esx_3/r35u2/vi3_35_25_u2_3_server_config.pdf

9.2.3 Assigning paths from an ESX 3.5 host to XIV

All information in this section relates to ESX 3.5 (and not other versions of ESX) unless otherwise specified. The procedures and instructions given here are based on code that was available at the time of writing this book.

VMware provides its own multipathing I/O driver for ESX. No additional drivers or software are required. As such, the XIV Host Attachment Kit only provides documentation, and no software installation is required.

The ESX 3.5 multipathing supports the following path policies:

- ▶ **Fixed:** Always use the preferred path to the disk. If preferred path is not available, an alternative path to the disk should be chosen. When the preferred path is restored, an automatic failback to the preferred path occurs.
- ▶ **Most Recently Used:** Use the path most recently used while the path is available. Whenever a path failure occurs, an alternate path is chosen. There is no automatic failback to the original path.
- ▶ **Round-Robin (ESX 3.5 experimental):** Multiple disk paths are used and balanced based on load. Round-Robin is not supported for production use in ESX version 3.5.

ESX Native Multipathing automatically detects IBM XIV and sets the path policy to Fixed by default. Users should not change this. Also, setting the preferred path or manually assigning LUNs to specific path should be monitored carefully so you do not overload the IBM XIV storage controller port. Monitoring can be done using **esxtop** to monitor outstanding queues pending execution.

XIV is an active/active storage system, and therefore it can serve I/Os to all LUNs using every available path. However, the driver with ESX 3.5 cannot perform the same function and by default cannot fully load balance. It is possible to partially overcome this limitation by setting the correct pathing policy and distributing the I/O load over the available HBAs and XIV ports. This could be referred to as ‘manually’ load balancing. To achieve this, follow these instructions:

1. The pathing policy in ESX 3.5 can be set to either **Most Recently Used (MRU)** or **Fixed**. When accessing storage on the XIV the correct policy is **Fixed**. In the VMware Infrastructure Client select the server then click the **Configuration** tab and select **Storage**. Refer to Figure 9-7.

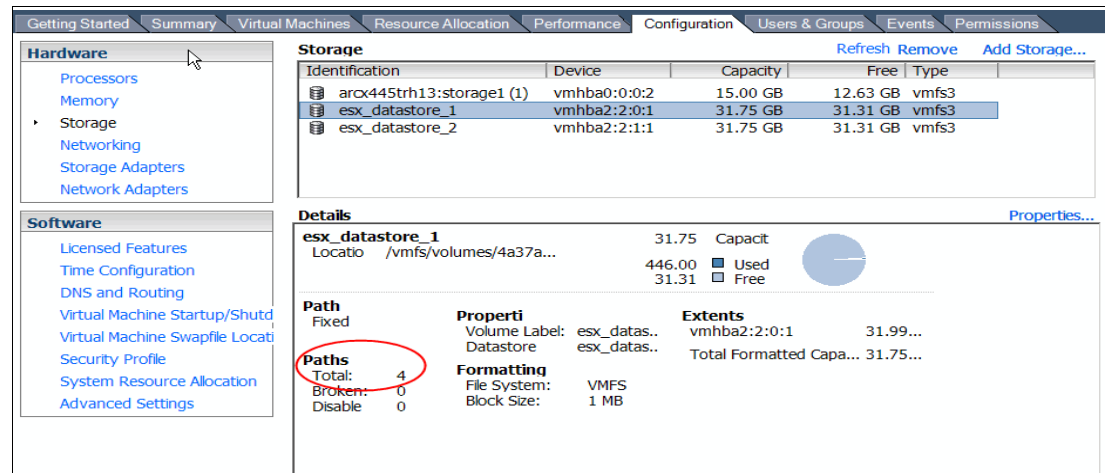


Figure 9-7 Storage paths

You can see the LUN highlighted (esx_datastore_1) and the number of paths is 4 (circled in red). Select **Properties** to view more details about the paths.

2. In the **Properties** window, you can see that the active path is vmhba2:2:0 as shown in Figure 9-8 on page 200.

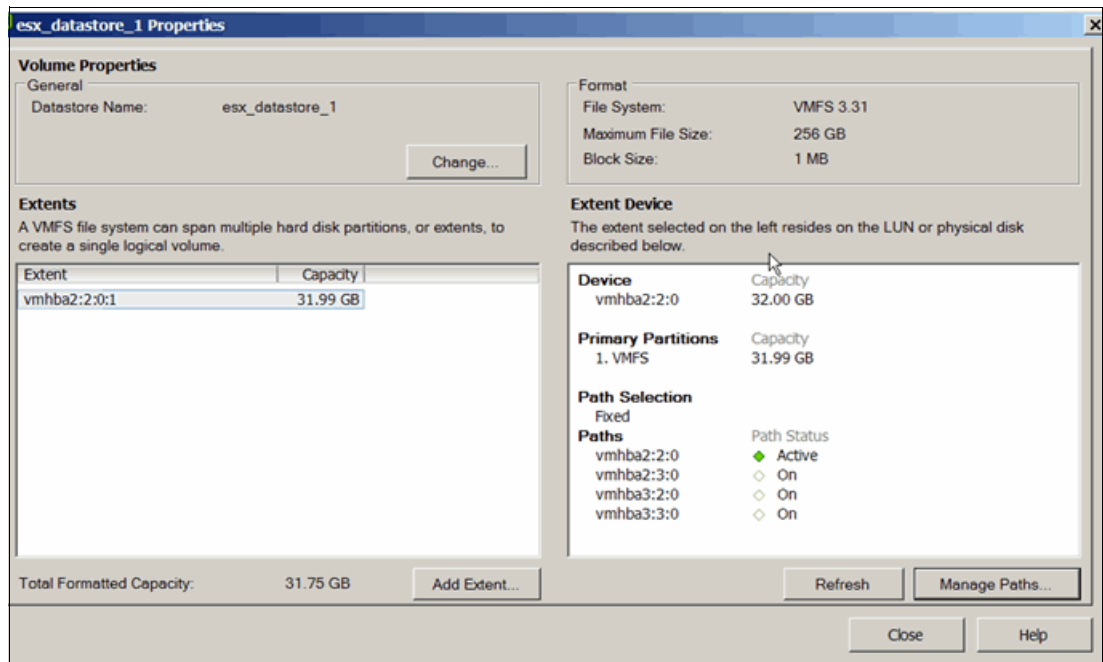


Figure 9-8 Storage path details

3. To change the current path, select **Manage Paths** and the Manage Paths window opens as shown in Figure 9-9. The pathing policy should be **Fixed**; if it is not, then select **Change** in the **Policy** pane and change it to **Fixed**.

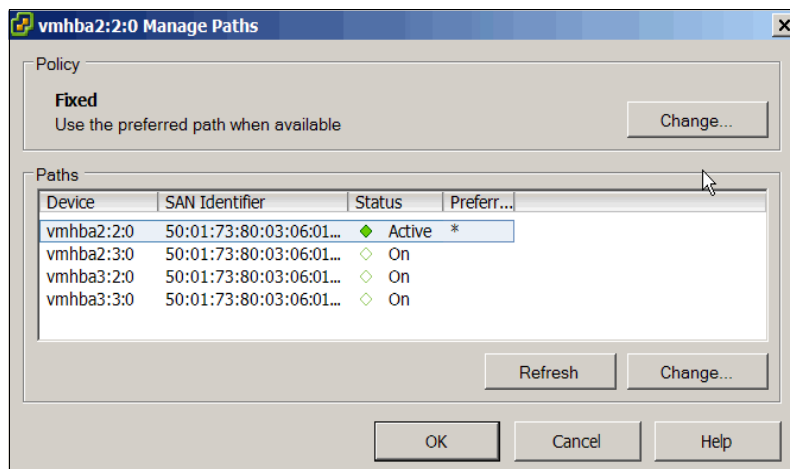


Figure 9-9 Change paths window

4. To manually load balance, highlight the preferred path in the **Paths** pane and click **Change**. Then assign an HBA and target port to the LUN. Refer to Figure 9-10 on page 201, Figure 9-11 on page 201, and Figure 9-12 on page 201.

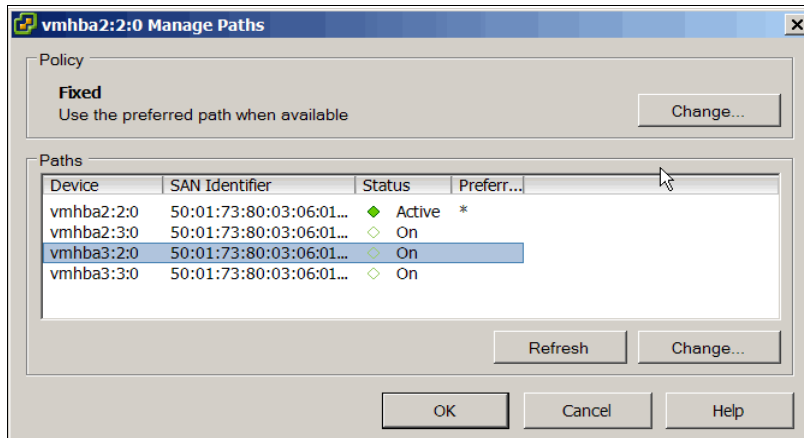


Figure 9-10 Change to new path

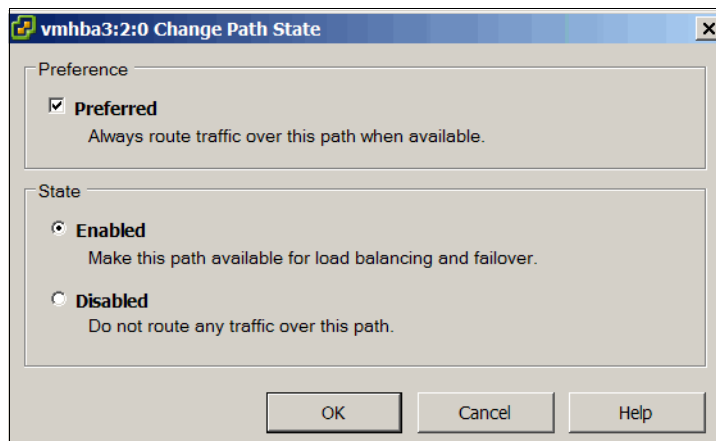


Figure 9-11 Set preferred

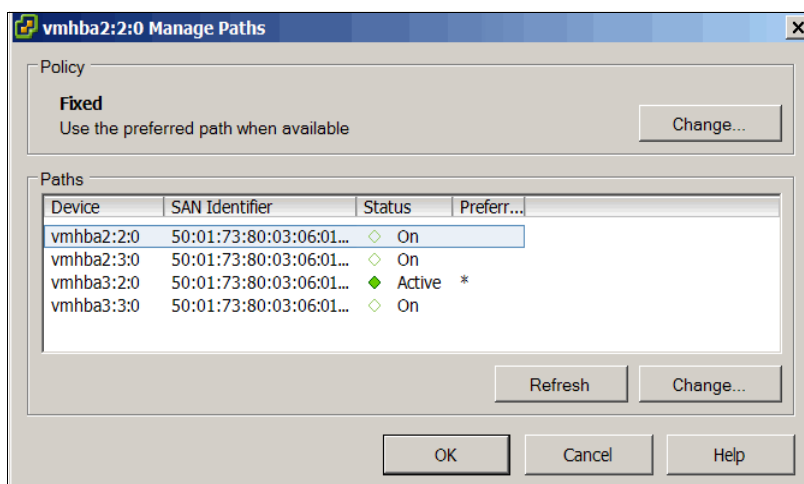


Figure 9-12 New preferred path set

- Repeat steps 1-4 to manually balance IOs across the HBAs and XIV target ports. Due to the manual nature of this configuration, it will need to be reviewed over time.

Important: When setting paths, if a LUN is shared by multiple ESX hosts, it should be accessed through the same XIV port, and therefore the same interface module.

Example 9-1 and Example 9-2 show the results of manually configuring two LUNs on separate preferred paths on two ESX hosts. Only two LUNs are shown for clarity, but this can be applied to all LUNs assigned to the hosts in the ESX datacenter.

Example 9-1 ESX Host 1 preferred path

```
[root@arcx445trh13 root]# esxcfg-mpath -l
Disk vmhba0:0:0 /dev/sda (34715MB) has 1 paths and policy of Fixed
  Local 1:3.0 vmhba0:0:0 On active preferred

Disk vmhba2:2:0 /dev/sdb (32768MB) has 4 paths and policy of Fixed
  FC 5:4.0 210000e08b0a90b5<->5001738003060140 vmhba2:2:0 On active preferred
  FC 5:4.0 210000e08b0a90b5<->5001738003060150 vmhba2:3:0 On
  FC 7:3.0 210000e08b0a12b9<->5001738003060140 vmhba3:2:0 On
  FC 7:3.0 210000e08b0a12b9<->5001738003060150 vmhba3:3:0 On

Disk vmhba2:2:1 /dev/sdc (32768MB) has 4 paths and policy of Fixed
  FC 5:4.0 210000e08b0a90b5<->5001738003060140 vmhba2:2:1 On
  FC 5:4.0 210000e08b0a90b5<->5001738003060150 vmhba2:3:1 On
  FC 7:3.0 210000e08b0a12b9<->5001738003060140 vmhba3:2:1 On
  FC 7:3.0 210000e08b0a12b9<->5001738003060150 vmhba3:3:1 On active preferred
```

Example 9-2 ESX host 2 preferred path

```
[root@arcx445bvkf5 root]# esxcfg-mpath -l
Disk vmhba0:0:0 /dev/sda (34715MB) has 1 paths and policy of Fixed
  Local 1:3.0 vmhba0:0:0 On active preferred

Disk vmhba4:0:0 /dev/sdb (32768MB) has 4 paths and policy of Fixed
  FC 7:3.0 10000000c94a0436<->5001738003060140 vmhba4:0:0 On active preferred
  FC 7:3.0 10000000c94a0436<->5001738003060150 vmhba4:1:0 On
  FC 7:3.1 10000000c94a0437<->5001738003060140 vmhba5:0:0 On
  FC 7:3.1 10000000c94a0437<->5001738003060150 vmhba5:1:0 On

Disk vmhba4:0:1 /dev/sdc (32768MB) has 4 paths and policy of Fixed
  FC 7:3.0 10000000c94a0436<->5001738003060140 vmhba4:0:1 On
  FC 7:3.0 10000000c94a0436<->5001738003060150 vmhba4:1:1 On
  FC 7:3.1 10000000c94a0437<->5001738003060140 vmhba5:0:1 On
  FC 7:3.1 10000000c94a0437<->5001738003060150 vmhba5:1:1 On active preferred
```

9.3 ESX 4.x Fibre Channel configuration

This section describes attaching ESX 4 hosts to XIV through Fibre Channel.

9.3.1 Installing HBA drivers

ESX includes drivers for all the HBAs that it supports. VMware strictly controls driver policy, and only drivers provided by VMware must be used. Any driver updates are normally included in service/update packs.

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

<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.

9.3.2 Identifying ESX host port WWN

You need to identify the host port WWN for FC adapters installed in the ESX Servers before you can start defining the ESX cluster and its host members.

To identify ESX host port WWNs, run the VMware vSphere client and connect to the ESX Server. In the VMware vSphere client, select the server, click the Configuration tab, and then select **Storage Adapters**. Refer to Figure 9-13, where you can see the port WWNs for the installed FC adapters (circled in red).

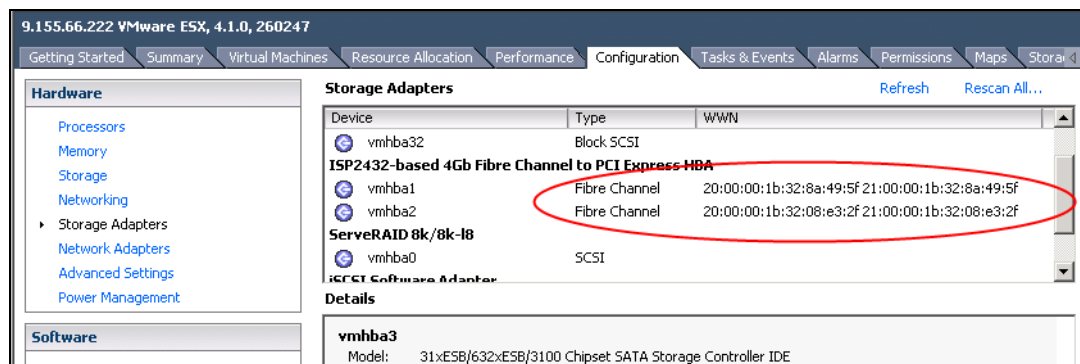


Figure 9-13 ESX host port WWNs

You need to repeat this process for all ESX hosts that you plan to connect to the XIV Storage System.

After identification of the ESX host ports WWNs, you are ready to define hosts and clusters for the ESX servers, create LUNs, and map them to defined ESX clusters and hosts on the XIV Storage System. Refer to Figure 9-2 on page 196 and Figure 9-3 on page 196 for how this might typically be set up.

Note: The ESX hosts that access the same LUNs should be grouped in a cluster (XIV cluster) and the LUNs assigned to the cluster.

The maximum LUN size usable by an ESX host is 2181GB.

9.3.3 Scanning for new LUNs

To scan and configure new LUNs, follow these instructions:

1. After the host definition and LUN mappings have been completed in the XIV Storage System, click the Configuration tab for your host, and select **Storage Adapters** as shown in Figure 9-14.

Here you can see vmhba1 highlighted but a rescan will scan across all adapters. The adapter numbers might be enumerated differently on the different hosts, this is not an issue.

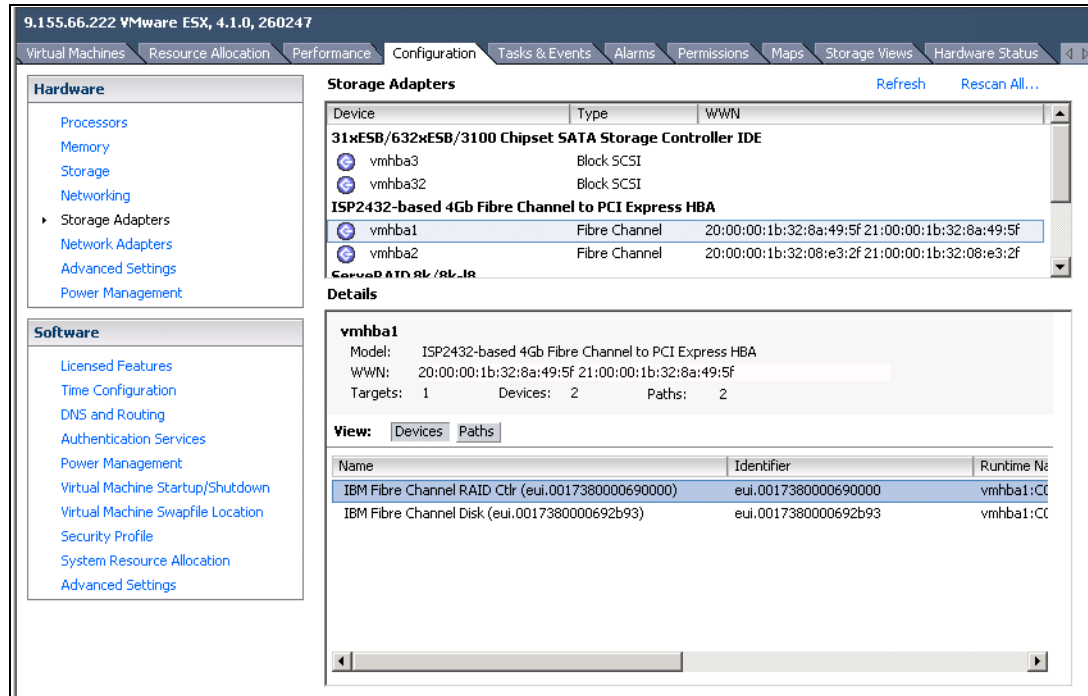


Figure 9-14 Select Storage Adapters

2. Select the **Scan for New Storage Devices** and **Scan for New VMFS Volumes** check boxes, and then click **OK** to scan for new storage devices as shown in Figure 9-15.

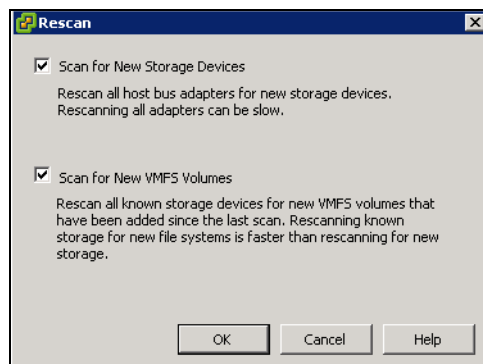
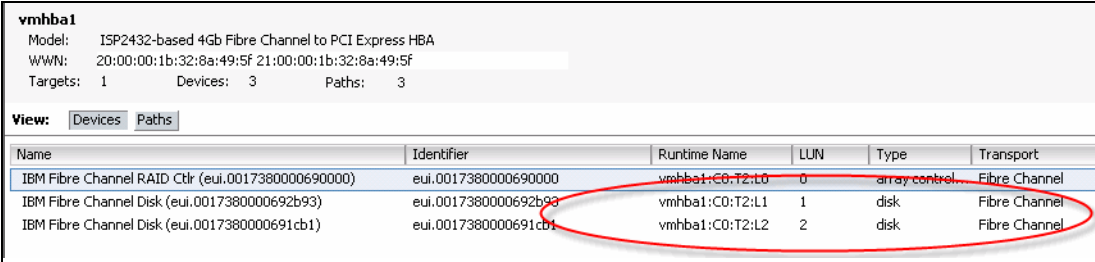


Figure 9-15 Rescan for New Storage Devices

3. The new LUNs assigned will appear in the **Details** pane as depicted in Figure 9-16.



Name	Identifier	Runtime Name	LUN	Type	Transport
IBM Fibre Channel RAID Ctr (eui.0017380000690000)	eui.0017380000690000	vmhba1:C0:T2:L0	0	array control	Fibre Channel
IBM Fibre Channel Disk (eui.0017380000692b93)	eui.0017380000692b93	vmhba1:C0:T2:L1	1	disk	Fibre Channel
IBM Fibre Channel Disk (eui.0017380000691cb1)	eui.0017380000691cb1	vmhba1:C0:T2:L2	2	disk	Fibre Channel

Figure 9-16 FC discovered LUNs on vmhba1

Here, you observe that controller vmhba1 can see two LUNs (LUN 1 and LUN 2) circled in red. The other controllers on the host will show the same path and LUN information.

9.3.4 Attaching an ESX 4.x host to XIV

This section describes the attachment of ESX 4 based hosts to the XIV Storage System. It provides specific instructions for Fibre Channel (FC) and Internet Small Computer System Interface (iSCSI) connections. All the information in this section relates to ESX 4 (and not other versions of ESX) unless otherwise specified.

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

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

By default ESX 4 supports the following types of storage arrays:

- ▶ Active/active storage systems: allows access to the LUN simultaneously through all storage ports without influence on performance. All the paths are active all the time. If one port fails, all of the other available ports would continue servicing access from servers to the storage system.
- ▶ Active/passive storage systems: systems where a LUN is accessible over a single storage port. The other storage ports act as backup for the active storage port.
- ▶ Asymmetrical storage systems (VMW_SATP_DEFAULT_ALUA): supports Asymmetrical Logical Unit Access (ALUA). ALUA-compliant storage systems provide different levels of access per port. This allows the SCSI Initiator port to make intelligent decisions for bandwidth. The host uses some of the active paths as primary while others as secondary.

With the release of VMware ESX 4 and VMware ESXi 4, VMware has introduced the concept of a Pluggable Storage Architecture (PSA), which in turn introduced additional concepts to its Native Multipathing (NMP). The PSA interacts at the VMkernel level. It is an open and modular framework that can coordinate the simultaneous operations across various multipathing solutions.

VMware NMP chooses the multipathing algorithm based on the storage system type. The NMP associates a set of physical paths with a specific storage device, or LUN. The NMP module works with sub-plug-ins such as a Path Selection Plug-In (PSP) and a Storage Array Type Plug-In (SATP). The SATP plug-ins are responsible for handling path failover for a given storage system and the PSPs plug-ins are responsible for determining which physical path is used to issue an I/O request to a storage device.

ESX 4 provides default SATPs that support non-specific active-active (VMW_SATP_DEFAULT_AA) and ALUA storage system (VMW_SATP_DEFAULT_ALUA). Each SATP accommodates special characteristics of a certain class of storage systems and can perform the storage system specific operations required to detect path state and activate an inactive path.

Note: Starting with XIV software Version 10.1, the XIV Storage System is a T10 ALUA-compliant storage system.

ESX 4 automatically selects the appropriate SATP plug-in for the IBM XIV Storage System based on the XIV Storage System software version. For versions prior to 10.1 and for ESX 4.0, the Storage Array Type is VMW_SATP_DEFAULT_AA. For XIV versions later than 10.1 and with ESX 4.1, the Storage Array Type is VMW_SATP_DEFAULT_ALUA.

Path Selection Plug-Ins (PSPs) run with the VMware NMP and are responsible for choosing a physical path for I/O requests. The VMware NMP assigns a default PSP to each logical device based on the SATP associated with the physical paths for that device.

VMware ESX 4 supports the following PSP types:

- ▶ Fixed (VMW_PSP_FIXED): Always use the preferred path to the disk. If the preferred path is not available, an alternate path to the disk is chosen. When the preferred path is restored, an automatic failback to the preferred path occurs.
- ▶ Most Recently Used (WMV_PSP_MRU): Use the path most recently used while the path is available. Whenever a path failure occurs, an alternate path is chosen. There is no automatic failback to the original path.
- ▶ Round-Robin (VMW_PSP_RR): Multiple disk paths are used and are load balanced.

ESX has built-in rules defining relations between SATP and PSP for the storage system.

Figure 9-17 illustrates the structure of VMware Pluggable Storage Architecture and relations between SATP and PSP.

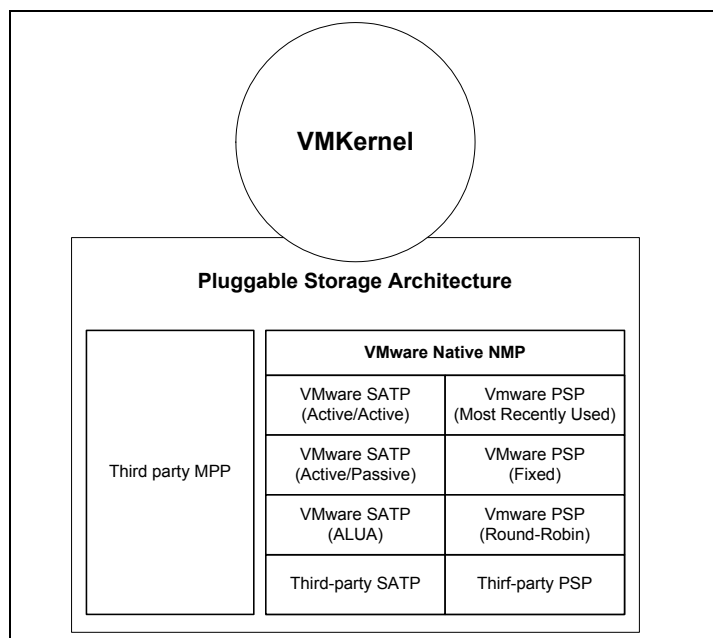


Figure 9-17 VMware multipathing stack architecture

The vStorage API Array Integration (VAAI) initiative

ESX 4 also brings a new level of integration with the storage systems through the vStorage API Array Integration (VAAI) initiative. VAAI helps reduce host overhead, and increases scalability and operational performance of storage systems. The traditional ESX operational model with storage systems forced the ESX host to issue a large number of identical commands to complete some types of operations, such as for a datastore full copy. With the use of VAAI the same task can be accomplished with just one command.

Starting with software version 10.2.4, the IBM XIV Storage System supports VAAI for ESX 4.

9.3.5 Configuring ESX 4 host for multipathing with XIV

With ESX Version 4, VMWare supports a round-robin multipathing policy for production environments. The round-robin multipathing policy is always preferred over other choices when attaching to the IBM XIV Storage System.

Before proceeding with the multipathing configuration, be sure that you completed the tasks described under 9.3.1, “Installing HBA drivers” on page 203, 9.3.2, “Identifying ESX host port WWN” on page 203, and 9.3.3, “Scanning for new LUNs” on page 204.

We start by illustrating the addition of a new datastore, then setting its multipathing policy.

To add a new datastore, follow these instructions:

1. Launch the VMware vSphere client and then connect to your vCenter server. Select the server that you plan to add a new datastore to.
2. In the vSphere client main window, click the **Configuration** tab for your host and select **Storage** as shown in Figure 9-18.

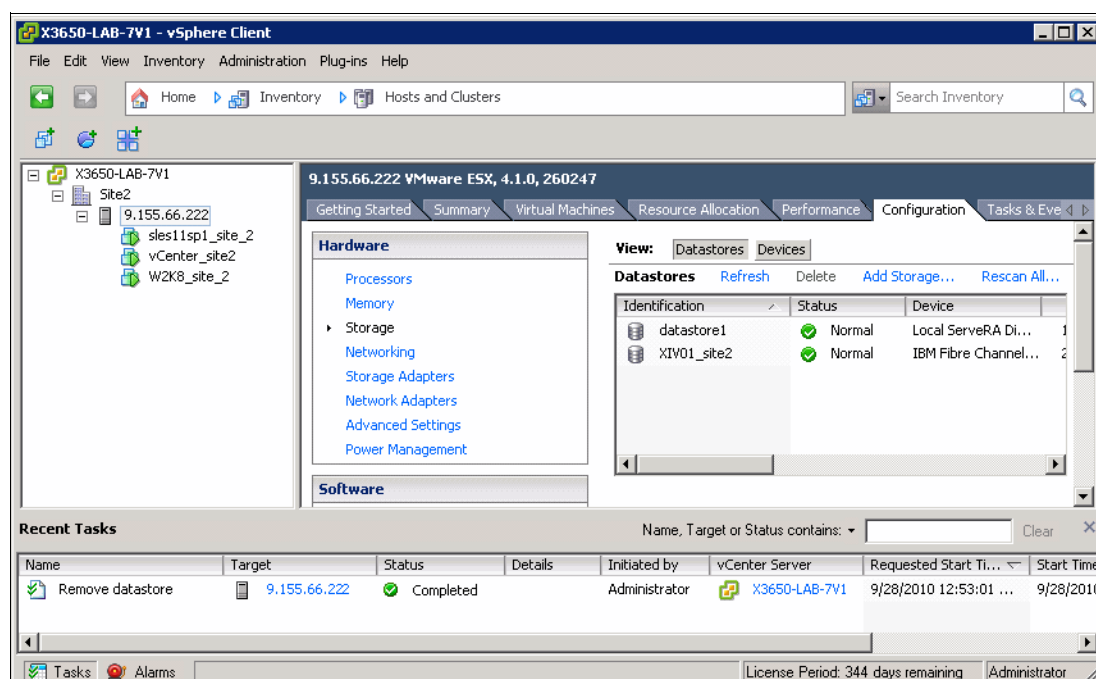


Figure 9-18 ESX 4 Defined datastores

Here you can see datastores currently defined for the ESX host.

- Click **Add Storage** to display the window shown in Figure 9-19.

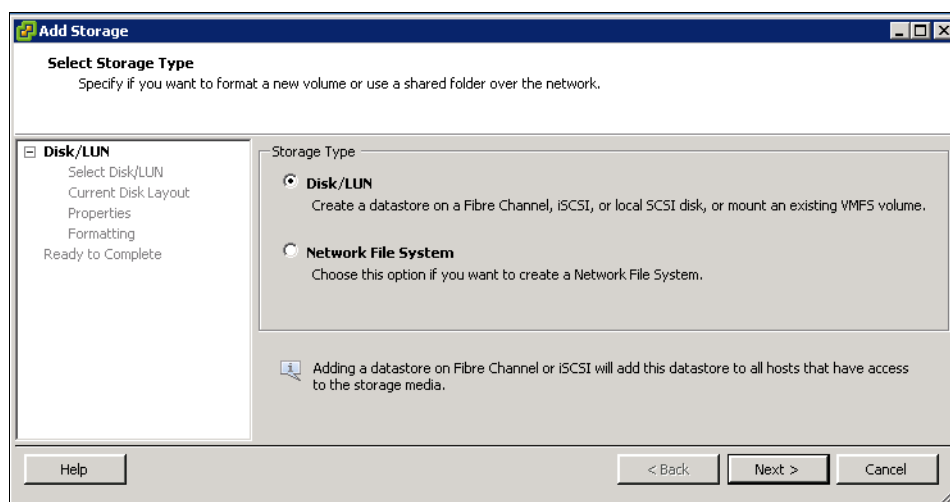


Figure 9-19 Add Storage dialog

- In the **Storage Type** box select **Disk/LUN** and click **Next** to get to the window shown in Figure 9-20. You can see listed the Disks and LUNs that are available to use as a new datastore for the ESX Server.

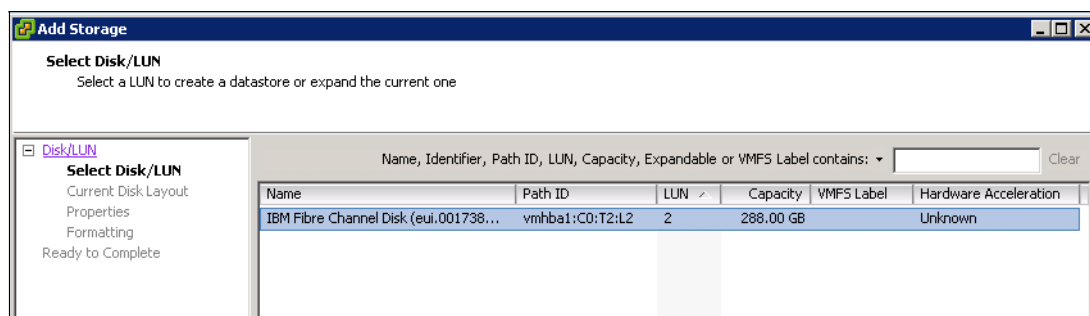


Figure 9-20 List of Disks/LUNs for use as a datastore

- Select the LUN that you want to use as a new datastore, then click **Next** (not shown) at the bottom of this window. A new window like Figure 9-21 displays.

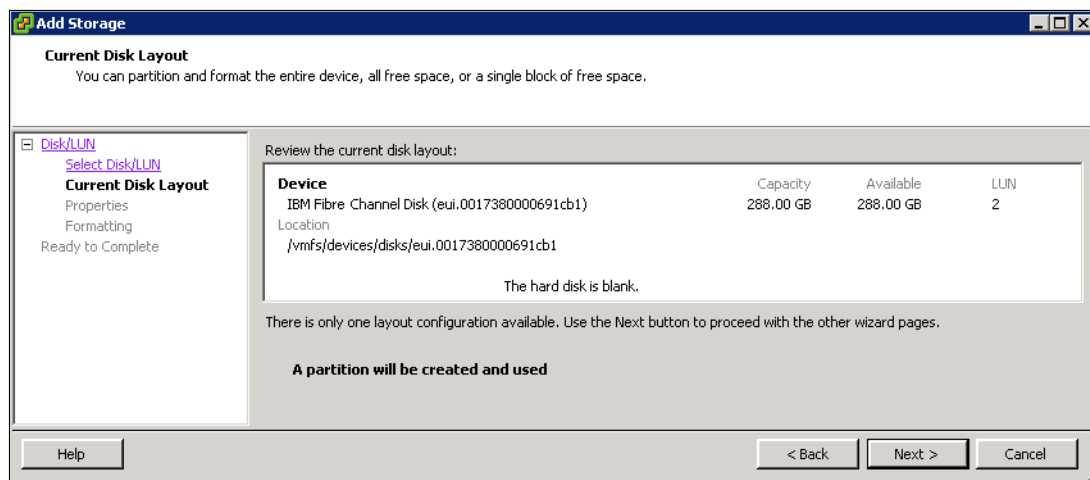


Figure 9-21 Partition parameters

In Figure 9-21 on page 208, you can observe the partition parameters that will be used to create the new partition. If you need to change the parameters, use the **Back** button. Otherwise, click **Next**.

6. The window shown in Figure 9-22 displays. Type a name for the new datastore. In our illustration, the name is XIV_demo_store.

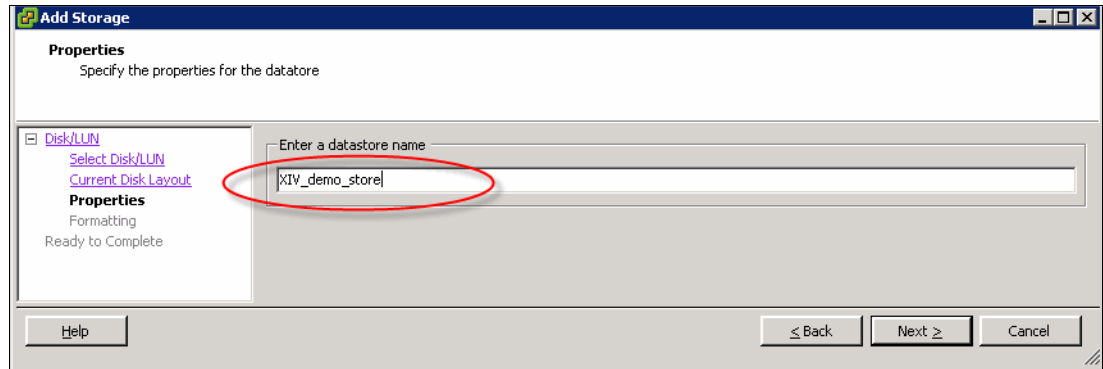


Figure 9-22 Datastore name

7. Click **Next** to display the window shown in Figure 9-23. Enter the file system parameters for your new datastore, then click **Next** to continue.

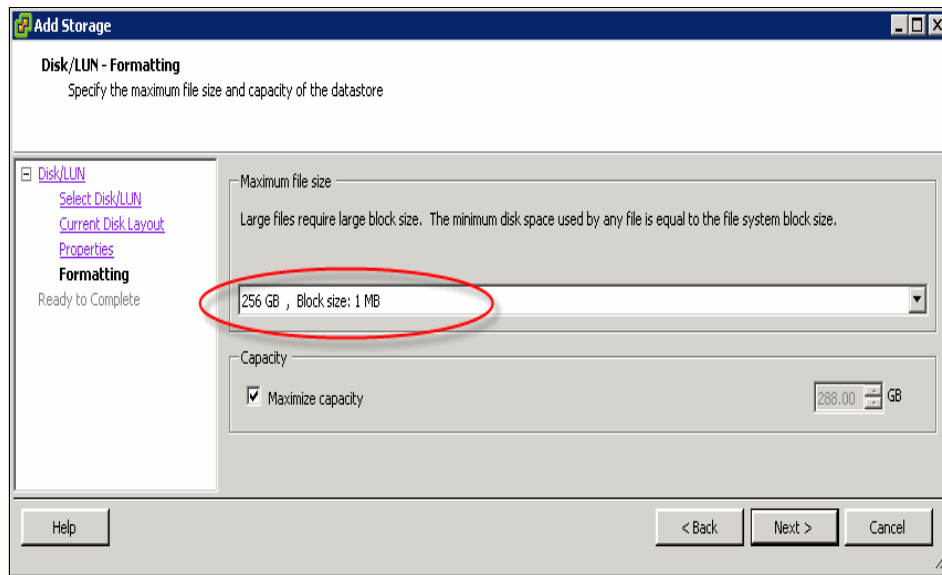


Figure 9-23 Choose the filesystem parameters for ESX datastore

Note: Refer to VMware ESX 4 documentation to choose the right values for file system parameters, in accordance with your specific environment.

8. The window shown in Figure 9-24 displays.

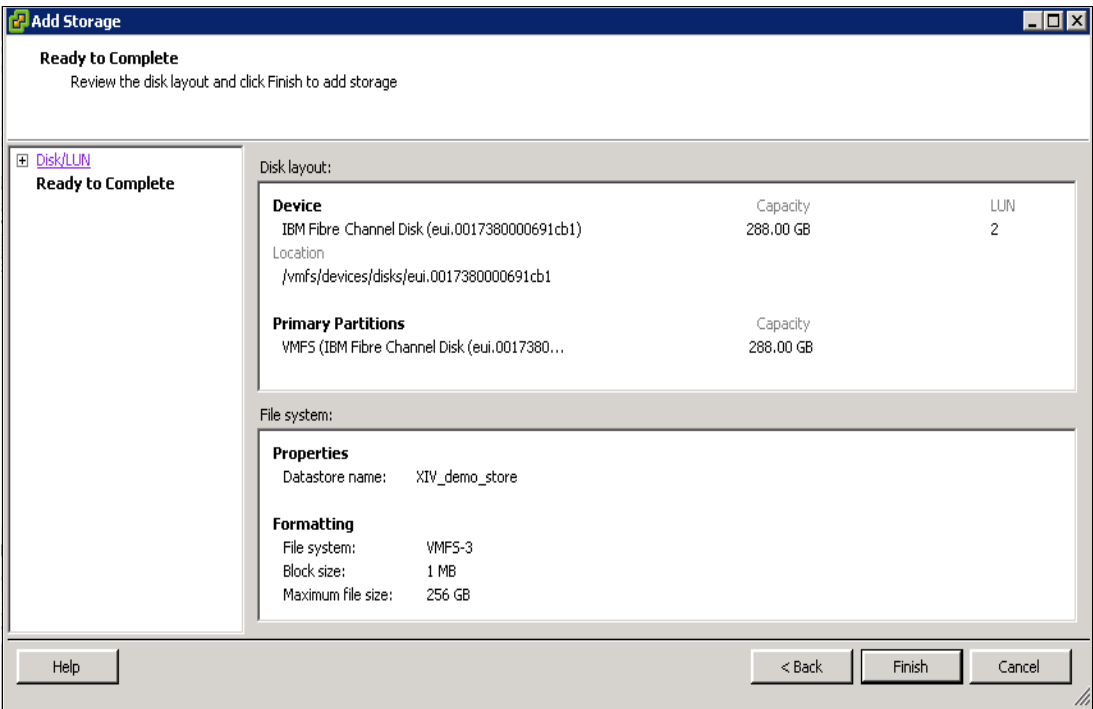


Figure 9-24 Summary on datastore selected parameters

- This window summarizes the complete set of parameters that you just specified. Make sure everything is correct and click **Finish**.
9. You are returned to the vSphere client main window, where you will see two new tasks displayed in the recent task pane, shown in Figure 9-25, indicating the completion of the new datastore creation.

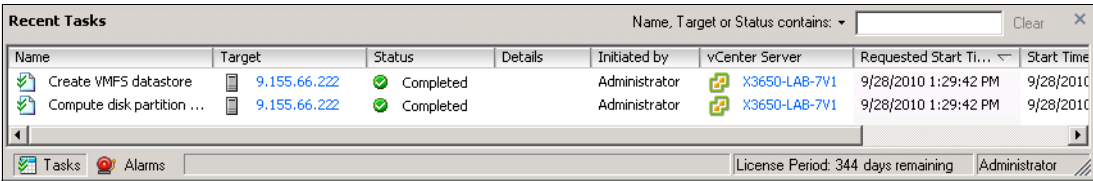


Figure 9-25 Tasks related to datastore creation

Set up the round-robin policy for the new datastore by following these steps:

1. From the vSphere client main window, as shown in Figure 9-26, you can see a list of all datastores, including the new one you just created.

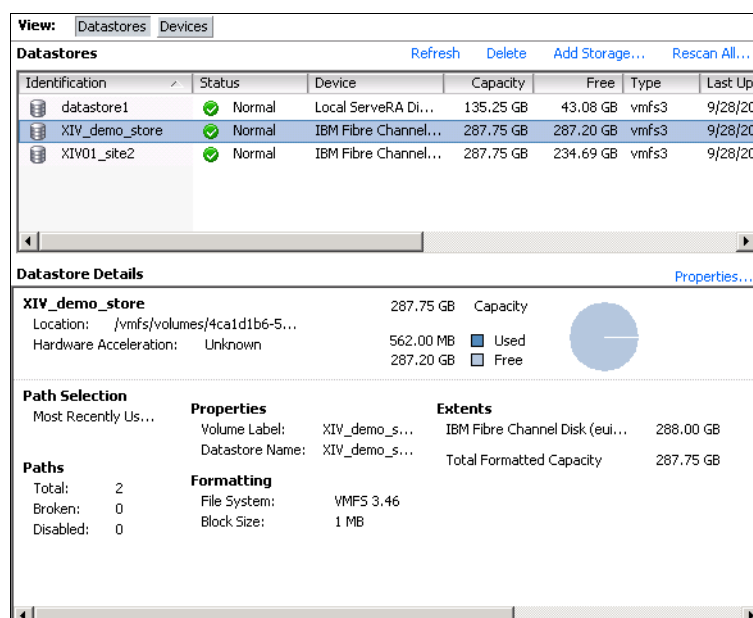


Figure 9-26 Datastores updated list

2. Select the datastore, then click **Properties** to display the Properties window shown in Figure 9-27. At the bottom of the datastore Properties window click **Manage Paths**.

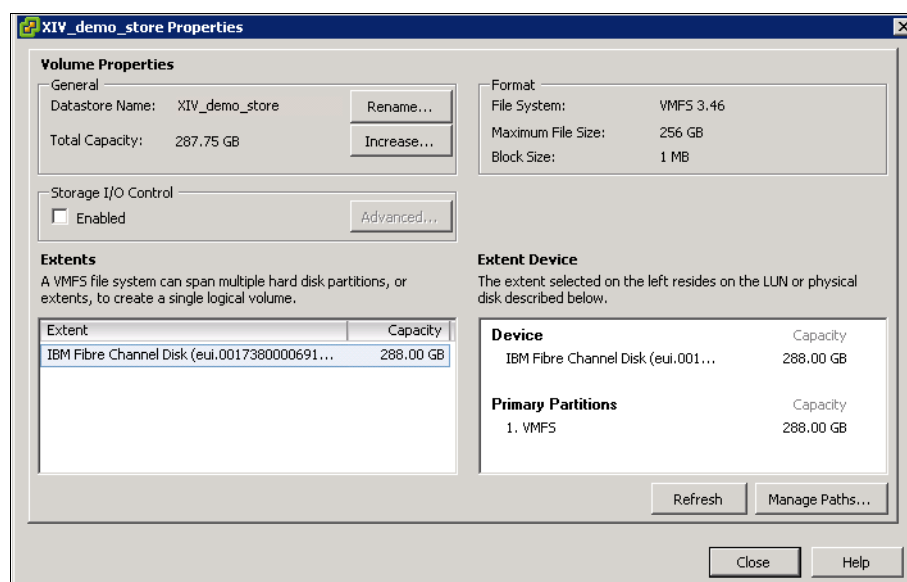


Figure 9-27 Datastore properties

3. The Manage Paths window shown in Figure 9-28 displays. Select any of the vmhba's listed.

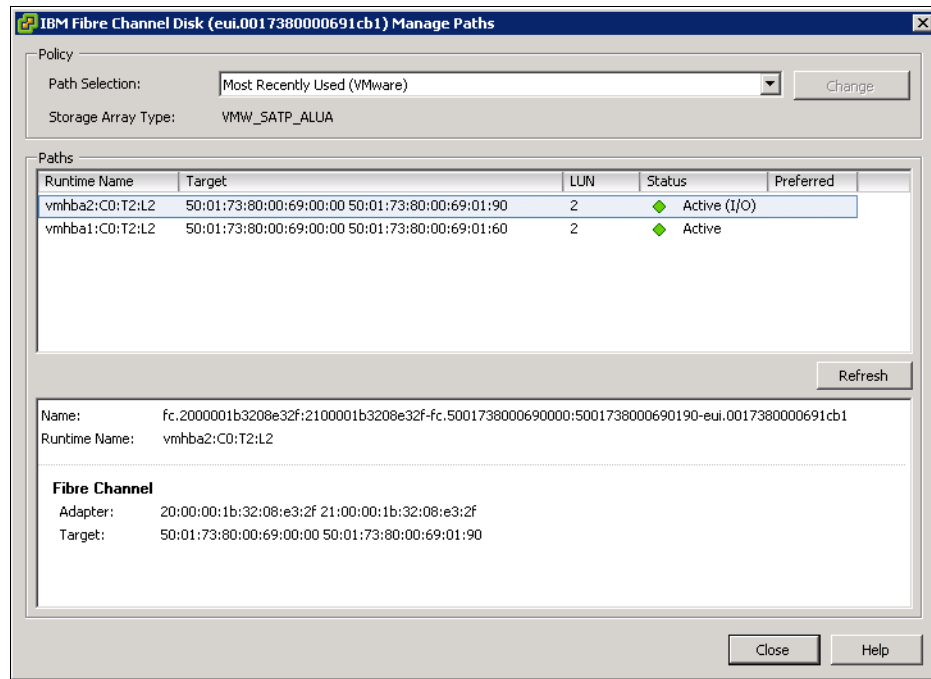


Figure 9-28 Manage path window

4. Click the **Path selection** pull-down as shown in Figure 9-29 and select **Round Robin (VMWare)**. Press the **Change** button.

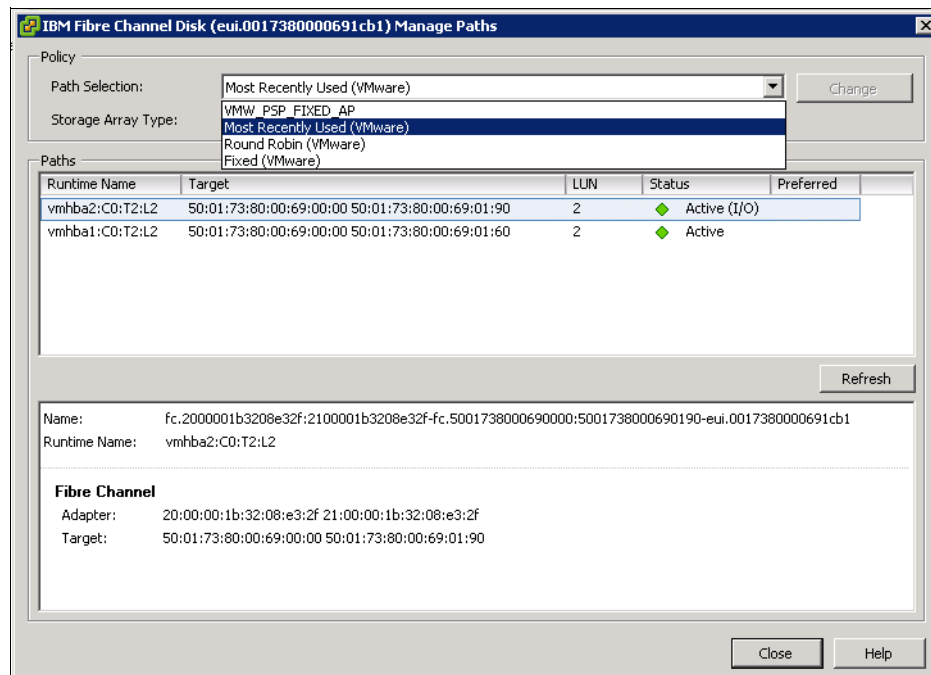


Figure 9-29 List of the path selection options

5. The Manage Paths window shown in Figure 9-30 is now displayed.

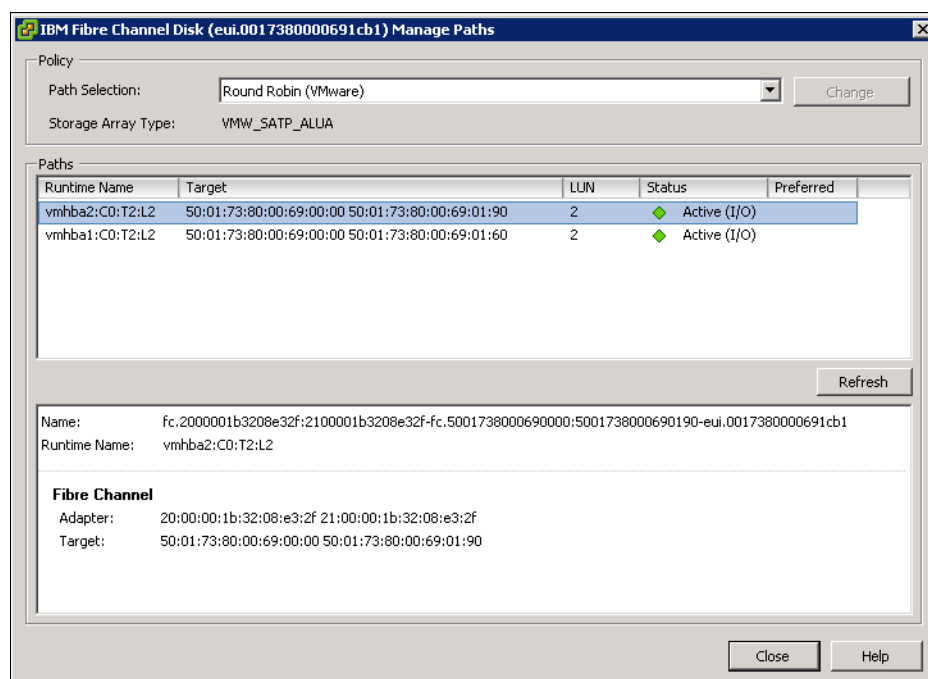


Figure 9-30 Datastore paths with selected round robin policy for multipathing

Your ESX host is now connected to the XIV Storage System with the proper settings for multipathing. If you have previously created datastores that are located on the IBM XIV Storage System but the round-robin policy on VMWare multipathing was not applied, you can apply the process presented above to those datastores.

9.3.6 Performance tuning tips for ESX 4 hosts with XIV

Review your ESX additional settings to see if they affect performance in your environment and with applications you normally use.

Settings to consider include:

- ▶ Leverage large LUNs for datastore.
- ▶ Do not use LVM extents instead of using large LUNs.
- ▶ Use a smaller number of large LUNs rather than many small LUNs.
- ▶ Increase the queue size for outstanding IOs on HBA and VMWare kernel levels.
- ▶ Use all available paths for round-robin.
- ▶ Decrease the number of IOs executed by one path.

Queue size for outstanding IOs

In general, it is not necessary to change the HBA queue depth and the corresponding **Disk.SchedNumReqOutstanding** VMWare kernel parameter.

If your workload is such that you need to change the queue depth, proceed as follows:

1. Log on to the service console as root
 - For Emulex HBAs:

- i. Verify which Emulex HBA module is currently loaded as shown in Example 9-3

Example 9-3 Emulex HBA module identification

```
#vmkload_mod -l|grep lpfc
lpfc820          0x418028689000    0x72000    0x417fe9499f80    0xd000 33 Yes
```

- ii. Set the new value for the **queue_depth** parameter and check that new values are applied as shown in Example 9-4.

Example 9-4 Setting new value for queue_depth parameter on Emulex FC HBA

```
# esxcfg-module -s lpfc0_lun_queue_depth=64 lpfc820
# esxcfg-module -g lpfc820
lpfc820 enabled = 1 options = 'lpfc0_lun_queue_depth=64'
```

– For Qlogic HBAs:

- i. Verify which Qlogic HBA module is currently loaded as shown in Example 9-5.

Example 9-5 Qlogic HBA module identification

```
#vmkload_mod -l|grep qla
qla2xxx          2    1144
```

- ii. Set the new value for **queue_depth** parameter and check that new value is applied as shown in Example 9-6.

Example 9-6 Setting new value for queue_depth parameter on Qlogic FC HBA

```
# esxcfg-module -s ql2xmaxqdepth=64 qla2xxx
# esxcfg-module -g qla2xxx
qla2xxx enabled = 1 options = 'ql2xmaxqdepth=64'
```

You can also change the **queue_depth** parameters on your HBA using tools or utilities provided by your HBA vendor.

To change the corresponding **Disk.SchedNumReqOutstanding** parameter in the VMWare kernel, after changing the HBA queue depth, proceed as follows:

1. Launch the VMWare vSphere client and choose the server for which you plan to change the settings.
2. Click the **Configuration** tab under **Software** section and click **Advanced Settings** to display the Advanced Settings window.
3. Select **Disk** (circled in green in Figure 9-31 on page 215) and set the new value for **Disk.SchedNumReqOutstanding** (circled in red on Figure 9-31 on page 215). Then click **OK** at the bottom of active window to save your changes.

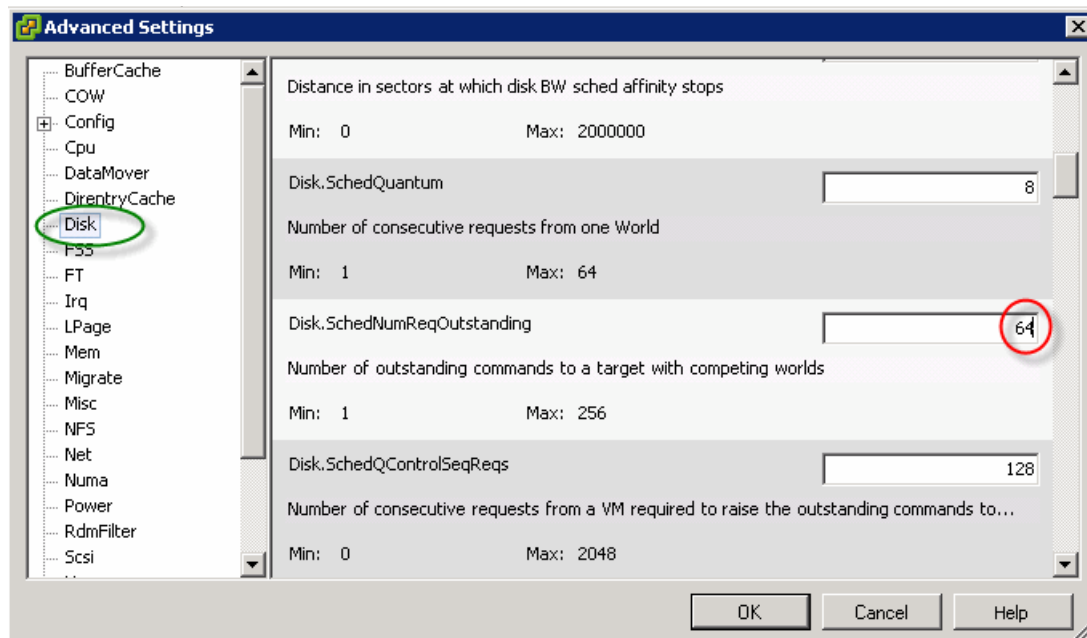


Figure 9-31 Changing `Disk.SchedNumReqOutstanding` parameter in VMware ESX 4

Tune multipathing settings for round-robin

Important: The default ESX VMware settings for round-robin are adequate for most workloads and should not normally be changed.

If, after observing your workload, you decide that the default settings need to be changed, you can enable the non-optimal use for round-robin and decrease the number of IOs going over each path. This can help the ESX host utilize more resources on the XIV Storage System.

If you determine that a change is required, follow these instructions:

1. Launch the VMware vSphere client and connect to the vCenter server.
2. From the vSphere Client select your server, click the **Configuration** tab and select **Storage** in the **Hardware** section as shown in Figure 9-32.

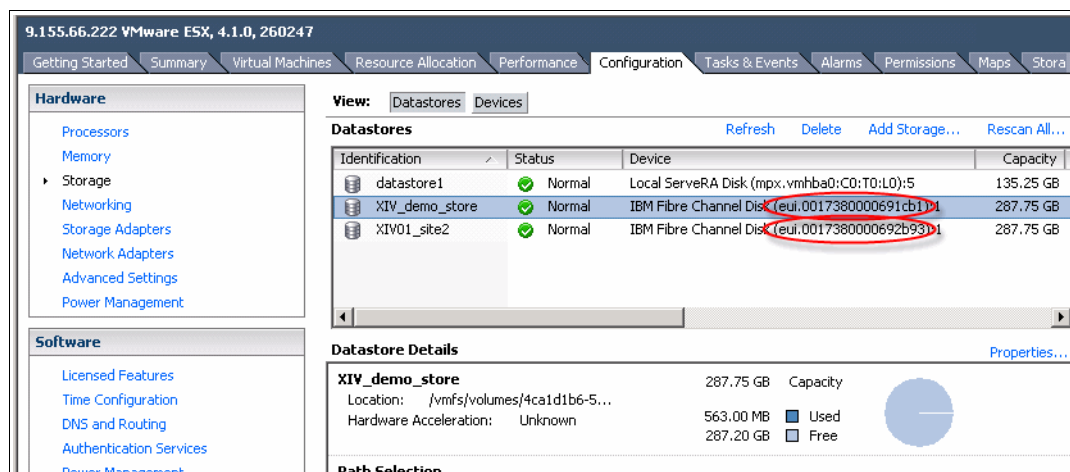


Figure 9-32 Identification of device identifier for your datastores

Here you can view the device identifier for your datastores (circled in red).

3. Log on to the service console as root.
4. Enable use of non-optimal paths for round-robin with the **esxcli** command as shown in Example 9-7.

Example 9-7 Enable use of non-optimal paths for round-robin on ESX 4 host

```
#esxcli nmp roundrobin setconfig --device eui.0017380000691cb1 --useANO=1
```

5. Change the number of I/Os executed over each path, as shown in Example 9-8. In this example we use a value of 10 for an extremely heavy workload. Leave the default (1000) for normal workloads.

Example 9-8 Change the number of I/Os executed over one path for round-robin algorithm

```
# esxcli nmp roundrobin setconfig --device eui.0017380000691cb1 --iops=10 --type "iops"
```

6. Check that your settings have been applied as illustrated in Example 9-9.

Example 9-9 Check the round-robin options on datastore

```
#esxcli nmp roundrobin getconfig --device eui.0017380000691cb1
Byte Limit: 10485760
Device: eui.0017380000691cb1
I/O Operation Limit: 10
Limit Type: Iops
Use Active Unoptimized Paths: true
```

If you have multiple datastores for which you need to apply the same settings, you can also use a script similar to the one shown in Example 9-10.

Example 9-10 Setting round-robin tweaks for all IBM XIV Storage System devices connected to ESX host

```
# for i in `ls /vmfs/devices/disks/ | grep eui.001738*|grep -v \:~ ; \
> do echo "Update settings for device" $i ; \
> esxcli nmp roundrobin setconfig --device $i --useANO=1;\
> esxcli nmp roundrobin setconfig --device $i --iops=10 --type "iops";\
> done
Update settings for device eui.0017380000691cb1
Update settings for device eui.0017380000692b93
```

9.3.7 Managing ESX 4 with IBM XIV Management Console for VMWare vCenter

The IBM XIV Management Console for VMware vCenter is a plug-in that integrates into the VMware vCenter Server and manages XIV systems.

The IBM XIV Management Console for VMware vCenter installs a service on the VMware vCenter Server. This service queries the VMware software development kit (SDK) and the XIV systems for information that is used to generate the appropriate views.

After you have configured the IBM XIV Management Console for VMware vCenter, new tabs are added to the VMware vSphere Client. You can access the tabs from the Datacenter, Cluster, Host, Datastore, and Virtual Machine inventory views.

From the XIV tab, you can view the properties for XIV volumes that are configured in the system as shown in Figure 9-33.

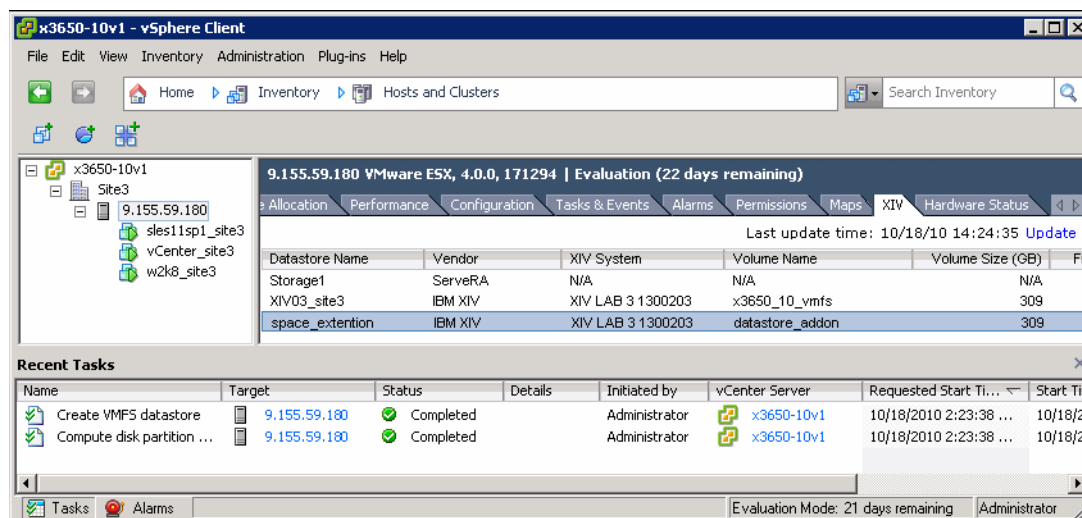


Figure 9-33 XIV tab view in VMWare vCenter Client console

The IBM XIV Management Console for VMWare vCenter is available for download from:

<http://www.ibm.com/support/docview.wss?uid=ssg1S4000884&myns=s028&mynp=familyind5368932&mync=E>

For installation instructions, refer to IBM XIV Management Console for VMWare vCenter User Guide at:

<http://publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/GA32-0820-01.pdf>

9.4 XIV Storage Replication Agent for VMware SRM

In normal production, the virtual machines (VMs) run on ESX hosts and storage devices located in the primary datacenter while additional ESX servers and storage devices are stand by in the backup datacenter.

Mirroring functions of the storage subsystems create a copy of the data on the storage device at the backup location. In a failover case, all VMs will be shut down at the primary site (if still possible/required) and will be restarted on the ESX hosts at the backup datacenter, accessing the data on the backup storage system. Doing all this requires a lot of steps including stopping any running VMs at the primary side, stopping the mirroring, making the copy accessible to the backup ESX servers, and registering and restarting the VMs on the backup ESX servers.

VMware SRM can automatically perform all these steps and failover complete virtual environments in just one click. This saves time, eliminates user errors, and provides a detailed documentation of the disaster recovery plan. SRM can also perform a test of the failover plan by creating an additional copy of the data at the backup site and starting the virtual machines from this copy without actually connecting them to any network. This enables administrators to test recovery plans without interfering with the running production.

A minimum setup for SRM will contain a two ESX servers (one at each site), a vCenter for each site, and two storage systems (one on each site). The storage systems need to be in a copy services relationship. Ethernet connectivity between the two datacenter is also required for the SRM to work.

Details on installing, configuring and using VMware Site Recovery Manager can be found on the VMware website at:

http://www.vmware.com/support/pubs/srm_pubs.html

Integration with storage systems require a Storage Replication Agent specific to the storage system. A Service Replication Agent is available for the IBM XIV Storage System.

At the time of writing this book, the IBM XIV Storage Replication Agent supports the following versions of VMware SRM server:

- ▶ 1.0
- ▶ 1.0 U1
- ▶ 4.0 and 4.1



Citrix XenServer connectivity

This chapter explains the basics of server virtualization with the Citrix XenServer, and discusses considerations for attaching XIV to the Citrix XenServer.

For the latest information about the Citrix XenServer visit:

<http://www.citrix.com/English/ps2/products/product.asp?contentID=683148>

The documentation is available in PDF format at:

http://docs.vmd.citrix.com/XenServer/5.6.0/1.0/en_gb/

10.1 Introduction

The development of virtualization technology continues to grow, offering new opportunities to use data center resources more and more effectively. Nowadays companies are using virtualization to minimize their Total Cost of Ownership (TCO), making it important to remain up-to-date with new technologies to reap the benefits of virtualization in terms of server consolidation, disaster recovery, or high availability.

The storage of the data is an important aspect of the overall virtualization strategy and it is critical to select an appropriate storage system that achieves a complete, complementary virtualized infrastructure. In comparison to other storage systems, the IBM XIV Storage System, with its grid architecture, automated load balancing, and exceptional ease of management, provides best-in-class virtual enterprise storage for virtual servers. IBM XIV and Citrix XenServer together can provide hot-spot-free server-storage performance with optimal resources usage. Together, they provide excellent consolidation, with performance, resiliency, and usability features that can help you reach your virtual infrastructure goals.

The Citrix XenServer consists of four editions:

- ▶ The Free edition is a proven virtualization platform that delivers uncompromised performance, scale, and flexibility.
- ▶ The Advanced edition includes high availability and advanced management tools that take virtual infrastructure to the next level.
- ▶ The Enterprise edition adds essential integration and optimization capabilities for deployments of virtual machines.
- ▶ The Platinum edition with advanced automation and cloud computing features can address the requirements for enterprise-wide virtual environments.

Figure 10-1 illustrates the editions and their corresponding features.

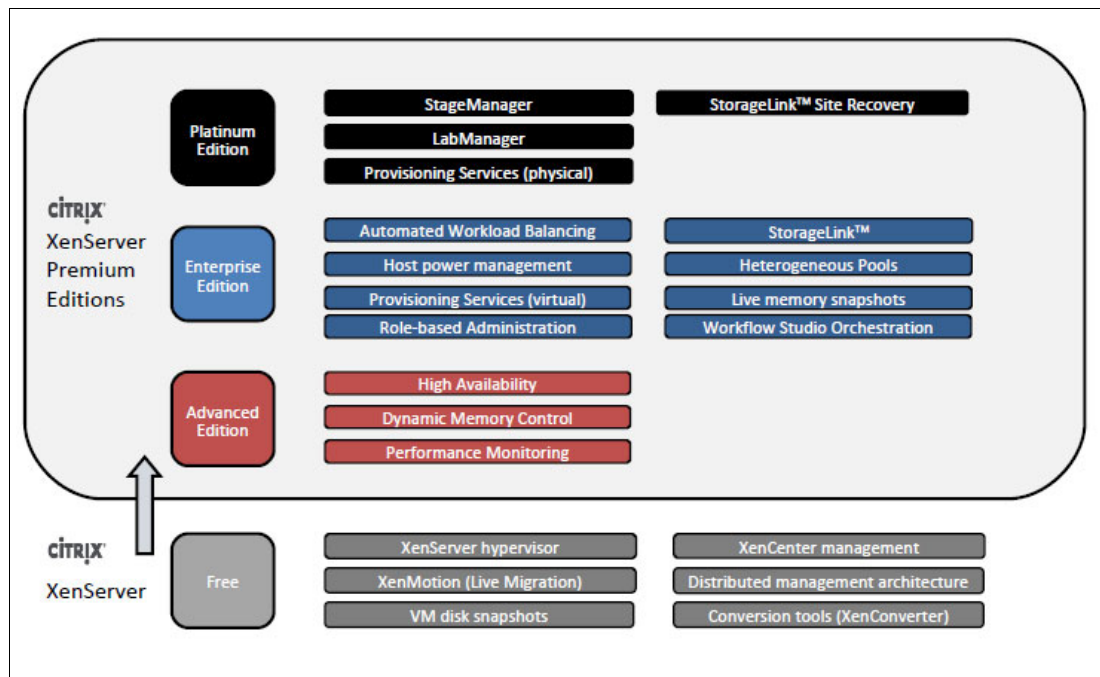


Figure 10-1 Citrix XenServer Family

Most of these features are similar to those of other hypervisors such as VMware, but there are also the following new and different features:

- XenServer hypervisor: hypervisors are installed directly onto a physical server, without requiring a host operating system as shown in Figure 10-2. The hypervisor controls the hardware and monitors guest operating systems that have to share specific physical resources.

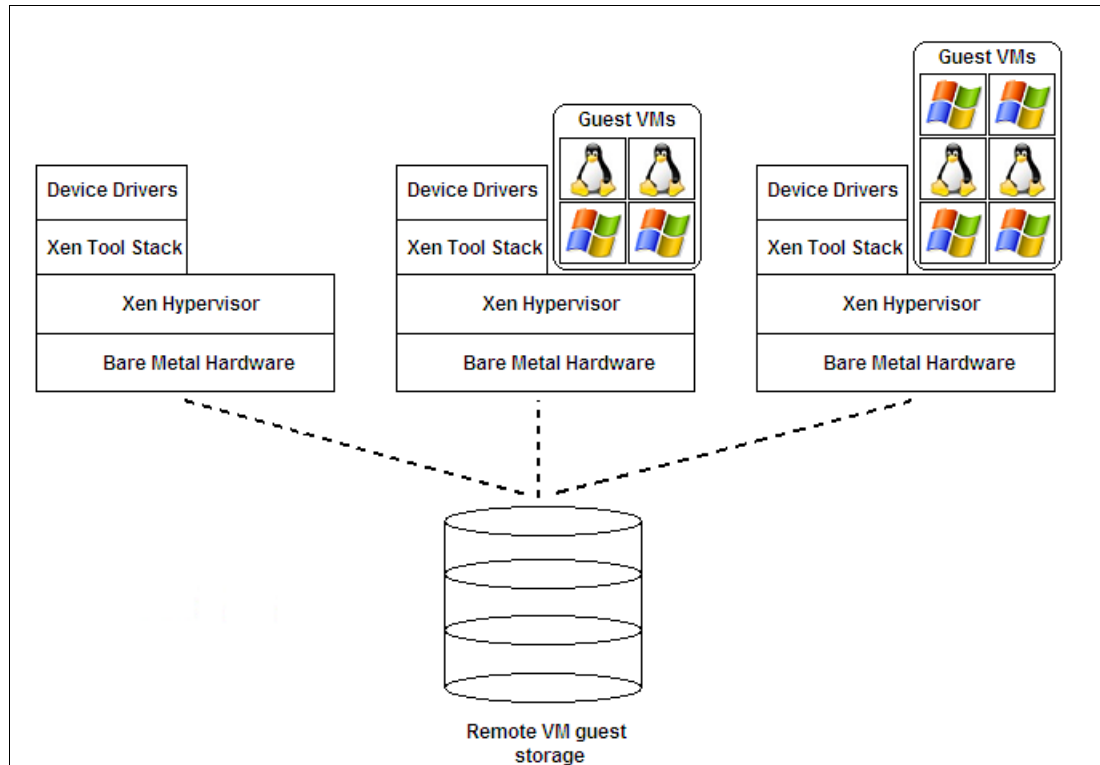


Figure 10-2 XenServer hypervisor

- XenMotion (Live migration): enables the live migration of running virtual machines from one physical server to another with zero downtime, continuous service availability, and complete transaction integrity.
- VMs disk snapshots: snapshots provide a “point in time” image of disk state and are useful for virtual machine backup.
- XenCenter management: Citrix XenCenter offers monitoring, management, and general administrative functions for VM from a single interface. This interface allows easy management of hundreds of virtual machines.
- Distributed management architecture: this architecture prevents a single point of failure from bring down all servers across an entire data center.
- Conversion Tools (Citrix XenConverter): XenConverter can convert a server or desktop workload to a XenServer virtual machine. It also allows migration of physical and virtual servers (P2V and V2V).
- High availability: this feature allows you to restart virtual machine after it was affected by a server failure. The auto-restart functionality enables the protection of all virtualized applications and increases the availability of business operations.
- Dynamic memory control: can change the amount of host physical memory assigned to any running virtual machine without rebooting it. It is also possible to start additional virtual

machine on a host whose physical memory is currently full by automatically reducing the memory of the existing virtual machines.

- ▶ Workload balancing: automatically places VMs on the most suitable host in the resource pool.
- ▶ Host power management: with fluctuating demand of IT services XenServer automatically adapts to changing requirements. VMs can be consolidated and under utilized Servers can be switched off.
- ▶ Provisioning services: reduces total cost of ownership, and improves manageability and business agility by virtualizing the workload of a data center server, streaming server workloads on demand to physical or virtual servers in the network.
- ▶ Role based administration: buser access rights XenServer role-based administration features improve the security and allow authorized access, control, and use of XenServer pools.
- ▶ Storage Link: allows easy integration of leading network storage systems. Data management tools can be used to maintain consistent management processes for physical and virtual environments.
- ▶ Site Recovery: offers cross-location disaster recovery planning and services for virtual environments.
- ▶ LabManager: this web-based application enables you to automate your virtual lab setup on virtualization platforms. LabManager automatically allocates infrastructure, provisions operating systems, sets up software packages, installs your development and testing tools, and downloads required scripts and data to execute an automated job or manual testing jobs.
- ▶ StageManager: automates the management and deployment of multi-tier application environments and other IT services.

The Citrix XenServer supports the following operating systems as VMs:

Windows

- ▶ Windows Server 2008 64-bit & 32-bit & R2
- ▶ Windows Server 2003 32-bit SP0, SP1, SP2, R2; 64-bit SP2
- ▶ Windows Small Business Server 2003 32-bit SP0, SP1, SP2, R2
- ▶ Windows XP 32-bit SP2, SP3
- ▶ Windows 2000 32-bit SP4
- ▶ Windows Vista 32-bit SP1
- ▶ Windows 7

Linux

- ▶ Red Hat Enterprise Linux 32-bit 3.5-3.7, 4.1-4.5, 4.7 5.0-5.3; 64-bit 5.0-5.4
- ▶ Novell SUSE Linux Enterprise Server 32-bit 9 SP2-SP4; 10 SP1; 64-bit 10 SP1-SP3, SLES 11 (32/64)
- ▶ CentOS 32-bit 4.1-4.5, 5.0-5.3; 64-bit 5.0-5.4
- ▶ Oracle Enterprise Linux 64-bit & 32-bit 5.0-5.4
- ▶ Debian Lenny (5.0)

10.2 Attaching a XenServer host to XIV

This section includes general information and required tasks for attaching the Citrix XenServer to the IBM XIV Storage System.

10.2.1 Prerequisites

To successfully attach an XenServer host to XIV and assign storage, a number of prerequisites need to be met. The following is a generic list. Your environment might have additional requirements.

- ▶ Complete the cabling.
- ▶ Configure the SAN zoning.
- ▶ Install any service packs and updates if required.
- ▶ Create volumes to be assigned to the host.

Supported Hardware

Information about the supported hardware for XenServer can be found in the XenServer Hardware Compatibility List at:

<http://hcl.xensource.com/BrowsableStorageList.aspx>

Supported versions of Xenserver

At the time of writing, the XenServer 5.6.0 is supported for attachment with XIV.

10.2.2 Multi-path support and configuration

The Citrix XenServer supports dynamic multipathing with Fibre Channel and iSCSI storage back-ends. By default, it uses a round-robin mode for load balancing. Both paths will carry I/O traffic during normal operations. To enable multipathing you can use the xCLI or the XenCenter. In this section we illustrate how to enable multipathing using the XenCenter GUI.

To enable multipathing using XenCenter you have to differentiate between two cases:

- ▶ If there are only local Storage Repositories (SR), follow these steps:
 - a. Enter maintenance mode on the chosen server as shown in Figure 10-3 on page 224. Entering maintenance mode will migrate all running VMs from this server. If this server is the pool master a new master will be nominated for the pool and XenCenter will temporarily lose its connection to the pool.
 - b. Right-click the server that is in maintenance mode, select **Properties**, click the **Multipathing** tab, and select **Enable multipathing on this server**. See Figure 10-4 on page 224.
 - c. Exit Maintenance Mode the same way like you have entered it. You will be asked to restore your VMs to their previous host. Restore the VMs.
 - d. Repeat first 3 steps on each XenServer in the pool.
 - e. Now you can create your Storage Repositories, which will go over multiple paths automatically.
- ▶ If there are existing SRs on the host running in single path, follow these steps:
 - a. Migrate or suspend all virtual machines running out of the SRs.
 - b. To find and unplug the Physical Block Devices you will need the SR uuid. Open the console tab and type **#xe sr-list**; This command will display all SRs and the corresponding uuids.
 - c. Find the Physical Block Devices (PBDs) that represent the interface between a physical server and an attached Storage Repository using the following command:

```
# xe sr-list uuid=<sr-uuid> params=all
```
 - d. Unplug the Physical Block Devices (PBDs) using the following command:

```
# xe pbd-unplug uuid=<pbid_uuid>
```

- e. Enter Maintenance Mode on the server as shown in Figure 10-3.

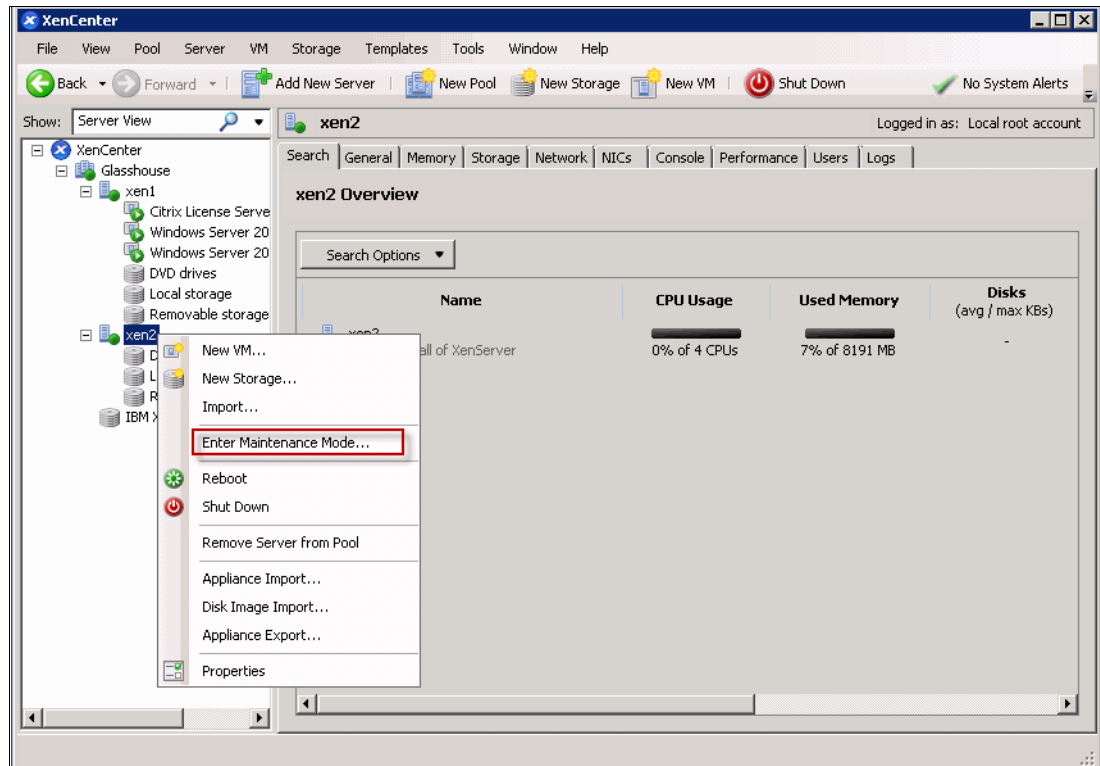


Figure 10-3 Set the XenServer to the maintenance mode.

- f. Enable multipathing by opening the server's Properties page, click the Multipathing tab, and select the **Enable multipathing on this server** check box as shown in Figure 10-4.

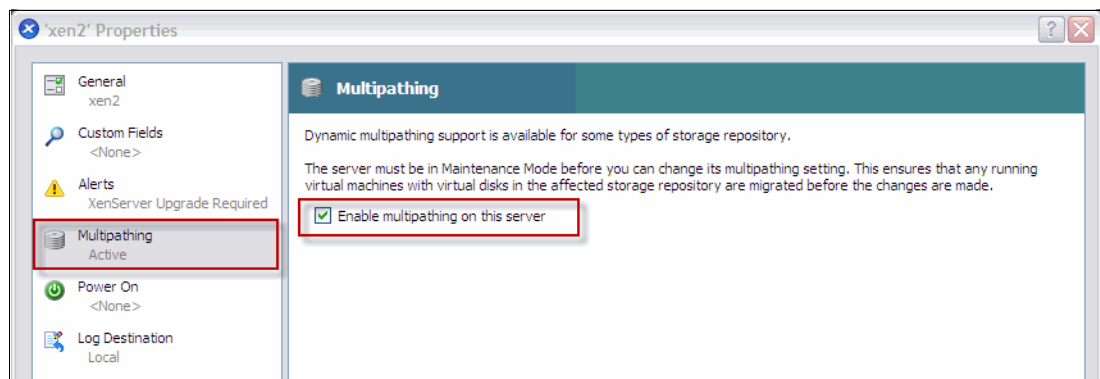


Figure 10-4 Enable multipathing for the chosen XenServer

- g. Exit Maintenance Mode.
h. Repeat steps 5, 6, and 7 on each XenServer in the pool.

10.2.3 Attachment tasks

This section describes the attachment of XenServer based hosts to the XIV Storage System. It provides specific instructions for Fibre Channel (FC) connections. All information in this section relates to XenServer 5.6 exclusively, unless otherwise specified.

Scanning for new LUNs

To scan for new LUNs, the XenServer host needs to be added and configured in XIV (see Chapter 1, “Host connectivity” on page 1 for information about setup).

The XenServer hosts that need to access the same shared LUNs must be grouped in a cluster (XIV cluster) and the LUNs must be assigned to the cluster. Refer to Figure 10-5 and Figure 10-6 for how this might typically be set up.

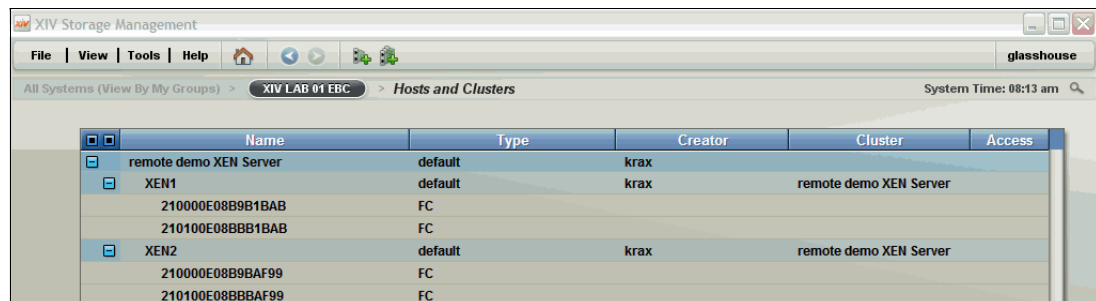


Figure 10-5 XenServer host cluster setup in XIV GUI

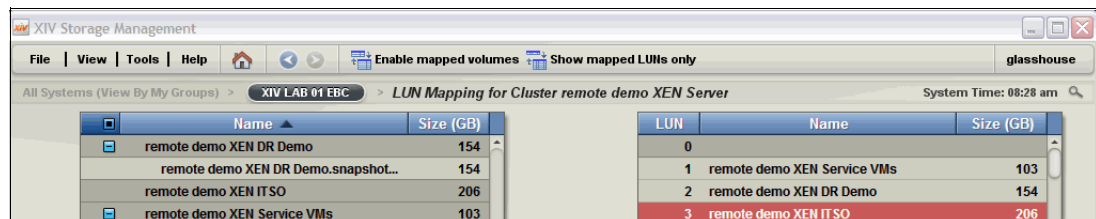


Figure 10-6 XenServer LUN mapping to the cluster

To create a new Storage Repository (SR), follow these instructions:

1. After the host definition and LUN mappings have been completed in the XIV Storage System, open XenCenter and choose a pool or a host to attach the new SR. As you can see in Figure 10-7 on page 226, you have two possibilities (highlighted with a red rectangle in the picture) to create new Storage Repository. It does not matter which button you click, the result will be the same.

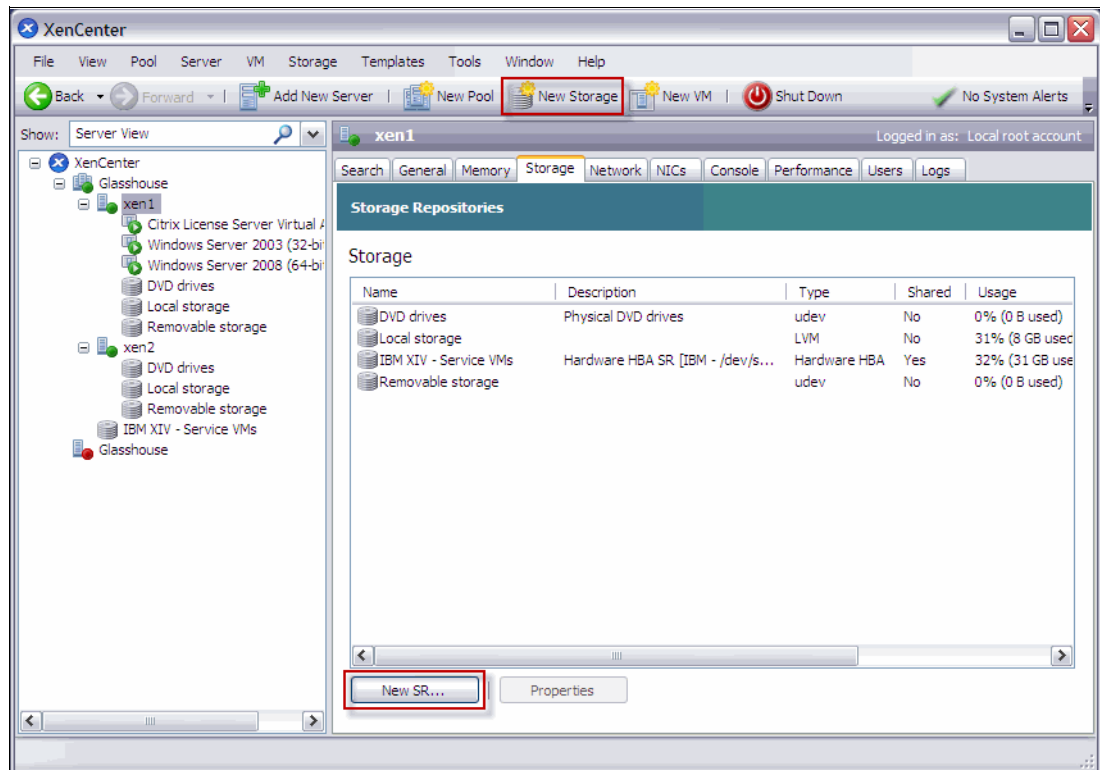


Figure 10-7 Attaching new Storage Repository

- The Choose the type of new storage window opens as shown in Figure 10-8. Select **Hardware HBA** and click **Next**. The XenServer will be probing for LUNs, and open a new window with the LUNs that were found.

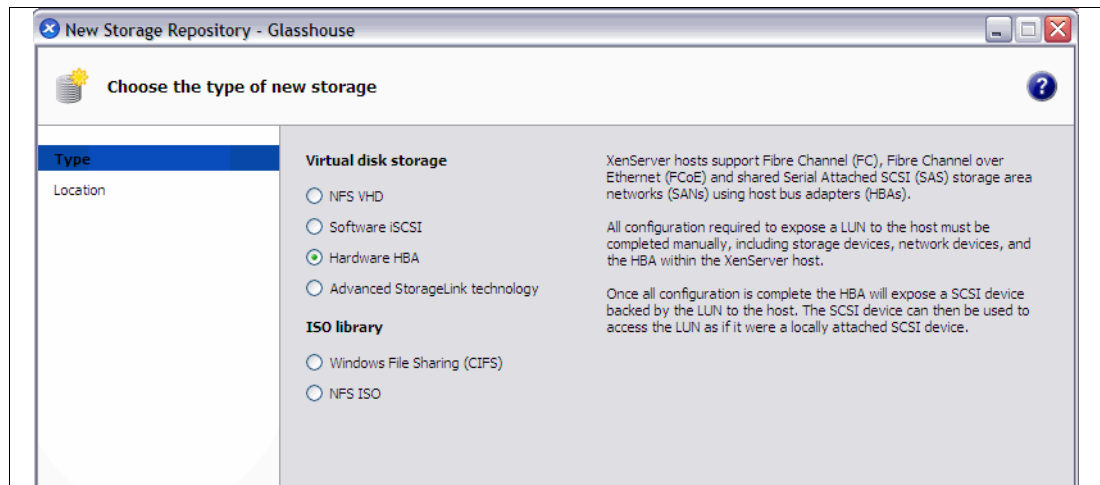


Figure 10-8 Choosing storage type

- The Select the LUN to reattach or create a new SR on window displays as shown in Figure 10-9 on page 227. In the **Name** field, type a meaningful name for your new SR. That will help you differentiate and identify SRs if you will have more of them in the future. In the box below the Name, you can see the LUNs that were recognized as a result of the LUN probing. The first one is the LUN we have added to the XIV. Select the LUN and click **Finish** to complete the configuration. XenServer will start to attach the SR, creating Physical Block Devices (PBDs) and plugging PBDs to each host in the pool.

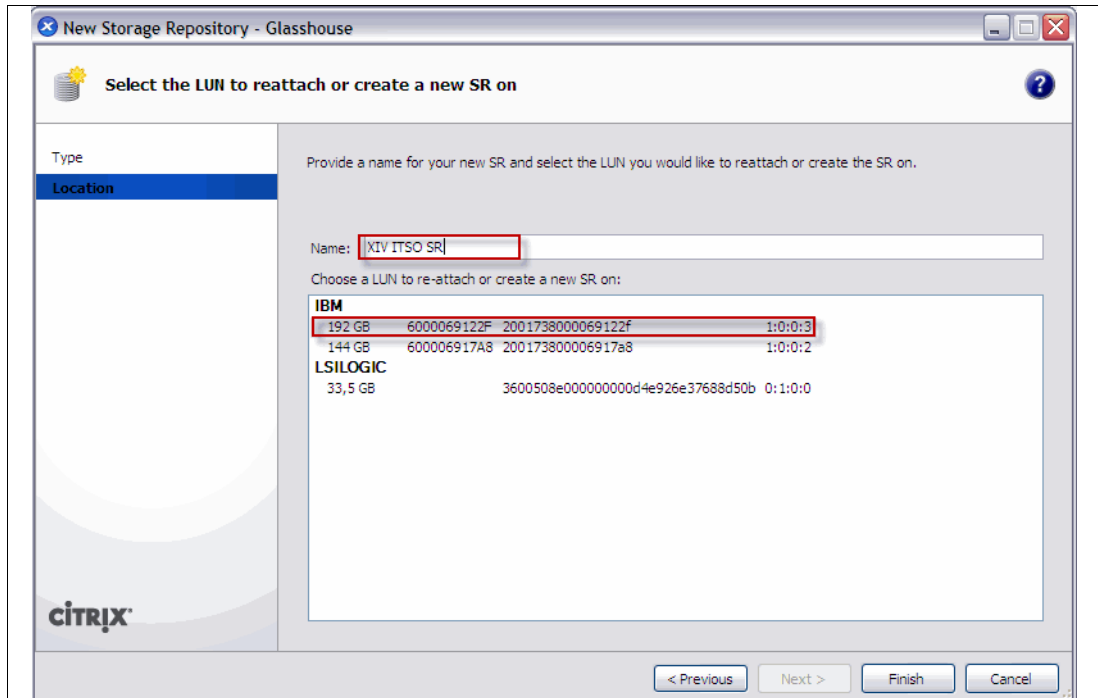


Figure 10-9 Selecting LUN to create or reattach new SR

4. In order to validate your configuration see the Figure 10-10 (the attached SR is marked red here).

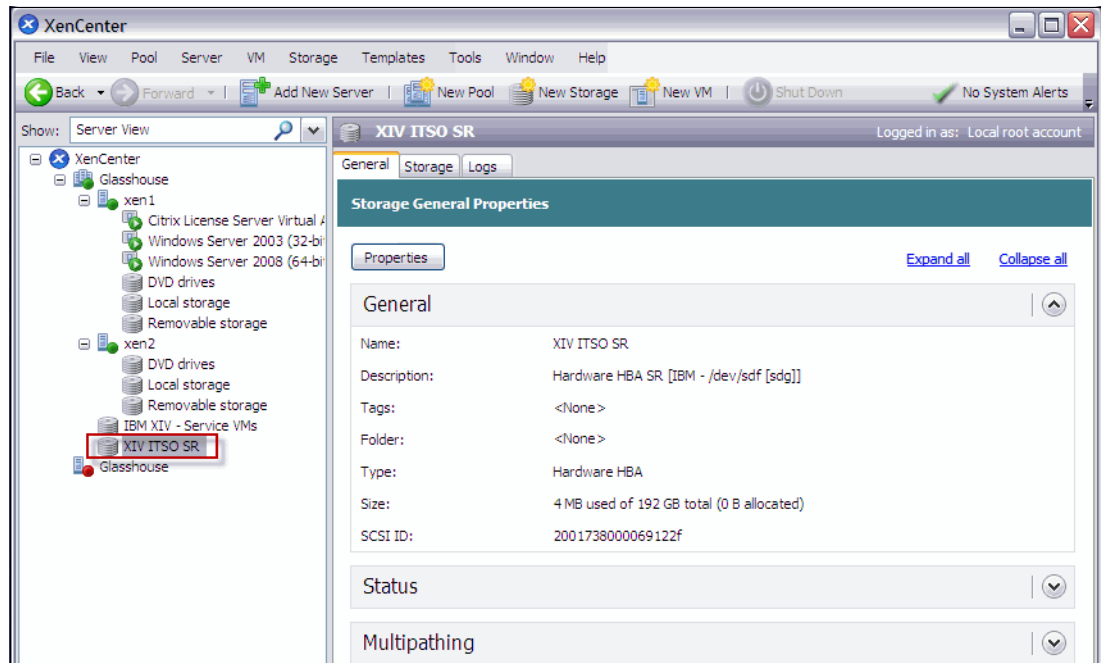


Figure 10-10 Attached SR



SVC specific considerations

This chapter discusses specific considerations for attaching the XIV Storage System to a SAN Volume Controller (SVC).

11.1 Attaching SVC to XIV

When attaching the SAN Volume Controller (SVC) to XIV, in conjunction with connectivity guidelines already presented in Chapter 1, “Host connectivity” on page 1, the following considerations apply:

- ▶ Supported versions of SVC
- ▶ Cabling considerations
- ▶ Zoning considerations
- ▶ XIV Host creation
- ▶ XIV LUN creation
- ▶ SVC LUN allocation
- ▶ SVC LUN mapping
- ▶ SVC LUN management

11.2 Supported versions of SVC

At the time of writing, currently, SVC code v4.3.0.1 and later are supported when connecting to the XIV Storage System. For up-to-date information, refer to:

<http://www.ibm.com/support/docview.wss?rs=591&uid=ssg1S1003277#XIV>

For specific information regarding SVC code, refer to the SVC support page located at:

<http://www.ibm.com/systems/support/storage/software/sanvc>

The SVC supported hardware list, device driver, and firmware levels for the SAN Volume Controller can be viewed at:

<http://www.ibm.com/support/docview.wss?rs=591&uid=ssg1S1003277>

Information about the SVC 4.3.x Recommended Software Levels can be found at:

<http://www.ibm.com/support/docview.wss?rs=591&uid=ssg1S1003278>

Although SVC supports the IBM XIV System with a minimum SVC Software level of 4.3.0.1, use SVC software v4.3.1.4 or later for best results.

At the time of writing this book, a special edition of IBM System Storage SAN Volume Controller for XIV version 6.1 introduces a new licensing scheme for IBM XIV Storage System, based on the number of XIV modules installed, rather than the available capacity.

Cabling considerations

The IBM XIV supports both iSCSI and Fibre Channel protocols but when connecting to SVC, only Fibre Channel ports can be utilized.

To take advantage of the combined capabilities of SVC and XIV, you should connect two ports from every interface module into the fabric for SVC use. You need to decide which ports you to use for the connectivity. If you do not use and do not have plans to use XIV functionality for remote mirroring or data migration, you **MUST** change the role of port 4 from initiator to target on all XIV interface modules and use ports 1 and 3 from every interface module into the fabric for SVC use. Otherwise, you **MUST** use ports 1 and 2 from every interface modules instead of ports 1 and 3.

Figure 11-1 on page 231 shows a two node cluster connected using redundant fabrics.

In this configuration:

- ▶ Each SVC node is equipped with four FC ports. Each 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.

This configuration has no single point of failure:

- ▶ If a module fails, each SVC host remains connected to 5 other modules.
- ▶ If an FC switch fails, each node remains connected to all modules.
- ▶ If an SVC HBA fails, each host remains connected to all modules.
- ▶ If an SVC cable fails, each host remains connected to all modules.

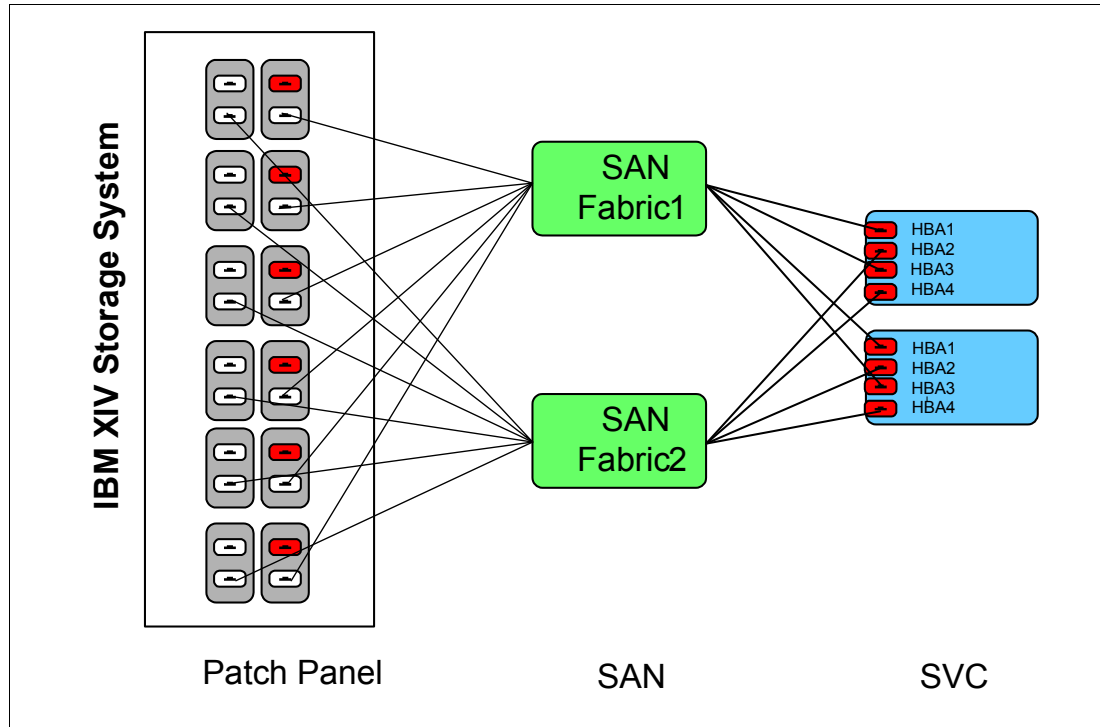


Figure 11-1 Two node SVC configuration with XIV

SVC supports a maximum of 16 ports from any disk system. The IBM XIV System supports from 8 to 24 FC ports, depending on the configuration (from 6 to 15 modules). Figure 11-2 indicates port usage for each IBM XIV System configuration.

Number of IBM XIV Modules	IBM XIV System Modules with FC Ports	Number of FC ports available on IBM XIV	Ports Used per Card on IBM XIV	Number of SVC ports utilized
6	Module 4, 5	8	1	4
9	Module 4, 5, 7, 8	16	1	8
10	Module 4, 5, 7, 8	16	1	8
11	Module 4, 5, 7, 8, 9	20	1	10
12	Module 4, 5, 7, 8, 9	20	1	10
13	Module 4, 5, 6, 7, 8, 9	24	1	12
14	Module 4, 5, 6, 7, 8, 9	24	1	12
15	Module 4, 5, 6, 7, 8, 9	24	1	12

Figure 11-2 Number of SVC ports and XIV Modules

SVC and IBM XIV system port naming conventions

The port naming convention for the IBM XIV System ports is:

- ▶ WWPN: 5001738NNNNNRRMP
 - 001738 = Registered identifier for XIV
 - NNNNN = Serial number in hex
 - RR = Rack ID (01)
 - M = Module ID (4-9)
 - P = Port ID (0-3)

The port naming convention for the SVC ports is:

- ▶ WWPN: 5005076801X0YYZZ
 - 076801 = SVC
 - X0 = first digit is the port number on the node (1-4)
 - YY/ZZ = node number (hex value)

Zoning considerations

As a best practice, a single zone containing all 12 XIV Storage System FC ports (in an XIV System with 15 modules) along with all SVC node ports (a minimum of eight) should be enacted when connecting the SVC into the SAN with the XIV Storage System. This any-to-any connectivity allows the SVC to strategically multi-path its I/O operations according to the logic aboard the controller, again making the solution as a whole more effective:

- ▶ Depending on your XIV, decide which ports you would use for the connectivity. If you do not use and doesn't have a plans to use XIV functionality for remote mirroring or data migration you MUST to change the role of the port 4 from initiator to target on all XIV interface modules and use ports 1 and 3 from every interface module into the fabric for SVC use, in other cases you MUST use ports 1 and 2 from every interface modules instead of ports 1 and 3. SVC nodes should connect to all Interface Modules using port 1 and port 3 on every module.
- ▶ Zones for SVC nodes should include all the SVC HBAs and all the storage HBAs (per fabric). Further details on zoning with SVC can be found in *Implementing the IBM System Storage SAN Volume Controller V4.3*, SG24-6423.

The zoning capabilities of the SAN switch are used to create distinct zones. The SVC in release 4 supports 1 Gbps, 2 Gbps, or 4 Gbps Fibre Channel fabrics. The SVC in release 5.1 and later adds support for 8 Gbps Fibre Channel Fabrics. This depends on the hardware platform and on the switch where the SVC is connected.

Connect the SVC and the disk subsystem to the switch operating at the highest speed in an environment where you have a fabric with multiple speed switches. All SVC nodes in the SVC cluster are connected to the same SAN, and present virtual disks to the hosts.

There are two distinct zones in the fabric:

- ▶ Host zones: These zones allow host ports to see and address the SVC nodes. There can be multiple host zones.
- ▶ Disk zone: There is one disk zone in which the SVC nodes can see and address the LUNs presented by XIV.

Creating a host object for SVC

Although a single host instance can be created for use in defining and then implementing the SVC, the ideal host definition for use with SVC is to consider each node of the SVC (a minimum of two) an instance of a cluster.

When creating the SVC host definition, first select **Add Cluster** and give the SVC host definition a name. Next, select **Add Host** and give the first node instance a **Name** making sure to click the **Cluster** drop-down box and select the SVC cluster you just created.

After these have been added, repeat the steps for each instance of a node in the cluster.

From there, right-click a node instance and select **Add Port**. In Figure 11-3, note that four ports per node can be added by referencing almost identical World Wide Port Names (WWPN) to ensure the host definition is accurate.

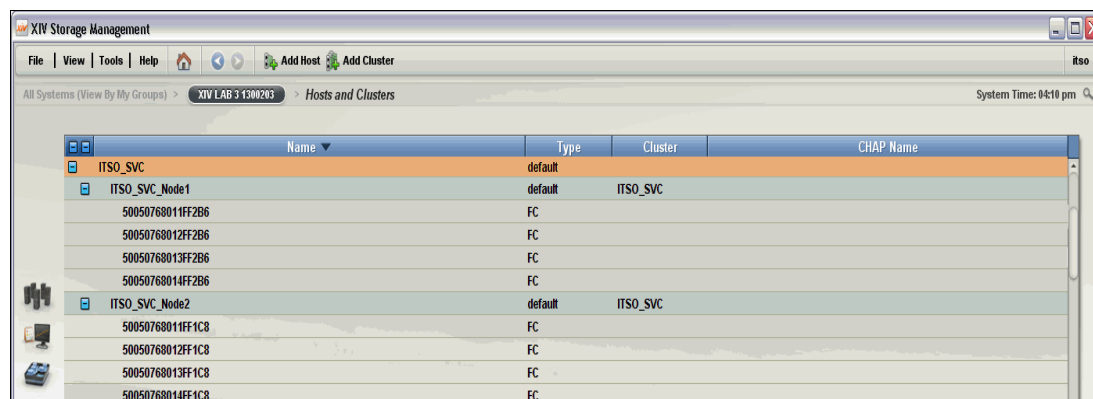


Figure 11-3 SVC host definition on XIV Storage System

By implementing the SVC as listed above, host management will ultimately be simplified and statistical metrics will be more effective because performance can be determined at the node level instead of the SVC cluster level.

For example, after the SVC is successfully configured with the XIV Storage System, if an evaluation of the VDisk management at the I/O Group level is needed to ensure efficient utilization among the nodes, a comparison of the nodes can be achieved using the XIV Storage System statistics as documented in *XIV Architecture, Implementation*, SG24-7659.

See Figure 11-4 on page 234 for a sample display of node performance statistics.

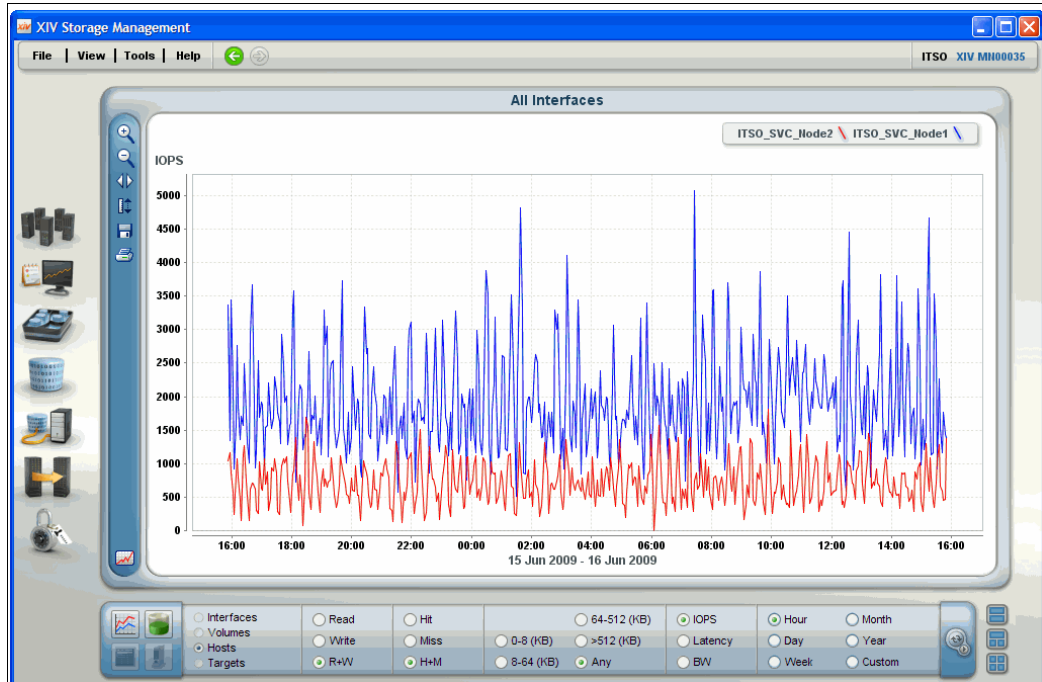


Figure 11-4 SVC node performance statistics on XIV Storage System

Volume creation for use with SVC

The IBM XIV System currently supports from 27 TB to 79 TB of usable capacity when using 1 TB drives, or from 55 TB to 161 TB when using 2 TB disks. The minimum volume size is 17 GB. Although smaller LUNs can be created, LUNs should be defined on 17 GB boundaries to maximize the physical space available.

SVC has a maximum LUN size of 2 TB that can be presented to it as a Managed Disk (MDisk). It has a maximum of 511 LUNs that can be presented from the IBM XIV System. SVC does not currently support dynamically expanding the size of the MDisk.

Note: At the time of this writing, a maximum of 511 LUNs from the XIV Storage System can be mapped to an SVC cluster.

For a fully populated rack, with 12 ports, you should create 48 volumes of 1632 GB each. This takes into account that the largest LUN that SVC can use is 2 TB.

Because the IBM XIV System configuration grows from 6 to 15 modules, use the SVC rebalancing script to restripe VDisk extents to include new MDisk. The script is located at:

<http://www.ibm.com/alphaworks>

From there, click the **All downloads** link and search for “svctools.”

Tip: Always use the largest volumes possible without exceeding the 2 TB limit of SVC.

Figure 11-5 shows the number of 1632 GB LUNs created, depending on the XIV capacity:

Number of LUNs (MDisks) at 1632GB each	IBM XIV System TB used	IBM XIV System TB Capacity Available
16	26.1	27
26	42.4	43
30	48.9	50
33	53.9	54
37	60.4	61
40	65.3	66
44	71.8	73
48	78.3	79

Figure 11-5 Values using 1632 GB LUNs

Restriction: The use of any XIV Storage System copy services functionality on LUNs presented to the SVC is not supported. Snapshots, thin provisioning, and replication are not allowed on XIV Volumes managed by SVC (MDisks).

LUN allocation using the SVC

The best use of the SVC virtualization solution with the XIV Storage System can be achieved by executing LUN allocation using these basic parameters:

- ▶ Allocate all LUNs, known to the SVC as MDisks, to one Managed Disk Group (MDG). If multiple IBM XIV Storage Systems are being managed by SVC, there should be a separate MDG for each physical IBM XIV System. We recommend that you do not include multiple disk subsystems in the same MDG because the failure of one disk subsystem will make the MDG go offline, and thereby all VDisks belonging to the MDG will go offline.
SVC supports up to 128 MDGs.
- ▶ When creating one MDG per XIV Storage System, use 1 GB or larger extent sizes because this large extent size ensures that data is striped across all XIV Storage System drives.

Figure 11-6 illustrates these two parameters, number of managed disks, and extent size, used in creating the MDG.

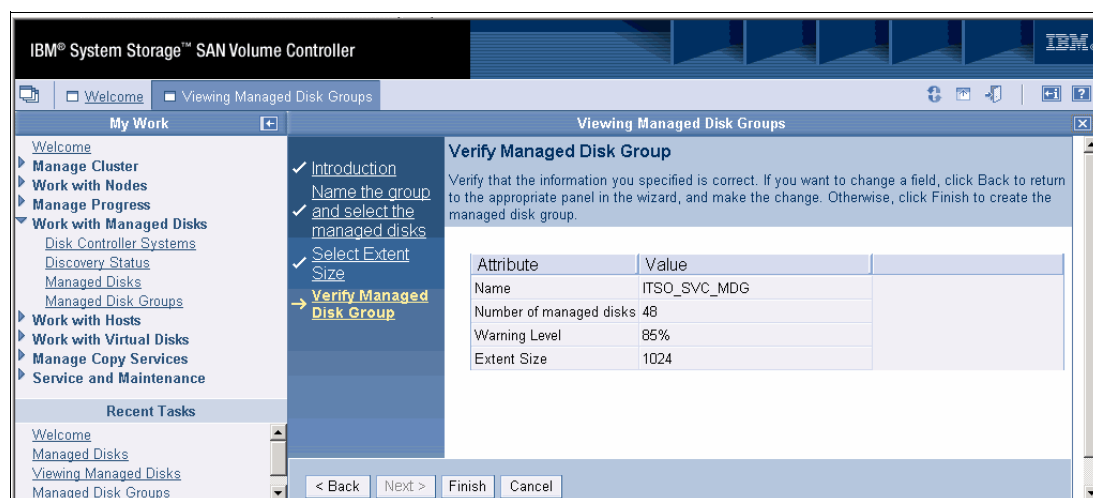


Figure 11-6 SVC Managed Disk Group creation

Doing so will drive I/O to the 4 MDisk/LUN for each of the 12 XIV Storage System Fibre Channel ports, resulting in an optimal queue depth on the SVC to adequately use the XIV Storage System.

Finalize the LUN allocation by creating striped VDIs for use by employing all 48 MDisk in the newly created MDG.

Queue depth

SVC submits I/O to the back-end storage (MDisk) in the same fashion as any direct-attached host. For direct-attached storage, the queue depth is tunable at the host and is often optimized based on specific storage type and various other parameters, such as the number of initiators. For SVC, the queue depth is also tuned. The optimal value used is calculated internally. The current algorithm used with SVC4.3 to calculate queue depth follows.

There are two parts to the algorithm: a per MDisk limit and a per controller port limit.

$$Q = ((P \times C) / N) / M$$

Where:

Q = The queue depth for any MDisk in a specific controller.

P = Number of WWPNs visible to SVC in a specific controller.

N = Number of nodes in the cluster.

M = Number of MDisk provided by the specific controller.

C = A constant. C varies by controller type:

- DS4100, and EMC Clarion = 200
- DS4700, DS4800, DS6K, DS8K and XIV = 1000
- Any other controller = 500

- If a 2 node SVC cluster is being used with a 6 module XIV system, 4 ports on the IBM XIV System and 16 MDisk, this will yield a queue depth that would be:

$$Q = ((4 \text{ ports} \times 1000) / 2 \text{ nodes}) / 16 \text{ MDisk} = 125$$

The maximum Queue depth allowed by SVC is 60 per MDisk

- If a 4 node SVC cluster is being used with a 15 module XIV system, 12 ports on the IBM XIV System and 48 MDisk, this will yield a queue depth that would be:

$$Q = ((12 \text{ ports} \times 1000) / 4 \text{ nodes}) / 48 \text{ MDisk} = 62$$

The maximum Queue depth allowed by SVC is 60 per MDisk.

SVC4.3.1 has introduced dynamic sharing of queue resources based on workload. MDisk with high workload can now borrow unused queue allocation from less busy MDisk on the same storage system. Although the values are calculated internally and this enhancement provides for better sharing, it is important to consider queue depth in deciding how many MDisk to create. In these examples, when SVC is at the maximum queue depth of 60 per MDisk, dynamic sharing does not provide additional benefit.

Striped, sequential or image mode VDisk guidelines

When creating a VDisk for host access, it can be created as Striped, Sequential, or Image Mode.

Striped VDIs provide for the most straightforward management. With Striped VDIs, they will be mapped to the number of MDisk in a MDG. All extents are automatically spread across all ports on the IBM XIV System. Even though the IBM XIV System already stripes the data across the entire back-end disk, we recommend that you configure striped VDIs.

Do not use Image Mode Disks unless it is for temporary purposes. Utilizing Image Mode disks creates additional management complexity with the one-to-one VDisk to MDisk mapping.

Each node presents a VDisk to the SAN through four ports. Each VDisk is accessible from the two nodes in an I/O group. Each HBA port can recognize up to eight paths to each LUN that is presented by the cluster. The hosts must run a multipathing device driver before the multiple paths can resolve to a single device. You can use fabric zoning to reduce the number of paths to a VDisk that are visible by the host. The number of paths through the network from an I/O group to a host must not exceed eight: configurations that exceed eight paths are not supported. Each node has four ports and each I/O group has two nodes. A VDisk should be seen in the SAN by four paths.

Guidelines for SVC extent size

SVC divides the managed disks (MDisks) that are presented by the IBM XIV System into smaller chunks that are known as extents. These extents are then concatenated to make virtual disks (VDisks). All extents that are used in the creation of a particular VDisk must all come from the same Managed Disk Group (MDG).

SVC supports extent sizes of 16, 32, 64, 128, 256, 512, 1024, and 2048 MB, and the IBM Storage System SAN Volume Controller Software version 6.1 adds the support extent size of 8 GB. The extent size is a property of the Managed Disk Group (MDG) that is set when the MDG is created. All managed disks that are contained in the MDG have the same extent size, so all virtual disks associated with the MDG must also have the same extent size.

Figure 11-7 depicts the relationship of an MDisk to MDG to a VDisk.

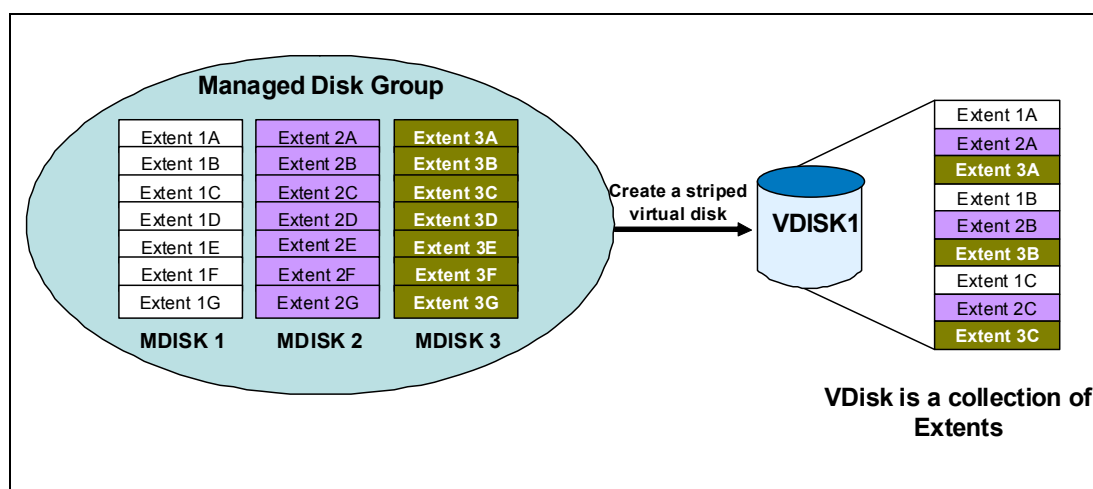


Figure 11-7 MDisk to VDisk mapping

The recommended extent size is 1 GB. Although smaller extent sizes can be used, this will limit the amount of capacity that can be managed by the SVC Cluster.



IBM SONAS Gateway connectivity

This chapter discusses specific considerations for attaching the IBM XIV Storage System to an IBM Scale Out Network Attached Storage Gateway (SONAS Gateway).

12.1 IBM SONAS Gateway

The Scale Out Network Attached Storage (SONAS) leverages mature technology from IBM High Performance Computing experience, and is based on the IBM General Parallel File System (GPFS™). It is aimed at delivering the highest level of reliability, availability, and scalability in Network Attached Storage (NAS).

SONAS is configured as a multi-system gateway device built from standard components and IBM software. The SONAS Gateway configurations are shipped in pre-wired racks made up of internal switching components along with Interface Nodes, Management Node, and Storage Nodes.

With the use of customer Fibre Channel switches or direct connected Fibre Channel, IBM SONAS Gateways can now be attached to IBM XIV Storage Systems of varying capacities for additional advantages in ease of use, reliability, performance, and Total Cost of Ownership (TCO).

Figure 12-1 is a schematic view of the SONAS gateway and its components, attached to two XIV systems.

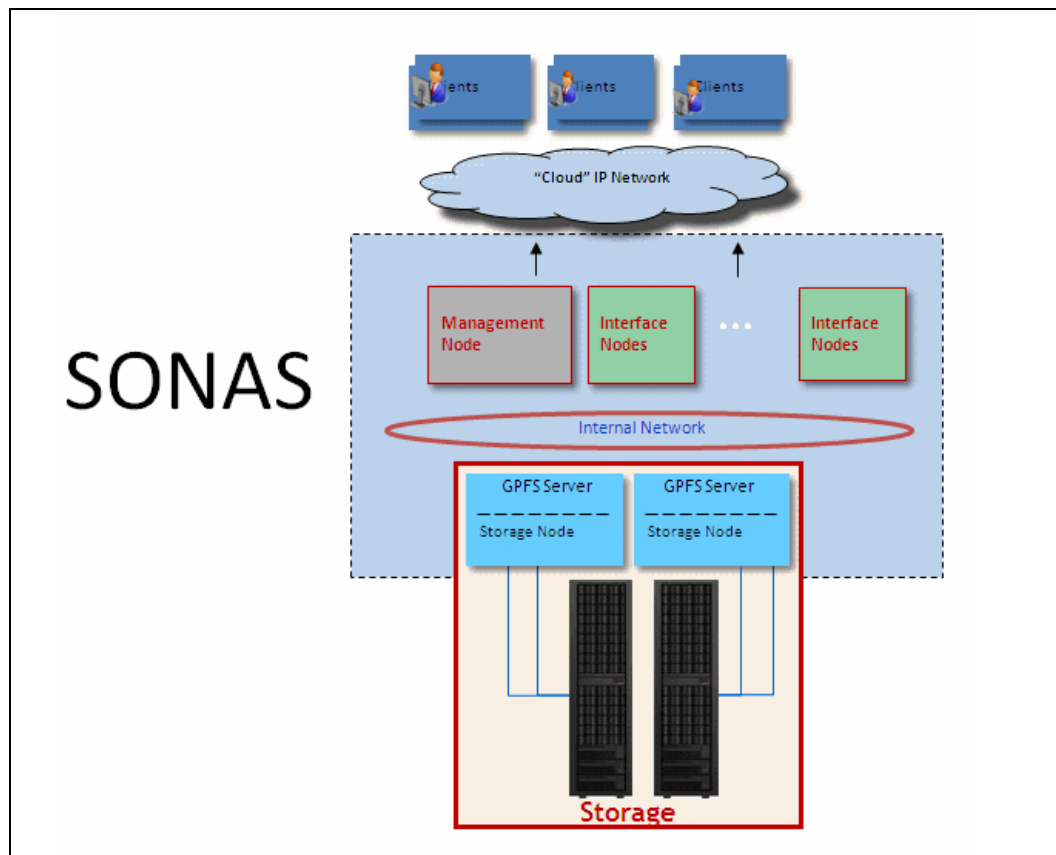


Figure 12-1 IBM SONAS with two IBM XIV Storage Systems

The SONAS Gateway is built on several components that are connected with InfiniBand:

- ▶ The Management Node handles all internal management and code load functions.
- ▶ Interface Nodes are the connection to the customer network. They provide file shares from the solution. They can be expanded after scalability demands.

- Storage Nodes are the components that deliver the General Parallel File System (GPFS) to the interface nodes and have the fiber connection to the IBM XIV Storage System. When more storage is required, more IBM XIV Storage Systems can be added.

12.2 Preparing an XIV for attachment to a SONAS Gateway

If you will attach your own storage device (such as the XIV Storage System) to SONAS, the SONAS system has to be ordered as a Gateway (without storage included). Note also that an IBM SONAS that was ordered with its own storage included can not be reinstalled as a gateway easily. Mixing different types of storages with an IBM SONAS or IBM SONAS Gateway is not currently supported.

When attaching an IBM SONAS Gateway to an IBM XIV Storage System, the following checks and preparations are necessary, in conjunction with connectivity guidelines already presented in Chapter 1, “Host connectivity” on page 1:

- Check for supported versions and prerequisite
- Do/verify the cabling
- Define zoning
- Create Regular Storage Pool for SONAS volumes
- Create XIV volumes
- Create SONAS Cluster
- Add SONAS Storage Nodes to cluster
- Add Storage Node Fibre Channel port (WWPN) to nodes in the cluster
- Map XIV volumes to the cluster

12.2.1 Supported versions and prerequisites

An IBM SONAS Gateway will work with an XIV only when specific prerequisites are fulfilled. These will be checked during the Technical Delivery Assessment (TDA) meeting that must take place before any installation.

SONAS Gateway has the following requirements:

- IBM SONAS version 1.1.1.0-19 or later.
- XIV Storage System software version 10.2 or later.
- XIV must be installed, configured, and functional before installing and attaching the IBM SONAS Gateway.
- One of the following:
 - Direct fiber attachment between XIV and IBM SONAS Gateway Storage Nodes
 - Fiber attachment from XIV to IBM SONAS Gateway Storage Nodes through redundant Fibre Channel Switch fabrics, either existing or newly installed.
 - These switches must be in the list of switches that are supported by the IBM XIV Storage System Refer to the IBM System Storage Interoperation Center, SSIC at: <http://www.ibm.com/systems/support/storage/config/ssic>
 - Each switch must have 4 available ports for attachment to the SONAS Storage Nodes (each switch will have 2 ports connected to each SONAS Storage Node).

12.2.2 Direct attached connection to XIV

For a direct attachment to XIV, connect Fibre Channel cables between the two IBM SONAS Gateway Storage Nodes and the XIV patch panel as depicted in Figure 12-2.

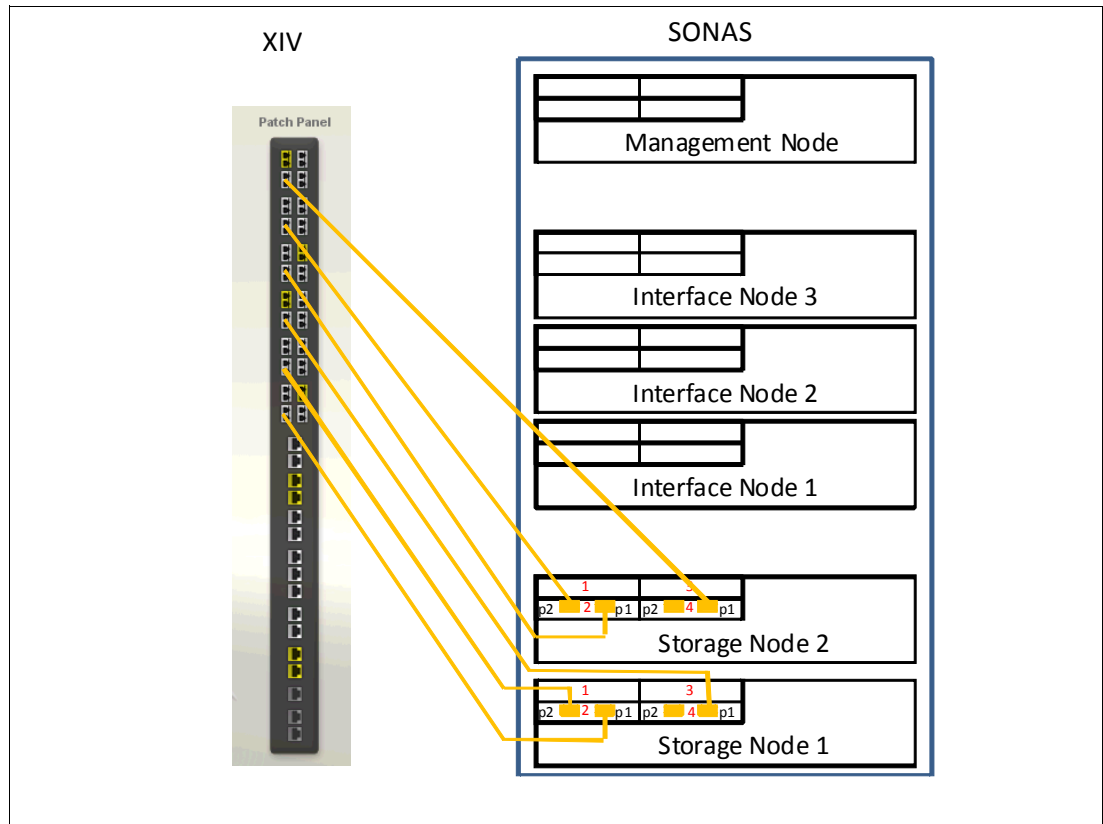


Figure 12-2 Direct connect cabling

The cabling is realized as follows:

- ▶ Between the SONAS Storage Node 1 HBA and XIV Storage, connect:
 - PCI Slot 2 Port 1 to XIV Interface Module 4 Port 1
 - PCI Slot 2 Port 2 to XIV Interface Module 5 Port 1
 - PCI Slot 4 Port 1 to XIV Interface Module 6 Port 1
- ▶ Between SONAS Storage Node 2 HBA and XIV Storage, connect:
 - PCI Slot 2 Port 1 to XIV Interface Module 7 Port 1
 - PCI Slot 2 Port 2 to XIV Interface Module 8 Port 1
 - PCI Slot 4 Port 1 to XIV Interface Module 9 Port 1

12.2.3 SAN connection to XIV

For maximum performance with an XIV system, its important to have many paths. Connect Fibre Channel cables from IBM SONAS Gateway Storage Nodes to two switched fabrics.

If a single IBM XIV Storage System is being connected, each switch fabric must have 6 available ports for Fibre Channel cable attachment to XIV: one for each interface module in XIV. Typically XIV interface module port 1 is used for switch fabric 1, and port 3 for switch fabric 2, as depicted in Figure 12-3.

If two IBM XIV Storage Systems are to be connected, each switch fabric must have 12 available ports for attachment to the XIV (6 ports for each XIV).

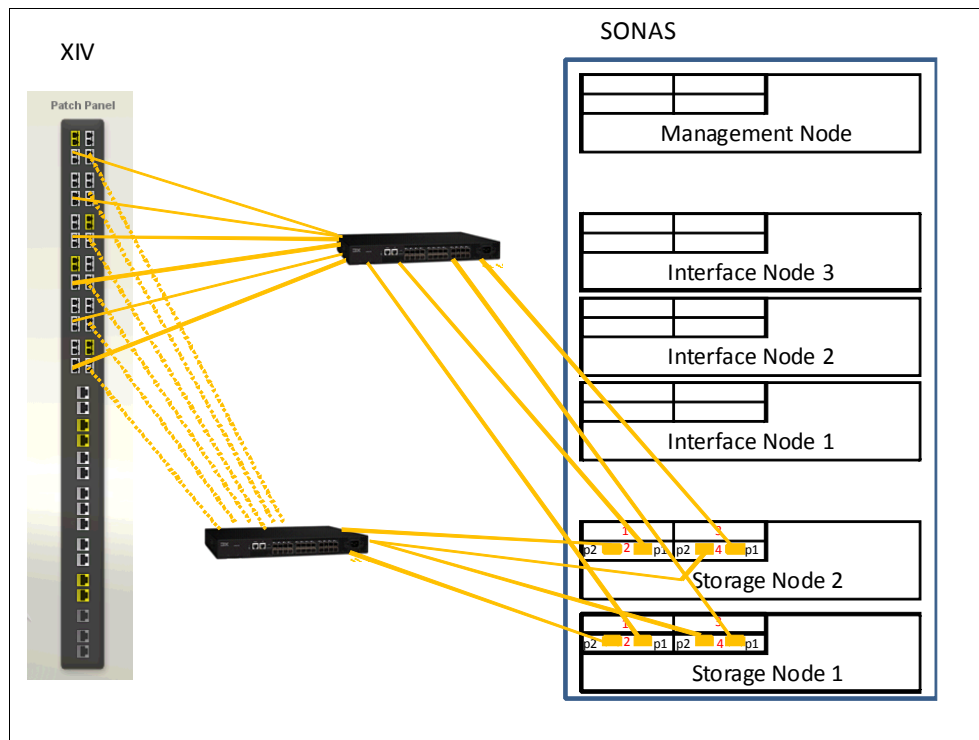


Figure 12-3 SAN Cabling diagram for SONAS Gateway to XIV

Zoning

Attaching the SONAS Gateway to XIV over a switched fabric requires an appropriate zoning of the switches.

Configure zoning on the Fibre Channel switches using single initiator zoning. That means only one host (in our case Storage Node) HBA port in every zone and multiple targets (in our case XIV ports).

Zone each HBA port from the IBM SONAS Gateway Storage Nodes to all six XIV interface modules; If you have two XIV systems, you have to zone to all XIV interface modules as shown in Example 12-2 on page 244. This is to get a maximum number of available paths to the XIV.

An IBM SONAS gateway connected to XIV uses multipathing with round-robin feature enabled, which means that all IOs to the XIV will be spread over all available paths.

Example 12-1 shows the zoning definitions for “SONAS Storage node 1 hba 1 port 1” (initiator) to all XIV interface modules (targets). The zoning is such that each HBA port has 6 possible paths to one XIV (and 12 possible paths to two XIV systems as shown in Example 12-2).

Following the same pattern for each HBA port in the IBM SONAS Storage Nodes will create 24 paths per IBM SONAS Storage Node.

Example 12-1 Zoning for one XIV Storage

Switch1 Zone1:

“SONAS Storage node 1 hba 1 port 1”, “XIV module 4 port 1”, “XIV module 5 port 1”, “XIV module 6 port 1”, “XIV module 7 port 1”, “XIV module 8 port 1”, “XIV module 9 port 1”

Switch1 Zone2:

“SONAS Storage node 1 hba 2 port 1”, “XIV module 4 port 1”, “XIV module 5 port 1”, “XIV module 6 port 1”, “XIV module 7 port 1”, “XIV module 8 port 1”, “XIV module 9 port 1”

Switch1 Zone3:

“SONAS Storage node 2 hba 1 port 1”, “XIV module 4 port 1”, “XIV module 5 port 1”, “XIV module 6 port 1”, “XIV module 7 port 1”, “XIV module 8 port 1”, “XIV module 9 port 1”

Switch1 Zone4:

“SONAS Storage node 2 hba 2 port 1”, “XIV module 4 port 1”, “XIV module 5 port 1”, “XIV module 6 port 1”, “XIV module 7 port 1”, “XIV module 8 port 1”, “XIV module 9 port 1”

Switch2 Zone1:

“SONAS Storage node 1 hba 1 port 2”, “XIV module 4 port 3”, “XIV module 5 port 3”, “XIV module 6 port 3”, “XIV module 7 port 3”, “XIV module 8 port 3”, “XIV module 9 port 3”

Switch2 Zone2:

“SONAS Storage node 1 hba 2 port 2”, “XIV module 4 port 3”, “XIV module 5 port 3”, “XIV module 6 port 3”, “XIV module 7 port 3”, “XIV module 8 port 3”, “XIV module 9 port 3”

Switch2 Zone3:

“SONAS Storage node 2 hba 1 port 2”, “XIV module 4 port 3”, “XIV module 5 port 3”, “XIV module 6 port 3”, “XIV module 7 port 3”, “XIV module 8 port 3”, “XIV module 9 port 3”

Switch2 Zone4:

“SONAS Storage node 2 hba 2 port 2”, “XIV module 4 port 3”, “XIV module 5 port 3”, “XIV module 6 port 3”, “XIV module 7 port 3”, “XIV module 8 port 3”, “XIV module 9 port 3”

Example 12-2 Zoning for two XIV storage systems

Switch1 Zone1:

“SONAS Storage node 1 hba 1 port 1”, “XIV1 module 4 port 1”, “XIV1 module 5 port 1”, “XIV1 module 6 port 1”, “XIV1 module 7 port 1”, “XIV1 module 8 port 1”, “XIV1 module 9 port 1”, “XIV2 module 4 port 1”, “XIV2 module 5 port 1”, “XIV2 module 6 port 1”, “XIV2 module 7 port 1”, “XIV2 module 8 port 1”, “XIV2 module 9 port 1”

Switch1 Zone2:

“SONAS Storage node 1 hba 2 port 1”, “XIV1 module 4 port 1”, “XIV1 module 5 port 1”, “XIV1 module 6 port 1”, “XIV1 module 7 port 1”, “XIV1 module 8 port 1”, “XIV1 module 9 port 1”, “XIV2 module 4 port 1”, “XIV2 module 5 port 1”, “XIV2 module 6 port 1”, “XIV2 module 7 port 1”, “XIV2 module 8 port 1”, “XIV2 module 9 port 1”

Switch1 Zone3:

“SONAS Storage node 2 hba 1 port 1”, “XIV1 module 4 port 1”, “XIV1 module 5 port 1”, “XIV1 module 6 port 1”, “XIV1 module 7 port 1”, “XIV1 module 8 port 1”, “XIV1 module 9 port 1”, “XIV2 module 4 port 1”, “XIV2 module 5 port 1”, “XIV2 module 6 port 1”, “XIV2 module 7 port 1”, “XIV2 module 8 port 1”, “XIV2 module 9 port 1”

Switch1 Zone4:

“SONAS Storage node 2 hba 2 port 1”, “XIV1 module 4 port 1”, “XIV1 module 5 port 1”, “XIV1 module 6 port 1”, “XIV1 module 7 port 1”, “XIV1 module 8 port 1”, “XIV1 module 9 port 1”, “XIV2 module 4 port 1”, “XIV2 module 5 port 1”, “XIV2 module 6 port 1”, “XIV2 module 7 port 1”, “XIV2 module 8 port 1”, “XIV2 module 9 port 1”

Switch2 Zone1:

“SONAS Storage node 1 hba 1 port 2”, “XIV1 module 4 port 3”, “XIV1 module 5 port 3”, “XIV1 module 6 port 3”, “XIV1 module 7 port 3”, “XIV1 module 8 port 3”, “XIV1 module 9 port 3”, “XIV2 module 4 port 3”, “XIV2 module 5 port 3”, “XIV2 module 6 port 3”, “XIV2 module 7 port 3”, “XIV2 module 8 port 3”, “XIV2 module 9 port 3”

Switch2 Zone2:

“SONAS Storage node 1 hba 2 port 2”, “XIV1 module 4 port 3”, “XIV1 module 5 port 3”, “XIV1 module 6 port 3”, “XIV1 module 7 port 3”, “XIV1 module 8 port 3”, “XIV1 module 9 port 3”, “XIV2 module 4 port 3”, “XIV2 module 5 port 3”, “XIV2 module 6 port 3”, “XIV2 module 7 port 3”, “XIV2 module 8 port 3”, “XIV2 module 9 port 3”

Switch2 Zone3:

“SONAS Storage node 2 hba 1 port 2”, “XIV1 module 4 port 3”, “XIV1 module 5 port 3”, “XIV1 module 6 port 3”, “XIV1 module 7 port 3”, “XIV1 module 8 port 3”, “XIV1 module 9 port 3”, “XIV2 module 4 port 3”, “XIV2 module 5 port 3”, “XIV2 module 6 port 3”, “XIV2 module 7 port 3”, “XIV2 module 8 port 3”, “XIV2 module 9 port 3”

Switch2 Zone4:

“SONAS Storage node 2 hba 2 port 2”, “XIV1 module 4 port 3”, “XIV1 module 5 port 3”, “XIV1 module 6 port 3”, “XIV1 module 7 port 3”, “XIV1 module 8 port 3”, “XIV1 module 9 port 3”, “XIV2 module 4 port 3”, “XIV2 module 5 port 3”, “XIV2 module 6 port 3”, “XIV2 module 7 port 3”, “XIV2 module 8 port 3”, “XIV2 module 9 port 3”

Zoning is also described in the *IBM Scale Out Network Attached Storage - Installation Guide for iRPQ 8S1101: Attaching IBM SONAS to XIV*, GA32-0797 available at:

http://publib.boulder.ibm.com/infocenter/sonasic/sonaslic/topic/com.ibm.sonas.doc/xiv_installation_guide.pdf

12.3 Configuring an XIV for IBM SONAS Gateway

Configuration of an XIV Storage System to be used by an IBM SONAS Gateway should be done before SONAS Gateway is installed by IBM service representative.

- ▶ In XIV GUI, configure one *regular* storage pool for an IBM SONAS Gateway. You can set the corresponding Snapshot™ reserve space to zero because snapshots on an XIV are not required, nor supported with SONAS. See Figure 12-4 on page 246.
- ▶ In XIV GUI, define XIV volumes in the storage pool previously created. All capacity that will be used by the IBM SONAS Gateway must be configured into LUNs where each volume is 4 TB in size. Refer to Figure 12-5 on page 247.
- ▶ The volumes are named sequentially SONAS_1, SONAS_2, SONAS_3 and so on. When the volumes are imported as Network Shared Disks (NSD), they are named XIV<serial>SONAS_#, where <serial> is the serial number of the XIV Storage System,

and SONAS_# is the name automatically assigned by XIV. Refer to Figure 12-5 on page 247.

- Volumes that are used by the IBM SONAS Gateway must be mapped to the IBM SONAS Gateway cluster so that they are accessible to all IBM SONAS Storage nodes. See Figure 12-12 on page 249.

12.3.1 Sample configuration

We illustrate in this section a sample configuration.

1. We create a Regular Storage Pool of 8 TB (the pool size should be a multiple of 4 TB because each Volume in the pool will have to be 4 TB exactly), as illustrated in Figure 12-4.

Note: You must use Regular Storage Pool. Thin provisioning is not supported when attaching to the IBM SONAS Gateway.

The screenshot shows the 'Add Pool' dialog box. At the top, there is a 'Select Type' dropdown menu with 'Regular Pool' selected. Below this, a progress bar shows the 'Total System Capacity: 79113 GB'. The progress bar is divided into three segments: 'Allocated' (58102 GB, green), 'Pool Size' (8005 GB, yellow), and 'Free' (13005 GB, grey). Below the progress bar, there are three radio buttons: 'Allocated' (selected), 'Pool Size', and 'Free'. Below the radio buttons, there are input fields for 'Pool Size' (8005 GB), 'Snapshots Size' (0 GB), and 'Pool Name' (*SONAS). At the bottom of the dialog are 'Add' and 'Cancel' buttons.

Figure 12-4 Create Storage Pool

2. We create volumes for the IBM SONAS Gateway in the storage pool. Two 4 TB volumes are created as shown in Figure 12-5 on page 247.

Note: The volumes will be 4002 GB, because the XIV Storage System uses 17 GB capacity increments.

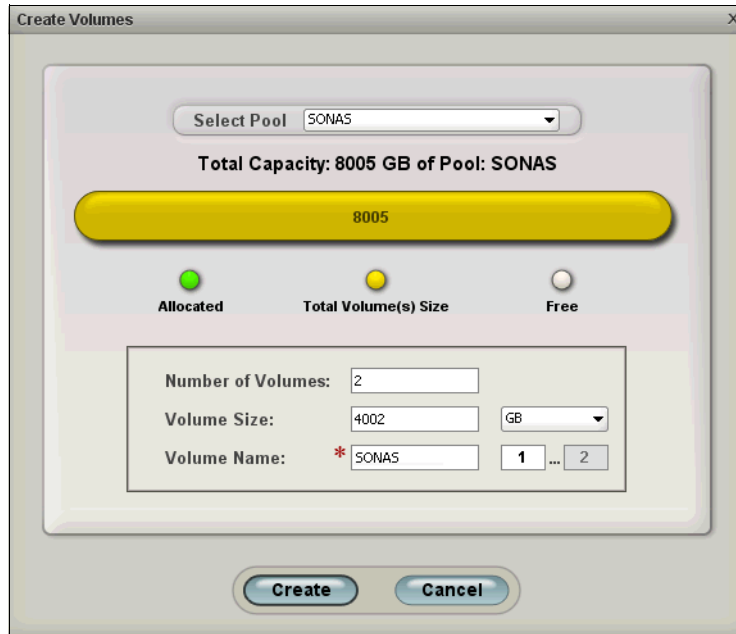


Figure 12-5 Volume creation

Figure 12-6 shows the two volumes created in the pool.

0 %	SONAS	8005
	SONAS_1	4002
	SONAS_2	4002

Figure 12-6 Volumes created

- Now, we define in XIV a cluster for the SONAS Gateway as we have multiple SONAS Storage Nodes that need to see the same volumes. See Figure 12-7.

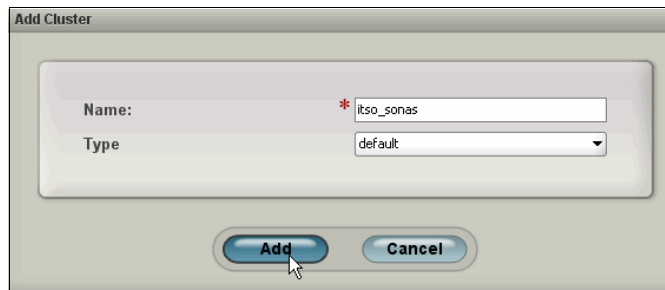


Figure 12-7 Cluster creation

4. Create hosts for each of the SONAS Storage Nodes as illustrated in Figure 12-8.

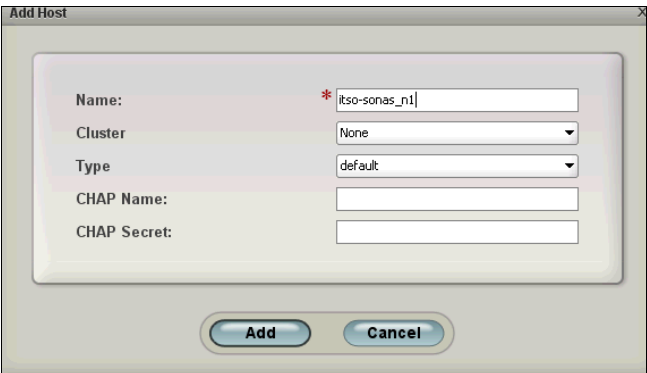


Figure 12-8 Host creation IBM SONAS Storage Node 1

We create another host for IBM SONAS Storage Node 2. Figure 12-9 shows both hosts in the cluster.

☐	itso_sonas	default
☐	itso_sonas_n1	default
☐	itso_sonas_n2	default

Figure 12-9 Create a host for both Storage Nodes

Given a correct zoning, we should now be able to add ports to their storage nodes. You can get the WWPN for each SONAS Storage Node in the name server of the switch (or for direct attached by looking look at the back of the IBM SONAS Storage Nodes PCI slot 2 and 4 which have a label indicating the WWPNs). To add the ports, right-click the host name and select **Add Port** as illustrated in Figure 12-10.

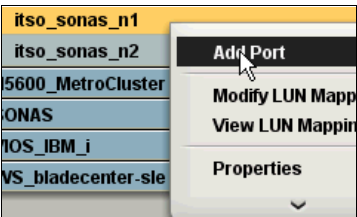


Figure 12-10 adding ports

5. After adding all four ports on each node, all the ports are listed as depicted in Figure 12-11.

☐	itso_sonas	default
☐	itso_sonas_n1	default
	2100001B329C2F21	FC
	2100001B329C4F20	FC
	2101001B32BC2F21	FC
	2101001B32BC4F20	FC
☐	itso_sonas_n2	default
	2100001B329C4C21	FC
	2100001B329C6A21	FC
	2101001B32BC4C21	FC
	2101001B32BC6A21	FC

Figure 12-11 SONAS Storage Node Cluster port config

6. Now we map the 4 TB volumes to the cluster, so both storage nodes can see the same volumes. Refer to Figure 12-12.

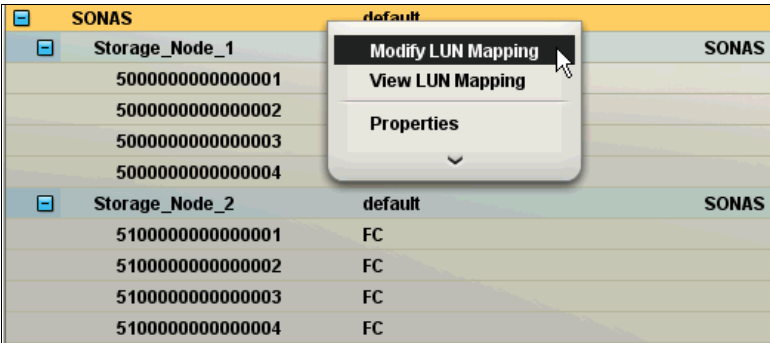


Figure 12-12 Modify LUN mapping

The two volumes are mapped as LUN id 1 and LUN id 2 to the IBM SONAS Gateway cluster, as shown in Figure 12-13.

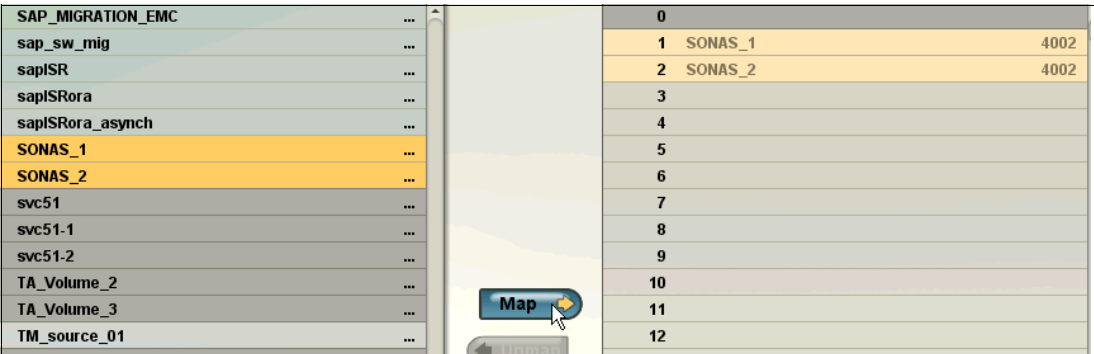


Figure 12-13 Mappings

12.4 IBM Technician can now install SONAS Gateway

An IBM Technician will now install the IBM SONAS Gateway code to all the IBM SONAS Gateway components. It will include the loading code and configuring basic settings.

Note: It is essential that the IBM SONAS Gateway be ordered as a Gateway and not a normal SONAS. XIV must be the only storage that the IBM SONAS Gateway will handle.

Follow instructions in Installation guide for SONAS Gateway and XIV found at the following URL:

http://publib.boulder.ibm.com/infocenter/sonasic/sonas1ic/topic/com.ibm.sonas.doc/xiv_installation_guide.pdf



N series Gateway connectivity

This chapter discusses specific considerations for attaching a N series Gateway to an IBM XIV Storage System.

13.1 Overview

The IBM 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 with software level 10.1 and later. Exact details on currently supported levels can be found in the N series interoperability matrix at:

ftp://public.dhe.ibm.com/storage/nas/nseries/nseries_gateway_interoperability.pdf

Figure 13-1 illustrates the attachment and possible multiple use of the XIV Storage System with the N Series gateway.

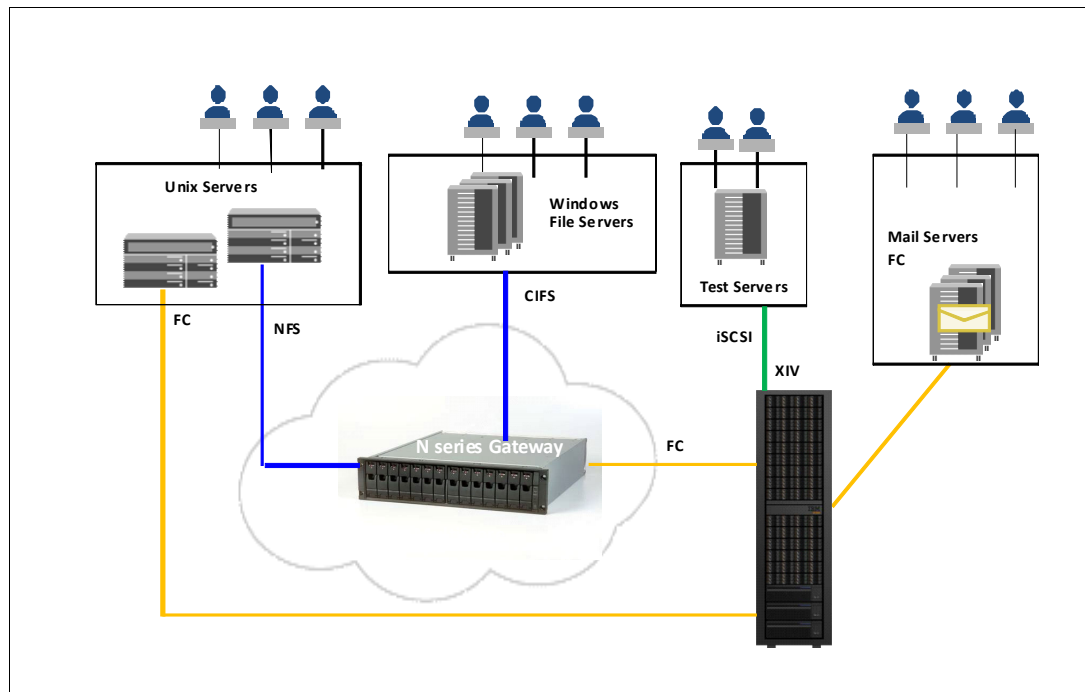


Figure 13-1 N series Gateway with IBM XIV Storage System

13.2 Attaching N series Gateway to XIV

When attaching the N series Gateway to an XIV, in conjunction with connectivity guidelines already presented in Chapter 1, “Host connectivity” on page 1, the following considerations apply:

- ▶ Check for supported versions and other considerations
- ▶ Do the cabling
- ▶ Define zoning
- ▶ Create XIV volumes
- ▶ Make XIV host definitions
- ▶ Map XIV volumes to corresponding host
- ▶ Install Data Ontap

13.2.1 Supported versions

At the time of writing this book, Data Ontap 7.3.3 and XIV code level 10.2.0.a, 10.2.1 and 10.2.1.b are supported as listed in the interoperability matrix extract shown in Figure 13-2.

For the latest information and supported versions, always verify the N series Gateway interoperability matrix at:

ftp://public.dhe.ibm.com/storage/nas/nseries/nseries_gateway_interoperability.pdf

Storage Array	IBM N series Gateway cluster or single head models	Data ONTAP versions									
		7.1									
		7.1.0.1									
		7.1.1									
		7.1.2		7.2.1.1							7.3.2
		7.1.2.1		7.2.2			7.2.6.1			7.3.1.1	7.3.3
		7.1.3	7.2.1	7.2.3	7.2.4	7.2.5.1	7.2.7	7.3	7.3.1.1L1		7.3.4
XIV(MT 2810) Restrictions on supported fabrics. See notes.											
Microcode FW:	N7700, N7900										7.3.3
10.2.1b	N6040, N6060, N6070										and later
10.2.1	N5300, N5600										
10.2.0.a											
See notes: 4, 11, 12											
4. Non Disruptive Upgrade of Storage Array Firmware is not supported.											
11. Use Implementation Guide for IBM XIV Storage.											
12. Supported with Brocade 4G & 8G fabrics from this support matrix, firmware 6.1.2a or later. Supported with Cisco 4G fabrics from this support matrix, firmware 4.1(3a) or later.											

Figure 13-2 Currently supported N series models and Data Ontap versions; Extract from interoperability matrix

Other considerations

- ▶ Only FC connections between N series Gateway and an XIV system are allowed.
- ▶ A volume needs to be mapped as LUN 0: create a 17 GB dummy volume for this. Refer to 13.5.5, “Mapping the root volume to the host in XIV gui” on page 261.
- ▶ N series can only handle two paths per LUN. Refer to 13.4, “Zoning” on page 255.
- ▶ N series can only handle up to 2 TB LUNs. Refer to 13.6.4, “Adding data LUNs to N series Gateway” on page 264.

13.3 Cabling

This section shows how to layout the cabling when connecting the XIV Storage System, either to a single N series Gateway or to a N series cluster gateway.

13.3.1 Cabling example for single N series Gateway with XIV

The N series Gateway should be cabled such 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 they need to be set as initiators. The reason for using 0a and 0c is that they are on separate Fibre Channel chips, thus providing better resiliency. Refer to Figure 13-3 on page 254.

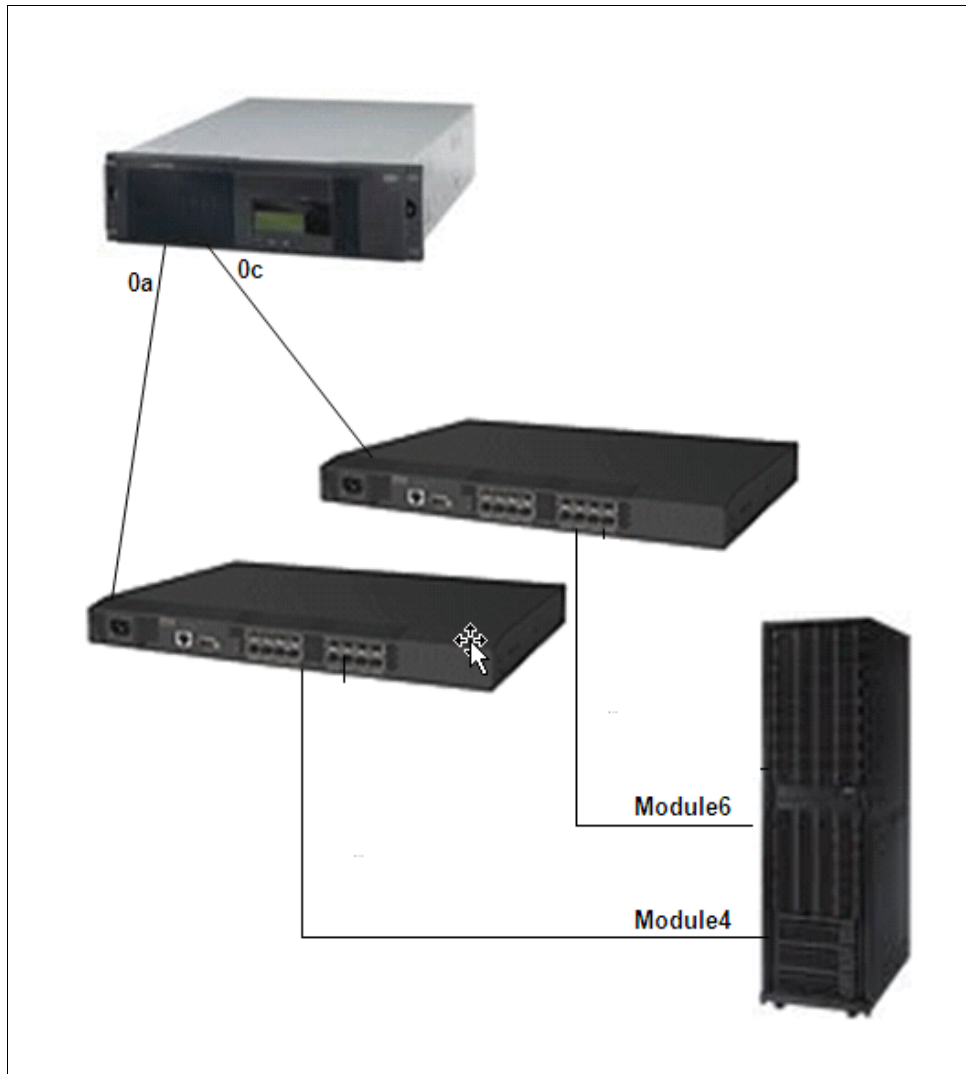


Figure 13-3 Single N series to XIV cable overview

13.3.2 Cabling example for N series Gateway cluster with XIV

An N series Gateway cluster should be cabled such 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 they need to be set as initiators. The reason for using 0a and 0c is that they are on separate Fibre Channel chips, which provides better resiliency.

The link between the N series Gateways is the cluster interconnect. Refer to Figure 13-4 on page 255.

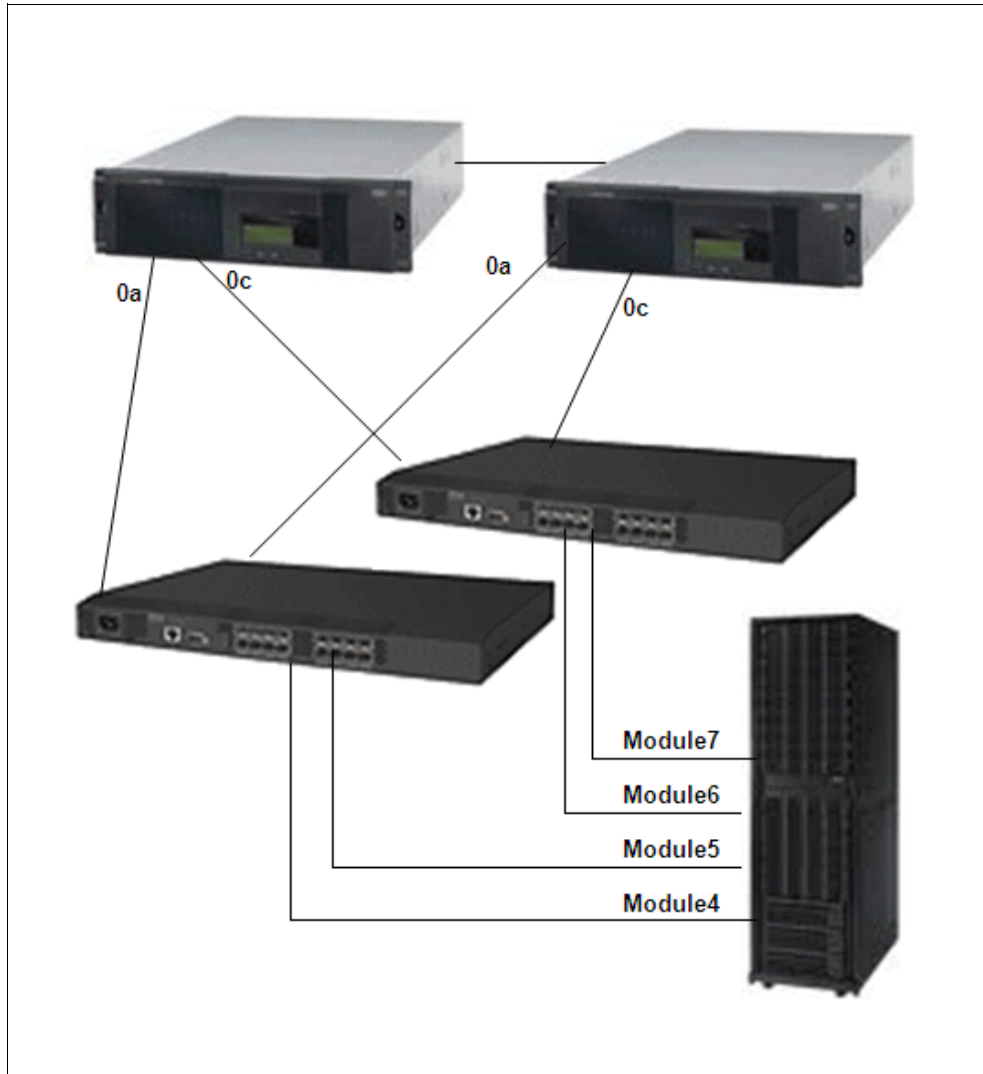


Figure 13-4 Clustered N series to XIV cable overview

13.4 Zoning

Zones have to be created such 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 only has two paths. It also limits the RSCN (Registered State Change Notification) traffic in the switch.

13.4.1 Zoning example for single N series Gateway attachment to XIV

A possible zoning definition could be as follows:

- ▶ Switch 1
 - Zone 1
 - NSeries_port_0a, XIV_module4_port1
- ▶ Switch 2
 - Zone 1

NSeries_port_0c, XIV_module6_port1

13.4.2 Zoning example for clustered N series Gateway attachment to XIV

A possible zoning definition could be as follows:

- 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

13.5 Configuring the XIV for N series Gateway

N series Gateway boots from an XIV volume. Consequently, before you can configure an XIV for a N series Gateway, the proper root sizes must be chosen. Figure 13-5 is a capture of the recommended minimum root volume sizes from N series Gateway interoperability matrix.

Root Volume Limits		
	Prior to 8.0	Starting in 8.0.x
Minimum root volume size		
N5300	16 GB	160 GB
N5600	23 GB	230 GB
N6040	16 GB	160 GB
N6060	23 GB	230 GB
N6070	37 GB	250 GB
N7700	37 GB	250 GB
N7900	69 GB	250 GB

Figure 13-5 Recommended minimum root volume sizes on different N series hardware

13.5.1 Create a Storage Pool in XIV

When N series Gateway is attached to XIV, it will not support XIV snapshots, synchronous mirror, asynchronous mirror, or thin provisioning features. If such features are needed they must be used from the corresponding functions than N series Data Ontap natively offers.

To prepare the XIV Storage System for use with N series Gateway, first create a Storage Pool using the XIV GUI as shown in Figure 13-6 on page 257.

Tip: No snapshot space reservation is needed because the XIV snapshots are not supported with N series Gateways.

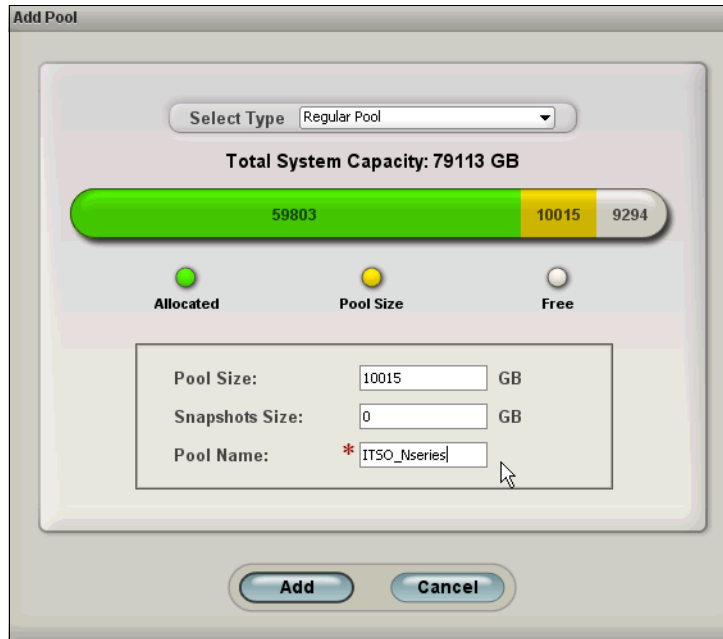


Figure 13-6 Create a Regular Storage Pool in XIV gui

13.5.2 Create the root volume in XIV

The N series interoperability matrix displayed (in part) in Figure 13-5 on page 256 shows that for the N5600 model, the recommended minimum size for a root volume is 957 GB.

N series calculates GB differently from XIV, and you will need to do some adjustments to get the right size. N series GB are expressed as 1000 x 1024 x 1024 bytes, and XIV GB are expressed as 1000 x 1000 x 1000 bytes.

Thus, for the case considered, we have:

$$957\text{GB} \times (1000 \times 1024 \times 1024) / (1000 \times 1000 \times 1000) = 1003 \text{ GB}$$

But, because XIV is using capacity increments of about 17 GB, it will automatic set our size to 1013 GB.

As shown in Figure 13-7 on page 258, create a volume of 1013 GB for the root volume in the storage pool previously created. Create also a 17 GB dummy volume to be mapped as LUN 0.

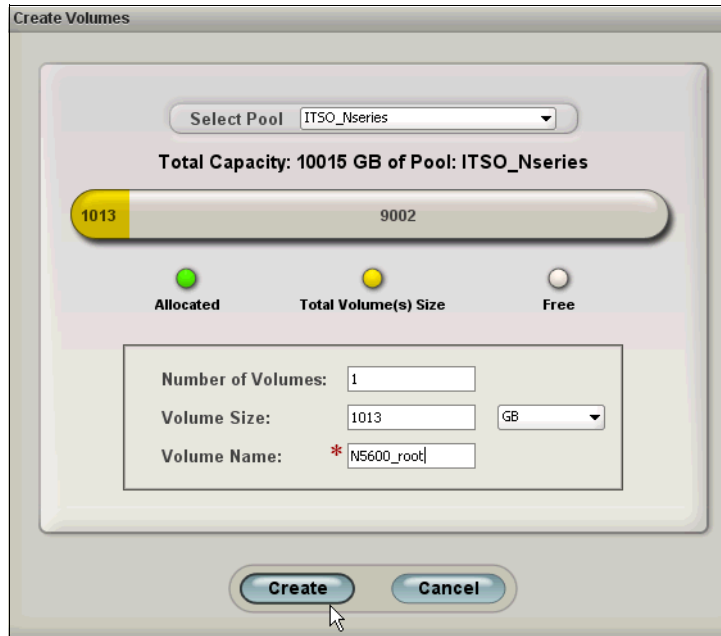


Figure 13-7 Volume creation

13.5.3 N series Gateway Host create in XIV

Now, you must create the host definitions in XIV.

Our example, as illustrated in Figure 13-8, is for a single N series Gateway and we can just create a host. Otherwise, for a 2 node cluster Gateway, you would have to create a cluster in XIV first, and then add the corresponding hosts to the cluster.



Figure 13-8 Single N series Gateway host create

Note: If you are deploying a N series Gateway cluster, you need to create an XIV cluster group and add both N series Gateways to the XIV cluster group

13.5.4 Add the WWPN to the host in XIV

Obtain the WWPN of the N series Gateway.

One way is to boot the N series Gateway into 'Maintenance mode'. That will make it log in to the switches if zoning is correct. To get the N series into Maintenance mode, you need to access the N series console. That can be done with using the null modem cable that came with the system or through the Remote Login Module (RLM) interface.

Here is a short description for the RLM method:

1. Power on your N series Gateway.
2. Connect to RLM ip using ssh, and log in as naroot.
3. Type system console.
4. Observe the boot process as illustrated in Example 13-1, and when you see Press CTRL-C for special boot menu, press Ctrl+C.

Example 13-1 N series Gateway booting

```
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. Now you can 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 has logged in. Refer to Figure 13-9.

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 13-9 N series Gateway logged into switch as Network appliance

7. Now you are ready to add the WWPN to the host in the XIV GUI, as depicted in Figure 13-10.

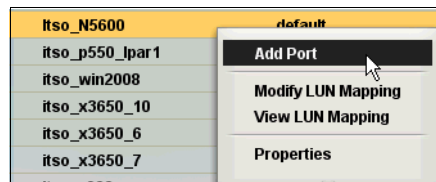


Figure 13-10 Right-click and add port to the Host

Make sure you add both ports. If your zoning is right they should show up in the list. If they do not show up, check your zoning. Refer to Figure 13-11.

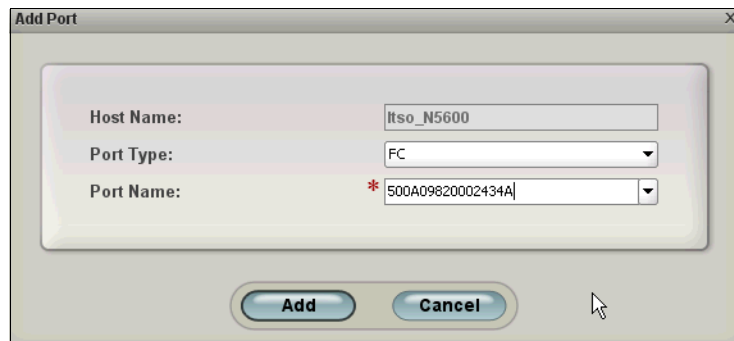


Figure 13-11 Add both ports, 0a and 0c

8. Verify that both ports are connected to XIV by checking the Host Connectivity view in the XIV GUI, as shown in Figure 13-12 on page 261.

Name	1:4	1:5	1:6
Itso_N5600			
500A09800002434A			
500A09820002434A	✓1		✓1

Figure 13-12 Host connectivity verify

13.5.5 Mapping the root volume to the host in XIV gui

To map the root volume to the host now defined in XIV, proceed as follows:

1. In the XIV GUI host view, right-click the host name and select **Modify LUN Mapping** from the pop-up menu as shown in Figure 13-13.

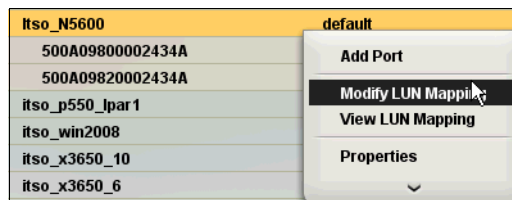


Figure 13-13 Select modify mapping

2. As shown in Figure 13-14, right-click LUN 0 and select **Enable** from the pop-up menu.

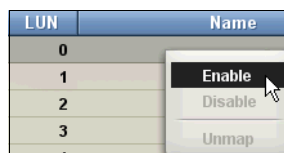


Figure 13-14 LUN 0 enable

3. Click the 17 GB dummy volume for LUN 0 and your root volume as LUN 1 then click **Map**, as illustrated in Figure 13-15.



Figure 13-15 Mapping view

Tip: Map the dummy XIV volume to LUN 0 and the N series root volume to LUN 1.

Note: If you are deploying a N series Gateway cluster, you need to map both N series Gateway root volumes to the XIV cluster group.

13.6 Installing Data Ontap

For completeness, we briefly document how Data Ontap is installed on the XIV volume.

13.6.1 Assigning the root volume to N series Gateway

In the N series Gateway ssh shell type **disk show -v** to see the mapped disk, as illustrated in Example 13-2.

Example 13-2 disk show -v

```
*> 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

```
*>
```

Note: If you do not see any disk, make sure you have Data Ontap 7.3.3. If you need to upgrade, follow the N series documentation to perform a netboot update.

Assign the root LUN to the N series Gateway with **disk assign <disk name>**, ex. **disk assign Primary_SW2:6.126L1** as shown in Example 13-3.

Example 13-3 Execute command: disk assign all

```
*> disk assign Primary_SW2:6.126L1
Wed Oct disk 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 13-4.

Example 13-4 disk show

```
*> disk show
Local System ID: 118054991
```

DISK	OWNER	POOL	SERIAL NUMBER
Primary_SW2:6.126L1	(118054991)	Poo10	13000CB11A4

13.6.2 Installing Data Ontap

To proceed with the Data Ontap installation, follow the steps:

1. Stop the 'Maintenance mode' with **halt** as illustrated in Example 13-5.

Example 13-5 Stop 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 13-6.

Example 13-6 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 4a to install Data Ontap.

13.6.3 Data Ontap update

After Data Ontap installation is finished, and you have entered all the relevant information, an update of Data Ontap needs to be performed because the install from special boot menu is just a limited install. Follow normal N series update procedures to update Data Ontap to perform a full installation.

Tip: Transfer the right code package to the root volume in directory `/etc/software`. One way to do this from Windows is to start cifs and map c\$ of the N series Gateway, go to the `/etc` directory and create a folder called `software`, then copy the code package there. When copy is finished run **software update <package name>** from the N series Gateway shell.

Note: A second LUN should always be assigned afterwards as core dump LUN. The size depends on the hardware. The interoperability matrix should be consulted to find the appropriate size.

13.6.4 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 size of a LUN that Data Ontap can handle is 2 TB. To reach the maximum of 2 TB we need to consider the 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 $1000 \times 1024 \times 1024$ bytes, while XIV uses GB as $1000 \times 1000 \times 1000$ bytes.

For Data Ontap 2TB LUN, the XIV size express in GB should be $2000 \times (1000 \times 1024 \times 1024) / (1000 \times 1000 \times 1000) = 2097$ GB

The biggest LUN size that can effectively be used in XIV is 2095 GB as XIV capacity is based on 17 GB increments.



ProtecTIER Deduplication Gateway connectivity

This chapter discusses specific considerations for using the IBM XIV Storage System as storage for a TS7650G ProtecTIER® Deduplication Gateway (3958-DD3).

For details on TS7650G ProtecTIER Deduplication Gateway (3958-DD3), refer to *IBM System Storage TS7650, TS7650G and TS7610*, SG24-7652.

14.1 Overview

The ProtecTIER Deduplication Gateway is used to provide virtual tape library functionality with deduplication features. Deduplication means that only the unique data blocks will be stored on the attached storage. The ProtecTIER will present virtual tapes to the backup software, making the process transparent to the backup software. The backup software will perform backups as usual, but the backups will be deduplicated before they are stored on the attached storage. In Figure 14-1 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.

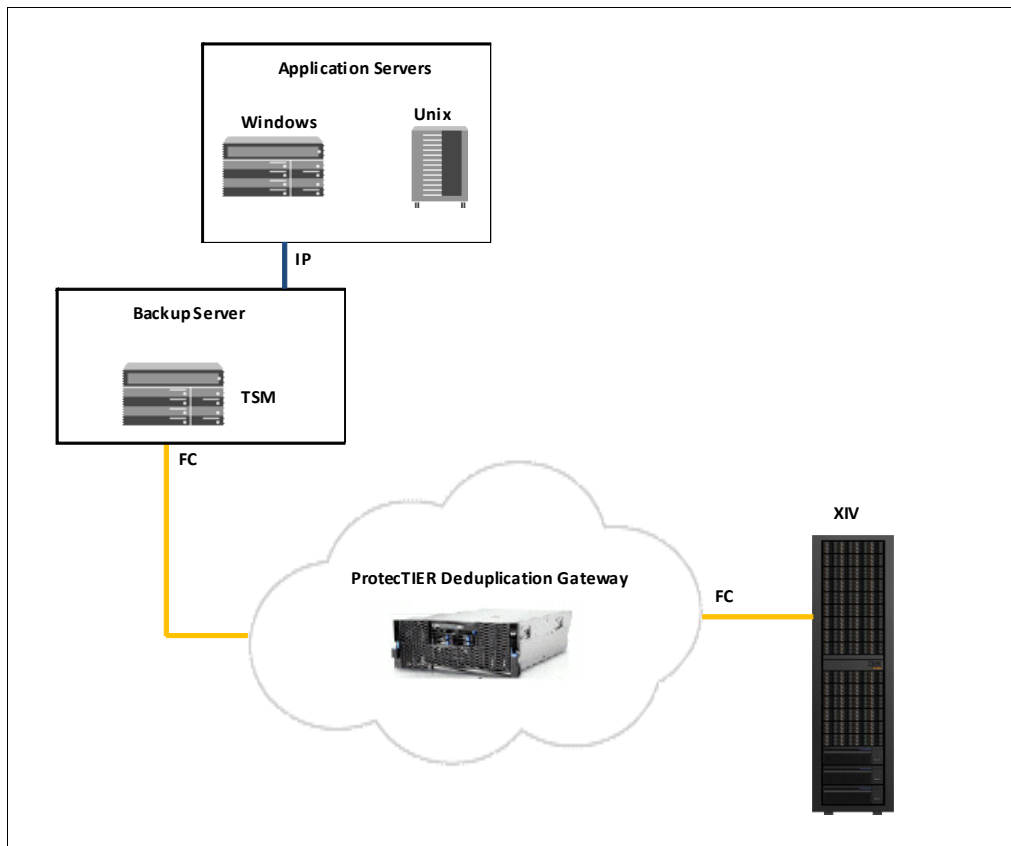


Figure 14-1 Single ProtecTIER Deduplication Gateway

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 will include two ProtecTIER nodes. The TS7650G ProtecTIER Deduplication Gateway offers:

- ▶ Inline data deduplication powered by HyperFactor® technology
- ▶ Multicore virtualization and deduplication engine
- ▶ Clustering support for higher performance and availability
- ▶ Fibre Channel ports for host and server connectivity
- ▶ Performance: 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, and 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 details on ProtecTier, refer to *IBM System Storage TS7650 and TS7610 with ProtecTIER*, SG24-7652, at:

<http://www.redbooks.ibm.com/redbooks/pdfs/sg247652.pdf>

14.2 Preparing an XIV for ProtecTIER Deduplication Gateway

The TS7650G ProtecTIER Deduplication Gateway is ordered together with ProtecTIER Software, but the ProtecTIER Software is shipped separately.

When attaching the TS7650G ProtecTIER Deduplication Gateway to IBM XIV Storage System, preliminary conditions must be met and preparation tasks must be performed in conjunction with 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 volume
- ▶ Make XIV host definitions for the ProtecTier Gateway
- ▶ Map XIV volume to corresponding host

14.2.1 Supported versions and prerequisites

A TS7650G ProtecTIER Deduplication Gateway will work 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.
- ▶ XIV Storage System must be functional before installing the TS7650G ProtecTIER Deduplication Gateway.
- ▶ Fiber attachment through existing Fibre Channel switched fabrics, or at least one Fibre Channel switch, needs to be installed to allow connection of the TS7650G ProtecTIER Deduplication Gateway to IBM XIV Storage System.
- ▶ These Fibre Channel switches must be in the list of Fibre Channel switches that are supported by the IBM XIV Storage System, as per the IBM System Storage Interoperation Centre at:

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

Note: Direct attachment between TS7650G ProtecTIER Deduplication Gateway and IBM XIV Storage System is not supported.

14.2.2 Fibre Channel switch cabling

For maximum performance with an IBM XIV Storage System it is important to connect all available XIV Interface Modules. For redundancy connect Fibre Channel cables from TS7650G ProtecTIER Deduplication Gateway to two Fibre Channel switched fabrics.

If a single IBM XIV Storage System is being connected, each Fibre Channel switched fabric must have 6 available ports for Fibre Channel cable attachment to IBM XIV Storage System. One for each interface module in XIV. Typically XIV interface module port 1 is used for Fibre Channel switch one and XIV interface module port 3 for Fibre Channel switch 2, as depicted in Figure 14-2.

When using a partially configured XIV rack, refer to Figure 1-1 on page 2 to locate the available FC ports.

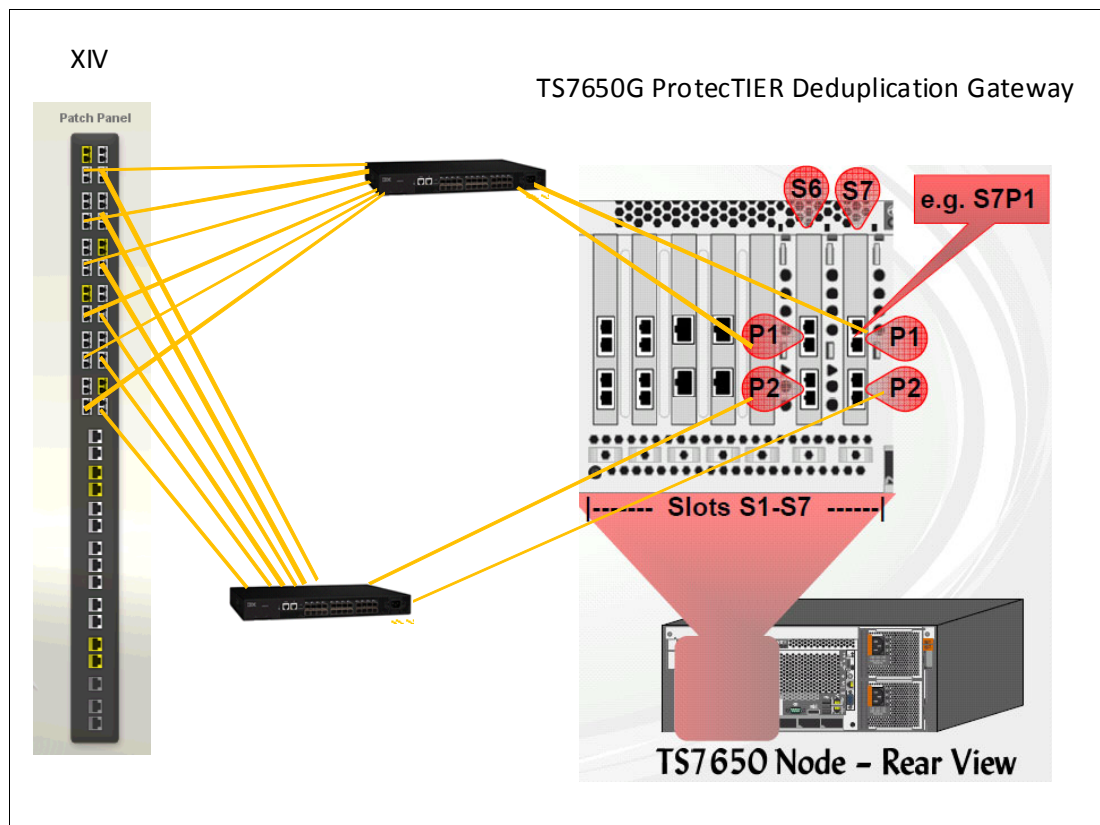


Figure 14-2 Cable diagram for connecting a TS7650G to IBM XIV Storage System

14.2.3 Zoning configuration

For each TS7650G disk attachment port, multiple XIV host ports are configured into separate zones:

- ▶ All XIV Interface Modules port 1 get zoned to the ProtecTIER HBA in slot 6 port 1 and HBA in slot 7 port 1
- ▶ All XIV Interface Modules port 3 get zoned to the ProtecTIER HBA in slot 6 port 2 and HBA in slot 7 port 2

Each interface module in IBM XIV Storage System has connection with both TS7650G HBAs. Best practice for ProtecTIER tells to use 1-1 zoning (one initiator and one target in each zone)

to create zones for connection of a single ProtecTIER node with a 15 module IBM XIV Storage System with all six Interface Modules. Refer to Example 14-1.

Example 14-1 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

- ▶ Each ProtecTIER Gateway backend HBA port sees three XIV interface modules.
- ▶ Each XIV interface module is connected redundantly to two separate ProtecTIER backend HBA ports.
- ▶ There are 12 paths (4 x 3) to one volume from a single ProtecTIER Gateway node.

14.2.4 Configuring XIV Storage System for ProtecTIER Deduplication Gateway

An IBM representative will use the ProtecTIER Capacity Planning Tool to size the ProtecTIER repository (Metadata and User Data). Every capacity planning is different because it depends heavily on your type of data and expected deduplication ratio. The planning tool output will include the detailed information about all volume sizes and capacities for your specific ProtecTIER installation. If you do not have this information, contact your IBM representative to get it. An example for XIV can be seen in Figure 14-3.

The screenshot shows the 'Meta Data Planner' application window. The title bar reads 'Meta Data Planner'. The main window has a blue header with the title 'Meta Data Planner'. Below the header, there are input fields for 'Company' (set to 'IBM'), 'Workload' (set to 'XIV Best Practice'), 'Created' (set to '23-Nov-10'), 'Updated' (empty), and 'By' (set to 'Admin').

The configuration section includes the following fields:

- Model: TS7650G
- Repository Size: 79
- Raid Type: FC-15K4+4
- Include Growth: ☐
- Release: 2.3.2.0
- Factoring Ratio: 12:1
- Drive Capacity: 600
- Planner: 2.3.1
- Max Throughput: 1.000

Below the configuration section, there is a 'Meta Data Configuration' section with a note: 'Note! First 1GB File system does not require a separate raid group'. To the right of this section is a 'Results: Meta Data:' section with the following values:

- Capacity (GB): 2.884
- Spindles: 32

At the bottom, there is a table of file system sizes in GB:

File System	Size (GB)
1-15:	1, 860, 1.144, 451, 428, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0

Figure 14-3 ProtecTIER Capacity Planning Tool example

Note: 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.

Configuration of IBM XIV Storage System to be used by ProtecTIER Deduplication Gateway should be done before the ProtecTIER Deduplication Gateway is installed by an IBM service representative.

- Configure one storage pool for ProtecTIER Deduplication Gateway. You can set snapshot space to zero because creating snapshots on IBM XIV Storage System is not supported with ProtecTIER Deduplication Gateway.
- Configure the IBM XIV Storage System into volumes. Follow the ProtecTIER Capacity Planning Tool output. The capacity planning tool output will give you the Metadata volume size and the size of the 32 Data volumes. A Quorum volume of a minimum of 1 GB should always be configured as well, in case solution needs to grow to include 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.

Sample illustration of configuring an IBM XIV Storage System

First you have to create a Storage pool for the capacity you want to use for ProtecTIER Deduplication Gateway. Use the XIV GUI as shown in Figure 14-4.

The screenshot shows the 'Add Pool' window in the IBM XIV GUI. At the top, there's a 'Select Type' dropdown menu with 'Regular Pool' selected. Below this, the 'Total System Capacity' is shown as 79113 GB, accompanied by a yellow progress bar. Three radio buttons are visible: 'Allocated' (which is selected and has a green dot), 'Pool Size' (with a yellow dot), and 'Free' (with a grey dot). Under the 'Pool Size' radio button, there's a text box for 'Pool Size' containing '79113' followed by 'GB'. Below that is a 'Snapshots Size' field with '0' and 'GB'. The 'Pool Name' field contains '* ProtecTIER'. At the bottom of the window are two buttons: 'Add' and 'Cancel'. A mouse cursor is pointing at the 'Add' button.

Figure 14-4 Storage Pool create in XIV gui

Note: Use a Regular Pool and zero out the snapshot reserve space, as snapshots and thin provisioning are not supported when XIV is used with ProtecTIER Deduplication Gateway.

For a 79 TB IBM XIV Storage System

In the example from Figure 14-3 on page 269 with a 79TB XIV Storage System and a deduplication Factoring Ratio of 12, the volumes sizes are as follows:

- ▶ 4x 1202GB volumes for Metadata
- ▶ 1x 17G volume for Quorum (enough with 1 GB but due to XIV min size it will be 17 GB)
- ▶ $32 \times (79113 - (4 \times 1202) - 17) / 32 = 2321$ GB, due to XIV architecture 2319 GB volume size for Data volumes will be used.

For Metadata see Figure 14-5.

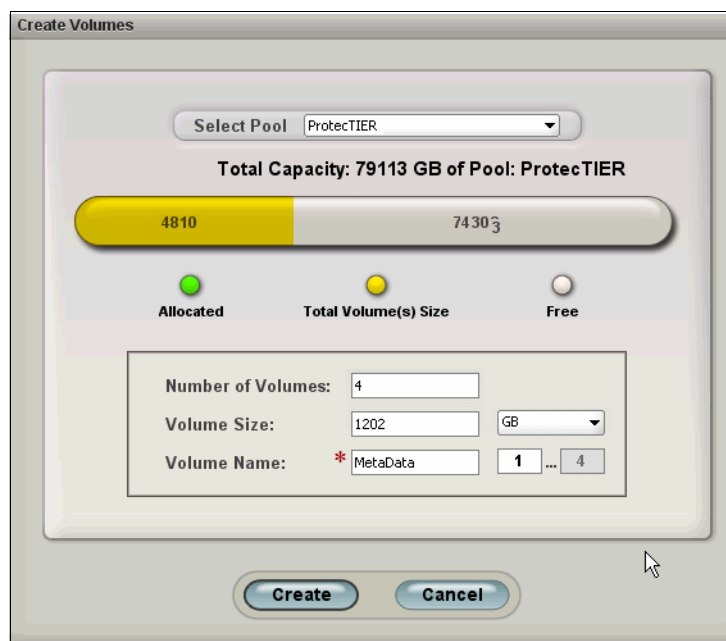


Figure 14-5 Metadata volumes create

For Quorum see Figure 14-6 on page 272.

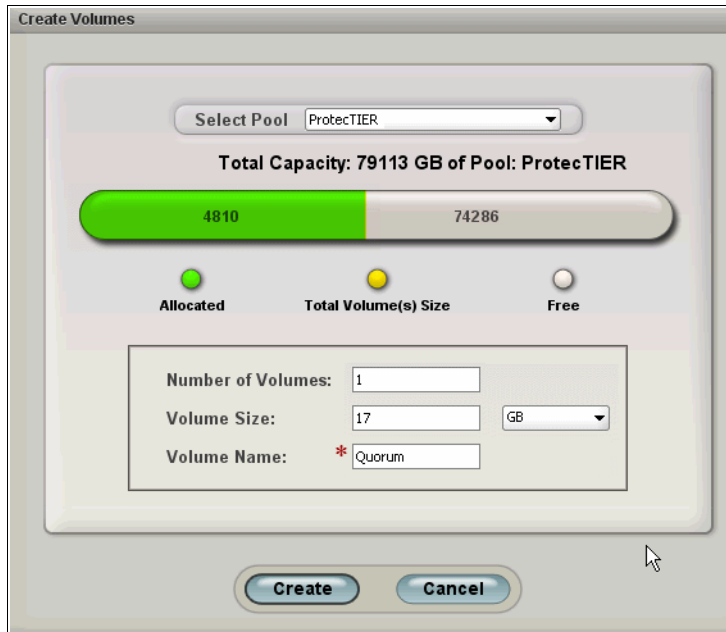


Figure 14-6 Quorum volume create

For User Data see Figure 14-7.

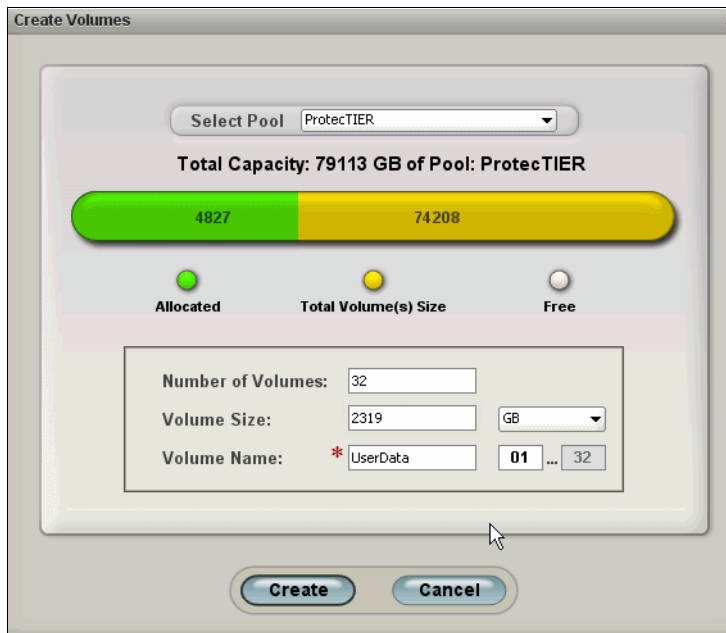
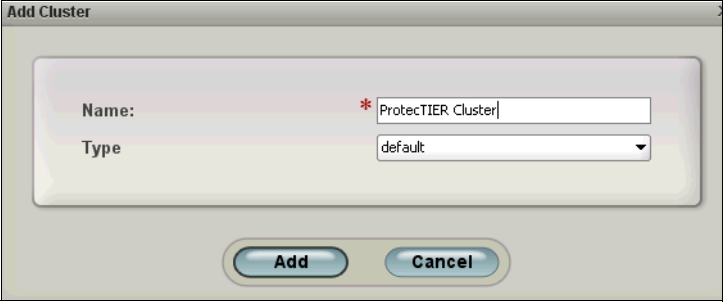


Figure 14-7 User Data volumes create

The next step is to create host definitions in XIV GUI, as shown in Figure 14-9 on page 273. If you have a ProtecTIER Gateway cluster (two ProtecTIER nodes in High Availability solution) you would first need to create a cluster group, and then add host defined for each node to the cluster group. Refer to Figure 14-8 on page 273, Figure 14-9 on page 273, and Figure 14-10 on page 273.

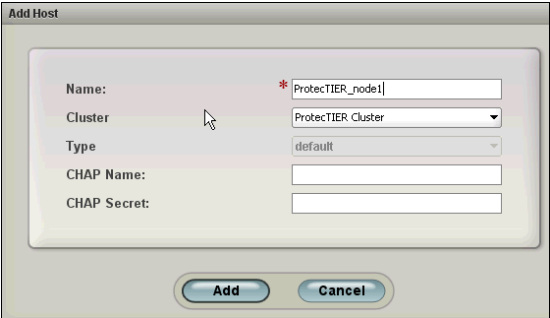


The 'Add Cluster' dialog box contains the following fields:

- Name:** A text input field containing 'ProtectTIER Cluster' with a red asterisk indicating it is required.
- Type:** A dropdown menu currently set to 'default'.

At the bottom of the dialog are two buttons: 'Add' and 'Cancel'.

Figure 14-8 ProtectTIER cluster definition

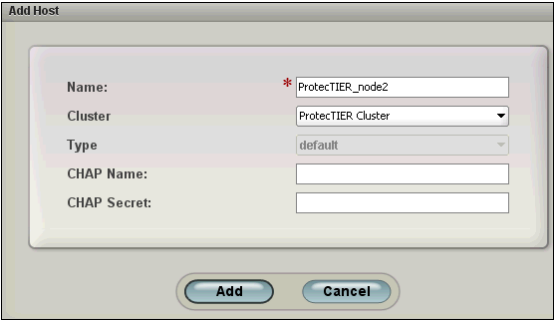


The 'Add Host' dialog box for node 1 contains the following fields:

- Name:** A text input field containing 'ProtectTIER_node1' with a red asterisk.
- Cluster:** A dropdown menu set to 'ProtectTIER Cluster'.
- Type:** A dropdown menu set to 'default'.
- CHAP Name:** An empty text input field.
- CHAP Secret:** An empty text input field.

At the bottom are 'Add' and 'Cancel' buttons.

Figure 14-9 Add ProtectTIER cluster node 1



The 'Add Host' dialog box for node 2 contains the following fields:

- Name:** A text input field containing 'ProtectTIER_node2' with a red asterisk.
- Cluster:** A dropdown menu set to 'ProtectTIER Cluster'.
- Type:** A dropdown menu set to 'default'.
- CHAP Name:** An empty text input field.
- CHAP Secret:** An empty text input field.

At the bottom are 'Add' and 'Cancel' buttons.

Figure 14-10 Add ProtectTIER cluster node 2

Now you need to find the WWPNs of the ProtectTIER nodes. The WWPNs can be found in the name server of the Fibre Channel switch, or if zoning is in place they should be selectable from the drop down list. Alternatively they can also be found in the BIOS of the HBA cards and then entered by hand in the XUICV GUI.

After you have identified the WWPNs, add them to the ProtectTIER Gateway hosts as shown in Figure 14-11.

[-] ProtectTIER Cluster	default
[-] ProtectTIER_node1	default
5000000000000001	FC
5000000000000002	FC
5000000000000003	FC
5000000000000004	FC
ProtectTIER_node2	default

Figure 14-11 WWPNs added to ProtectTIER cluster node 1

Last step is to map the volumes to the ProtecTIER Gateway cluster. In the XIV GUI, right-click the cluster name or on the host if you only have one ProtecTIER node, and select **Modify LUN Mapping**. Figure 14-12 shows you what the mapping view looks like.

Note: If you only have one ProtecTIER Gateway node, map the volumes directly to the ProtecTIER gateway node.

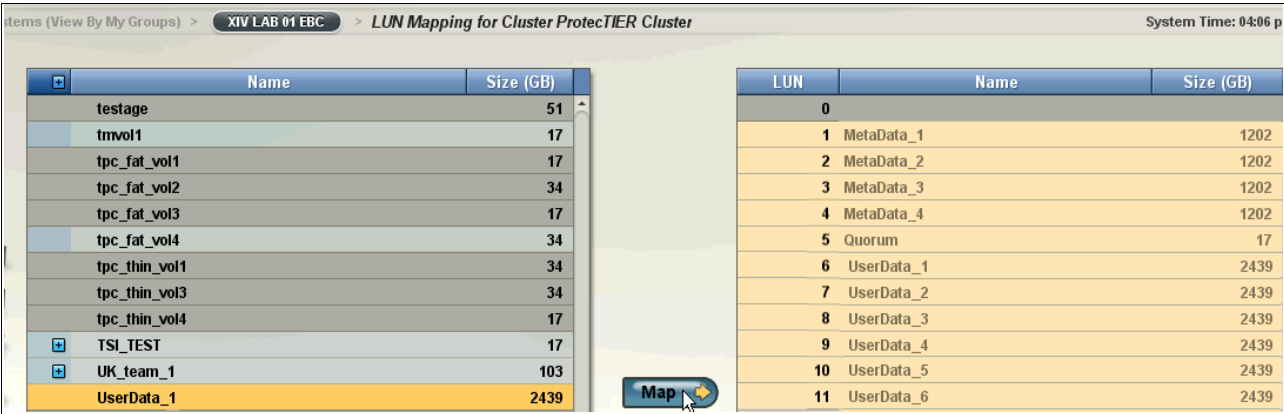


Figure 14-12 Mapping LUNs to ProtecTIER cluster

14.3 Ready for ProtecTIER software install

The IBM Technician can now install ProtecTIER software on the ProtecTIER Gateway nodes following the install instructions.



XIV in database application environments

The purpose of this chapter is to provide guidelines and recommendations on how to use the IBM XIV Storage System in Oracle and DB2® database application 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 looking at database engine, file systems, operating system, and storage to name a few. Regarding the non-storage components of the environment, the chapter gives hints, tips, and web links for additional information.

15.1 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. For example, the detailed volume layout requirements to allocate database space for optimal performance that are required for conventional storage systems are not required for XIV storage system.

Most storage vendors publish recommendations on how to distribute data across the resources of the storage system to achieve optimal I/O performance. Unfortunately the original setup and tuning cannot usually be kept over the lifetime of an application environment because as applications change and storage landscapes grow, traditional storage systems need to be constantly *re-tuned* to maintain optimal performance. One common less-than-optimal solution is that additional storage capacity is often provided on a best effort level and I/O performance tends to deteriorate.

This aging process that affects application environments on many storage subsystems does not occur with the XIV architecture:

- ▶ 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.

Consequently it is not necessary to develop a performance-optimized volume layout for database application environments with XIV. However, it is worth considering some configuration aspects during setup.

Common best practices

The most unique aspect of XIV is its inherent ability to utilize all resources (drives, cache, CPU) within its storage subsystem regardless of the layout of the data. However, to achieve maximum performance and availability, there are a few best practices:

- ▶ For data, use a small number of large XIV volumes (typically 2 - 4 volumes). Each XIV volume should be between 500 GB and 2 TB in size, depending on the database size. Using a small number of large XIV volumes takes better advantage of XIV's aggressive caching technology and simplifies storage management.
- ▶ When creating the XIV volumes for the database application, make sure to plan for the extra capacity required. XIV shows volume sizes in base 10 (KB = 1000 B), and operating systems can show them in base 2 (KB = 1024 B). In addition, file system overhead will also claim some storage capacity.
- ▶ Place your data and logs on separate volumes to be able to recover to a certain point-in-time instead just going back to the to the last consistent snapshot image after database corruption occurs. In addition, certain backup management and automation tools, for example Tivoli® Flash Copy Manager, require separate volumes for data and logs.
- ▶ If more than one XIV volume is used, implement an XIV consistency group in conjunction with XIV snapshots. This implies that the volumes are in the same storage pool.
- ▶ XIV offers thin provisioning storage pools. If the operating system's volume manager fully supports thin provisioned volumes, consider creating larger volumes than needed for the database size.

Oracle

Oracle database server without the ASM option (see below) does not stripe table space data across the corresponding files or storage volumes. Thus the common recommendations still apply.

Asynchronous I/O is recommended 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, it is best practice to 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 details about Oracle asynchronous I/O, refer to the Oracle manuals *Oracle Database High Availability Best Practices 11g Release 1* and *Oracle Database Reference 11g Release 1* available at the following URL:

http://www.oracle.com/pls/db111/portal.all_books

Oracle ASM

Oracle Automatic Storage Management (ASM) is Oracle's alternative storage management solution to conventional volume managers, file systems, and raw devices.

The main components of Oracle ASM are disk groups, each of which 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: data files, online and offline redo logs, control files, data file copies, Recovery Manager (RMAN) backups, and so on. Oracle binary and ASCII files, for example 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 configuring Oracle database using ASM on XIV, to achieve better performance and create a configuration that is easy to manage and use:

- ▶ Use 1 or 2 XIV volumes to create an ASM disk group
- ▶ Set 8M or 16M Allocation Unit (stripe) size

Note that with Oracle ASM, asynchronous I/O is used by default.

DB2

DB2 offers two types of table spaces that can exist in a database: *system-managed space (SMS)* and *database-managed space (DMS)*.

SMS table spaces are managed by the operating system, which stores the database data into file systems 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 is stored.

The files or directories or devices where data is stored are also called *containers*.

To achieve optimum database performance and availability, it is important to take advantage of the unique capabilities of XIV and DB2. This section focuses on the physical aspects of XIV volumes and how these volumes are mapped to the host.

- ▶ When creating a database, consider using DB2 automatic storage (AS) technology as a simple and effective way to provision storage for a database. When more than one XIV volume is used for database or for a database partition, AS will distribute the data evenly

among the volumes. Avoid using other striping methods such as the operating system's logical volume manager. DB2 automatic storage is used by default when you create a database 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. The purpose of pooling all data volumes together per partition is to facilitate the use of XIV's ability to create a consistent snapshot of all volumes within an XIV consistency group. Do not place your database transaction logs in the same consistency group as data.
- ▶ 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 partitioned DB2 database environment, use separate XIV volumes per partition to enable independent backup and recovery of each partition.

DB2 parallelism options for Linux, UNIX, and Windows

When there are multiple containers for a table space, the database manager can exploit parallel I/O. *Parallel I/O* refers to 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* refers to 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* refers to the simultaneous processing of parts of a single query, using either intrapartition parallelism, interpartition parallelism, or both.

Prefetching is important to the performance of intrapartition parallelism. DB2 prefetching pages means that one or more data or index pages are retrieved from disk in the expectation that they will be required by an application.

Database environment variable settings for Linux, UNIX, and Windows

The DB2_PARALLEL_IO registry variable influences parallel I/O on 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 container multiplied by the value 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 15-1 shows how to configure DB2 parallel I/O for all table spaces with the **db2set** command on AIX operating system.

Example 15-1 Enable DB2 parallel I/O

```
# su - db2xiv
$ db2set DB2_PARALLEL_IO=*:64
```

More details about DB2 parallelism options can be found in the *DB2 for Linux, UNIX, and Windows Information Center* available at the following URL:

<http://publib.boulder.ibm.com/infocenter/db2luw/v9r7>

15.2 Database Snapshot backup considerations

IBM offers the *Tivoli Storage FlashCopy® Manager* software product to create consistent database snapshots backups and to offload the data from the snapshot backups to an external backup/restore system like Tivoli Storage Manager (TSM).

Even without an appropriate product it is possible to create a consistent snapshot of a database. Consistency must be created on several layers: database, file systems and storage.

This section gives hints and tips to create consistent storage-based snapshots in database environments.

An example of a storage-based snapshot backup of a DB2 database on AIX operating system is shown in the *Snapshot* chapter of *IBM XIV Storage System: Copy Services and Migration*, SG24-7759.

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 creating the snapshot pairs. If a database cannot be stopped for the snapshot, pre- and post-processing actions have to be performed to create a consistent copy. The processes to create consistency in an XIV environment are outlined below.

An XIV Consistency Group comprises multiple volumes so that a snapshot can be taken of all the volumes at the same moment in time. This action creates a synchronized snapshot of all the volumes and is ideal for applications that span multiple volumes such as a database application that has the transaction logs on one set of volumes and the database on another set of volumes.

When creating a backup of the database, it is important to synchronize the data so that it is consistent at the database level as well. If the data is inconsistent, a database restore will not be possible because the log and the data are different, and therefore part of the data might be lost.

If XIV's consistency groups and snapshots are used to back up the database, database consistency can be established without shutting-down the application by following the steps in the procedure outlined below:

- ▶ Suspend database I/O. In case of Oracle, an I/O suspend is not required if the backup mode is enabled. Oracle handles the resulting inconsistencies during database recovery.
- ▶ If the database resides in file systems, write all modified file system data back to the storage system and thus flush the file systems buffers before creating the snapshots i.e. to perform a so-called "file system sync".
- ▶ Optionally perform 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 will be required before mounting the file systems allocated to the snapshots copies.
- ▶ Use snapshot-specific consistency groups.

Transaction log file handling

Considerations for transaction log files handling include the following:

- ▶ 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 and it might not be possible to recover the database to the most current point-in-time (using the forward recovery process of the database).

Snapshot restore

An XIV snapshot is performed 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 possible with certain databases, but combining partial restores with storage-based snapshots can require an 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 almost impractical. Therefore only full database restores are discussed with regard to storage-based snapshots.

A full database restore - with and without snapshot technology - requires a downtime. The database must be shut down, in case the file systems must be un-mounted and the volume groups deactivated (if file systems or a volume manager are used on the operating system level).

Following is a high-level description of the tasks required to perform a full database restore from a storage-based snapshot:

1. Stop application and shutdown database
2. Un-mount file systems (if applicable) and deactivate volume group(s)
3. Restore the XIV snapshots
4. Activate volume groups and mount file systems
5. Recover database (complete forward recovery or incomplete recovery to a certain point in time)
6. Start database and application



Snapshot Backup/Restore Solutions with XIV and Tivoli Storage Flashcopy Manager

This chapter explains how FlashCopy Manager leverages the XIV Snapshot function to backup and restore applications in a Unix and Windows environment. The chapter will contain three main parts:

- ▶ An overview of IBM Tivoli Storage FlashCopy Manager for Windows and Unix.
- ▶ The installation and configuration of FlashCopy Manager 2.2 for Unix together with an example of a disk-only backup and restore in an SAP/DB2 environment running on the AIX platform.
- ▶ The installation and configuration of IBM Tivoli Storage FlashCopy Manager 2.2 for Windows and Microsoft Volume Shadow Copy Services (VSS) for backup and recovery of Microsoft Exchange.

16.1 IBM Tivoli FlashCopy Manager Overview

In today's IT world, where application servers are operational 24 hours a day, the data on these servers must be fully protected. With the rapid increase in the amount of data on these servers, their critical business needs and the shrinking backup windows, traditional backup and restore methods are reaching their limits in meeting these challenging requirements. Snapshot operations can help minimize the impact caused by backups and provide near-instant restore capabilities. Because a snapshot operation typically takes much less time than the time for a tape backup, the window during which the data is being backed up can be reduced. This allows more frequent backups and increases the flexibility of backup scheduling and administration because the time spent for forward recovery through transaction logs after a restore is minimized.

IBM Tivoli Storage FlashCopy Manager software provides fast application-aware backups and restores, allowing using advanced snapshot technologies in IBM storage systems.

16.1.1 Features of IBM Tivoli Storage FlashCopy Manager

Figure 16-1 gives an overview of the supported applications and storage systems that are supported by FlashCopy Manager.

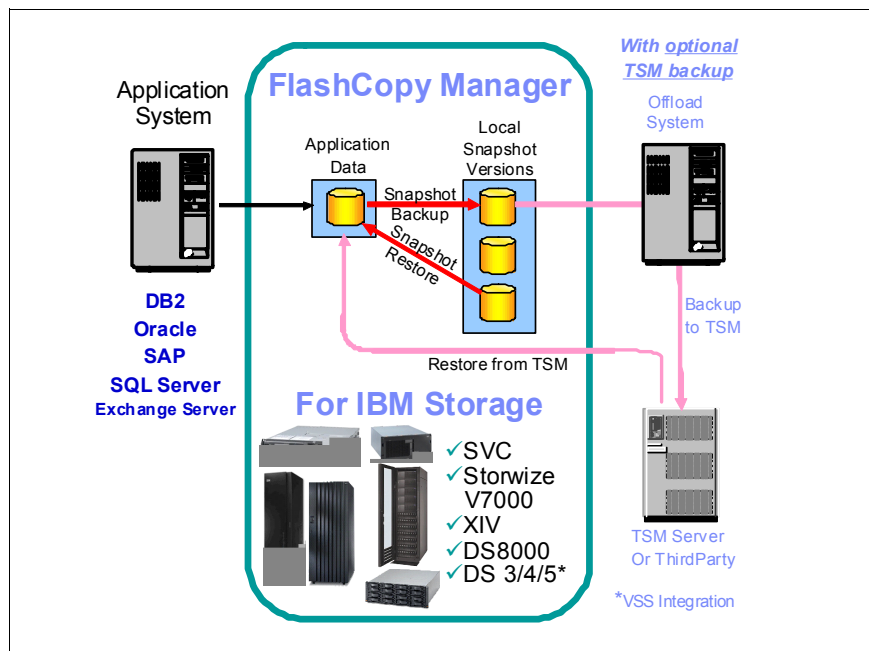


Figure 16-1 IBM Tivoli Storage FlashCopy Manager overview

IBM Tivoli Storage FlashCopy Manager uses the data replication capabilities of intelligent storage subsystems to create point-in-time copies. These are application-aware copies (FlashCopy or snapshot) of the production data. This copy is then retained on disk as a backup, allowing for a fast restore operation (flashback). FlashCopy Manager also allows mounting the copy on an auxiliary server (backup server) as a logical copy. This copy (instead of the original production-server data) is made accessible for further processing. This processing includes creating a backup to Tivoli Storage Manager (disk or tape) or doing backup verification functions (for example, the Database Verify Utility). If a backup to Tivoli Storage Manager fails, IBM Tivoli Storage FlashCopy Manager can restart the backup after

the cause of the failure is corrected. In this case, data already committed to Tivoli Storage Manager is not resent.

Highlights of IBM Tivoli Storage FlashCopy Manager include:

- ▶ Performs near-instant application-aware snapshot backups, with minimal performance impact for IBM DB2, Oracle, SAP, and Microsoft SQL Server and Exchange.
- ▶ Improve application availability and service levels through high-performance, near-instant restore capabilities that reduce downtime.
- ▶ Integrate with IBM System Storage DS8000, IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM XIV Storage System on AIX, Linux, Solaris, and Microsoft Windows.
- ▶ Protect applications on IBM System Storage DS3000, DS4000®, and DS5000 on Windows using VSS.
- ▶ Satisfy advanced data protection and data reduction needs with optional integration with IBM Tivoli Storage Manager.
- ▶ Operating systems IBM Tivoli Storage FlashCopy Manager supports are Windows, AIX, Solaris, and Linux

FlashCopy Manager for Unix and Linux supports the cloning of an SAP database since release 2.2. In SAP terms, this is called a Homogeneous System Copy: that is, the system copy runs the same database and operating system as the original environment. Again, FlashCopy Manager leverages the FlashCopy or Snapshot features of the IBM storage system to create a point-in-time copy of the SAP database. This feature is explained in more detail in 16.3.2, “SAP Cloning” on page 289.

For more information about IBM Tivoli Storage FlashCopy Manager, refer to:

<http://www-01.ibm.com/software/tivoli/products/storage-flashcopy-mgr/>

For detailed technical information, visit the IBM Tivoli Storage Manager Version 6.2 Information Center at:

http://publib.boulder.ibm.com/infocenter/tsminfo/v6r2/index.jsp?topic=/com.ibm.itsm.fcm.doc/c_fcm_overview.html

16.2 FlashCopy Manager 2.2 for Unix

The chapter describes the installation and configuration on AIX. FlashCopy Manager 2.2 for Unix can be installed on AIX, Solaris, and Linux. Before installing FlashCopy Manager, check the hardware and software requirements for your specific environment at the following URL:

<http://www-01.ibm.com/support/docview.wss?uid=swg21428707>

The pre-installation checklist defines hardware and software requirements and describes the volume layout for the SAP environment. To ensure a smooth installation of FlashCopy Manager, it is absolutely necessary that all requirements are fulfilled. For a list of considerations and decisions before installing IBM Tivoli Storage FlashCopy Manager, refer to the Installation Planning Worksheet also available at the above link.

16.2.1 FlashCopy Manager prerequisites

Volume group layout for DB2 and Oracle

FlashCopy Manager requires a well-defined volume layout on the storage subsystem and a resulting volume group structure on AIX and Linux. The FlashCopy Manager pre-installation checklist specifies the volume groups. FlashCopy Manager only processes table spaces, the local database directory, and log files. The volume group layout in Table 16-1 is the best practice for DB2.

Table 16-1 Volume group layout for DB2

Type of Data	Location of data	Contents of data	Comments
Table space volume groups	XIV	Table spaces	One or more dedicated volume groups
Log volume groups	XIV	Log files	One or more dedicated volume groups
DB2 instance volume group	XIV or local storage	DB2 instance directory	One dedicated volume group

For Oracle the volume group layout also has to follow certain layout requirements that are shown in Table 16-2. The table space data and redo log directories must reside on separate volume groups.

Table 16-2 Volume group layout for Oracle

Type of data	Location of data	Contents of data	Comments
Table space volume groups	XIV	Table space files	One or more dedicated volume groups
Online redo log volumes groups	XIV	Online redo logs, control files	One or more dedicated volume groups

For further details on the database volume group layout, see the pre-installation checklist at 16.2, “FlashCopy Manager 2.2 for Unix” on page 283.

XCLI configuration for FlashCopy Manager

IBM Tivoli Storage FlashCopy Manager for Unix requires the XIV command-line interface (XCLI) to be installed on all hosts where IBM Tivoli Storage FlashCopy Manager is installed. A CIM server or VSS provider is not required for an XIV connection. The path to the XCLI is specified in the FlashCopy Manager profile, and has to be identical for the production server and the optional backup or clone server. You can download the XCLI software at the following URL:

http://www-01.ibm.com/support/docview.wss?rs=1319&context=STJTAG&context=HW3E0&dc=D400&q1=s&uid=s&loc=en_US&cs=utf-8&lang=en

16.3 Installing and configuring FlashCopy Manager for SAP/DB2

IBM Tivoli Storage FlashCopy Manager has to be installed on the production system. For off-loaded backups to a Tivoli Storage Manager server, it must also be installed on the backup system. To install FlashCopy Manager with a graphical wizard, an X server has to be installed on the production system. The main steps of the FlashCopy Manager installation are shown in Figure 16-3 on page 286.

The following steps describe the installation process:

1. Log on to the production server as root user.
2. Using the GUI mode, type `./2.2.x.x-TIV-TSFCM-AIX.bin`.
3. Follow the instructions that are displayed.
4. Check the summary of the install wizard as shown in Figure 16-2. Be sure to enter the correct instance ID of the database.

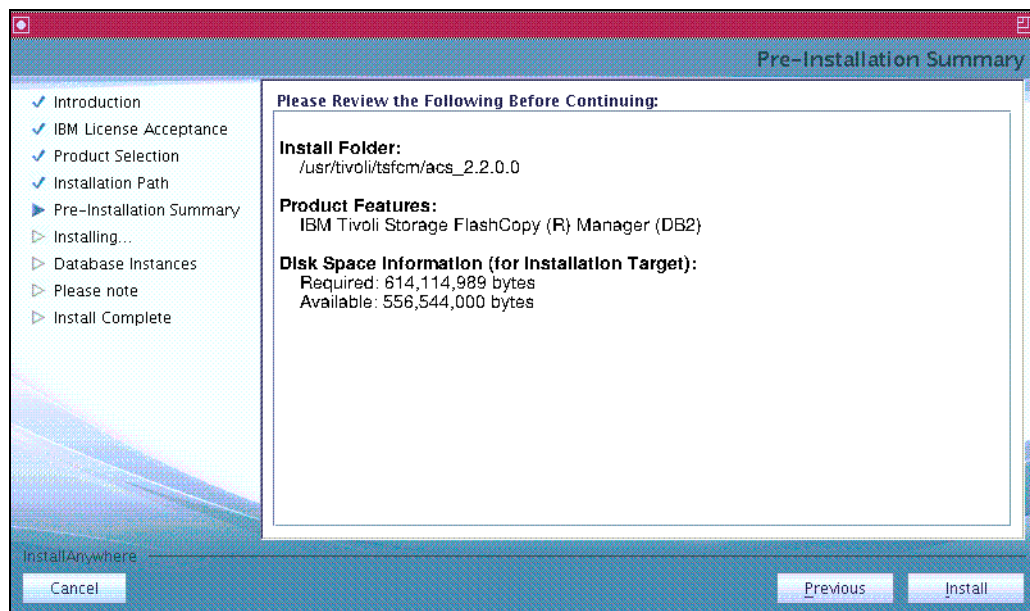


Figure 16-2 Pre-installation summary

5. After the installation completes, log in to the server as the database owner and start the `setup_db2.sh` script, which asks specific setup questions about the environment.
6. The configuration of the `init<SID>.utl` and `init<SID>.sap` file is only necessary if Tivoli Storage Manager for Enterprise Resource Planning is installed.

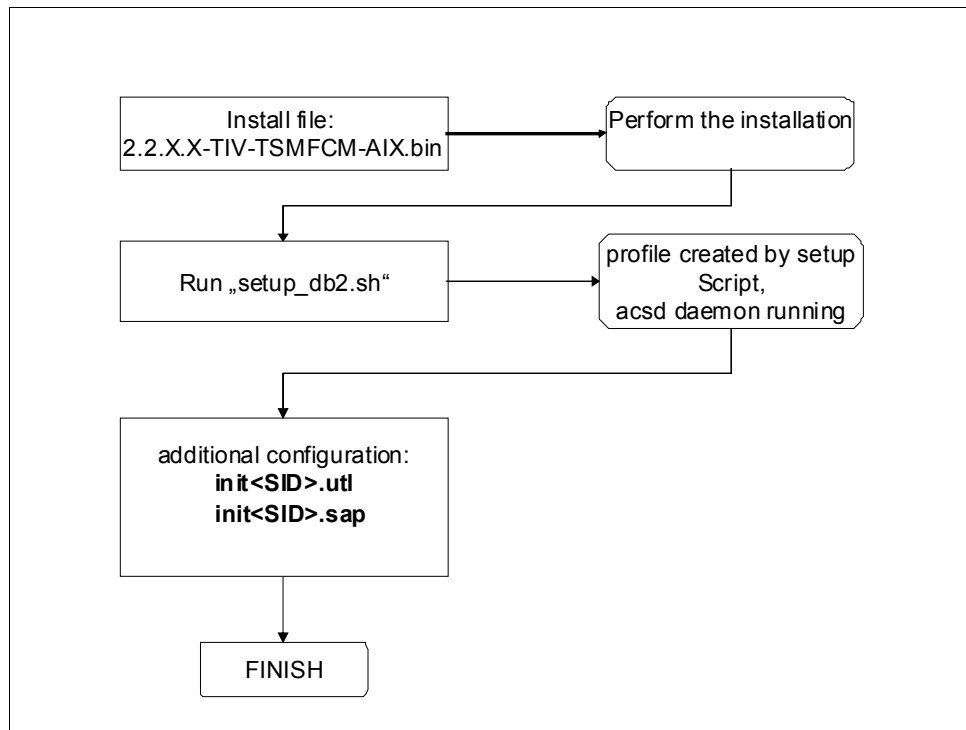


Figure 16-3 FlashCopy Manager installation workflow

16.3.1 FlashCopy Manager disk-only backup

The following section covers configuration and use of a FlashCopy Manager disk-only backup.

Disk-only backup configuration

A disk-only backup uses the point-in-time copy function of the storage subsystem to create copies of the LUNs that host the database. A disk-only backup requires no backup server or Tivoli Storage Manager server.

After installing FlashCopy Manager, a profile is required to run FlashCopy Manager properly. In the following example for a DB2 environment, FlashCopy Manager is configured to back up to disk only. To create the profile, log in as the db2 database instance owner and run the `setup_db2.sh` script on the production system. The script asks several profile-content questions. The main questions and answers for the XIV storage system are displayed in Example 16-1 on page 287. When starting the `setup_db2.sh` script you will be asked, what configuration you want:

- (1) backup only
- (2) cloning only
- (3) backup and cloning

For a disk-only configuration enter “1” to configure FlashCopy Manager for backup only. In Example 16-1 on page 287 the part of the XIV configuration is shown and the user input is indicated in **bold font type**.

In this example the device type is “XIV” and the `xcli` is installed in the `/usr/cli` directory on the production system. Specify the IP-address of the XIV storage system and enter a valid

XIV-user. The password for the XIV-user has to be specified at the end. The connection to the XIV will be checked while the script is running.

Example 16-1 FlashCopy Manager XIV configuration

```

***** Profile parameters for section DEVICE_CLASS DISK_ONLY: *****

Type of Storage system {COPYSERVICES_HARDWARE_TYPE} (DS8000|SVC|XIV) = XIV
Storage system ID of referred cluster {STORAGE_SYSTEM_ID} =      []
Filepath to XCLI command line tool {PATH_TO_XCLI} =      *input mandatory*
/usr/xcli
Hostname of XIV system {COPYSERVICES_SERVERNAME} =      *input mandatory*
9.155.90.180
Username for storage device {COPYSERVICES_USERNAME} = itso
Hostname of backup host {BACKUP_HOST_NAME} =      [NONE]
Interval for reconciliation {RECON_INTERVAL} (<hours> ) =      [12]
Grace period to retain snapshots {GRACE_PERIOD} (<hours> ) =      [24]
Use writable snapshots {USE_WRITABLE_SNAPSHOTS} (YES|NO|AUTO) = [AUTO]
Use consistency groups {USE_CONSISTENCY_GROUPS} (YES|NO) =      [YES]
.
.
Do you want to continue by specifying passwords for the defined sections? [Y/N]
y
Please enter the password for authentication with the ACS daemon:      [***]
Please re-enter password for verification:
Please enter the password for storage device configured in section(s) DISK_ONLY:
<< enter the password for the XIV >>

```

A disk-only backup is initiated with the **db2 backup** command and the **use snapshot** clause. DB2 creates a time stamp for the backup image ID that is displayed in the output of the **db2 backup** command and can also be read out with the FlashCopy Manager **db2acsutil** utility or the **db2 list history** DB2 command. This time stamp is required to initiate a restore. For a disk-only backup, no backup server or Tivoli Storage Manager server is required as shown in Figure 16-4.

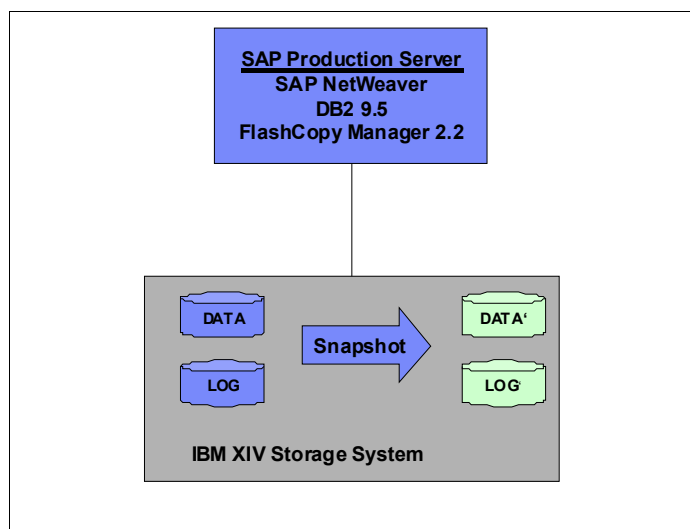


Figure 16-4 FlashCopy Manager with disk-only backup

The advantage of the XIV storage system is that snapshot target volumes do not have to be predefined. FlashCopy Manager creates the snapshots automatically during the backup or cloning processing.

Disk-only backup

A disk-only backup is initiated with the **db2 backup** command and the **use snapshot** clause. Example 16-2 shows how to create a disk-only backup with FlashCopy Manager. The user has to login as the db2 instance owner and can start the disk-only backup from the command line. The SAP system can stay up and running while FlashCopy Manager does the online backup.

Example 16-2 FlashCopy Manager disk-only backup

```
db2t2p> db2 backup db T2P online use snapshot
```

Backup successful. The timestamp for this backup image is : 20100315143840

Restore from snapshot

Before a restore can happen, shut down the application. A disk-only backup can be restored and recovered with DB2 commands. Snapshots are done on a volume group level. In other words: the storage-based snapshot feature is not aware of the database and file-systems structures and cannot perform restore operations on a file or table space level. FlashCopy manager backs up and restores only the volume groups.

The following command-line example shows restore, forward recovery, and activation of the database with the appropriate DB2 commands: **db2 restore**, **db2 rollforward**, and **db2 activate**.

Example 16-3 FlashCopy Manager snapshot restore

```
db2t2p> db2 restore database T2P use snapshot taken at 20100315143840
SQL2539W Warning! Restoring to an existing database that is the same as the
backup image database. The database files will be deleted.
Do you want to continue ? (y/n) y
DB20000I The RESTORE DATABASE command completed successfully.
```

```
db2od3> db2 start db manager
DB20000I The START DATABASE MANAGER command completed successfully.
```

```
db2od3> db2 rollforward database T2P complete
DB20000I The ROLLFORWARD command completed successfully.
```

```
db2od3> db2 activate db T2P
DB20000I The ACTIVATE DATABASE command completed successfully.
```

The XIV-GUI window in Figure 16-5 on page 289 shows multiple sequenced XIV snapshots created by FlashCopy Manager. XIV allocates snapshot space at the time it is required.

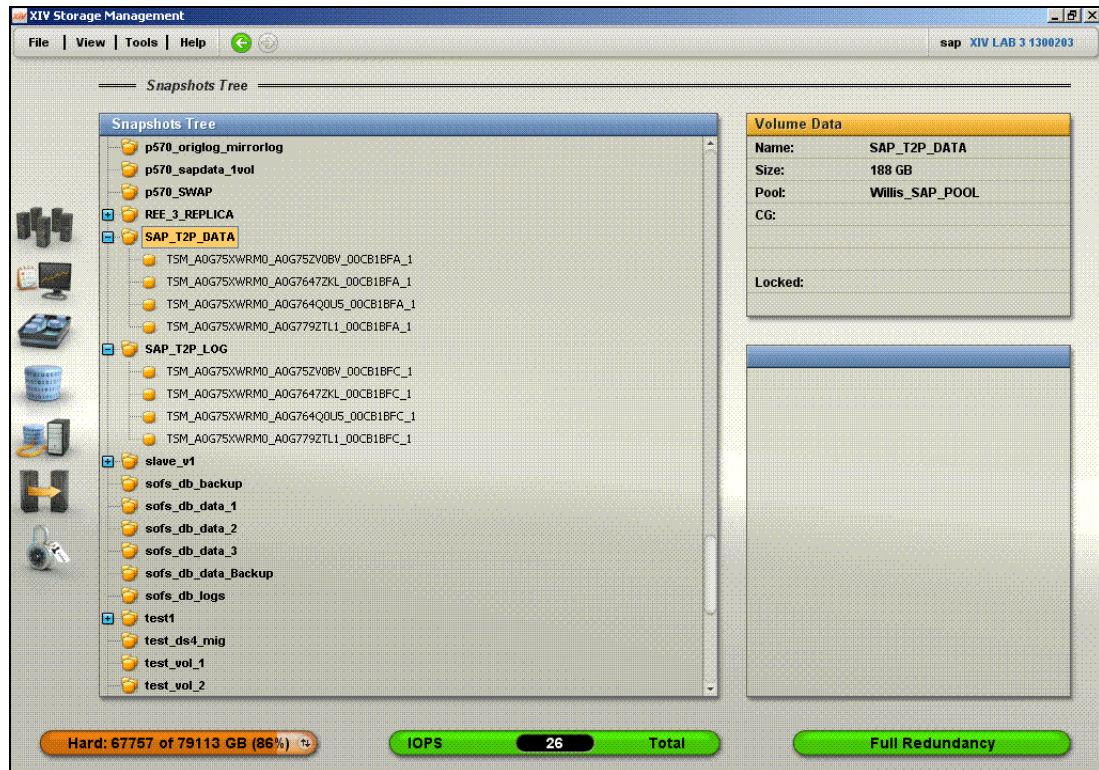


Figure 16-5 XIV snapshots view

Note: Check that enough XIV snapshot space is available for the number of snapshot versions to keep. If snapshot space is not sufficient, XIV will delete older snapshot versions.

Snapshot deletions are not immediately reflected in the FlashCopy Manager repository. FlashCopy Manager's interval for reconciliation is specified during FlashCopy Manager setup, and can be checked and updated in the FlashCopy Manager profile. The current default of the `RECON_INTERVAL` parameter is 12 hours (see Example 16-1 on page 287).

16.3.2 SAP Cloning

A productive SAP environment consists of multiple systems: production, quality assurance (QA), development and more. SAP recommends that you perform a system copy if you plan to set up a test system, demo system, or training system.

Possible reasons to perform system copies are as follows:

- ▶ To create test and quality-assurance systems that are recreated regularly from the production systems to test new developments with the most recent, actual production data
- ▶ To create migration or upgrade systems from the production system prior to phasing in new releases or functions into production
- ▶ To create education systems from a master training set to reset before a new course starts
- ▶ To create dedicated reporting systems to off-load workload from production

SAP defines a system copy as the duplication of an SAP system. Certain SAP parameters might change in a copy. When you perform a system copy, the SAP `SAPinst` procedure installs all the instances again, but instead of the database export delivered by SAP, it uses a

copy of the customer's source system database to set up the database. Commonly, a backup of the source system database is used to perform a system copy.

SAP differentiates between two system-copy modes:

- ▶ A *Homogeneous System Copy* uses the same operating system and database platform as the original system.
- ▶ A *Heterogeneous System Copy* changes either the operating system or the database system, or both. Heterogeneous system copy is a synonym for migration.

Performing an SAP system copy to back up and restore a production system is a long task (two or three days). Changes to the target system are usually applied either manually or supported by customer-written scripts. SAP strongly recommends that you only perform a system copy if you have experience in copying systems and good knowledge of the operating system, the database, the ABAP Dictionary, and the Java™ Dictionary.

Starting with version 2.2, Tivoli FlashCopy Manager supports the cloning (in SAP terms: the heterogeneous system copy) of an SAP database. The product leverages the FlashCopy or Snapshot features of IBM storage systems to create a point-in-time copy of the SAP source database in minutes instead of hours. The cloning process of an SAP database is shown in Figure 16-6.

FlashCopy Manager automatically performs these tasks:

- ▶ Creating a consistent snapshot of the volumes on which the production database resides
- ▶ Configuring, importing, and mounting the snapshot volumes on the clone system
- ▶ Recovering the database on the clone system
- ▶ Renaming the database to match the name of the clone database that resides on the clone system
- ▶ Starting the clone database on the clone system

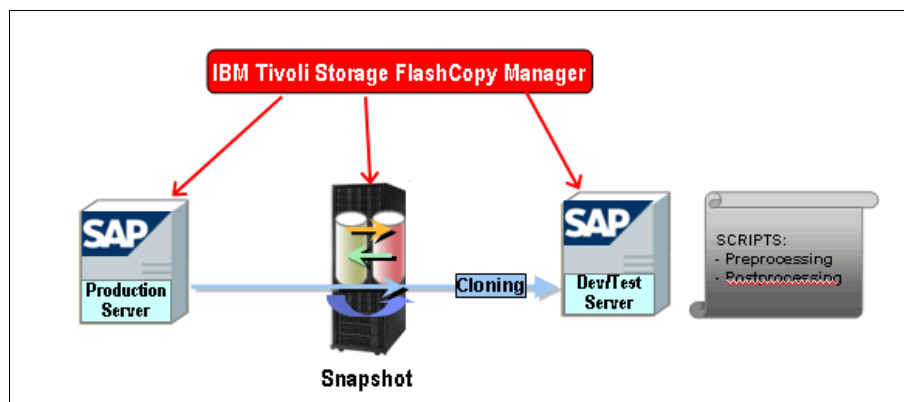


Figure 16-6 SAP cloning overview

The cloning function is useful to create quality assurance (QA) or test systems from production systems, as shown in Figure 16-7 on page 291. The renamed clone system can be integrated into the SAP Transport System that an SAP customer defines for his SAP landscape. Then updated SAP program sources and other SAP objects can be transported to the clone system for testing purpose.

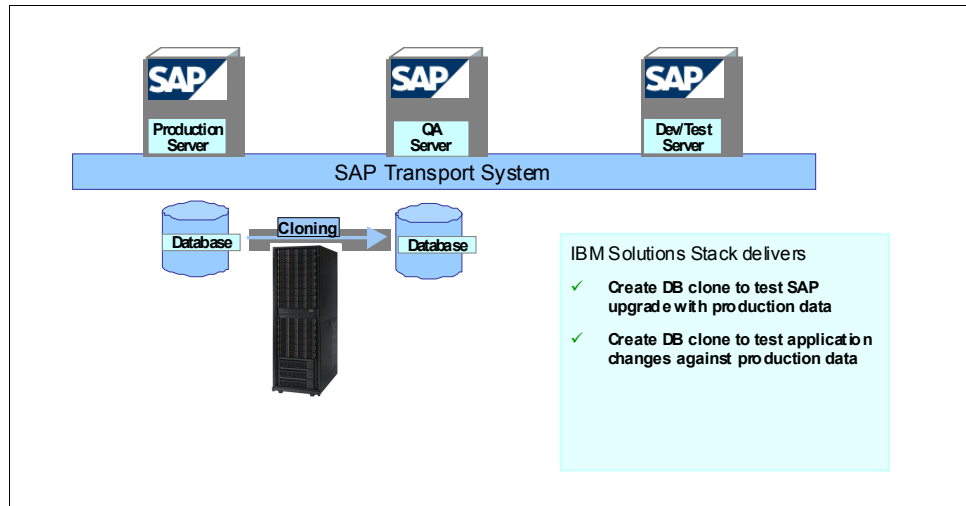


Figure 16-7 SAP Cloning example - upgrade and application test

IBM can provide a number of preprocessing and postprocessing scripts that automate important actions. FlashCopy Manager provides the ability to automatically run these scripts before and after clone creation, and before the cloned SAP system is started. The pre- and postprocessing scripts are not part of the FlashCopy Manager software package.

For more detailed information about backup/restore and SAP Cloning with FlashCopy Manager on Unix, refer to the following documents:

Quick Start Guides to FlashCopy Manager for SAP on IBM DB2 or Oracle Database with IBM XIV Storage System at the following URL:

<http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101703>

Tivoli Storage FlashCopy Manager Version 2.2 Installation and User's Guide for Unix and Linux at the following URL:

http://publib.boulder.ibm.com/infocenter/tsminfo/v6r2/topic/com.ibm.itsm.fcm.unx.doc/b_fcm_unx_guide.pdf

16.4 Tivoli Storage FlashCopy Manager for Windows

The XIV Snapshot function can be combined with the Microsoft Volume Shadow Copy Services (VSS) and IBM Tivoli Storage FlashCopy Manager to provide efficient and reliable application or database backup and recovery solutions.

After a brief overview of the Microsoft VSS architecture, we cover the requirements, configuration, and implementation of the XIV VSS Provider with the Tivoli Storage FlashCopy Manager for backing up Microsoft Exchange Server data. The product provides the tools and information needed to create and manage volume-level snapshots of Microsoft SQL Server and Microsoft Exchange server data. Tivoli Storage FlashCopy Manager uses Microsoft Volume Shadow Copy Services in a Windows environment. VSS relies on a VSS hardware provider. See 16.5.1, "VSS architecture and components" on page 293.

We explain in subsequent sections the installation of the XIV VSS Provider and provide detailed installation and configuration information for the IBM Tivoli Storage FlashCopy Manager. We have also included usage scenarios.

Tivoli Storage FlashCopy Manager for Windows is a package that is easy to install, configure, and deploy, and integrates in a seamless manner with any storage system that has a VSS provider like the IBM System Storage DS3000, DS4000, DS5000, DS8000, IBM SAN Volume Controller, IBM Storwize V7000, and IBM XIV Storage System.

Figure 16-8 shows the Tivoli Storage FlashCopy Manager Management Console in the windows environment.

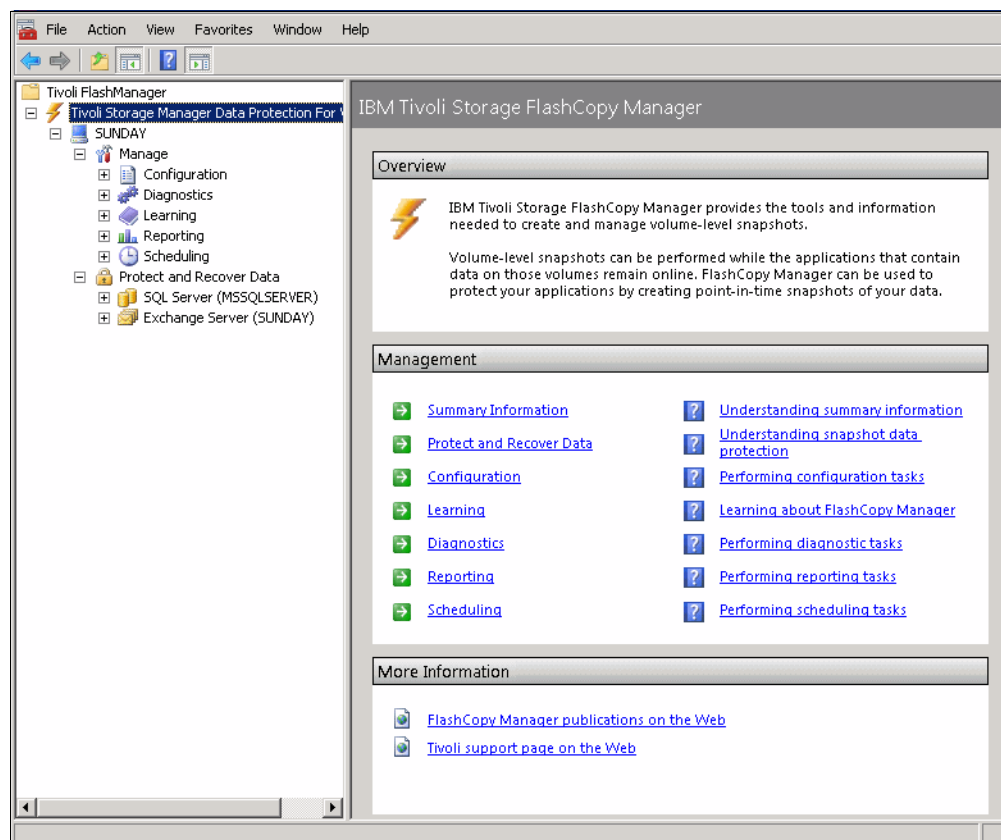


Figure 16-8 Tivoli Storage FlashCopy Manager: Management Console

16.5 Windows Server 2008 Volume Shadow Copy Service

Microsoft first introduced Volume Shadow Copy Services in Windows 2003 Server and has included it in all subsequent releases. VSS provides a framework and the mechanisms to create consistent point-in-time copies (known as shadow copies) of databases and applications data. It consists of a set of Microsoft COM APIs that enable volume-level snapshots to be performed while the applications that contain data on those volumes remain online and continue to write. This enables third-party software like FlashCopy Manager to centrally manage the backup and restore operation.

Without VSS, if you do not have an online backup solution implemented, you either must stop or quiesce applications during the backup process, or live with the side effects of an online backup with inconsistent data and open files that could not be backed up.

With VSS, you can produce consistent shadow copies by coordinating tasks with business applications, file system services, backup applications, fast recovery solutions, and storage hardware such as the XIV Storage System.

16.5.1 VSS architecture and components

Figure 16-9 shows the VSS architecture and how the VSS service interacts with the other components to create a *shadow copy* of a volume, or, when it pertains to XIV, a *volume snapshot*.

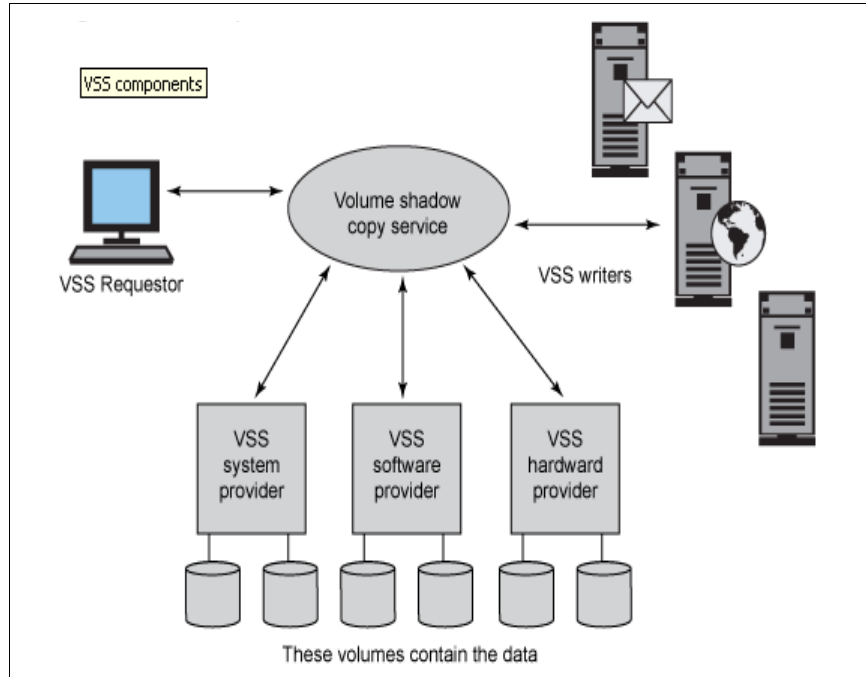


Figure 16-9 VSS architecture

The components of the VSS architecture are:

- VSS Service

The VSS Service is at the core of the VSS architecture. It is the Microsoft Windows service that directs all of the other VSS components that are required to create the volume's shadow copies (snapshots). This Windows service is the overall coordinator for all VSS operations.

- Requestor

This is the software application that commands that a shadow copy be created for specified volumes. The VSS requestor is provided by Tivoli Storage FlashCopy Manager and is installed with the Tivoli Storage FlashCopy Manager software.

- Writer

This is a component of a software application that places the persistent information for the shadow copy on the specified volumes. A database application (such as SQL Server or Exchange Server) or a system service (such as Active Directory) can be a writer.

Writers serve two main purposes by:

- Responding to signals provided by VSS to interface with applications to prepare for shadow copy.
- Providing information about the application name, icons, files, and a strategy to restore the files.

Writers prevent data inconsistencies.

For exchange data, the Microsoft Exchange Server contains the writer components and requires no configuration.

For SQL data, Microsoft SQL Server contains the writer components (SqlServerWriter). It is installed with the SQL Server software and requires the following minor configuration tasks:

- Set the SqlServerWriter service to automatic. This enables the service to start automatically when the machine is rebooted.
- Start the SqlServerWriter service.

► Provider

This is the application that produces the shadow copy and also manages its availability. It can be a system provider (such as the one included with the Microsoft Windows operating system), a software provider, or a hardware provider (such as the one available with the XIV storage system).

For XIV, you must install and configure the IBM XIV VSS Provider.

VSS uses the following terminology to characterize the nature of volumes participating in a shadow copy operation:

► Persistent

This is a shadow copy that remains after the backup application completes its operations. This type of shadow copy also survives system reboots.

► Non-persistent

This is a temporary shadow copy that remains only as long as the backup application needs it to copy the data to its backup repository.

► Transportable

This is a shadow copy volume that is accessible from a secondary host so that the backup can be off-loaded. Transportable is a feature of hardware snapshot providers. On an XIV you can mount a snapshot volume to another host.

► Source volume

This is the volume that contains the data to be shadow copied. These volumes contain the application data.

► Target or snapshot volume

This is the volume that retains the shadow-copied storage files. It is an exact copy of the source volume at the time of backup.

VSS supports the following shadow copy methods:

► Clone (full copy/split mirror)

A clone is a shadow copy volume that is a full copy of the original data as it resides on a volume. The source volume continues to take application changes while the shadow copy volume remains an exact read-only copy of the original data at the point-in-time that it was created.

► Copy-on-write (differential copy)

A copy-on-write shadow copy volume is a differential copy (rather than a full copy) of the original data as it resides on a volume. This method makes a copy of the original data before it is overwritten with new changes. Using the modified blocks and the unchanged blocks in the original volume, a shadow copy can be logically constructed that represents the shadow copy at the point-in-time at which it was created.

- Redirect-on-write (differential copy)

A redirect-on-write shadow copy volume is a differential copy (rather than a full copy) of the original data as it resides on a volume. This method is similar to copy-on-write, without the double-write penalty, and it offers storage-space- and performance-efficient snapshots. New writes to the original volume are redirected to another location set aside for the snapshot. The advantage of redirecting the write is that only one write takes place, whereas with copy-on-write, two writes occur (one to copy original data onto the storage space, the other to copy changed data). The XIV storage system supports redirect-on-write.

16.5.2 Microsoft Volume Shadow Copy Service function

Microsoft VSS accomplishes the fast backup process when a backup application (the requestor, which is Tivoli Storage FlashCopy Manager in our case) initiates a shadow copy backup. The VSS service coordinates with the VSS-aware writers to briefly hold writes on databases, applications, or both. VSS flushes the file system buffers and asks a provider (such as the XIV provider) to initiate a snapshot of the data. When the snapshot is logically completed, VSS allows writes to resume and notifies the requestor that the backup has completed successfully. The backup volumes are mounted, but hidden and read-only, ready to be used when a rapid restore is requested. Alternatively, the volumes can be mounted to a separate host and used for application testing or backup to tape.

The Microsoft VSS FlashCopy process includes the following steps:

1. The requestor notifies VSS to prepare for a shadow copy creation.
2. VSS notifies the application-specific writer to prepare its data for making a shadow copy.
3. The writer prepares the data for that application by completing all open transactions, flushing cache and buffers, and writing in-memory data to disk.
4. When the application data is ready for shadow copy, the writer notifies VSS, which in turn relays the message to the requestor to initiate the commit copy phase.
5. VSS temporarily quiesces application I/O write requests for a few seconds, and the VSS hardware provider performs the snapshot on the storage system.
6. After the storage snapshot has completed, VSS releases the quiesce, and the database or application writes resume.
7. VSS queries the writers to confirm that write I/Os were successfully held during the Volume Shadow Copy.

16.6 XIV VSS provider

A VSS hardware provider, such as the XIV VSS Provider, is used by third-party software to act as an interface between the hardware (storage system) and the operating system. The third-party application (which can be IBM Tivoli Storage FlashCopy Manager) uses XIV VSS Provider to instruct the XIV storage system to perform a snapshot of a volume attached to the host system.

16.6.1 XIV VSS Provider installation

This section illustrates the installation of the XIV VSS Provider. First, make sure that your Windows system meets the minimum requirements listed below.

At the time of writing, the XIV VSS Provider 2.2.3 version was available. We used a Windows 2008 64bit host system for our tests. To obtain the system requirements, please refer to the *IBM VSS Provider - Xprov Release Notes*. There is a chapter about the system requirements.

The XIV VSS Hardware Provider 2.2.3 version and release notes can be downloaded at:

<http://www.ibm.com/systems/storage/disk/xiv/index.html>

The installation of the XIV VSS Provider is a straightforward Windows application installation:

1. Locate the XIV VSS Provider installation file, also known as the *xProv* installation file. If the XIV VSS Provider 2.2.3 is downloaded from the Internet, the file name is *xProvSetup-x64-2.2.3.exe*. Execute the file to start the installation.

Tip: Uninstall any previous versions of the XIV VSS xProv driver if installed. An upgrade is not allowed with the 2.2.3 release of XIV VSS provider.

2. A Welcome window opens as shown in Figure 16-10. Click **Next**.

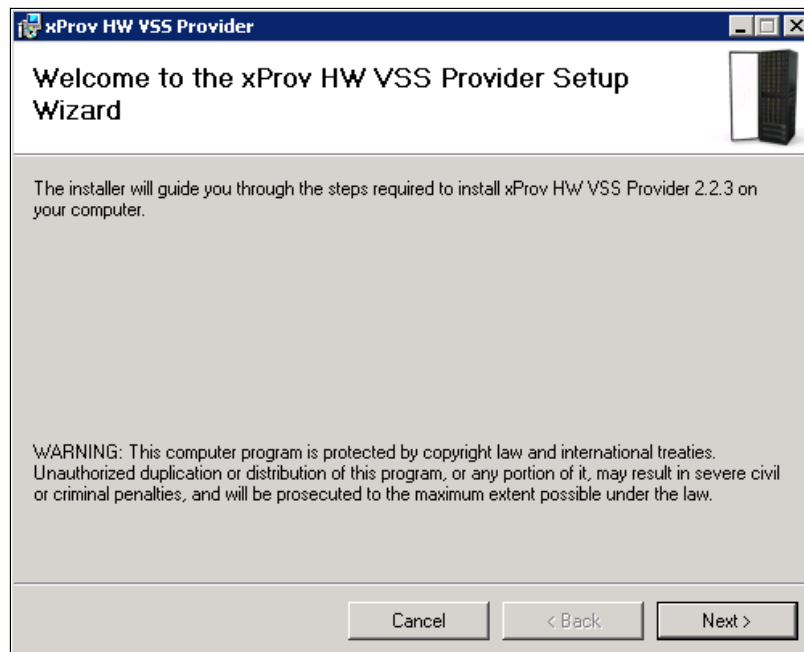


Figure 16-10 XIV VSS provider installation: Welcome window

3. The License Agreement window is displayed. To continue the installation you must accept the license agreement.
4. In the next step you can specify the XIV VSS Provider configuration file directory and the installation directory. Keep the default directory folder and installation folder or change it to meet your needs.

5. The next dialog window is for post-installation operations, as shown in Figure 16-11. Perform a post-installation configuration during the installation process. The configuration can, however, be performed at later time. When done, click **Next**.

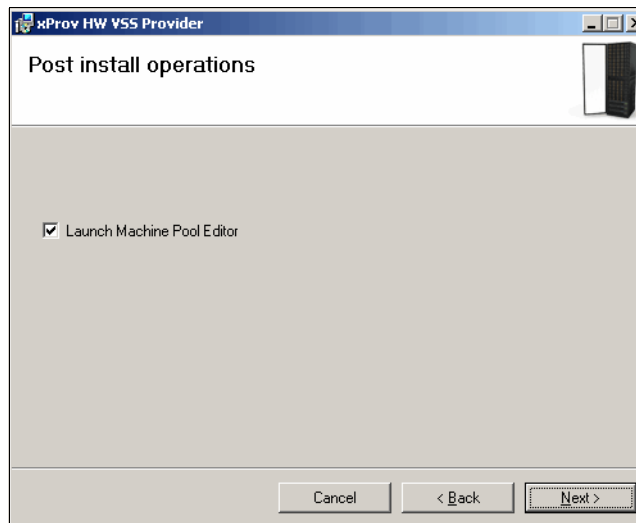


Figure 16-11 Installation: post-installation operation

6. A Confirm Installation window is displayed. If required you can go back to make changes or confirm the installation by clicking **Next**.
7. After the installation is complete click **Close** to exit.

16.6.2 XIV VSS Provider configuration

The XIV VSS Provider must now be configured using the following steps:

1. If the post installation check box was selected during the installation (Figure 16-11 on page 297), the XIV VSS Provider configuration window shown in Figure 16-13 is now displayed. If the post-installation check box had not been selected during the installation, it must be manually invoked by selecting **Start** → **All Programs** → **XIV** and starting the MachinePool Editor, as shown in Figure 16-12.



Figure 16-12 Configuration: XIV VSS Provider setup

2. Right-click the Machine Pool Editor. The New System window displays. Provide specific information regarding the XIV Storage System IP addresses and user ID and password with admin privileges. Have that information available.
3. In the dialog shown in Figure 16-13 on page 298, click **New System**.

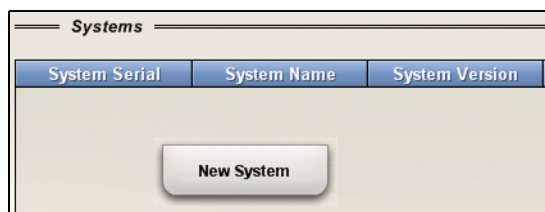


Figure 16-13 XIV Configuration: Machine Pool Editor

4. The Add System Management window shown in Figure 16-14 is displayed. Enter the user name and password of an XIV user with administrator privileges (storageadmin role) and the primary IP address of the XIV Storage System, then click **Add**.

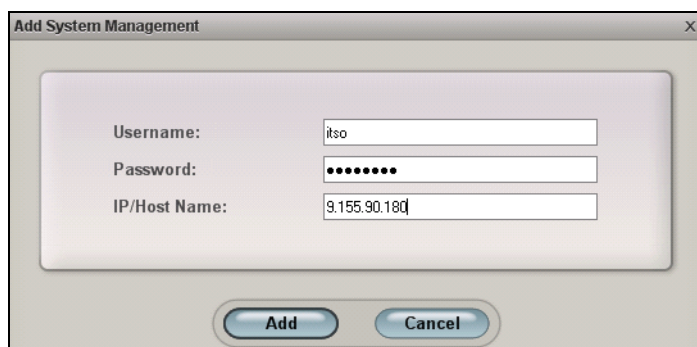


Figure 16-14 XIV configuration: add machine

5. You are now returned to the VSS MachinePool Editor window. The VSS Provider collected additional information about the XIV storage system, as illustrated in Figure 16-15.

Systems						
System Serial	System Name	System Ver...	IP/Host Na...	IP/Host Name 2	IP/Host Name 3	Username
1300203	XIV LAB 3 1300203	10.2.2	9.155.90.180	9.155.90.181	9.155.90.182	itso

Figure 16-15 XIV Configuration: Machine Pool Editor

At this point the XIV VSS Provider configuration is complete and you can close the Machine Pool Editor window. If you must add other XIV Storage Systems, repeat steps 3 to 5.

After the XIV VSS provider has been configured as just explained, ensure that the operating system can recognize it. For that purpose, launch the **vssadmin list providers** command from the operating system command line.

Make sure that IBM XIV VSS HW Provider appears among the list of installed VSS providers returned by the **vssadmin** command as shown in Example 16-4.

Example 16-4 Output of vssadmin command

```
C:\Users\Administrator>vssadmin list providers
vssadmin 1.1 - Volume Shadow Copy Service administrative command-line tool
(C) Copyright 2001-2005 Microsoft Corp.
```

```
Provider name: 'Microsoft Software Shadow Copy provider 1.0'
Provider type: System
Provider Id: {b5946137-7b9f-4925-af80-51abd60b20d5}
Version: 1.0.0.7
```

Provider name: 'IBM XIV VSS HW Provider'
Provider type: Hardware
Provider Id: {d51fe294-36c3-4ead-b837-1a6783844b1d}
Version: 2.2.3

Tip: The XIV VSS Provider log file is located in *C:\Windows\Temp\xProvDotNet*.

The Windows server is now ready to perform snapshot operations on the XIV Storage System. Refer to your application documentation for information on completing the VSS setup.

16.7 Installing Tivoli Storage FlashCopy Manager for Microsoft Exchange

To install Tivoli Storage FlashCopy Manager, insert the product media into the DVD drive and the installation starts automatically. If this does not occur, or if you are using a copy or downloaded version of the media, locate and execute the SetpFCM.exe file. During the installation, accept all default values.

The Tivoli Storage FlashCopy Manager installation and configuration wizards will guide you through the installation and configuration steps. After you run the setup and configuration wizards, your computer is ready to take snapshots.

Tivoli Storage FlashCopy Manager provides the following wizards for installation and configuration tasks:

- ▶ Setup wizard
Use this wizard to install Tivoli Storage FlashCopy Manager on your computer.
- ▶ Local configuration wizard
Use this wizard to configure Tivoli Storage FlashCopy Manager on your computer to provide locally-managed snapshot support. To manually start the configuration wizard, double-click **Local Configuration** in the results pane.
- ▶ Tivoli Storage Manager configuration wizard
Use this wizard to configure Tivoli Storage FlashCopy Manager to manage snapshot backups using a Tivoli Storage Manager server. This wizard is only available when a Tivoli Storage Manager license is installed.

After it is installed, Tivoli Storage FlashCopy Manager must be configured for VSS snapshot backups. Use the local configuration wizard for that purpose. These tasks include selecting the applications to protect, verifying requirements, provisioning, and configuring the components required to support the selected applications.

To configure for Microsoft Exchange Server, perform these steps:

1. Start the Local Configuration Wizard from the Tivoli Storage FlashCopy Manager Management Console, as shown in Figure 16-16 on page 300.

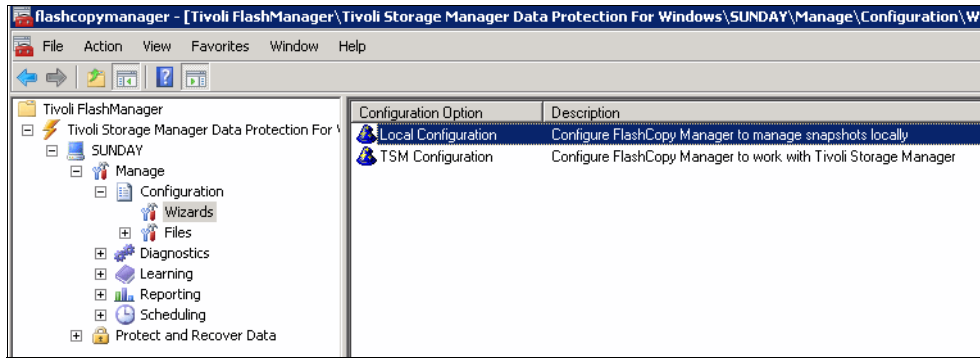


Figure 16-16 Tivoli FlashCopy Manager: local configuration wizard for Exchange Server

2. A dialog window is displayed, as shown in Figure 16-17. Select the Exchange Server to configure and click **Next**.

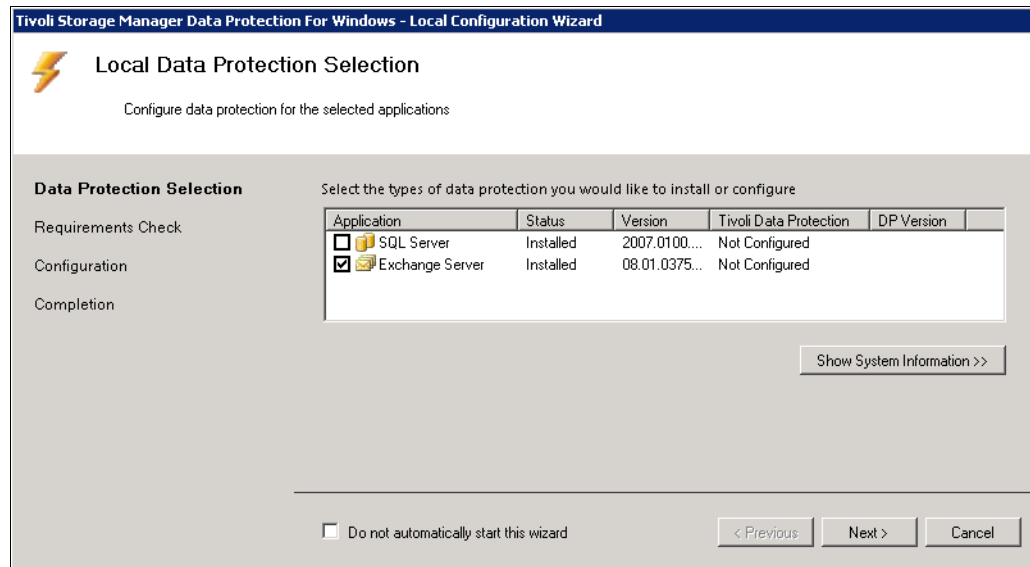


Figure 16-17 Local configuration wizard: local data protection selection

Note: The Show System Information button shows the basic information about your host system.

Tip: Select the check box at the bottom if you do not want the local configuration wizard to start automatically the next time that the Tivoli Storage FlashCopy Manager Management Console windows starts.

3. The Requirements Check window displays as shown in Figure 16-18 on page 301. At this stage, the systems checks that all prerequisites are met.

If any requirement is not met, the configuration wizard does not proceed to the next step. You might have to upgrade components to fulfill the requirements. The requirements check can be run again by clicking **Re-run** after they are fulfilled. When the check completes successfully, click **Next**.

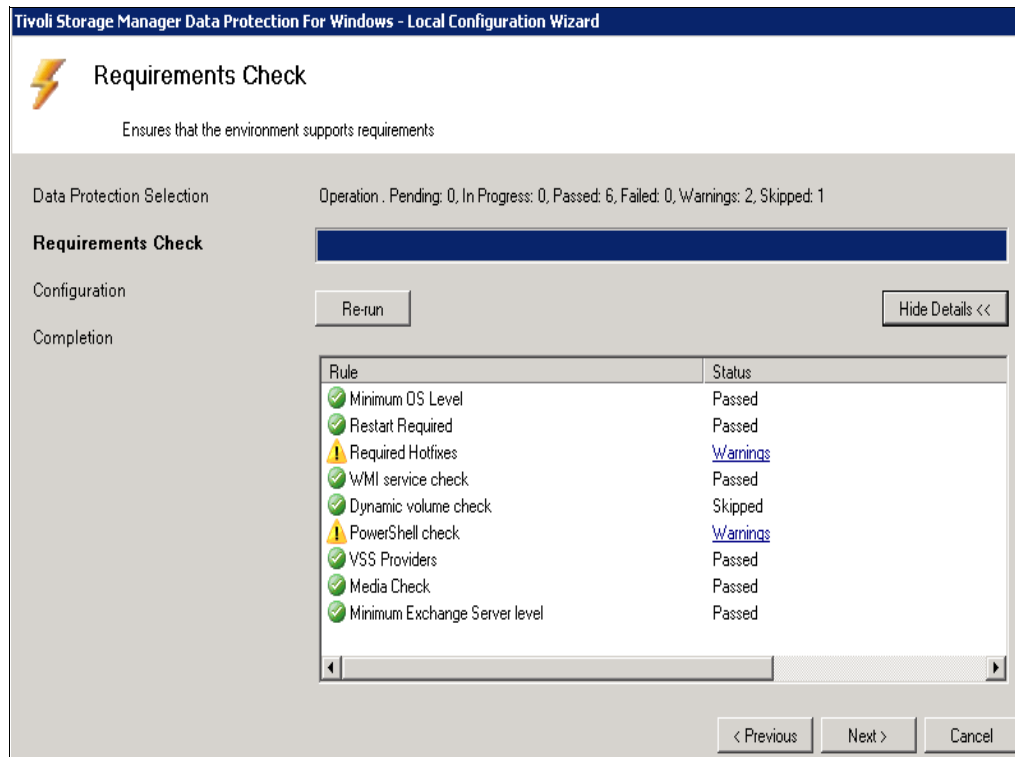


Figure 16-18 Local Configuration for exchange: requirements check

- The Local Configuration wizard now performs all necessary configuration steps as shown in Figure 16-19. The steps include provisioning and configuring the VSS Requestor, provisioning and configuring data protection for the Exchange Server, and configuring services. When done, click **Next**.

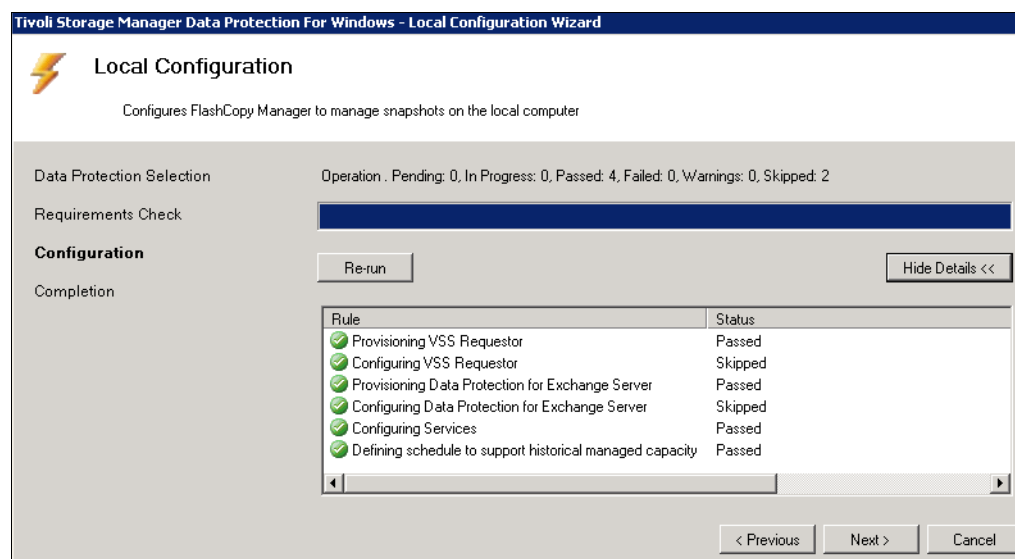


Figure 16-19 Local configuration for exchange: configuration

Note: By default, details are hidden. Details can be shown or hidden by clicking **Show Details** or **Hide Details**.

5. The completion window shown in Figure 16-20 is displayed. To run a VSS diagnostic check, ensure that the corresponding check box is selected and click **Finish**.

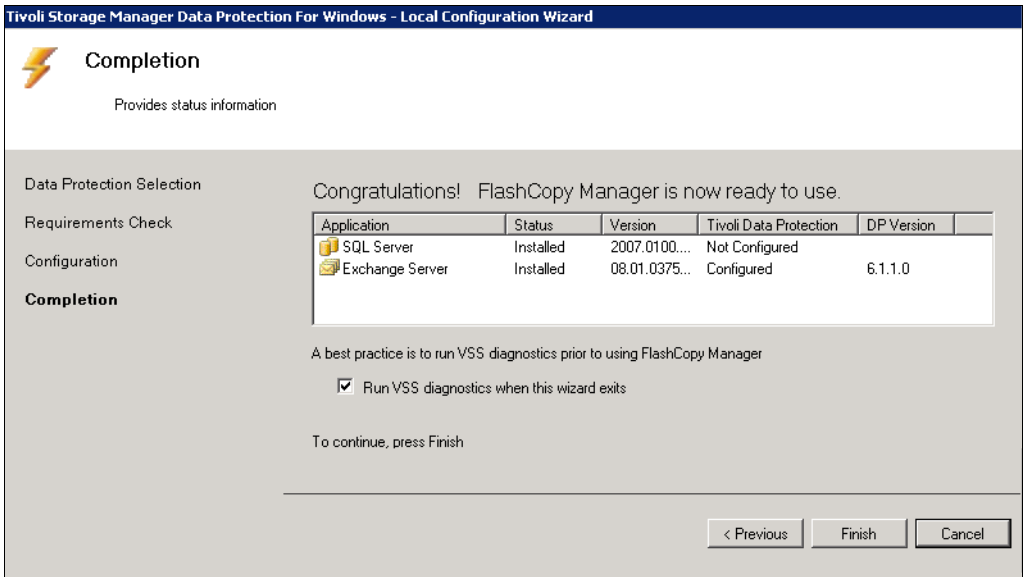


Figure 16-20 Local configuration for exchange: completion

6. The VSS Diagnostic window is displayed. The goal of this step is to verify that any volume that you select is capable of performing an XIV snapshot using VSS. Select the XIV mapped volumes to test, as shown in Figure 16-21, and click **Next**.

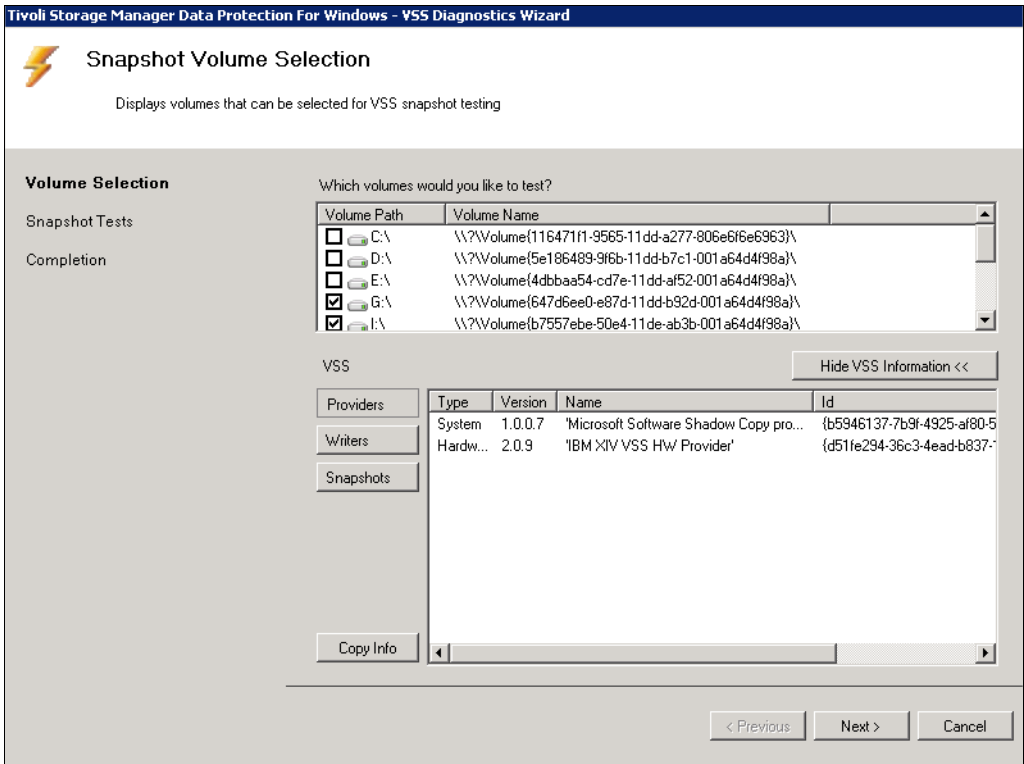


Figure 16-21 VSS Diagnostic Wizard: Snapshot Volume Selection

Tip: Any previously taken snapshots can be seen by clicking **Snapshots**. Clicking the button refreshes the list and shows all of the existing snapshots.

7. The VSS Snapshot Tests window is displayed, showing a status for each of the snapshots. This dialog also displays the event messages when you click **Show Details** as shown in Figure 16-22 on page 303. When done, click **Next**.

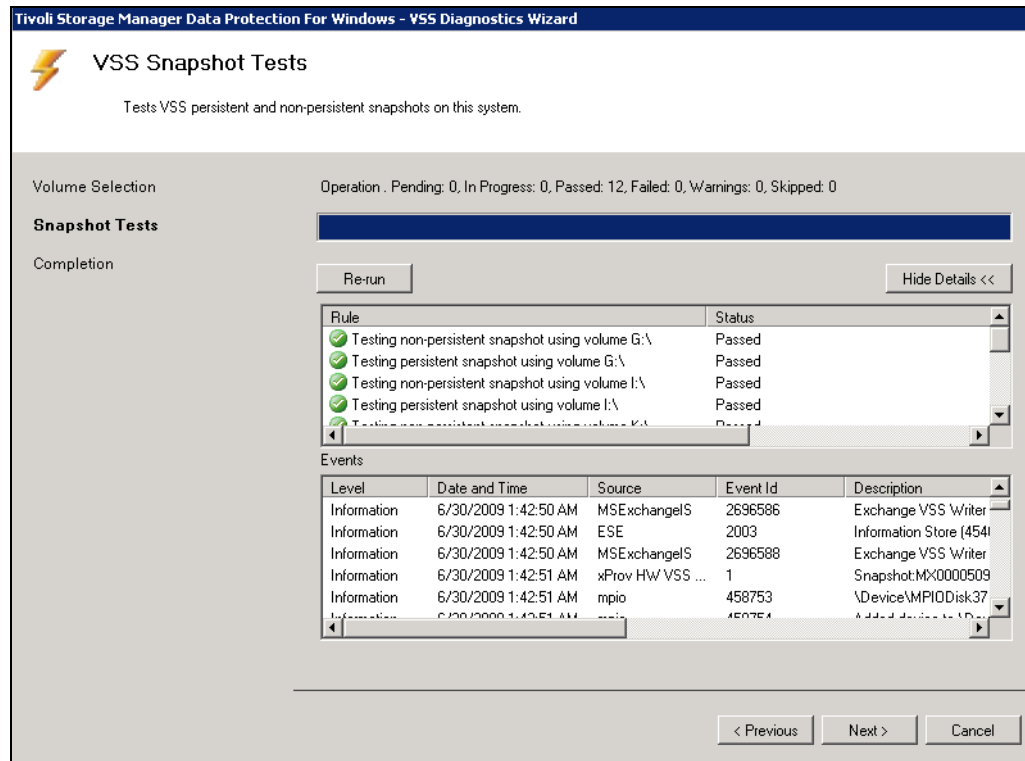


Figure 16-22 VSS Diagnostic Wizard: Snapshot tests

8. A completion window is displayed with the results. When done, click **Finish**.

Note: Microsoft SQL Server can be configured the same way as Microsoft Exchange Server to perform XIV VSS snapshots for Microsoft SQL Server using Tivoli Storage FlashCopy Manager.

16.8 Backup scenario for Microsoft Exchange Server

Microsoft Exchange Server is a Microsoft server-line product that provides the capability of messaging and collaboration software. The main features of Exchange Server are email exchange, contacts, and calendar functions.

To perform a VSS snapshot backup of Exchange data, we used the following setup:

- ▶ Windows 2008 64bit
- ▶ Exchange 2007 Server
- ▶ XIV Host Attachment Kit 1.0.4
- ▶ XIV VSS Provider 2.0.9
- ▶ Tivoli Storage FlashCopy Manager 2.0

Microsoft Exchange Server XIV VSS Snapshot backup

On the XIV Storage System a single volume has been created and mapped to the host system, as illustrated in Figure 16-23. On the Windows host system, the volume has been initialized as a basic disk and assigned the drive letter G. The G drive has been formatted as NTFS, and we created a single Exchange Server storage group with mailboxes on that drive.

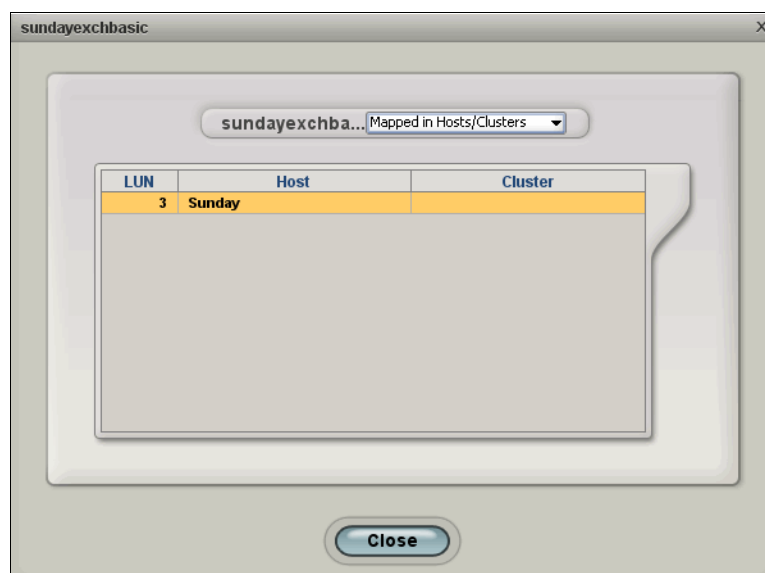


Figure 16-23 Mapped volume to the host system

Tivoli Storage FlashCopy Manager was already configured and tested for XIV VSS snapshot, as shown in 16.7, “Installing Tivoli Storage FlashCopy Manager for Microsoft Exchange” on page 299. To review the Tivoli Storage FlashCopy Manager configuration settings, use the command shown in Example 16-5.

Example 16-5 Tivoli Storage FlashCopy Manager for Mail: query DP configuration

```
C:\Program Files\Tivoli\TSM\TDPEXchange>tdpexcc query tdp
```

```
IBM FlashCopy Manager for Mail:
FlashCopy Manager for Microsoft Exchange Server
Version 6, Release 1, Level 1.0
(C) Copyright IBM Corporation 1998, 2009. All rights reserved.
```

```
FlashCopy Manager for Exchange Preferences
```

```
-----
BACKUPDESTination..... LOCAL
BACKUPMETHod..... VSS
BUFFers ..... 3
BUFFERSize ..... 1024
DATEformat ..... 1
LANGuage ..... ENU
LOCALDSMAgentnode..... sunday
LOGFile ..... tdpexc.log
LOGPrune ..... 60
MOUNTWait ..... Yes
NUMBERformat ..... 1
REMOTEDSMAgentnode.....
```

```

RETRies..... 4
TEMPDBRestorepath.....
TEMPLOGRestorepath.....
TIMEformat ..... 1

```

As explained earlier, Tivoli Storage FlashCopy Manger does not use (or need) a TSM server to perform a snapshot backup. You can see this when you execute the **query tsm** command, as shown in Example 16-6. The output does not show a TSM service but FLASHCOPYMANAGER instead for the NetWork Host Name of Server field. Tivoli Storage FlashCopy Manager creates a virtual server instead of using a TSM Server to perform a VSS snapshot backup.

Example 16-6 Tivoli FlashCopy Manager for Mail: query TSM

```
C:\Program Files\Tivoli\TSM\TDPEXchange>tdpexcc query tsm
```

```

IBM FlashCopy Manager for Mail:
FlashCopy Manager for Microsoft Exchange Server
Version 6, Release 1, Level 1.0
(C) Copyright IBM Corporation 1998, 2009. All rights reserved.

```

FlashCopy Manager Server Connection Information

```

Nodename ..... SUNDAY_EXCH
NetWork Host Name of Server ..... FLASHCOPYMANAGER
FCM API Version ..... Version 6, Release 1, Level 1.0

Server Name ..... Virtual Server
Server Type ..... Virtual Platform
Server Version ..... Version 6, Release 1, Level 1.0
Compression Mode ..... Client Determined
Domain Name ..... STANDARD
Active Policy Set ..... STANDARD
Default Management Class ..... STANDARD

```

Example 16-7 shows what options have been configured and used for TSM Client Agent to perform VSS snapshot backups.

Example 16-7 TSM Client Agent: option file

```

*=====*
*
* IBM Tivoli Storage Manager for Databases
*
* dsm.opt for the Microsoft Windows Backup-Archive Client Agent
*=====*
Nodename          sunday
CLUSTERnode       NO
PASSWORDAccess    Generate

*=====*
* TCP/IP Communication Options
*=====*
COMMMethod        TCPip
TCPSEVERADDRESS   FlashCopymanager
TCPPort           1500

```

TCPWindowSize 63
TCPBuffSize 32

Before we can perform any backup, we must ensure that VSS is properly configured for Microsoft Exchange Server and that the DSMagent service is running (Example 16-8).

Example 16-8 Tivoli Storage FlashCopy Manager: Query Exchange Server

C:\Program Files\Tivoli\TSM\TDPEXchange>**tdpexcc query exchange**

IBM FlashCopy Manager for Mail:
FlashCopy Manager for Microsoft Exchange Server
Version 6, Release 1, Level 1.0
(C) Copyright IBM Corporation 1998, 2009. All rights reserved.

Querying Exchange Server to gather storage group information, please wait...

Microsoft Exchange Server Information

Server Name: SUNDAY
Domain Name: sunday.local
Exchange Server Version: 8.1.375.1 (Exchange Server 2007)

Storage Groups with Databases and Status

First Storage Group
Circular Logging - Disabled
Replica - None
Recovery - False
Mailbox Database Online
User Define Public Folder Online

STG3G_XIVG2_BAS
Circular Logging - Disabled
Replica - None
Recovery - False
2nd MailBox Online
Mail Box1 Online

Volume Shadow Copy Service (VSS) Information

Writer Name : Microsoft Exchange Writer
Local DSMagent Node : sunday
Remote DSMagent Node :
Writer Status : Online
Selectable Components : 8

Our test Microsoft Exchange Storage Group is on drive G:\ and it is called STG3G_XIVG2_BAS. It contains two mailboxes:

- ▶ Mail Box1
- ▶ 2nd MailBox

Now we can take a full backup of the storage group by executing the **backup** command, as shown in Example 16-9.

Example 16-9 Tivoli Storage FlashCopy Manager: full XIV VSS snapshot backup

```
C:\Program Files\Tivoli\TSM\TDPEXchange>tdpexcc backup STG3G_XIVG2_BAS full
```

```
IBM FlashCopy Manager for Mail:
FlashCopy Manager for Microsoft Exchange Server
Version 6, Release 1, Level 1.0
(C) Copyright IBM Corporation 1998, 2009. All rights reserved.
```

```
Updating mailbox history on FCM Server...
Mailbox history has been updated successfully.
```

```
Querying Exchange Server to gather storage group information, please wait...
```

```
Connecting to FCM Server as node 'SUNDAY_EXCH'...
Connecting to Local DSM Agent 'sunday'...
Starting storage group backup...
```

```
Beginning VSS backup of 'STG3G_XIVG2_BAS'...
```

```
Executing system command: Exchange integrity check for storage group
'STG3G_XIVG2_BAS'
```

```
Files Examined/Completed/Failed: [ 4 / 4 / 0 ] Total Bytes: 44276
```

```
VSS Backup operation completed with rc = 0
```

```
Files Examined : 4
Files Completed : 4
Files Failed : 0
Total Bytes : 44276
```

Note that we did not specify a disk drive here. Tivoli Storage FlashCopy Manager finds out which disk drives to copy with snapshot when doing a backup of a Microsoft Exchange Storage Group. This is the advantage of an application-aware snapshot backup process.

To see a list of the available VSS snapshot backups issue a **query** command, as shown in Example 16-10.

Example 16-10 Tivoli Storage FlashCopy Manager: query full VSS snapshot backup

```
C:\Program Files\Tivoli\TSM\TDPEXchange>tdpexcc query TSM STG3G_XIVG2_BAS full
```

```
IBM FlashCopy Manager for Mail:
FlashCopy Manager for Microsoft Exchange Server
Version 6, Release 1, Level 1.0
(C) Copyright IBM Corporation 1998, 2009. All rights reserved.
```

```
Querying FlashCopy Manager server for a list of database backups, please wait...
```

```
Connecting to FCM Server as node 'SUNDAY_EXCH'...
```

Backup List

Exchange Server : SUNDAY

Storage Group : STG3G_XIVG2_BAS

Backup Date	Size	S Fmt	Type	Loc	Object Name/Database Name
06/30/2009 22:25:57	101.04MB	A VSS	full	Loc	20090630222557
	91.01MB				Logs
	6,160.00KB				Mail Box1
	4,112.00KB				2nd MailBox

To show that a restore operation is working, we deleted the *2nd Mailbox* mail box, as shown in Example 16-11.

Example 16-11 Deleting the mailbox and adding a file

G:\MSEExchangeSvr2007\Mailbox\STG3G_XIVG2_BAS>**dir**

Volume in drive G is XIVG2_SJCVTPOOL_BAS

Volume Serial Number is 344C-09F1

```

06/30/2009  11:05 PM    <DIR>          .
06/30/2009  11:05 PM    <DIR>          ..
06/30/2009  11:05 PM               4,210,688 2nd MailBox.edb
                                     :

```

G:\MSEExchangeSvr2007\Mailbox\STG3G_XIVG2_BAS> **del "2nd MailBox.edb"**

To perform a restore, all the mailboxes must be unmounted first. A restore will be done at the volume level, called instant restore (IR). When that is complete, the recovery operation will run, applying all the logs and mounting the mail boxes. The process is shown in Example 16-12.

Example 16-12 Tivoli Storage FlashCopy Manager: VSS Full Instant Restore and recovery.

C:\Program Files\Tivoli\TSM\TDPEXchange>**tdpexcc Restore STG3G_XIVG2_BAS Full**

/RECOVer=APPL

YALLlogs /MOUNTDatabases=Yes

IBM FlashCopy Manager for Mail:

FlashCopy Manager for Microsoft Exchange Server

Version 6, Release 1, Level 1.0

(C) Copyright IBM Corporation 1998, 2009. All rights reserved.

Starting Microsoft Exchange restore...

Beginning VSS restore of 'STG3G_XIVG2_BAS'...

Starting snapshot restore process. This process may take several minutes.

VSS Restore operation completed with rc = 0

Files Examined : 0

Files Completed : 0

Files Failed : 0

Total Bytes : 0

Recovery being run. Please wait. This may take a while...

C:\Program Files\Tivoli\TSM\TDPEXchange>

Note: Instant restore is at the volume level. It does not show the total number of files examined and completed like a normal backup process does.

To verify that the restore operation worked, open the Exchange Management Console and check that the storage group and all the mailboxes have been mounted. Furthermore, verify that the 2nd Mailbox.edb file exists.

See the *Tivoli Storage FlashCopy Manager: Installation and User's Guide for Windows*, SC27-2504, or *Tivoli Storage FlashCopy Manager for AIX: Installation and User's Guide*, SC27-2503, for more and detailed information about Tivoli Storage FlashCopy Manager and its functions. The latest information about the Tivoli Storage FlashCopy Manager is available on the web at:

<http://www.ibm.com/software/tivoli>



A

Quick guide for VMware SRM

This appendix explains VMware SRM-specific installation considerations, including information related to XIV configurations.

The goal of this appendix is only to give the reader enough information to quickly install, configure, and experiment with SRM. It is not meant as a guide on how to deploy SRM in a real production environment,

Introduction

VMware SRM (Site Recovery Manager) provides disaster recovery management, nondisruptive testing, and automated failover functionality. It can also help manage the following tasks in both production and test environments:

- ▶ Manage failover from production data centers to disaster recovery sites
- ▶ Failover between two sites with active workloads
- ▶ Planned datacenter failovers such as datacenter migrations

The VMware Site Recovery Manager enables administrators of virtualized environments to automatically fail over the entire environment or parts of it to a backup site.

VMware Site Recovery Manager utilizes the replication (mirroring function) capabilities of the underlying storage to create a copy of the data to a second location (a backup data center). This ensures that at any given time, two copies of the data are available and if the one currently used by production fails, production could then be switched to the other copy.

In a normal production environment, the virtual machines (VMs) are running on ESX hosts and utilizing storage systems in the primary datacenter. Additional ESX servers and storage systems are standing by in the backup datacenter. Mirroring functions of the storage systems maintain a copy of the data on the storage device at the backup location. In a failover scenario, all VMs will be shut down at the primary site (if possible/required) and will be restarted on the ESX hosts at the backup datacenter, accessing the data on the backup storage system.

This process requires multiple steps:

- ▶ Stop any running VMs at the primary site
- ▶ Stop the mirroring between the storage systems
- ▶ Make the secondary copy of data accessible for the backup ESX servers
- ▶ Register and restart the VMs on the backup ESX servers

VMware SRM can automate these tasks and perform the necessary steps to failover complete virtual environments with just one click. This saves time, eliminates user errors, and helps to provide detailed documentation of the disaster recovery plan.

SRM can also perform a test of the failover plan by creating an additional copy of the data on the backup system and start the virtual machines from this copy without connecting them to any network. This enables administrators to test recovery plans without interrupting production systems.

At a minimum, an SRM configuration will consist of two ESX servers, two vCenters, and two storage systems, one each at the primary and secondary locations. The storage systems are configured as a mirrored pair relationship. Ethernet connectivity between the two locations is required for the SRM to function properly.

Detailed information about the concepts, installation, configuration and usage of VMware Site Recovery Manager is provided on the VMware product site at the following location:

http://www.vmware.com/support/pubs/srm_pubs.html

In this appendix we provide specific information about installing, configuring and administering VMware Site Recovery Manager in conjunction with IBM XIV Storage Systems.

At the time of this writing, the following versions of Storage Replication Agent for VMware SRM server are supported with XIV Storage Systems: Versions 1.0, 1.0 U1 and 4.0.

Pre-requisites

To successfully implement a continuity and disaster recovery solution with VMware SRM, several prerequisites need to be met. The following is a generic list. However, your environment might have additional requirements (refer to the VMware SRM documentation as previously noted, and in particular to the VMware vCenter SRM Administration guide, at http://www.vmware.com/pdf/srm_admin_4_1.pdf):

- ▶ Complete the cabling
- ▶ Configure the SAN zoning
- ▶ Install any service packs and updates if required
- ▶ Create volumes to be assigned to the host
- ▶ Install VMware ESX server on host
- ▶ Attach ESX hosts to the IBM XIV Storage System
- ▶ Install and configure database at each location
- ▶ Install and configure vCenter server at each location
- ▶ Install and configure vCenter Client at each location
- ▶ Install the SRM server
- ▶ Download and configure SRM plug-in
- ▶ Install IBM XIV Storage System Storage Replication Adapter (SRA) for VMware SRM
- ▶ Configure and establishing remote mirroring for LUNs that are used for SRM
- ▶ Configure the SRM server
- ▶ Create a protected group
- ▶ Create a recovery plan

Refer to Chapter 1, “Host connectivity” on page 1 and Chapter 9, “VMware ESX host connectivity” on page 191 for information about implementing the first six bullets above.

Note: Use single initiator zoning to zone ESX host to all available XIV interface modules

Steps to meet the pre-requisites are described in the next sections of this chapter. Following the information provided, you can set up a simple SRM server installation in your environment.

After you meet all of the pre-requisites, you are ready to test your recovery plan. After successful testing of the recovery plan, you can perform a fail-over scenario for your primary site. Be prepared to run the virtual machines at the recovery site for an indefinite amount of time because VMware SRM server does not currently support automatic fail-back operations.

There are two options if you need to execute a fail-back operation. The first option is to define all the reconfiguration tasks manually and the second option is to configure the SRM server in reverse direction and then perform another failover. Both of these options require downtime for the virtual machines involved.

The SRM server needs to have its own database for storing recovery plans, inventory information and similar data. SRM supports the following databases:

- ▶ IBM DB2
- ▶ Microsoft SQL
- ▶ Oracle

The SRM server has a set of requirements for the database implementation, some of which are general without dependencies on the type of database used but others are not. Refer to VMware SRM documentation to get more detailed information about specific database requirements.

The SRM server database can be located on the same server as vCenter, on the SRM server host, or on a separate host. The location depends on the architecture of your IT landscape and on the database that is used.

Information about compatibility for SRM server versions can be found at the following locations:

- ▶ Version 4.0 and later:
http://www.vmware.com/pdf/srm_compat_matrix_4_x.pdf
- ▶ Version 1.0 update 1:
http://www.vmware.com/pdf/srm_101_compat_matrix.pdf
- ▶ Version 1.0:
http://www.vmware.com/pdf/srm_10_compat_matrix.pdf

Install and configure the database environment

This section illustrates the step-by-step installation and configuration of the database environment for the VMware vCenter and SRM server needs. In the following example, we use Microsoft SQL Server 2005 Express and Microsoft SQL Server Management Studio Express as the database environment for the SRM server. We installed the Microsoft SQL Express database on the same host server as vCenter.

The Microsoft SQL Express database is available at no additional cost for testing and development purposes. It is available for download from the Microsoft website at the following location:

<http://www.microsoft.com/downloads/en/details.aspx?FamilyID=3181842A-4090-4431-ACDD-9A1C832E65A6&displaylang=en>

The graphical user interface for the database can be downloaded for at no additional cost from the Microsoft website at the following location:

<http://www.microsoft.com/downloads/details.aspx?FamilyId=C243A5AE-4BD1-4E3D-94B8-5A0F62BF7796&DisplayLang=en>

Note: For specific requirements and details on installing and configuring the database application, refer to the database vendor and VMware documentation for SRM.

Microsoft SQL Express database installation

After the Microsoft SQL Express software has been downloaded, start the installation process:

1. Double-click SQLEXPRESS.EXE in Windows Explorer as shown in Figure A-1 on page 315.

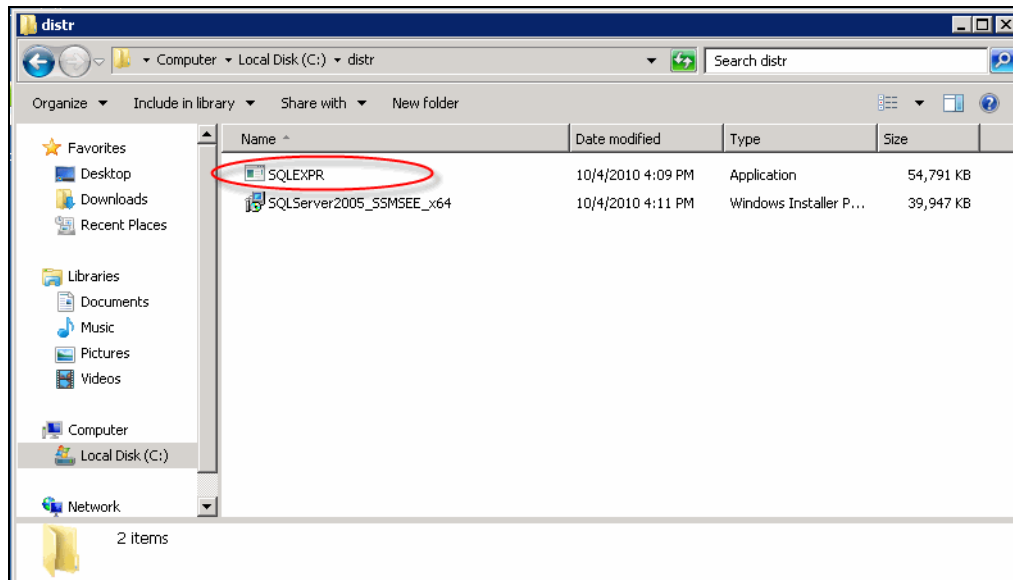


Figure A-1 Start Microsoft SQL Express installation

- After clicking the executable file, the installation wizard will start. Proceed through the prompts until you reach the **Feature Selection** dialog window shown in Figure A-2. Be aware that **Connectivity Components** must be selected for installation.

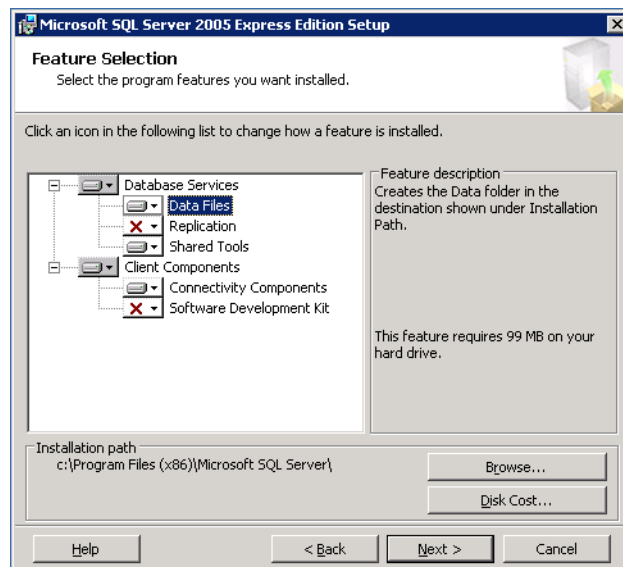


Figure A-2 List of components for install

- Proceed by clicking **Next**. The **Instance Name** window appears as shown in Figure A-3 on page 316.

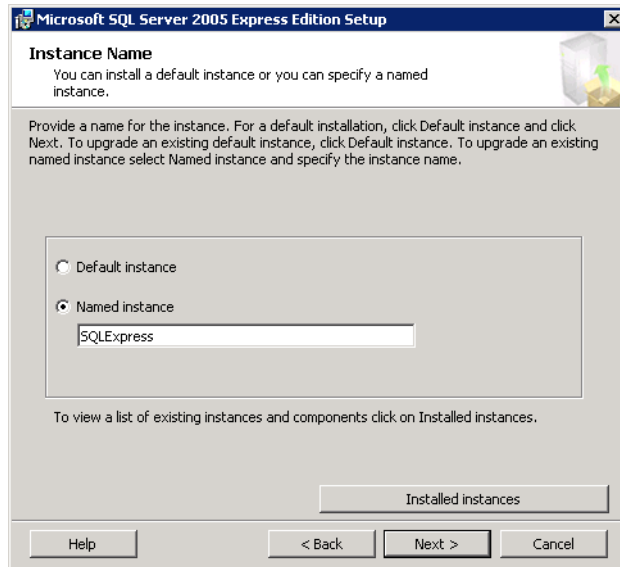


Figure A-3 Instance naming

4. Specify the name to use for the instance that will be created during the installation. This name will also be used for SRM server installation. Select the option **Named instance** and type **SQLExpress** as shown in Figure A-3. Click **Next** to display the **Authentication Mode** window shown in Figure A-4.

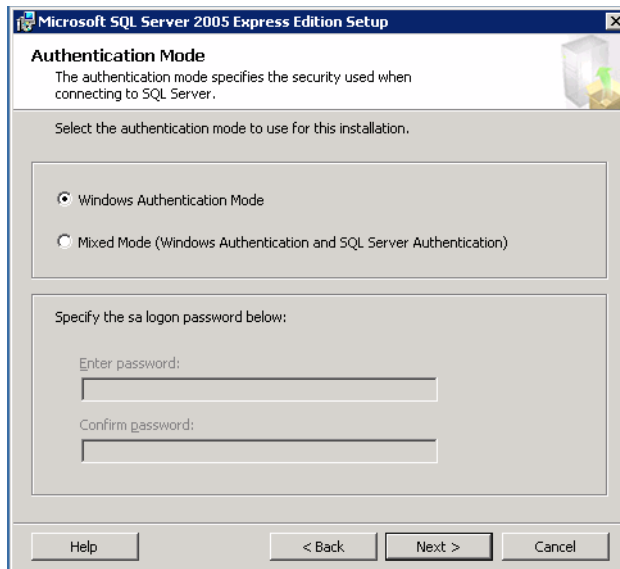


Figure A-4 Choose the type of authentication

5. Select **Windows Authentication Mode**. Use this setting for a simple environment. Depending on your environment and needs you might need to choose another option. Press **Next** to proceed to the **Configuration Options** dialog window as shown in Figure A-5 on page 317.

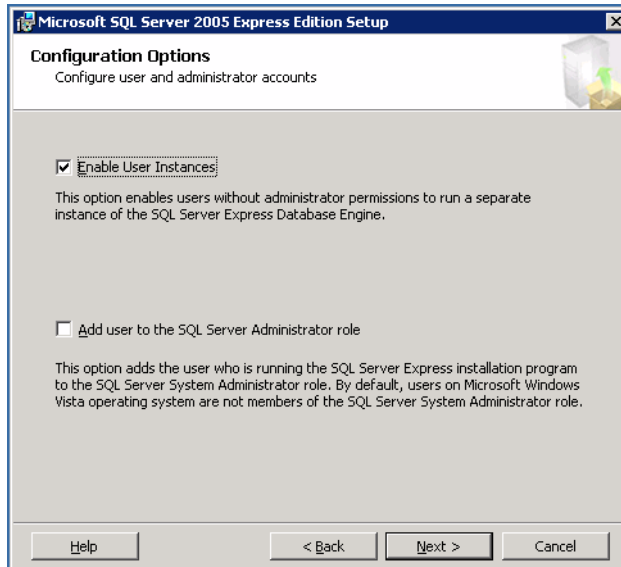


Figure A-5 Choose configuration options

6. For our simple example, select the option **Enable User Instances**. Click **Next** to display the **Error and Usage Report Settings** window as shown in Figure A-6. Here you are asked to choose the error reporting options. You can decide if you want to report errors to Microsoft Corporation by selecting the option which you prefer.

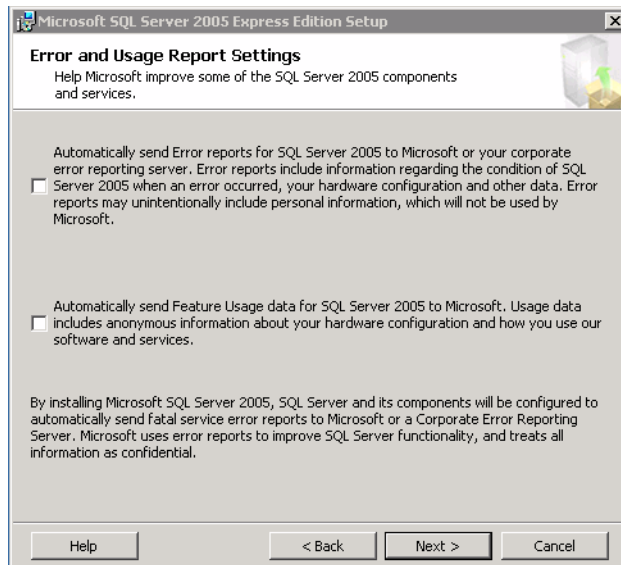


Figure A-6 Configuration on error reporting

7. Click **Next** to continue to the **Ready to Install** dialog window as shown in Figure A-7 on page 318.

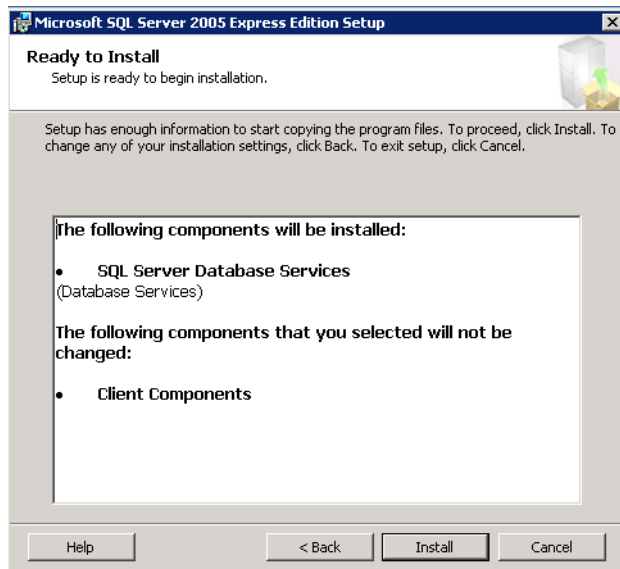


Figure A-7 Ready to install

8. You are now ready to start MS SQL Express 2005 installation process by clicking **Install**. If you decide to change previous settings, you can go backward using the **Back** button.
9. After the installation process is complete, the dialog window shown in Figure A-8 is displayed. Click **Next** to complete the installation procedure.

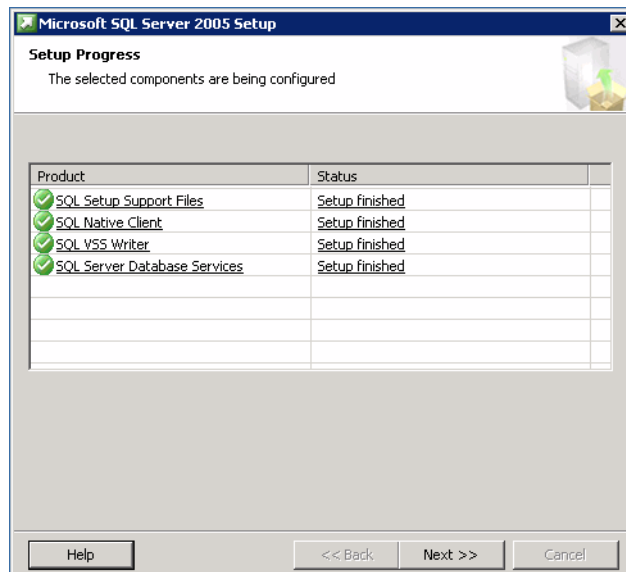


Figure A-8 Install finished

10. The final dialog window appears as shown in Figure A-9 on page 319.

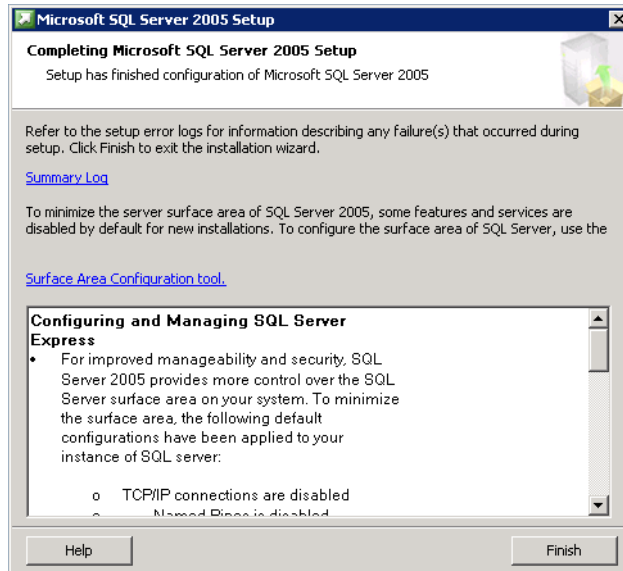


Figure A-9 Install completion

11. The final dialog window displays the results of the installation process. Click **Finish** to complete the process.

SQL Server Management Studio Express installation

Next we install the visual tools to configure the database environment:

1. After downloading the SQL Server Management Studio Express installation files from the Microsoft website, start the installation process by double clicking `SQLServer2005_SSMSEE_x64.msi` in Windows Explorer as shown in Figure A-10.

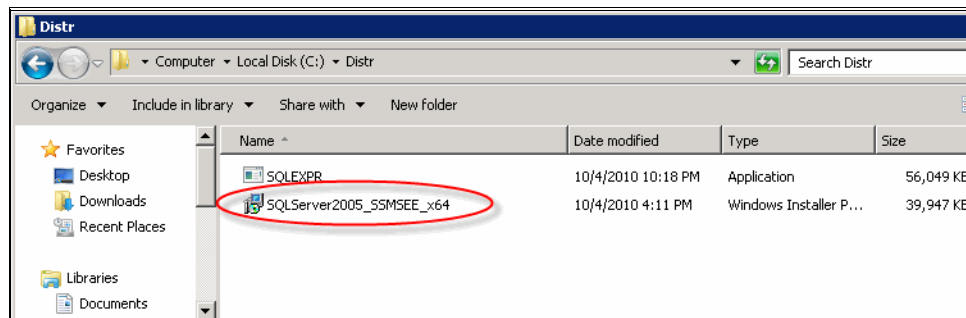


Figure A-10 Start the installation for the Microsoft SQL Server Management Studio Express

After clicking the file, the installation wizard starts. Proceed with the required steps to complete the installation.

The Microsoft SQL Server Management Studio Express software installation will need to be done at all locations which are chosen for your continuity and disaster recovery solution.

2. Before starting the configuration process for the database, you need to create additional local users on your host. To create users click **Start** in the task bar, then click **Administrative Tools->Computer Management** as shown in Figure A-11 on page 320.

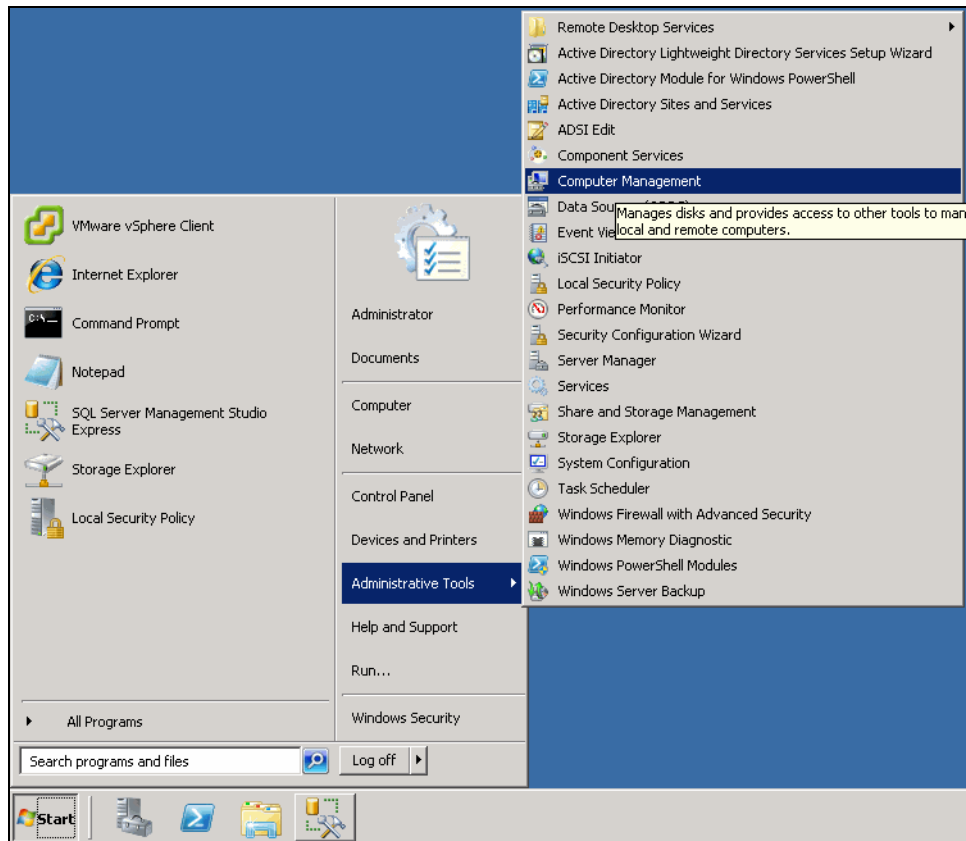


Figure A-11 Run the computer management

3. In the left pane navigate to the subfolder Computer Management (Local)\System Tools\Local Users and Groups then right-click **Users**.
4. Click **New User** in the pop-up window.
5. The **New User** dialog window is displayed as shown in Figure A-12. Enter details for the new user, then click **Create** and make sure in the main window that the new user was created. You need to add two users: one for the vCenter database and one for the SRM database.

Figure A-12 Add new user

Now you are ready to configure the databases.

6. Configure databases: one vCenter database and one SRM database for each site. In the examples below we provide the instructions for the vCenter database. Repeat the process for the SRM server database and the vCenter database at each site.
7. Start Microsoft SQL Server Management Studio Express by clicking **Start** → **All programs** → **Microsoft SQL Server 2005**, and then click **SQL Server Management Studio Express** as shown in Figure A-13.

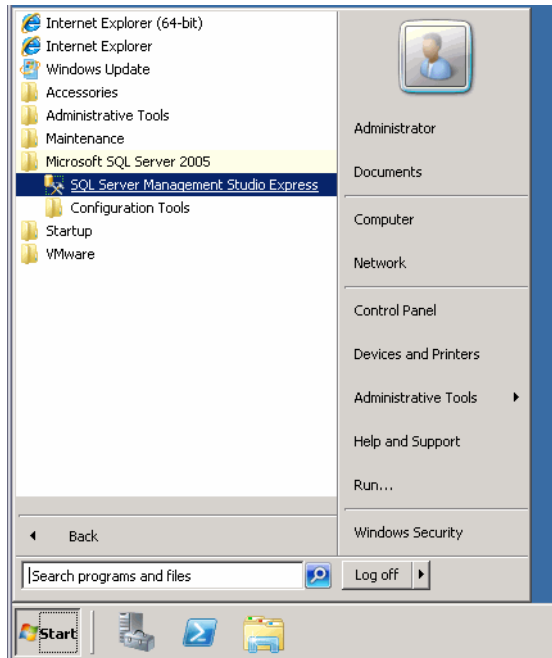


Figure A-13 Launch MS SQL Server Management Studio Express

8. The login window shown in Figure A-14 appears. Leave all values in this window unchanged and click **Connect**.

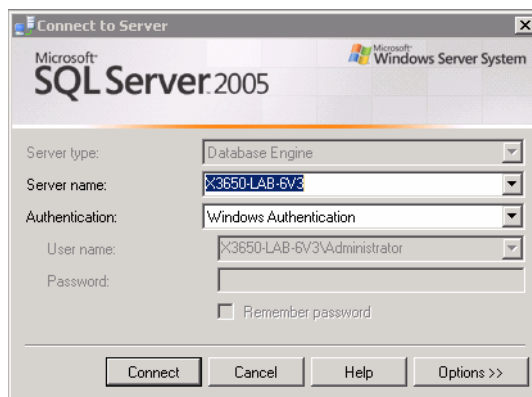


Figure A-14 Login window for the MS SQL Server Management Studio

9. After successful login, the MS SQL Server Management Suite Express main window is displayed (see Figure A-15 on page 322). In this window execute configuration tasks to create databases and logins.

To create databases, right-click **Databases**. In the pop-up window, click **New database** and a new window appears as shown in Figure A-15 on page 322.

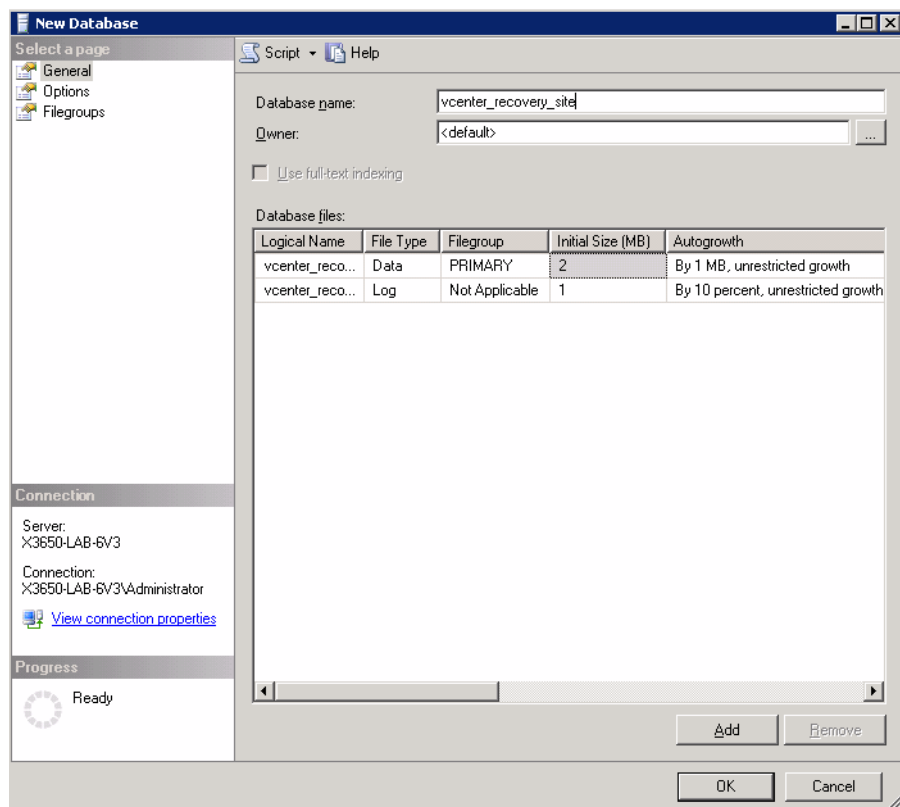


Figure A-15 Add database dialog

10. Enter the information for the database name, owner and database files.

In our example, we set up only the database name leaving all others parameters to their default values. Having done this, click **OK** and your database will be created.

11. Check to see if the new database was created by using the Object Explorer and expanding **Databases** → **System Databases** and verify that there is a database with the name you entered. See example in Figure A-16 on page 323 where the names of the created databases are circled in red. After creation of the required databases you need to create a login for them.

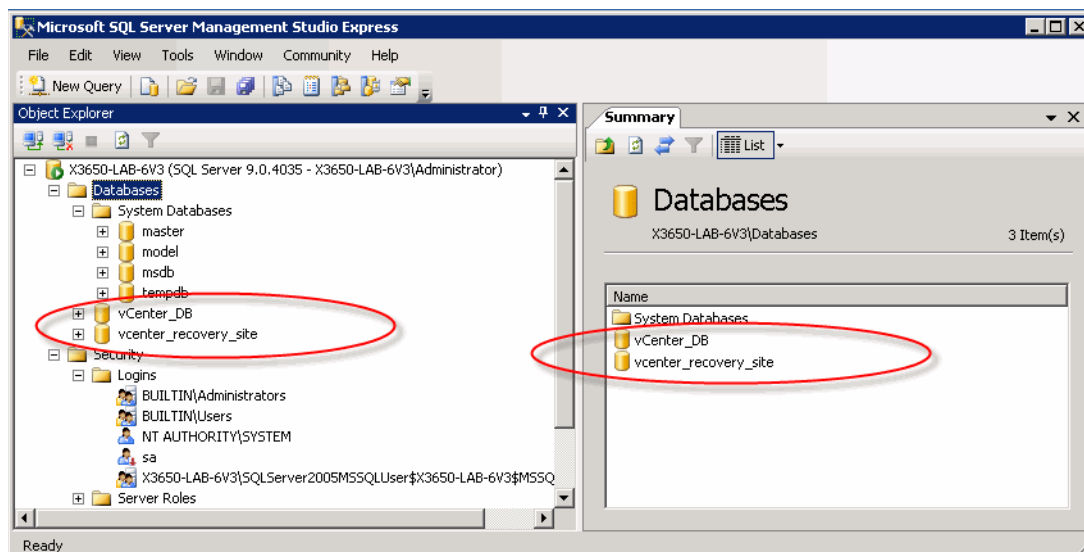


Figure A-16 Check that is new db's created

12. To create database logins, right-click the subfolder logins and select new login in the pop-up window as shown in Figure A-17. Enter the user name, type of authentication, default database, and default code page. For our simple example, we specify the user name, default database relative to user name, and leave all other parameters at their default values. Click **OK**.

Repeat this action for the vCenter and SRM servers databases.

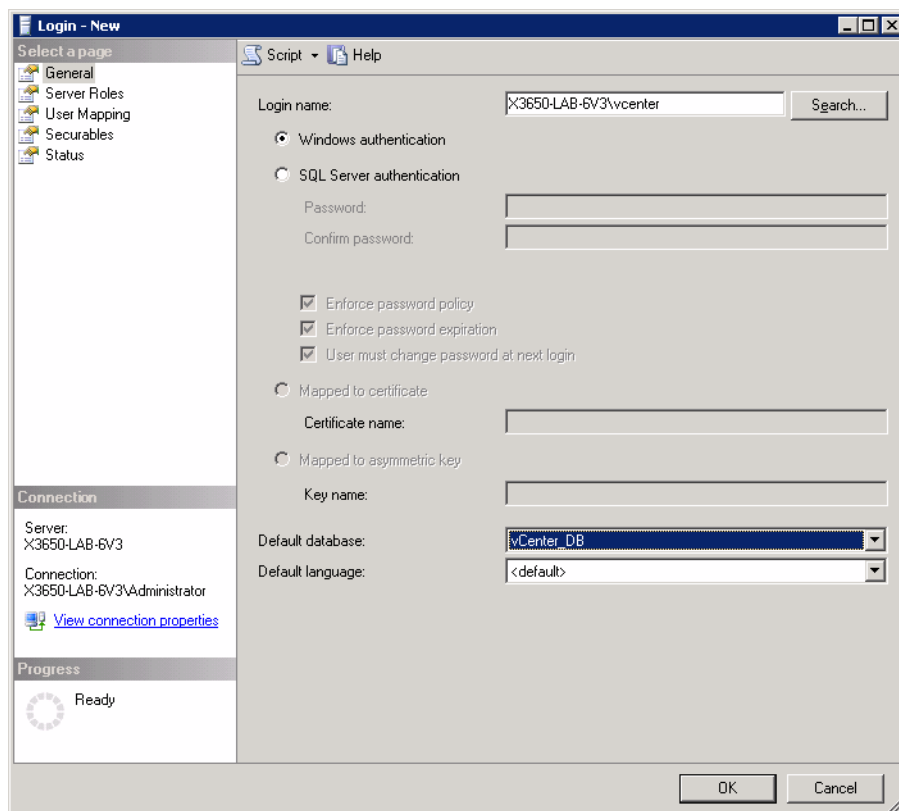


Figure A-17 Define database logins

13. Now you need to grant rights to the database objects for these logins as shown in Figure A-18

To grant rights to a database object for the created and associated logins you need to right-click **Logins** on the **vcenter** user login and select **Properties**.

14. A new window opens as shown in Figure A-18. In the top left pane, select **User Mappings** and select **vCenter** database in the top right pane. In the bottom right pane, select the **db_owner** and **public** roles. Finally, click **OK** and repeat those steps for the srmuser.

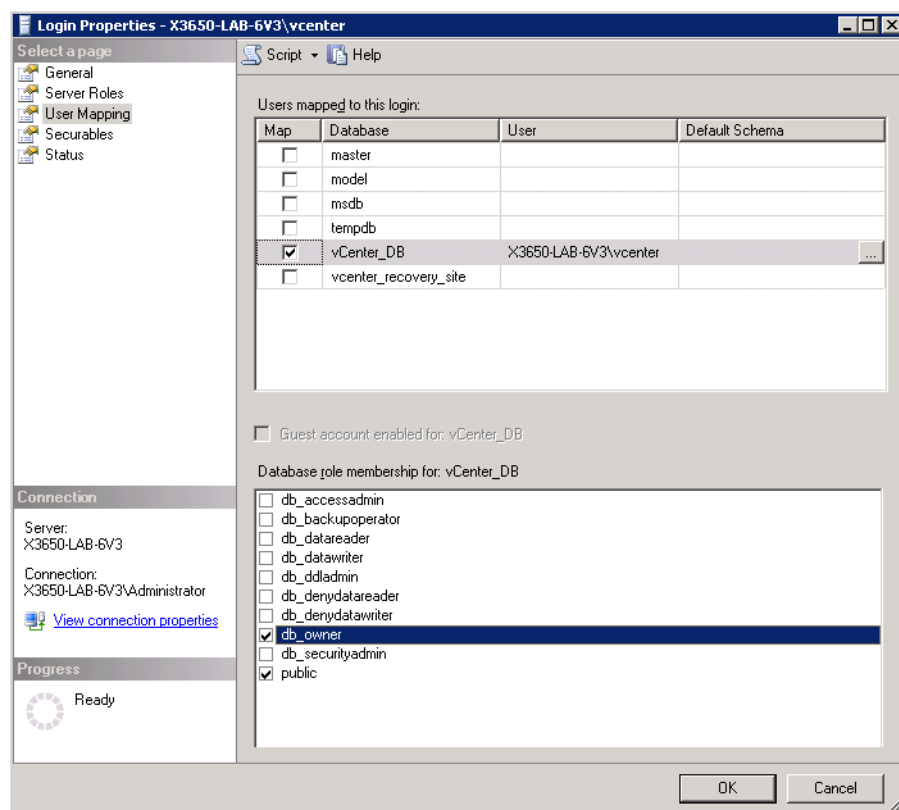


Figure A-18 Grant the rights on a database for the login created

Now, we are ready to start configuring ODBC data sources for the vCenter and SRMDB databases on a server where we plan install vCenter and SRM server.

15. To start configuring ODBC datastores click **Start** in the Windows desktop task bar, and select **Administrative Tools** → **Data Source (ODBC)**.

16. The ODBC Data Source Administrator window is now open as shown in Figure A-19 on page 325. Select **System DSN** tab and click **Add**.

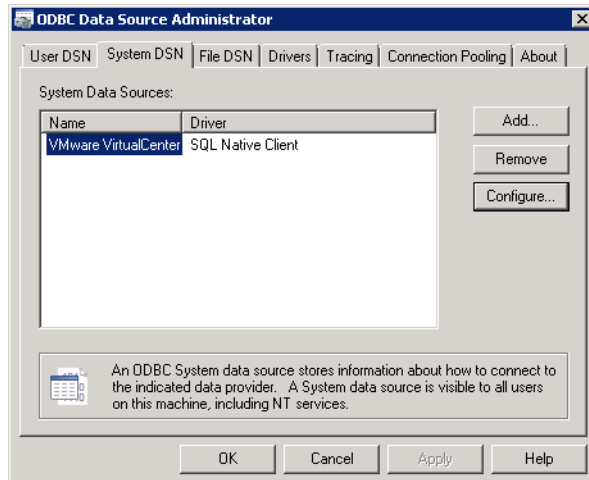


Figure A-19 Select system dsn

17. The Create New Data Source window opens as shown in Figure A-20. Select **SQL Native Client** and click **Finish**.

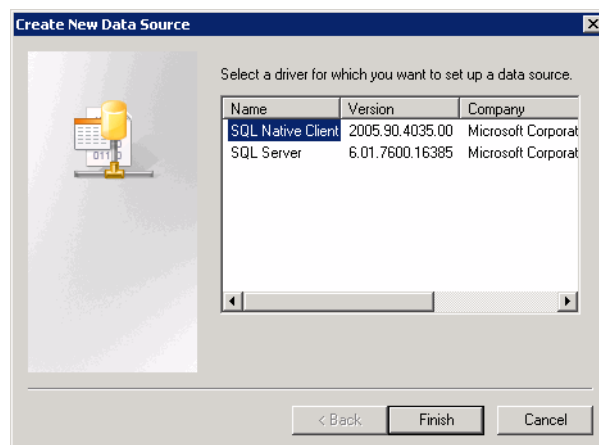


Figure A-20 Select SQL driver

18. The window shown in Figure A-21 on page 326 opens. Enter information for your data source like the name, description, server for the vcenter database. For our example, we set the name parameter to vcenter, description parameter to database for vmware vcenter, and server parameter to SQLEXPRESS as shown on Figure A-21 on page 326. Then click **Next**.

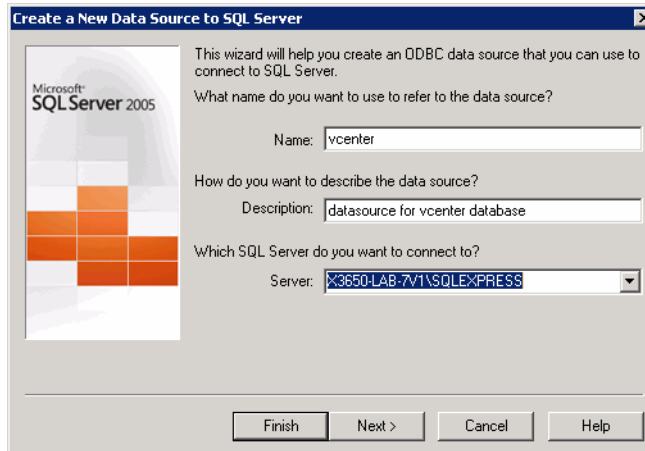


Figure A-21 Define data source name and server

19. The window shown in Figure A-22 opens. Select **With Integrated Windows Authentication** radio button, select the **Connect to SQL Server to obtain default settings to the additional configuration options** check box. Click **Next**.

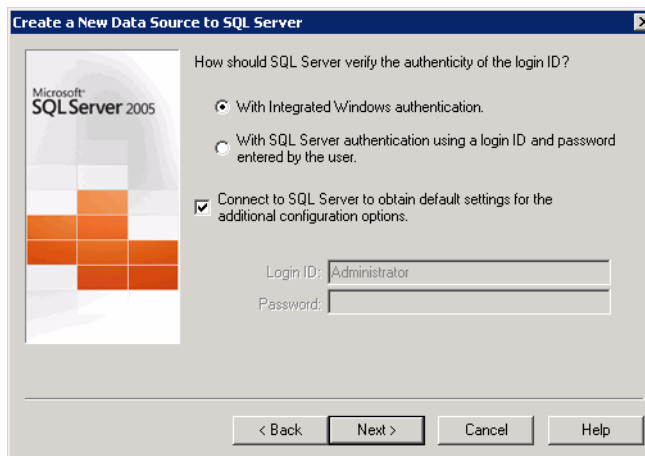


Figure A-22 Select authorization type

20. The window shown in Figure A-23 on page 327 opens. Select the **Change default database** check box, select **vCenter_DB** from the drop-down, and select the two check boxes at the bottom of the window. Click **Next**.

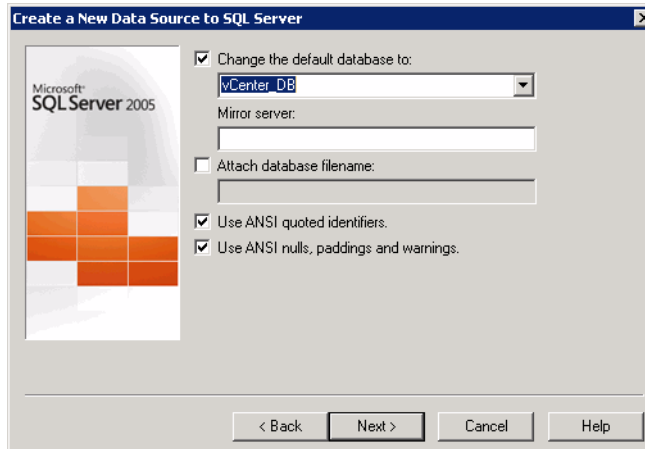


Figure A-23 Select default database for data source

21. The window shown in Figure A-24 is displayed. Select the **Perform translation for the character data** check box and then click **Finish**.

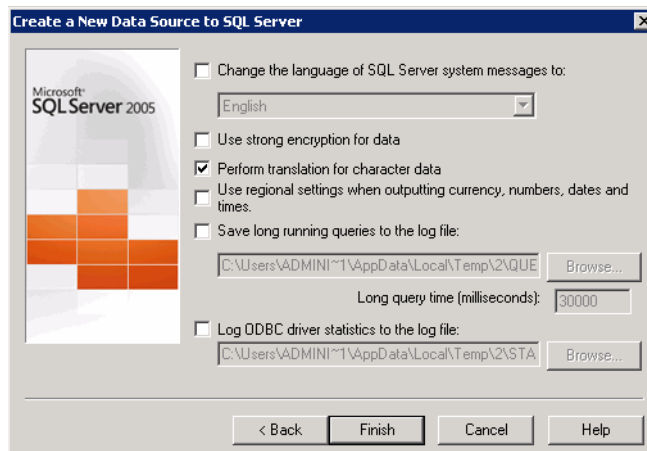


Figure A-24 SQL server database locale related settings

22. In the window shown in Figure A-25 inspect the information for your data source configuration and then click **Test Data Source**.

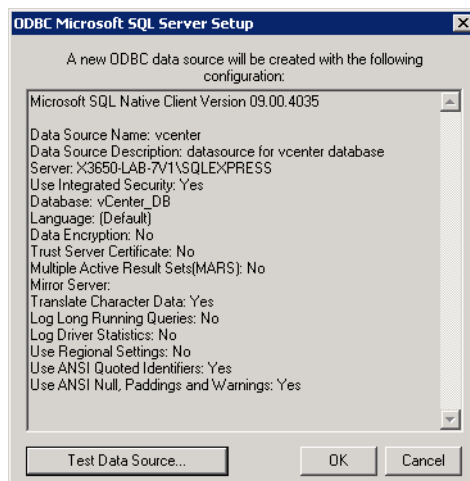


Figure A-25 Test data source and finish setup

23. The next window, shown in Figure A-26, indicates that the test completed successfully. Click **OK** to return to the previous window, and then click the **Finish**.



Figure A-26 Results on data source test

24. You are returned to the window shown in Figure A-27. You can see the list of Data Sources defined system-wide. Check the presence of **vcenter** data source.

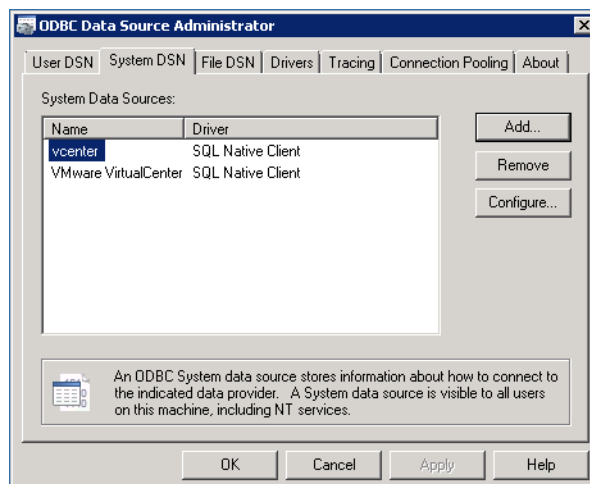


Figure A-27 Defined System Data Sources

You need to install and configure databases on all sites that you plan to include into your business continuity and disaster recovery solution.

Now you are ready to proceed with the installation of vCenter server, vCenter client, SRM server and SRA agent.

Installing vCenter server

This section illustrates the step-by-step installation of vCenter server under Microsoft Windows Server 2008 R2 Enterprise.

Note: For the detailed information about vCenter server installation and configuration, refer to the VMware documentation. This section includes only common, basic information for a simple installation used to demonstrate SRM server capabilities with the IBM XIV Storage System.

Perform the following steps to install the vCenter Server:

1. Locate the vCenter server installation file (either on the installation CD or a copy you downloaded from the Internet). Follow the installation wizard guidelines until you reach the step where you are asked to enter information about database options.
2. At this step, you will be asked to choose database for vCenter server. Select the **Using existing supported database** radio button, type vcenter into **Data Source Name** (the name of the DSN must be the same as the ODBC system DSN which was defined earlier), and click **Next** as seen in Figure A-28.

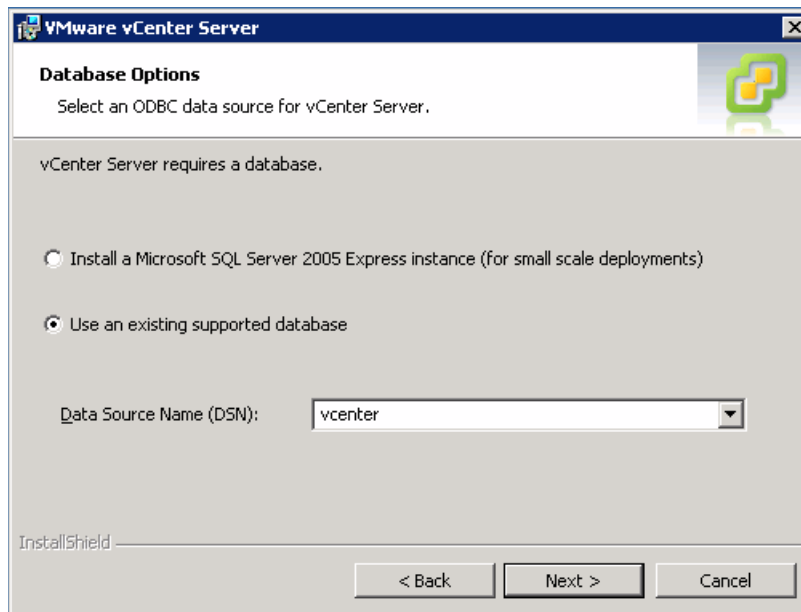


Figure A-28 Choosing database for vCenter server

3. In the window shown in Figure A-29 on page 330, type the password for the system account and click **Next**.

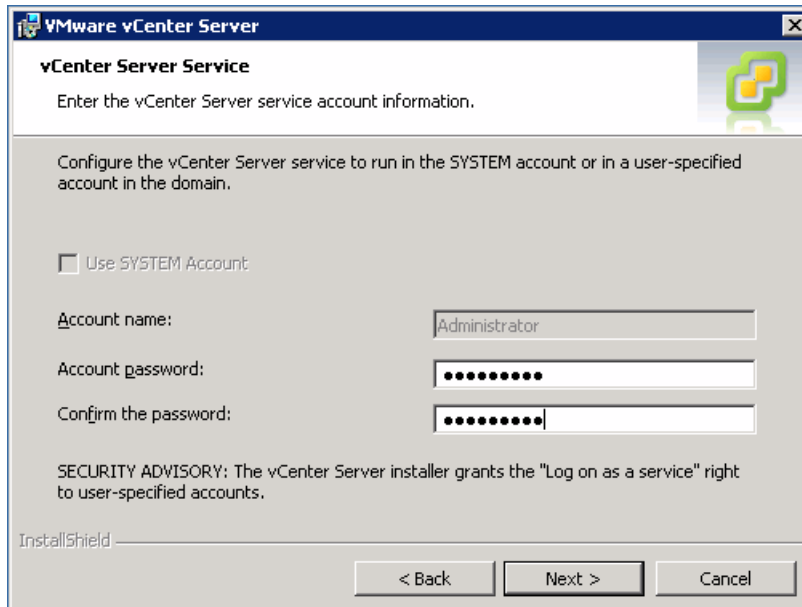


Figure A-29 Requesting password for the system account

4. In the next installation dialog shown in Figure A-30, you need to choose a Linked Mode for the installed server. For a first time installation, select **Create a standalone VMware vCenter server instance**. Click **Next**.

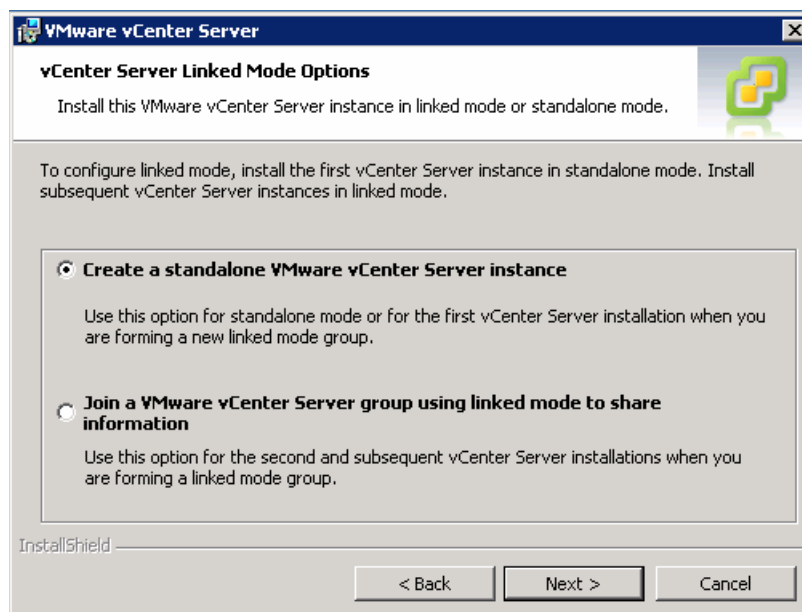


Figure A-30 Choosing Linked Mode options for the vCenter server

5. In the window shown in Figure A-31 on page 331, you can change default settings for ports used for communications by the vCenter server. For most implementations you should keep the default settings. Click **Next**.

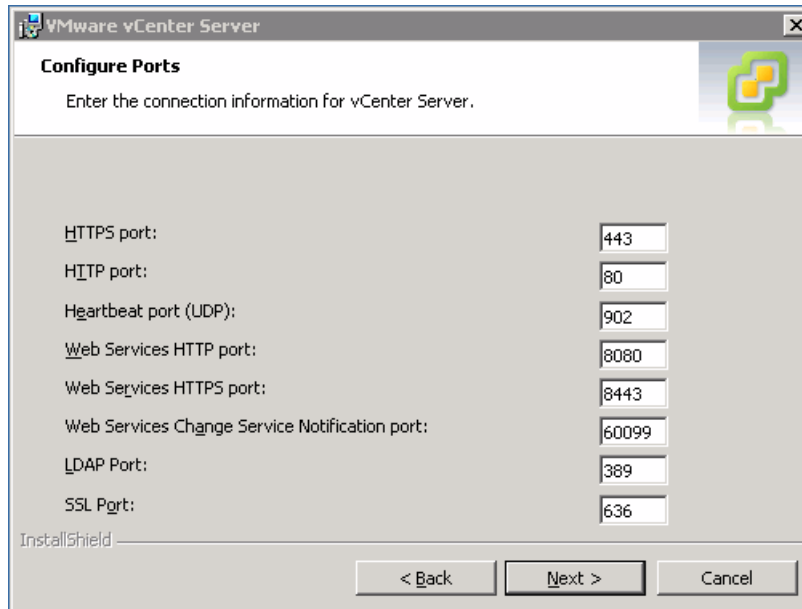


Figure A-31 Configure ports for the vCenter server

In the window shown in Figure A-32 select the required memory size for the JVM used by vCenter Web Services according to your environment. Click **Next**.

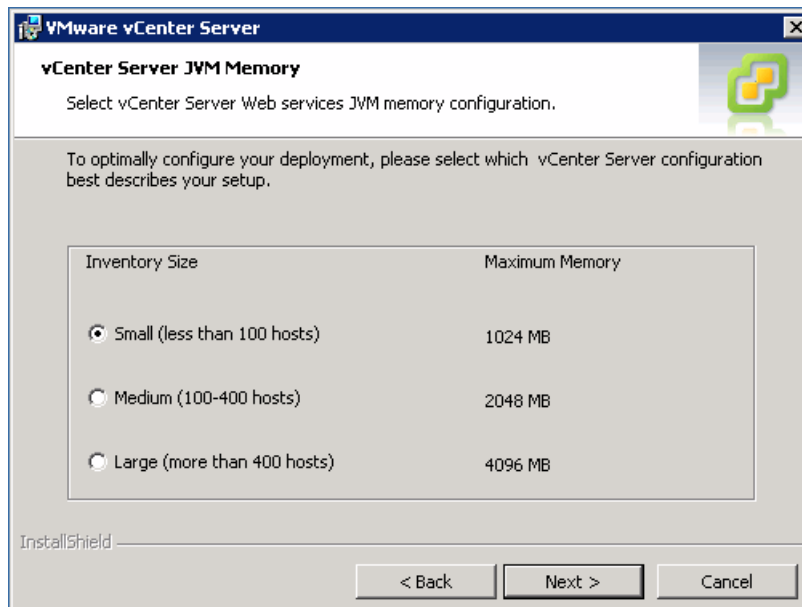


Figure A-32 Setting inventory size

- The window shown in Figure A-33 on page 332 indicates that the system is now ready to install vCenter. Click **Install**.

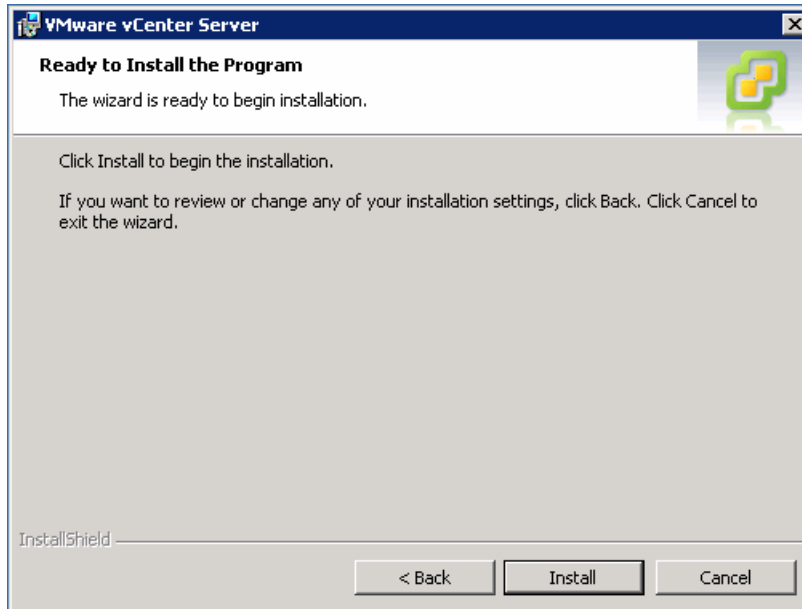


Figure A-33 Information about readiness to install

7. After the installation completes, the window shown in Figure A-34 displays. Click **Finish**.

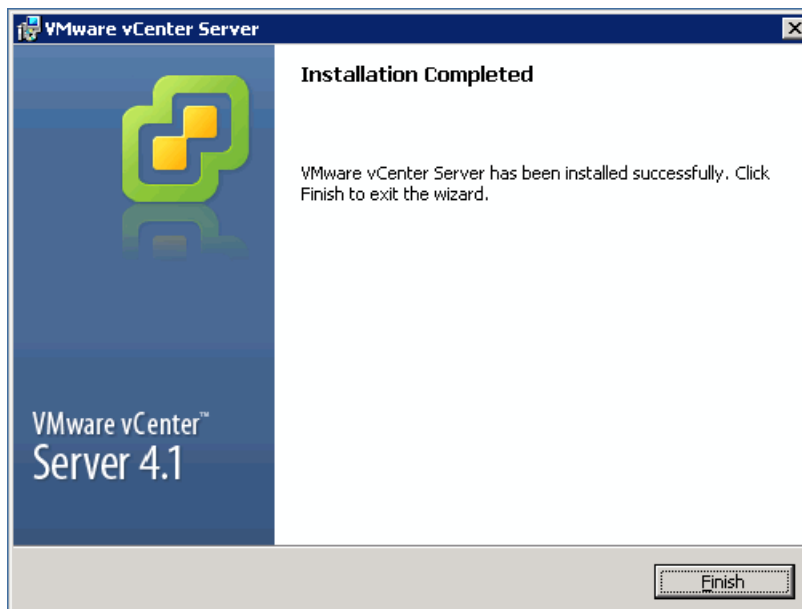


Figure A-34 The vCenter installation is completed

You need to install vCenter server on all sites that you plan to include as part of your business continuity and disaster recovery solution.

Installing and configuring vCenter client

This section illustrates the step-by-step installation of the vSphere client under Microsoft Windows Server 2008 R2 Enterprise.

Note: For detailed information about vSphere client and for complete installation and configuration instructions, refer to VMware documentation. This chapter includes only common and basic information required for installing the vSphere client and using it to manage the SRM server.

Installing the vCenter client is pretty straightforward: Locate the vCenter server installation file (either on the installation CD or a copy you downloaded from the Internet). Running the installation file displays the vSphere Client installation wizard welcome dialog. Just follow the installation wizard instructions to complete the installation.

You need to install vSphere client on all sites that you plan to include in your business continuity and disaster recovery solution.

Once you finish installing SQL Server 2005 Express, vCenter server, and vSphere Client, you can place existing ESX servers under control of the newly installed vCenter server. To perform this task follow these instructions:

1. Start the vSphere client.
2. In the login window shown in Figure A-35, type the IP address or machine name of your vCenter server, and a user name and password. Click **Login**.

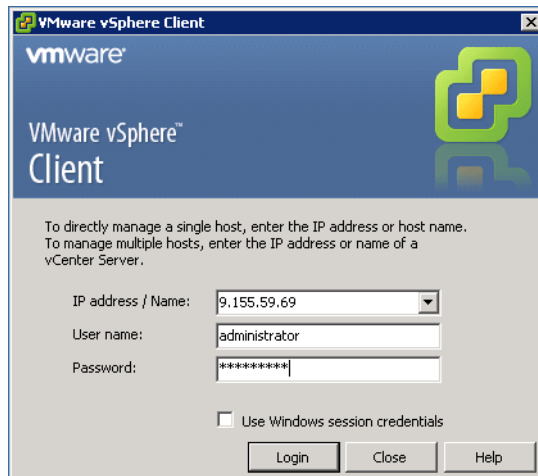


Figure A-35 vSphere client login window

3. Add the new datacenter under control of the newly installed vCenter server. In the main vSphere client window, right-click the server name and select **New Datacenter** as shown in Figure A-36 on page 334.

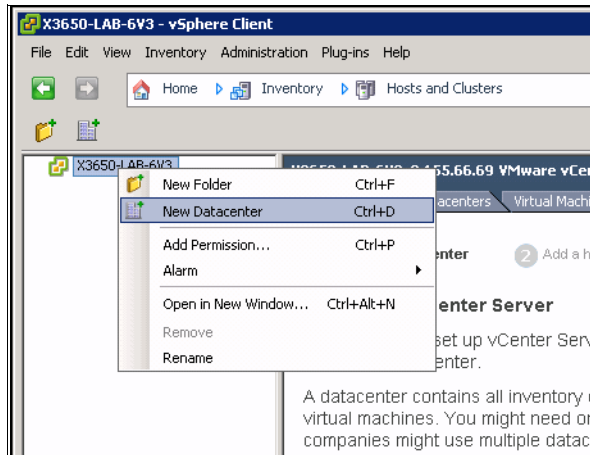


Figure A-36 Start to define the datacenter

4. Type a new name for the datacenter as shown in Figure A-37 and press Enter.

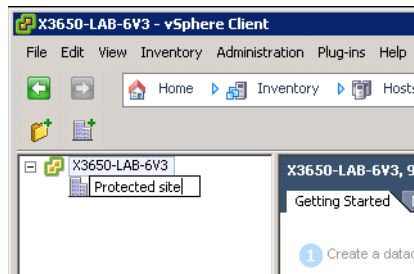


Figure A-37 Specify name of the datacenter

5. The Add Host wizard is started. Type the name or ip address of the ESX host, user name for the administrative account on this ESX server, and the account password as shown in Figure A-38. Click **Next**.

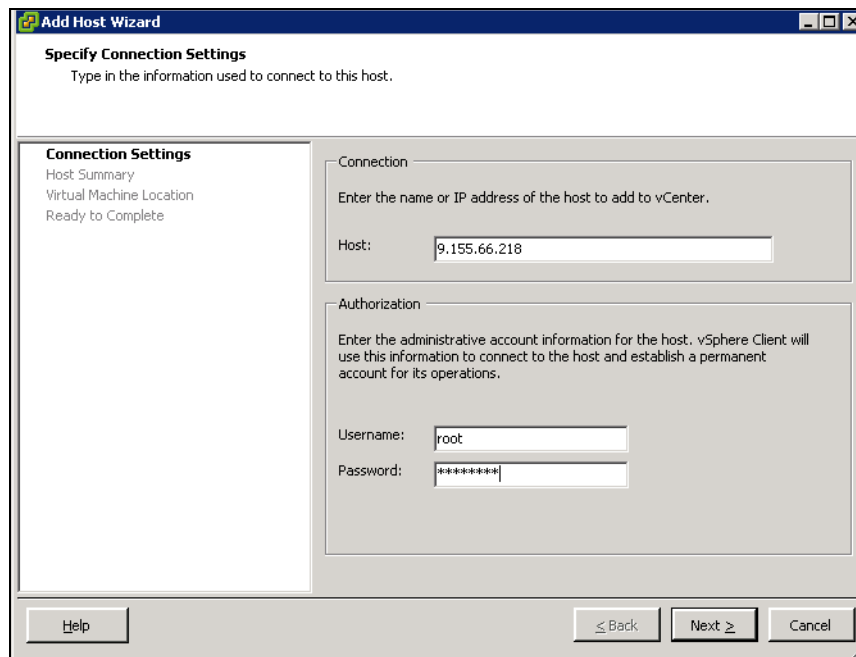


Figure A-38 Specifying host name, user name, and password

6. You must then verify the authenticity of the specified host as shown in Figure A-39. If correct, click **Yes** to continue to the next step.

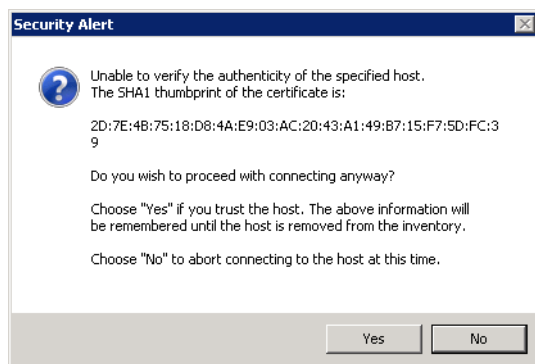


Figure A-39 Verifying the authenticity of the specified host

7. In the next window, you can observe the settings discovered for the specified ESX host as shown in Figure A-40. Check the information presented, and, if all is correct, click **Next**.

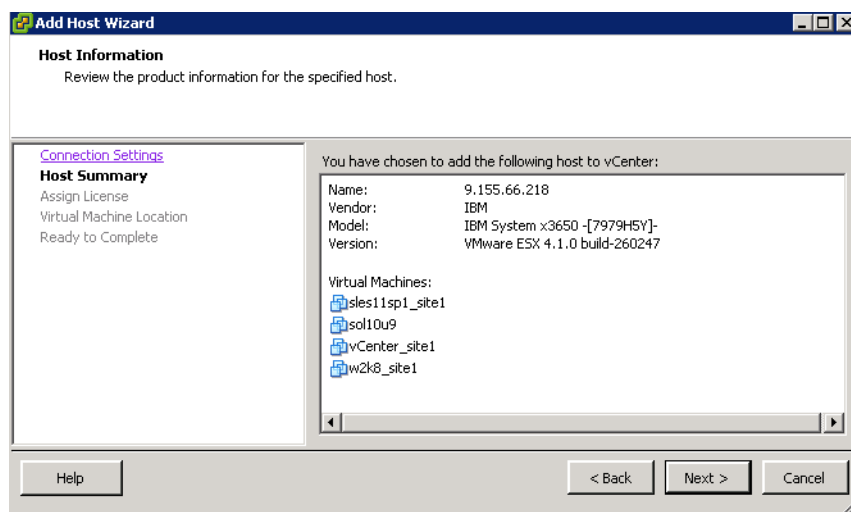


Figure A-40 Configuration summary on the discovered ESX host

8. In the window shown in Figure A-41 on page 336, you need to choose between ESX host in evaluation mode or enter a valid license key for the ESX server. Click **Next**.

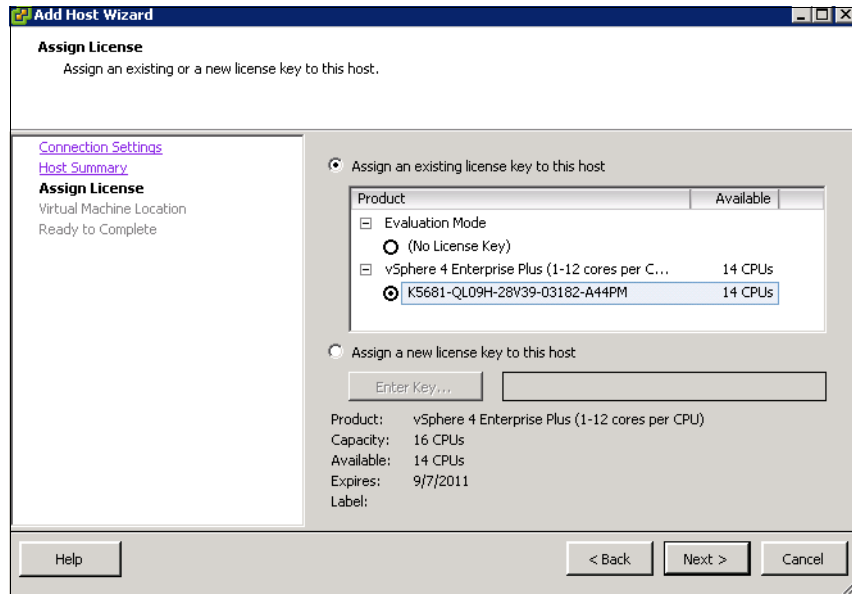


Figure A-41 Assign license to the host

9. Select a location for the newly added ESX server as shown in Figure A-42 and click **Next**.

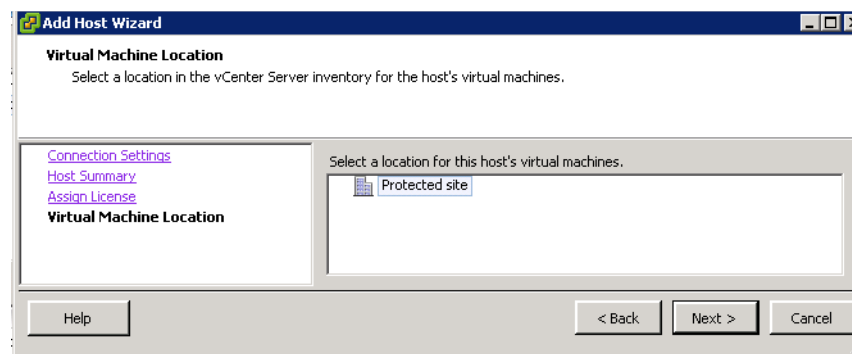


Figure A-42 Select the location in the vCenter inventory for the host's virtual machines

10. The window shown in Figure A-43 summarizes your settings. Check the settings and, if they are correct, click **Finish**.

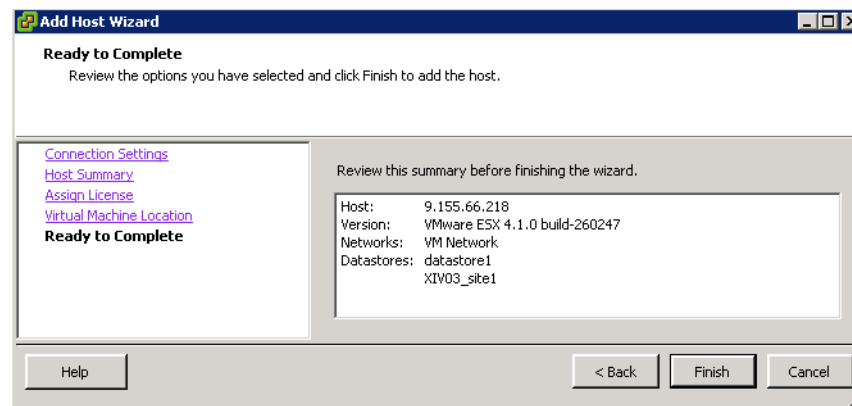


Figure A-43 Review summary

11. You will return to the vSphere Client main window as shown in Figure A-44 on page 337.

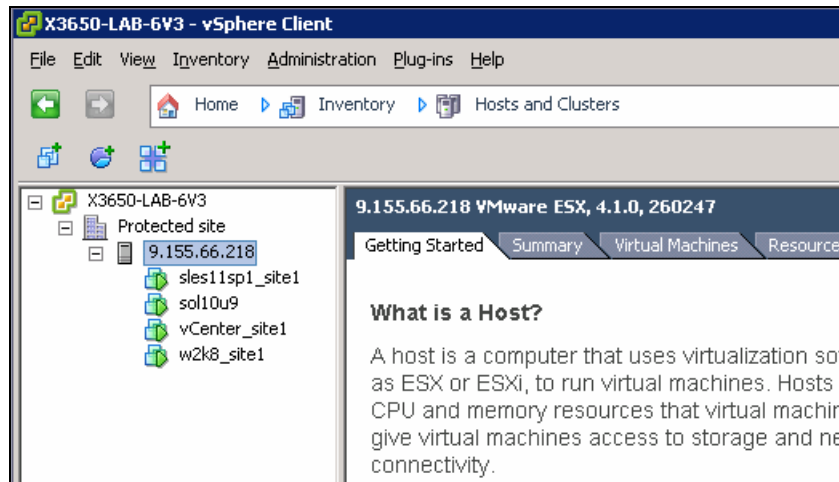


Figure A-44 Presenting inventory information about ESX server in the vCenter database

Repeat all the above steps for all the vCenter servers you want to include into your business continuity and disaster recovery solution.

Installing SRM server

This section describes the basic installation tasks for the VMware SRM server version 4 under Microsoft Windows Server 2008 R2 Enterprise. To install VMware SRM server, follow these instructions:

1. Locate the vCenter server installation file, either on the installation CD or a copy you downloaded from the Internet. Running the installation file launches the Welcome window for the vCenter Site Recovery Manager wizard as shown in Figure A-45. Click **Next** and follow the installation wizard guidelines.

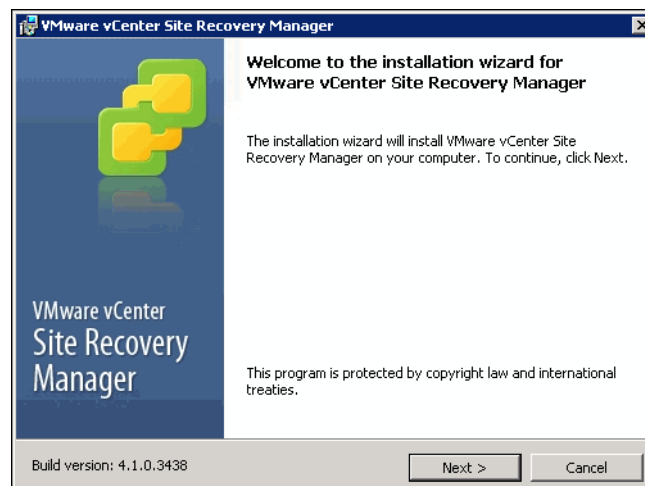


Figure A-45 SRM Installation wizard welcome message

2. In window shown in Figure A-46 on page 338, provide the vCenter server ip address, vCenter server port, vCenter administrator user name, and the password for the administrator account, then click **Next**.

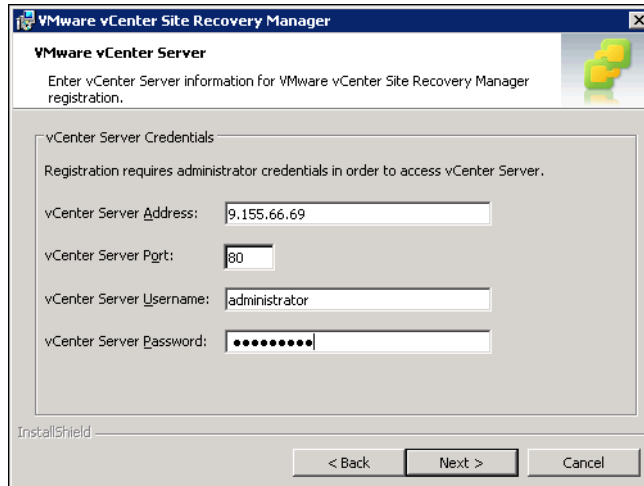


Figure A-46 SRM settings on paired vCenter server

3. You might can get a security warning like shown in Figure A-47. Check the vCenter server IP address and, if it is correct, click **OK**.

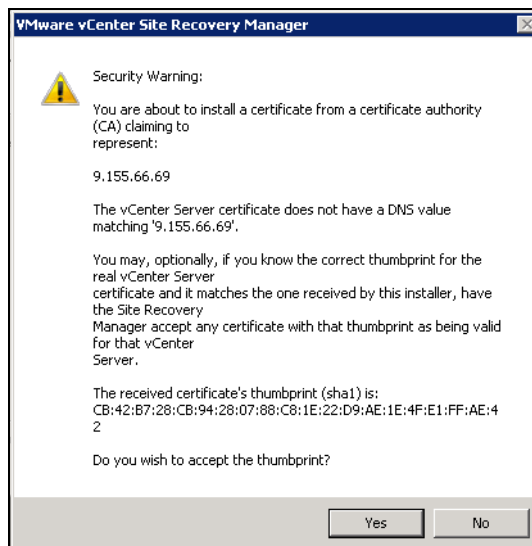


Figure A-47 Dialog on certificate acceptance

4. Select the **Automatically generate certificate** option as shown in Figure A-48 on page 339, and click **Next**.

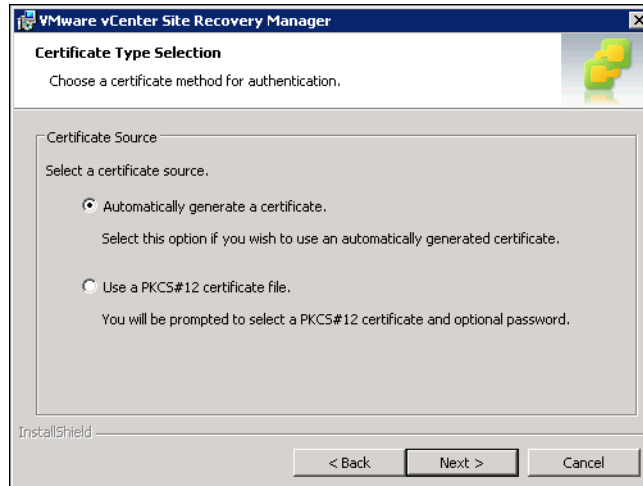


Figure A-48 Certificate type selection

Note: If your vCenter servers are using NON-default (that is, self signed) certificates, then you should choose the Use a PKCS#12 certificate file check box. For details, refer to the VMware vCenter SRM Administration guide at the following URL:

http://www.vmware.com/pdf/srm_admin_4_1.pdf

5. You must now type details such as organization name and organization unit that are used as parameters for certificates generation. See Figure A-49. When done, click **Next**.

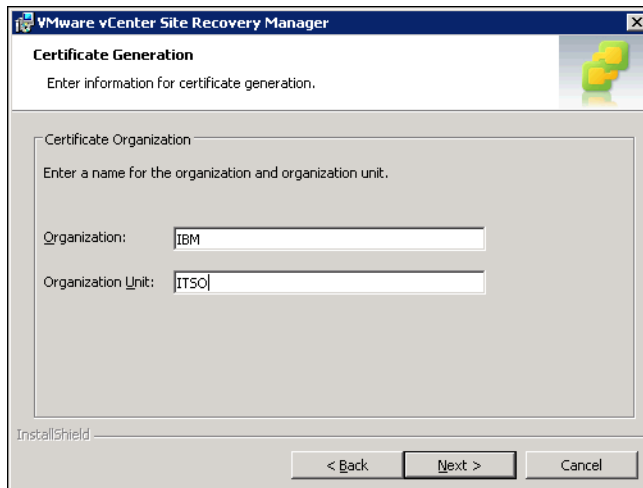


Figure A-49 Setting up certificate generation parameters

6. The window shown in Figure A-50 on page 340 asks for general parameters pertaining to your SRM installation. Provide the location name, administrator email, additional email, local host IP address or name, and the ports to be used for connectivity. When done, click **Next**.

Figure A-50 General SRM server setting for the installation location

7. Provide parameters related to the database that was previously installed as shown in Figure A-51. Type the following parameters: type of the database, ODBC System data source, user name and password, and connection parameters. Click **Next**.

Figure A-51 Specifying Database parameters for the SRM server

8. The next window informs you that the installation wizard is ready to proceed as shown in Figure A-51 on page 340. Click **Install** to start the installation process.

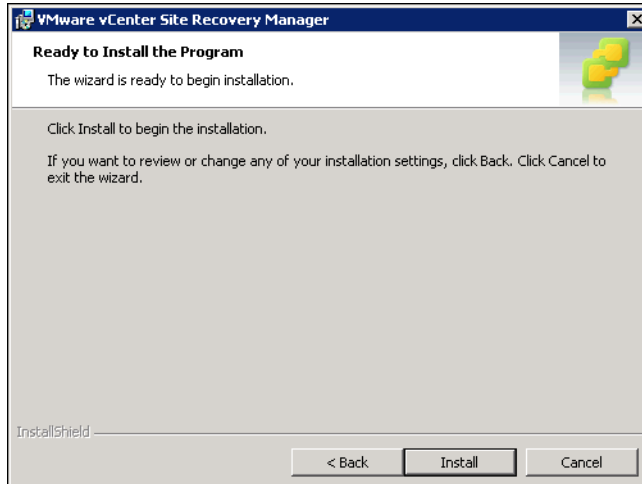


Figure A-52 Readiness of SRM server installation wizard to start the install

You need to install SRM server on each protected and recovery sites that you plan to include into your business continuity and disaster recovery solution.

Installing vCenter Site Recovery Manager plug-in

Now that you have installed the SRM server, you need to have the SRM plug-in installed on the system that is hosting your vSphere client. To do so, perform these steps:

1. Run the vSphere Client and connect to the vCenter server on the site where you installed the SRM server and are planning to install the SRM plug-in.
2. In the vSphere Client console, click **Plug-ins** → **Manage Plug-ins** as shown in Figure A-53.

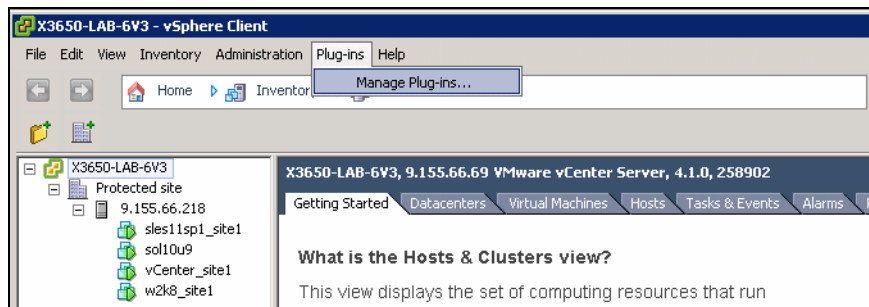


Figure A-53 Choosing the manage plug-in option

3. The Plug-in Manager window opens. Under the category **Available plug-ins**, right-click **vCenter Site Recovery Manager Plug-in** and select **Download and Install** as shown in Figure A-54 on page 342.

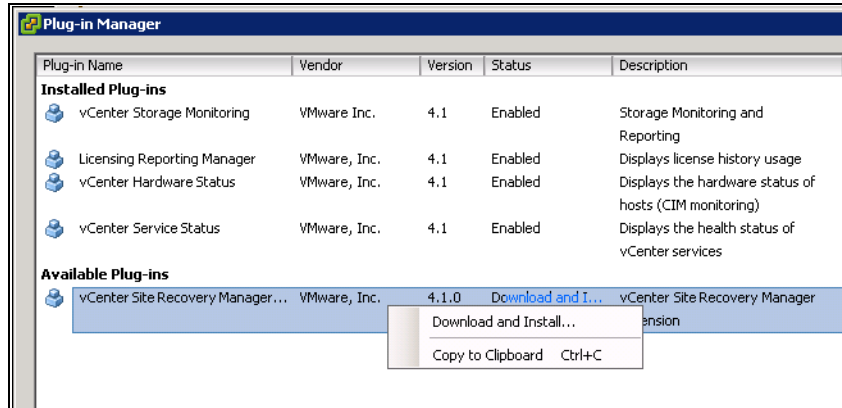


Figure A-54 Downloading and installing SRM plug-in

4. The vCenter Site Recovery Manager Plug-in wizard is launched. Follow the wizard guidelines to complete the installation.

You need to install SRM plug-in on each protected and recovery site that you plan to include in your business continuity and disaster recovery solution.

Installing XIV Storage Replication Adapter for VMware SRM

This section describes the tasks for installing the XIV SRA for VMware SRM server version 4 under Microsoft Windows Server 2008 R2 Enterprise. You need to download and install XIV SRA for VMware on each SRM server located at protected and recovery sites and that you plan to include into your business continuity and disaster recovery solution. To do so, perform these steps:

1. Locate the XIV Storage Replication Adapter installation file. Running the installation file launches vCenter Site Recovery Adapter installation wizard as shown in Figure A-55. Click **Next**.



Figure A-55 Welcome to SRA installation wizard

2. Follow the wizard guidelines to complete the installation.

Configure the IBM XIV System Storage for VMware SRM

Make sure that all virtual machines that you plan to protect reside on IBM XIV Storage System volumes. For any virtual machines that does not reside on IBM XIV Storage System you need to create volumes on XIV, add the data store to the ESX server, and migrate or clone that virtual machine to relocate it on XIV volumes.

For instructions on connecting ESX hosts to the IBM XIV Storage refer to Chapter 9, “VMware ESX host connectivity” on page 191.

Now, create a new storage pool on the IBM XIV Storage System at the recovery site. The new storage pool will contain the replicas of the ESX host datastores that are associated with virtual machines that you plan to protect.

Note: Configure a snapshot size of at least 20 percent of the total size of the recovery volumes in the pool. For testing failover operations that can last several days, increase the snapshot size to half the size of the recovery volumes in the pool. For even longer-term or I/O intensive tests, the snapshot size might have to be the same as the total size of the recovery volumes in the pool.

For the information about IBM XIV Storage System LUN mirroring refer to the IBM Redbooks publication *IBM XIV Storage System: Copy Services and Migration*, SG24-7759.

At least one virtual machine for the protected site must be stored on the replicated volume before you can start configuring SRM server and SRA adapter.

In addition, avoid replicating swap and paging files.

Configure SRM Server

You will need to configure the SRM server for the two sites.

To configure the SRM server on the protected site, follow these instructions:

1. Run the vCenter Client and connect to the vCenter server. In the vCenter Client main window, Click **Home** as shown in Figure A-56.

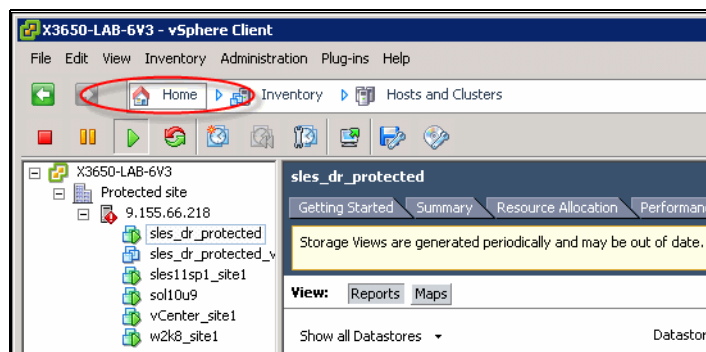


Figure A-56 Select the main vCenter Client window with applications

2. Go to the bottom of the main vSphere client window and click **Site Recovery** as shown in Figure A-57 on page 344.

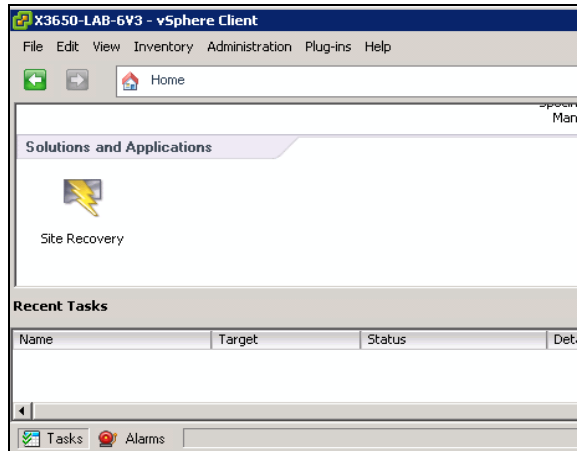


Figure A-57 Run the Site Recovery Manager from vCenter Client menu

3. The Site Recovery Project window displays. Click **Configure** for the Connections as shown circled in green in Figure A-58.

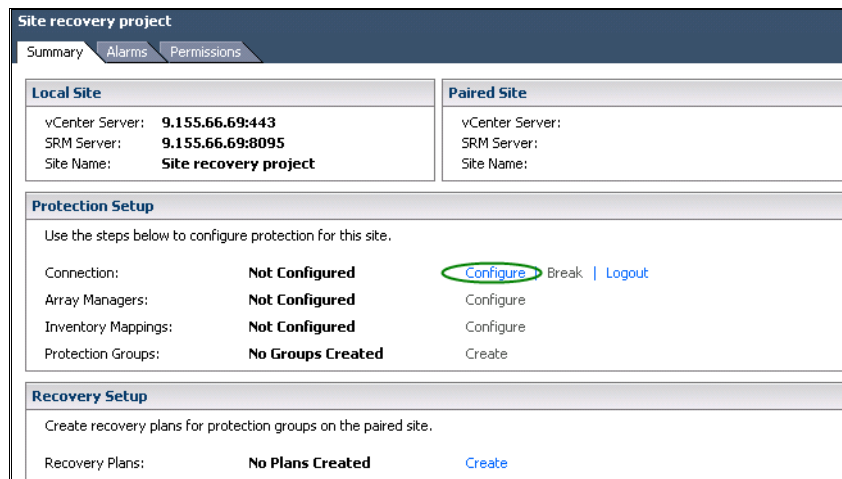


Figure A-58 SRM server configuration window at start point

4. The Connection to Remote Site window displays. Type the IP address and ports for the remote site which as shown in Figure A-59 on page 345. Click **Next**.

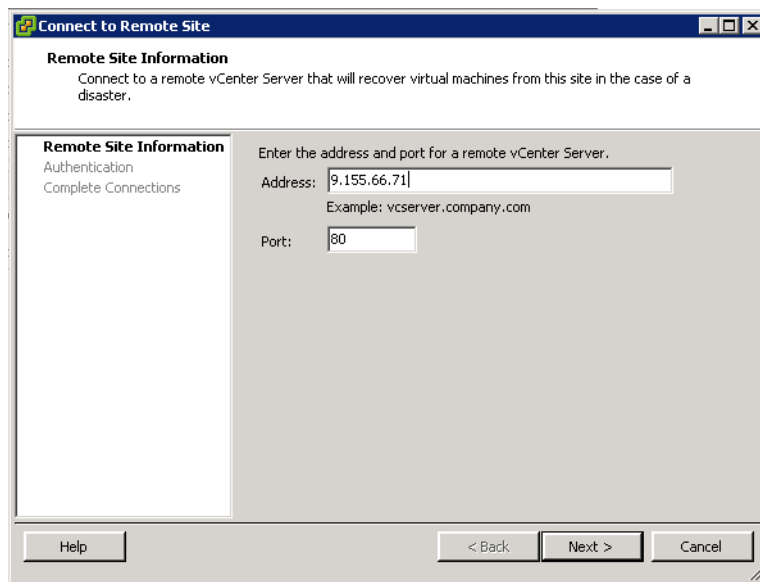


Figure A-59 Configure remote connection for the paired site

5. A remote vCenter server certificate error is displayed as shown in Figure A-60. Ignore the error message and click **OK**.

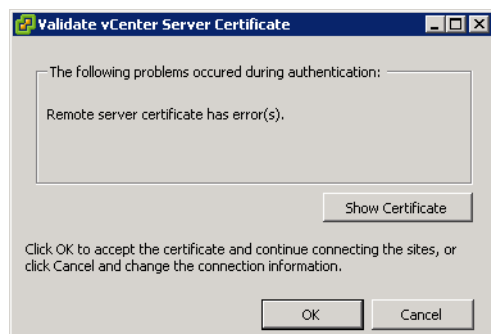


Figure A-60 Warning on vCenter Server certificate

6. In the window shown in Figure A-61 on page 346, type the user name and password to be used for connecting at the remote site. Click **Next**.

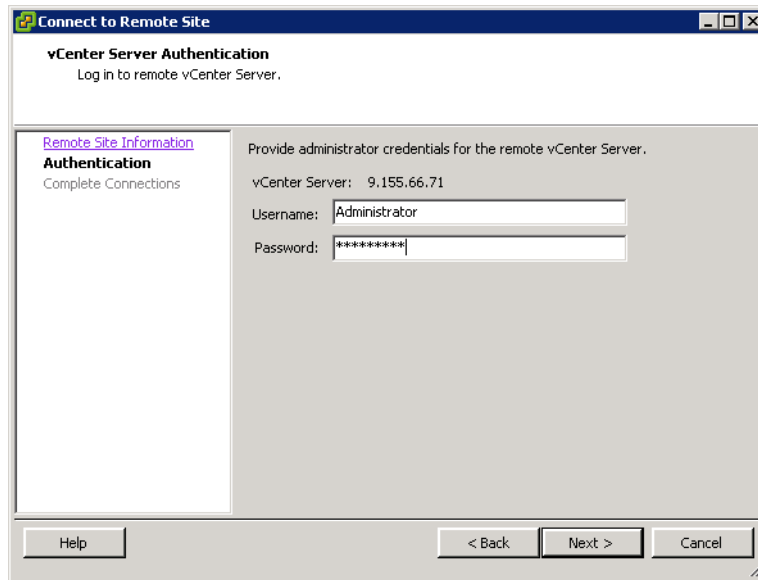


Figure A-61 Account details for the remote server

7. A remote vCenter server certificate error is displayed as shown in Figure A-62. Ignore this error message and click **OK**.

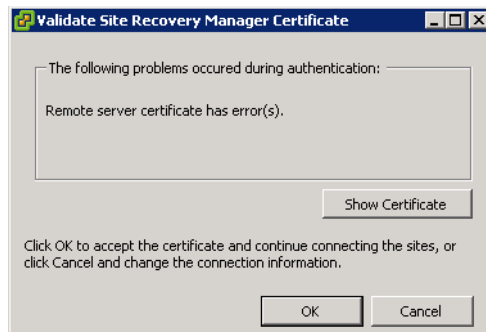


Figure A-62 SRM server certificate error warning

8. A configuration summary for the SRM server connection is now displayed as shown in Figure A-63 on page 347. Check that all is fine and click **Finish**.

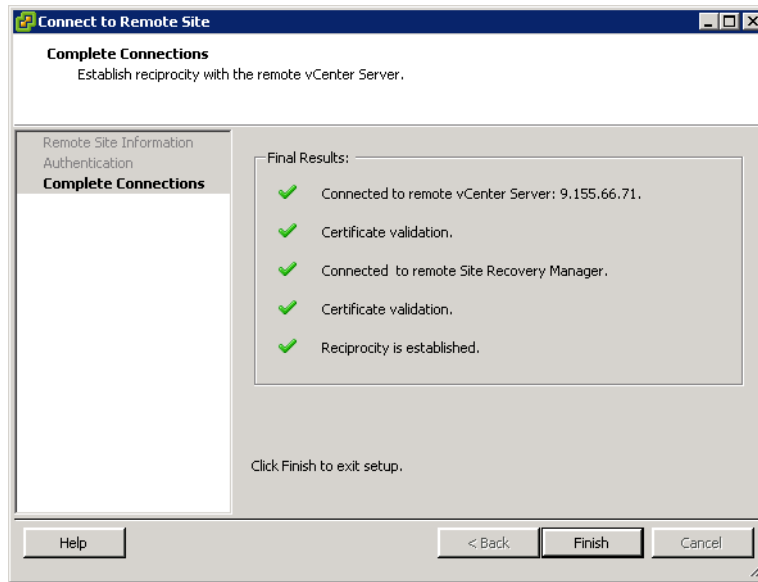


Figure A-63 Summary on SRM server connection configuration

9. In the main SRM server configuration window (see Figure A-58 on page 344) click **Configure** for Array Managers and the window shown in Figure A-64 opens. Click **Add**.

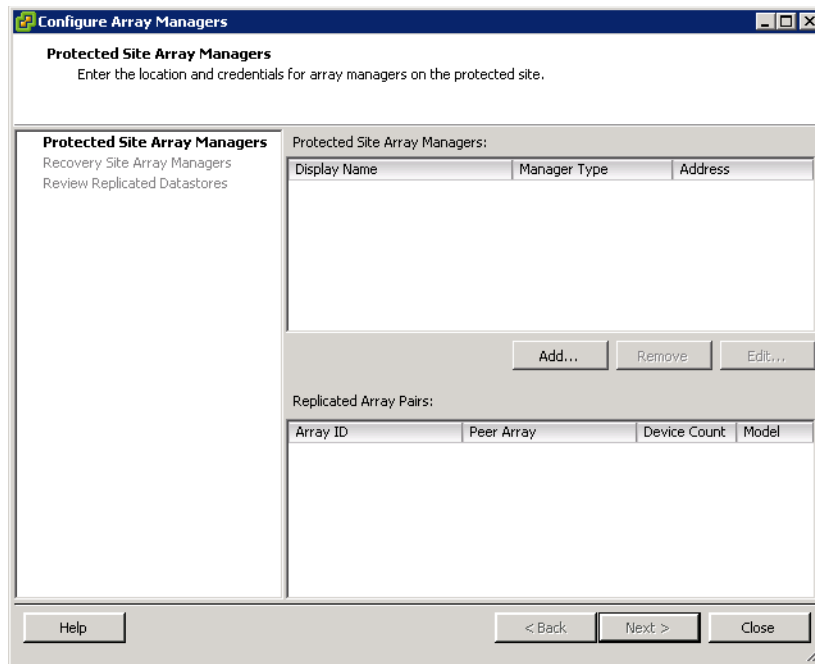


Figure A-64 Add the array manager to the SRM server

10. In the window now displayed, provide the information about the XIV Storage system located at the site that you are configuring at this moment, as shown in Figure A-65 on page 348. Click **Connect** to establish connection with XIV, then click **OK** to be returned to the previous window where you can observe the remote XIV paired with local XIV storage system. Click **Next**.

Add Array Manager

Array Manager Information

Display Name: XIV LAB 3 1300203

Manager Type: IBM XIV storage system

Address 1: 9.155.90.180

Address 2: 9.155.90.181

Address 3: 9.155.90.182

Username: itso

Password: *****

Connect

Array ID	Model
☒ XIV LAB 3 1300203	A14

Help OK Cancel

Figure A-65 XIV connectivity details for the protected site

11. Provide connectivity information for managing your secondary storage system as shown in Figure A-66. Click **Next**.

Configure Array Managers

Recovery Site Array Managers

Enter the location and credentials for array managers on the recovery site.

[Protected Site Array Managers](#)

Recovery Site Array Managers

Review Replicated Datastores

Display Name	Manager Type	Address
XIV LAB 01 EBC	IBM XIV storage syst...	9.155.53.250

Add... Remove Edit...

Replicated Array Pairs:

Array ID	Peer Array	Device Count	Model
☒ XIV LAB 3 1300203	XIV LAB 01 EBC	10	A14
☐ XIV LAB 3 1300203	XIV-02 7802959	1	A14

Help ≤ Back Next ≥ Close

Figure A-66 XIV connectivity details for the recovery site

12. The next window provides information about replicated datastores protected with remote mirroring on your storage system (refer to Figure A-67 on page 349). If all this information is correct, click **Finish**.

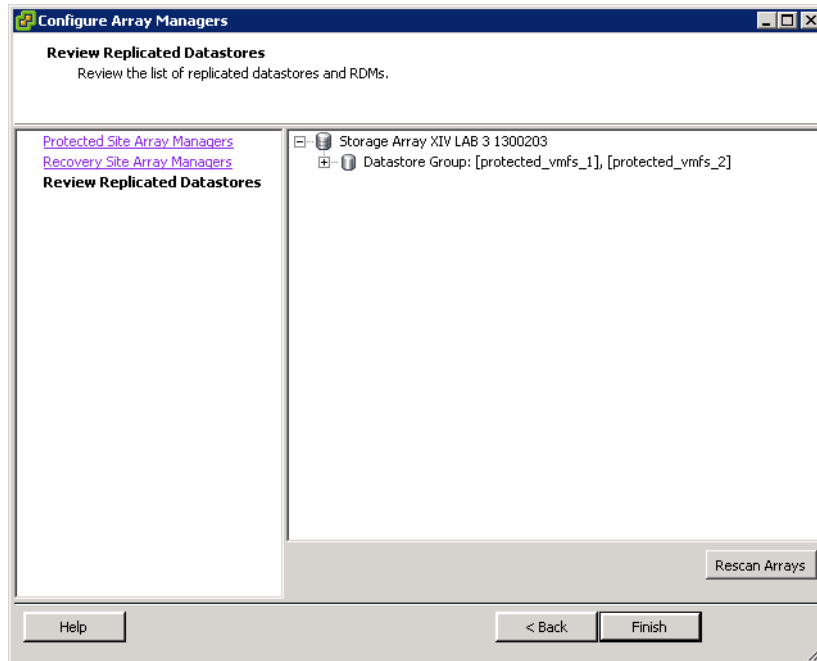


Figure A-67 Review replicated datastores

13. In the main SRM server configuration window click **Configure** and the window shown in Figure A-68 opens. Right-click sequentially on each category of resources (Networks, Compute Resources, Virtual Machine Folders) and select **Configure**. You will be asked to provide information about usage recovery site resources for the virtual machines at the protected site, in case of failure at the primary site.

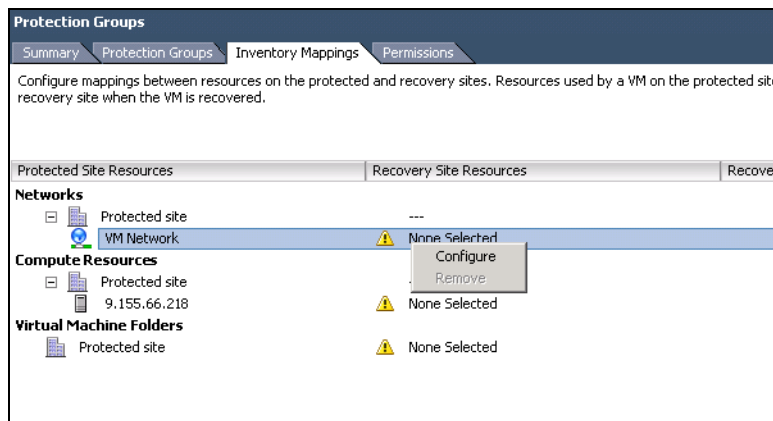


Figure A-68 Configure Inventory Mappings

14. To create a **Protection group**, from the main SRM server configuration window click **Create** next to the **Protection group** table. A window as shown in Figure A-69 on page 350 opens. Enter a name for the protection group, then click **Next**.

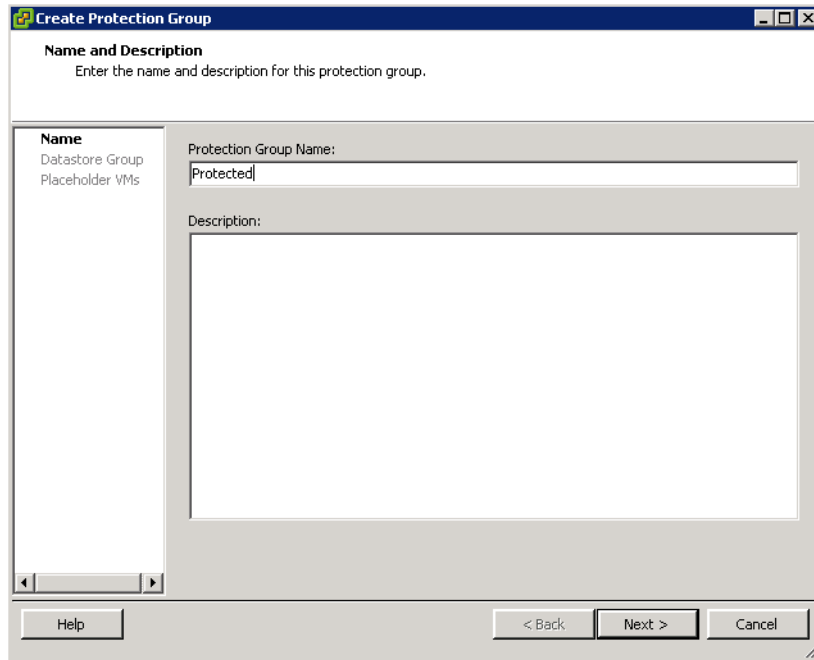


Figure A-69 Setting name for the protection group

15. Select datastores to be associated with Protection group created as shown in Figure A-70. Click **Next**.

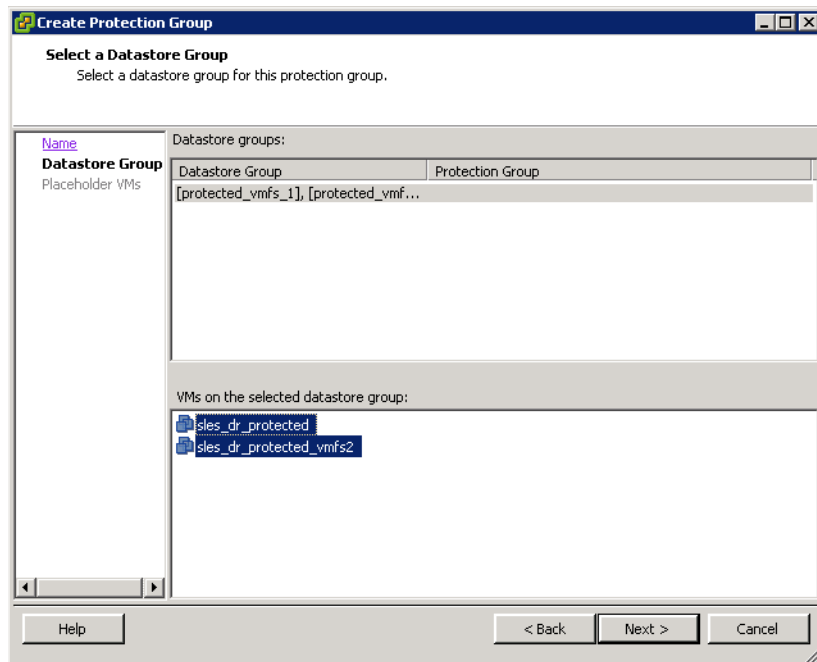


Figure A-70 Selection datastores for the protection group

16. Select placeholder to be used at the recovery site for the virtual machines included in the Protection Group, as shown in Figure A-71 on page 351. Then click **Finish**.

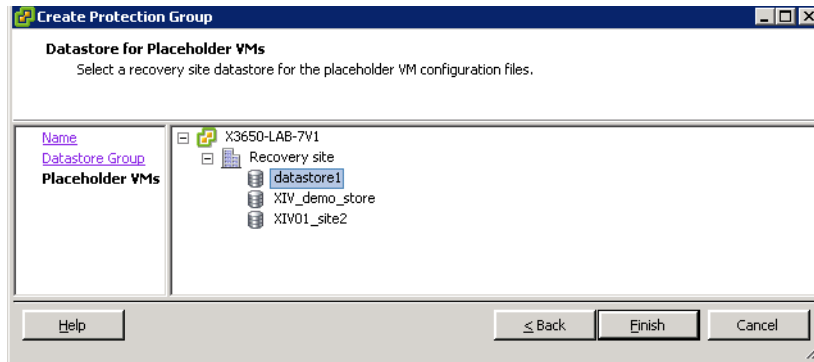


Figure A-71 Select datastore placeholder VMs

This completes the steps required at the protected site.

To configure the SRM server on the recovery site, follow these instructions:

1. Run the vCenter Client and connect to the vCenter server at the recovery site. From the main menu select **Home** and at the bottom of the next window, click **Site Recover** under the **Solutions and Applications** category. A window as shown in Figure A-72 is displayed. Select **Site Recovery** in the left pane and click **Create** (circled in red) in the right pane at the bottom of the window under the **Recovery Setup** subgroup.

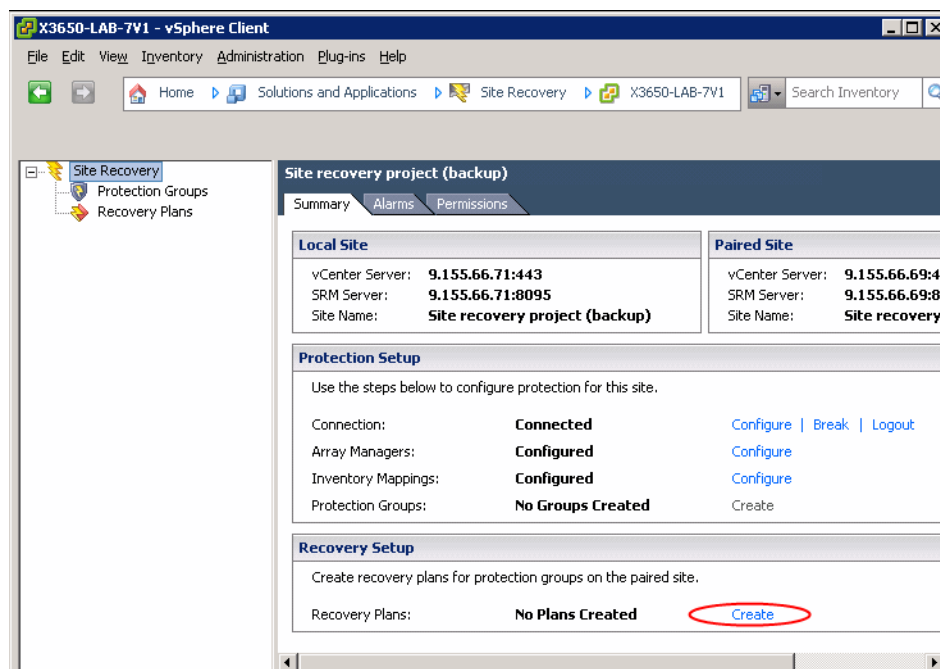


Figure A-72 Start the creating recovery plan

2. In the Create Recovery plan window now displayed, type a name for your recovery plan as shown in Figure A-73 on page 352, then click **Next**.

Create Recovery Plan

Recovery Plan Information
Enter the name and description for this recovery plan.

Recovery Plan Information

- Protection Groups
- Response Times
- Networks
- Suspend Local VMs

Name:

Description:

Help < Back Next > Cancel

Figure A-73 Setting name for your recovery plan

3. In the next window, select protection groups from you protected site for inclusion in the created recovery plan, as shown in Figure A-74, then click **Next**.

Create Recovery Plan

Protection Groups
Select the protection groups to recover with this plan.

Recovery Plan Information

- Protection Groups
- Response Times
- Networks
- Suspend Local VMs

Protection Groups at Site recovery project:

Name	Description
☒ Linux-1	

Help < Back Next > Cancel

Figure A-74 Select protection group which would be included into your recovery plan

4. The Response Times dialog is displayed as shown in Figure A-75. Type the desired values for your environment or leave the default values, then click **Next**.

Create Recovery Plan

Response Times
Set the response times for virtual machines in this plan.

Recovery Plan Information

- Protection Groups
- Response Times
- Networks
- Suspend Local VMs

VM Response Times

Additional time to wait for a response after the recovery step executes.

Change Network Settings: seconds

Wait for OS Heartbeat: seconds

Help < Back Next > Cancel

Figure A-75 Setting up response time settings

5. Select a network for use by virtual machines during a fail-over, as shown in Figure A-76 on page 353. You can specify networks manually or leave default settings, which imply that a new isolated network would be created when virtual machines start running at the recovery site. Click **Next**.

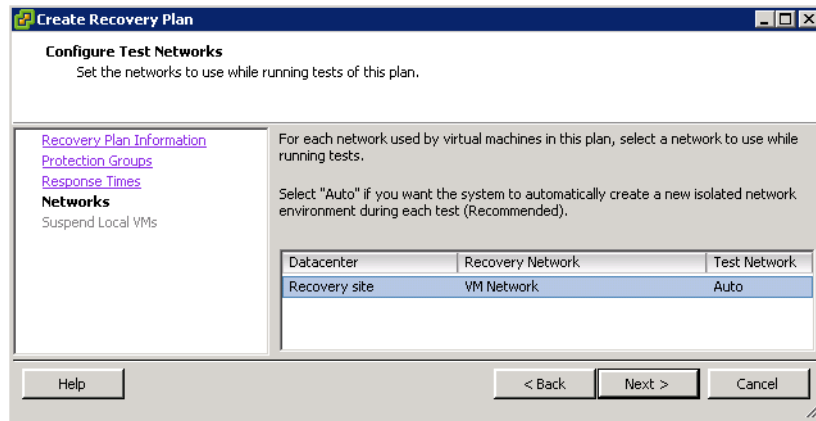


Figure A-76 Configure the networks which would be used for failover

- Finally, select the virtual machines that would be suspended at the recovery site when fail-over occurs as shown in Figure A-77. Make your selection and click **Finish**.

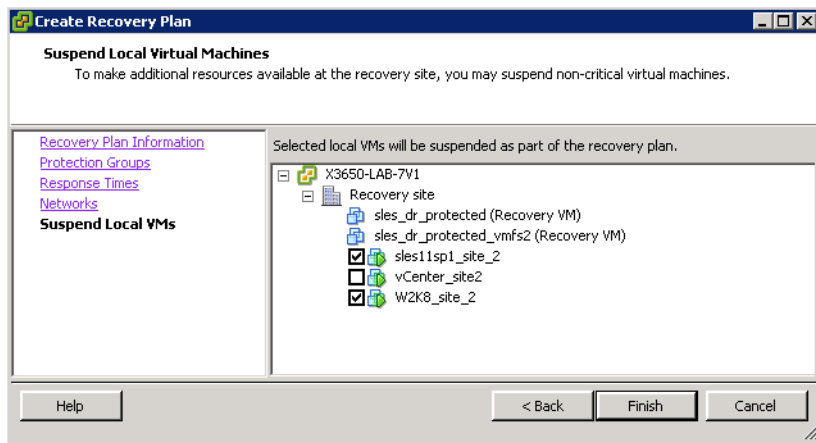


Figure A-77 Select virtual machines which would be suspended on recovery site during failover

Now, you have completed all steps required to install and configure a simple, proof-of-concept SRM server configuration.

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 356. Note that some of the documents referenced here may be available in softcopy only.

- ▶ *IBM XIV Storage System: Architecture and Implementation*, SG24-7659
- ▶ *IBM XIV Storage System: Copy Services and Data Migration*, SG24-7759
- ▶ *Introduction to Storage Area Networks*, SG24-5470
- ▶ *IBM System z Connectivity Handbook*, SG24-5444
- ▶ *PowerVM Virtualization on IBM System p: Introduction and Configuration*, SG24-7940
- ▶ *Implementing the IBM System Storage SAN Volume Controller V4.3*, SG24-6423
- ▶ *IBM System Storage TS7650, TS7650G and TS7610*, SG24-7652

Other publications

These publications are also relevant as further information sources:

- ▶ *IBM XIV Storage System Application Programming Interface*, GA32-0788
- ▶ *IBM XIV Storage System User Manual*, GC27-2213
- ▶ *IBM XIV Storage System: Product Overview*, GA32-0791
- ▶ *IBM XIV Storage System Planning Guide*, GA32-0770
- ▶ *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 HP/UX*, 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 Windows*, GA32-0652
- ▶ *IBM XIV Storage System Host Attachment Guide: Host Attachment Kit for Solaris*, GA32-0649
- ▶ *IBM XIV Storage System Pre-Installation Network Planning Guide for Customer Configuration*, GC52-1328-01

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 website:
<http://www.ibm.com/systems/storage/disk/xiv/index.html>
- ▶ System Storage Interoperability 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

Index

A

Active Directory 293
Agile 137
Agile View Device Addressing 136
AIX 43, 112, 133, 147, 157
ALUA 206
Array Support Library (ASL) 141–142, 158–160
Asymmetrical Logical Unit Access (ALUA) 205
Automatic Storage Management (ASM) 277

B

Block Zeroing 194
Brocade 74

C

cache 38
Capacity on Demand (COD) 179
cfgmgr 114
Challenge-Handshake Authentication Protocol (CHAP) 28–29, 51
chpath 118
client partition 177
clone 294
cluster 63, 196
Common Internet File System (CIFS) 252
connectivity 3–4, 6
context menu 27, 33, 35
Converged Network Adapters (CNA) 74
copy-on-write 294

D

Data Ontap 252, 256, 259
 7.3.3 253, 262
 installation 263–264
 update 264
 version 253
Data Protection for Exchange Server 301
database-managed space (DMS) 277
datastore 207–209
DB2 database 279
 storage-based snapshot backup 279
detailed information 198
Device Discovery Layer (DDL) 141
device node 71, 83–84
 additional set 91
 second set 91
Device Specific Module (DSM) 44
disk device 72–73, 82, 95, 163, 165–166, 174, 184, 186–187
 data area 100
disk group 163–164, 277
 name 165

 vgxiv 165
disk queue depth 40
Distributed Resource Scheduler (DRS) 193
DM-MP device 90–91
 new partitions 92
DS8000
 distinguish Linux from other operating systems 68
 existing reference materials 68
 Linux 68
 troubleshooting and monitoring 102
DSM 44
DSMagent 306

E

ESX 4 192, 202, 205
ESX host 202–203
ESX server 194, 203, 218
 new datastore 208
 virtual machine 194
esxtop 198
Ethernet switch 3
Exchange Server 293

F

fail_over 118
FB device 75
FC See *Fibre Channel (FC)*
fc_port_list 15
fcmsutil 134
FCP mode 73–75
fdisk 92
Fibre Channel (FC) 1–2, 9, 67–69, 94, 148, 173–174, 176, 195, 202, 205, 225, 266–268
 adapter 176, 184
 attachment 74, 79, 267
 card 74
 configuration 195
 connection 29, 195
 device 259
 fabric 72
 HBA 73–76
 HBA Adapter 94
 HBA driver 72, 75
 Host Bus Adapter 74
 interface 73
 port 2, 174, 176, 266
 Protocol 2, 70, 73
 SAN environment 174
 storage adapter 173
 storage attachment 70
 switch 3, 12, 31, 267–268
 switch one 268
file system 78, 91–92, 97, 99–100, 276–277, 279
FLASHCOPYMANAGER 305

Full Copy 194

G

General Parallel File System (GPFS) 240–241
given storage system
 path failover 205
Grand Unified Boot (GRUB) 105

H

HAK See *Host Attachment Kit (HAK)*
hard zone 12
Hardware Assisted Locking 194
Hardware Management Console (HMC) 172, 179, 181, 186
HBA See *host bus adapter (HBA)*
High Availability (HA) 193–194
HMC (Hardware Management Console) 172, 186
host
 transfer size 38
Host Attachment Kit (HAK) 7, 46, 67–68, 74, 78, 102, 148, 152, 160–161, 177, 186, 303
 package 79, 161
Host Attachment Wizard 48
host bus adapter (HBA) 2, 7–9, 72–74, 76, 82, 85, 97, 195–196, 199–201, 203, 213
 driver 9, 13, 76
 queue depth 38
host connectivity 1, 4, 6, 29, 67, 157, 178, 191, 260
 detailed view 6
 simplified view 4
host considerations
 distinguish Linux from other operating systems 68
 existing reference materials 68
 Linux 68
 support issues 68
 troubleshooting and monitoring 102
host definition 33, 36, 196, 204, 252, 258, 272
host HBA
 queue depth 38, 178
 side 40
host queue depth 38
host server 8, 29
 example power 36
hot-relocation use 164–166
HP Logical Volume Manager 138
HyperFactor 266

I

I/O operation 177–178
I/O request 38, 177–178
 maximum number 177
IBM i
 best practices 178
 queue depth 177
IBM i operating (I/O) 171–173
IBM SONAS
 Gateway 240–242
 Gateway cluster 246, 249

Gateway code 249
Gateway component 249
Gateway Storage Node 241, 243
Storage Node 244, 248
Storage Node 2 248
version 1.1.1.0-x 241
IBM System Storage
 Interoperability Center 8, 21
IBM Tivoli Storage FlashCopy Manager 291
IBM XIV 67, 78, 158–160, 192–194
 Array Support Library 159
 DMP multipathing 158
 end-to-end support 192
 engineering team 192
 FC HBA 31
 iSCSI IPs 31
 iSCSI IQN 31
 Management 216–217
 Serial Number 15
 Storage Replication Agent 218
 Storage System 14–15, 67, 167, 193–195, 239, 241, 251–252, 265, 267–268, 275, 277, 279
 Storage System device 216
 Storage System with VMware 192
 system 192
IBM XIV Storage
 System patch panel 30
Infocenter 112
Initial Program Load (IPL) 106
initRAMFS 76–77
Instant Restore 309
Instant Restore (IR) 308
Integrated Virtualization Manager (IVM) 172–173
Interface Module 2–4, 178, 242–243, 268–269
 iSCSI port 1 4
interface module 178
Interquery 278
Intraquery 278
inutoc 115
iopolicy 167
ioscan 134, 143
iostat 117
IP address 22, 56, 121
IP See *iSCSI port (IP)*
ipinterface_list 122
IPL See *Initial Program Load (IPL)*
IQN See *iSCSI Qualified Name (IQN)*
iSCSI 1
 boot 29
 configuration 22, 27
 connection 23, 26, 33, 37
 host specific task 32
 initiator 21
 name 24–25
 port (IP) 2, 26
 Qualified Name (IQN) 6, 22, 33, 53
 software initiator 2, 21
IVM (Integrated Virtualization Manager) 172–173

J

jumbo frame 22

L

LabManager 222
latency 24
left pane 35
legacy addressing 136
Legacy View 137
Level 1.0 304–306
link aggregation 22
Linux 68, 134, 148
 queue depth 45
Linux deal 75
Linux distribution 68, 71, 76
Linux kernel 68, 71, 86
Linux on Power (LOP) 72–73
Linux server 67–68, 78
 Host Attachment Kit 88
Linux system 72, 76–77, 82
load balancing policy 58
logical unit number (LUN) 2, 8–9, 82, 160, 162–163, 173, 176–178, 196–198, 200, 204–205, 225
 0 59
 id 35, 188
 id 1 188
 id 2 249
 large number 40
 scanning 196
logical volume 3, 69, 92, 95–96, 100
Logical Volume Manager (LVM) 38
logical volume manager (LVM) 71, 86, 158, 278
LOP See *Linux on Power (LOP)*
LPAR 72–73, 106, 172–173, 179
lsmmap 184
lsmod 75
lspath 118
LUN 0 257
LUN mapping 104, 162, 249, 261, 274
 view 35
 window 35
LUN See *logical unit number (LUN)*

M

MachinePool Editor 298
Managed Disk Group (MDG) 235
Master Boot Record (MBR) 105
Maximum Transmission Unit (MTU) 22
 default 22, 26
 maximum 22, 26
MBR 63
MBR See *Master Boot Record (MBR)*
MDG See *Managed Disk Group (MDG)*
menuing system 164, 166
metadata 99–100, 269–271
 capacity planning tool 270
Microsoft Exchange Server 294
Microsoft SQL Server 294
Microsoft Volume Shadow Copy Services (VSS) 281,

291–292
 architecture 293
 provider 291, 294
 requestor 293
 service 293
 writer 293
modinfo 75
modprobe 75
Most Recently Used (MRU) 199
MPIO See *multi-path I/O (MPIO)*
MRU See *Most Recently Used (MRU)*
MSDSM 45
MTU See *Maximum Transmission Unit (MTU)*
multipath 177
multipath device 69, 81, 87, 95, 98
 boot Linux 109
multi-path I/O (MPIO) 44, 116
 commands 118
multipathing 113, 115

N

N10116 293
NAA See *Network Address Authority (NAA)*
NAS See *Network Attached Storage (NAS)*
Native Multipathing (NMP) 205
Network Address Authority (NAA) 15
Network Attached Storage (NAS) 240, 252–253
Network File System (NFS) 252
NFS See *Network File System (NFS)*
NMP See *Native Multipathing (NMP)*
Node 174
Node Port ID Virtualization (NPIV) 172, 174
NPIV See *Node Port ID Virtualization (NPIV)*
NTFS 63

O

only specific (OS) 157–158, 160
operating system (OS) 6–7, 68, 159, 192, 275–277, 294–295, 298
 boot loader 105
 diagnostic information 102
original data 294
 exact read-only copy 294
 full copy 294
OS level 158
 unified method volume management 158
OS Type 31

P

parallelism 38
patch panel 3
Path Control Module (PCM) 116
Path Selection Plug-In (PSP) 205–206
PCM See *Path Control Module (PCM)*
performance 38
physical disk 177
 queue depth 177
Pluggable Storage Architecture (PSA) 205–206

- port 2 242, 244–245
- POST See *Power on Self Test (POST)*
- Power Blade servers 175
- Power on Self Test (POST) 104
- PowerVM 172
 - Enterprise Edition 173
 - Express Edition 172
 - Live Partition Mobility 173
 - Standard Edition 173
- ProtecTIER 266
- provider 291, 294
- PSA See *Pluggable Storage Architecture (PSA)*
- PSP See *Path Selection Plug-In (PSP)*
- pvlins 136
- Python 7
- python engine 47

Q

- QLogic
 - BIOS 105
- Qlogic
 - device driver 148
- queue depth 38–40, 45, 117, 177–178, 213
 - following types 177
- queue_depth 117
- quiesce 295

R

- Red Hat Enterprise Linux (RH-EL) 68
 - 5.2 148
- Redbooks Web site 356
 - Contact us xv
- redirect-on-write 295
- reference materials 68
- Registered State Change Notification (RSCN) 13, 255
- Remote Login Module (RLM) 259
- remote mirroring 8
- remote port 82, 84, 86, 95
 - sysfs structure 96
 - unit_remove meta file 96
- requestor 293, 295
- RH-EL See *Red Hat Enterprise Linux (RH-EL)*
- RLM See *Remote Login Module (RLM)*
- Round Robin 58
- round robin 118
- round_robin 118
- round-robin 0 89–90, 93
- RSCN See *Registered State Change Notification (RSCN)*

S

- same LUN 8, 168, 182
 - multiple snapshots 168
- SAN See *Storage Area Network (SAN)*
- SATP See *Storage Array Type Plug-In (SATP)*
- SCSI device 74, 82, 98, 103–104
 - dev/sgy mapping 104
- SCSI reservation 194
- SDK See *software development kit (SDK)*

- second HBA
 - port 14
 - WWPN 34
- series Gateway 251–252
 - fiber ports 253–254
- service 293
- sg_tools 104
- shadow copy 292
 - persistent information 293
- Site Recovery Manager (SRM) 193–195, 218, 312
- SLES 11 SP1 69
- SMIT See *System Management Interface Tool (SMIT)*
- SMP See *Symmetric Multi-Processing (SMP)*
- SMS See *system-managed space (SMS)*
- snapshot volume 294, 302
- soft zone 12
- software development kit (SDK) 193, 216
- software initiator 7
- Solaris 44, 112, 147
- SONAS Gateway 239–240
 - Installation guide 249
 - schematic view 240
- SONAS Storage 241–242, 244, 248
 - Node 244, 247–248
 - Node 1 HBA 244–245
 - Node 2 HBA 244–245
- SONAS Storage Node
 - 1 HBA 244–245
 - 2 HBA 242, 244–245
- SQL Server 293
- SqlServerWriter 294
- SRM See *Site Recovery Manager (SRM)*
- SSIC See *System Storage Interoperation Center (SSIC)*
- StageManager 222
- Storage Area Network (SAN) 3, 106, 182
 - boot 142, 145
 - switches 176
- Storage Array Type Plug-In (SATP) 205
- storage device 1, 69, 73, 80–81, 162, 173–174, 197, 204
 - physical paths 205
- storage pool 29, 36, 245–246, 256–257, 270, 276
 - IBM SONAS Gateway 245
- storage system 8, 29, 38, 67–68, 74, 82, 158–159, 162, 205–206, 218
 - operational performance 207
 - traditional ESX operational model 207
- storageadmin 298
- striping 38
- SUSE Linux
 - Enterprise Server 68–70
- SVC
 - LUN creation 234
 - LUN size 234
 - queue depth 236
 - zoning 232
- SVC cluster 232
- SVC nodes 232
- Symantec Storage Foundation
 - documentation 158, 163
 - installation 158

- version 5.0 159, 167–168
- Symmetric Multi-Processing (SMP) 192
- sysfs 78
- sysstat 104
- System Management Interface Tool (SMIT) 121
- System Service Tools (SST) 188
- System Storage Interoperation Center (SSIC) 6, 8, 21, 68, 176, 205, 241
- system-managed space (SMS) 277

T

- tar xvf 148
- Targets Portal 57
- TCO See *Total Cost of Ownership (TCO)*
- tdpexcc 304
- Technical Delivery Assessment (TDA) 241
- thread 40
- Tivoli Storage FlashCopy Manager 281, 291–292
 - detailed information 309
 - prerequisites 300
 - wizard 299
 - XIV VSS Provider 291
- Tivoli Storage Manager (TSM) 279, 304
 - Client Agent 305
- Total Cost of Ownership (TCO) 240
- transfer size 38
- troubleshooting and monitoring 102
- TS7650G ProtecTIER Deduplication Gateway 265–267
 - Direct attachment 267
- TSM See *Tivoli Storage Manager (TSM)*

U

- uname 78

V

- VAAI 207
- VAAI See *vStorage API Array Integration (VAAI)*
- vCenter 192–194
- VEA 140
- VERITAS Enterprise Administrator (VEA) 140
- VERITAS Volume Manager 138
- VIOS See *Virtual I/O Server (VIOS)*
- Virtual I/O Server (VIOS) 171–175, 177
 - client 171
 - logical partition 173
 - LVM mirroring 182
 - multipath 177
 - multipath capability 182
 - partition 175–176, 179, 183
 - queue depth 177
 - Version 2.1.1 175
 - XIV volumes 184
- virtual machine 74, 76, 107, 192–194
 - hardware resources 193
 - high-performance cluster file system 192
 - z/VM profile 107
- virtual SCSI
 - adapter 174, 176, 179–180, 183

- connection 174
- device 187–188
- HBA 72
- virtual tape 173, 266
- virtualization management (VM) 171–172, 174–175
- virtualization task 71, 86
- VM See *virtualization management (VM)*
- VMware ESX
 - 3.5 14
 - 3.5 host 195
 - 4 205, 215
 - server 192–193
 - server 3.5 195
- VMware Site Recovery Manager
 - IBM XIV Storage System 193
- Volume Group 120
- VSS See *Microsoft Volume Shadow Copy Services (VSS)*
- vssadmin 298
- vStorage API Array Integration (VAAI) 207
- vxdiskadm 138
- VxVM 138

W

- World Wide Node Name (WWNN) 83
- World Wide Port Name (WWPN) 6, 12, 14, 73–74, 78, 84, 86, 162, 248, 258, 260, 273
- writer 293, 295
- WWID 136
- WWNN See *World Wide Node Name (WWNN)*
- WWPN See *World Wide Port Name (WWPN)*

X

- XCLI 14, 25, 27, 36
- XCLI command 27, 29
- XenCenter 221
- XenConverter 221
- XenMotion 221
- XenServer
 - hypervisor 221
- XIV 1–3, 67–69, 79, 102, 133, 147–149, 153, 158–161, 191–193, 196, 198–199, 201, 225, 239–241, 251–252, 265–267, 281, 291–292
- XIV device 81, 102, 152
- XIV GUI 15, 25–27, 35, 97, 162, 196, 225, 245, 256, 260, 270, 272, 274
 - Regular Storage Pool 257
- XIV LUN 29, 177, 185, 264
 - exact size 264
- XIV storage
 - administrator 2
- XIV Storage System 1, 196, 204–205, 225, 294–295
 - architecture 178
 - I/O performance 178
 - LUN 177, 184–185
 - queue depth 177
 - main GUI window 32
 - point 36–37
 - primary IP address 298
 - serial number 24

- snapshot operations 299
- volume 38
- WWPN 15, 29
- XIV system 7, 40, 68, 80–81, 160–162, 173, 175–176, 192, 216, 240, 243–244, 253
 - maximum performance 243
 - now validate host configuration 80
- XIV volume 40, 72, 94–95, 99, 104–105, 107, 182, 217, 241, 245, 252, 256, 261–262, 276
 - direct mapping 72
 - iSCSI attachment 80
 - N series Gateway boots 256
- XIV VSS Provider
 - configuration 297
- XIV VSS provider 291, 295
- xiv_attach 79, 149, 161
- xiv_devlist 102, 152
- xiv_devlist command 81, 187
- xiv_diag 102, 153
- XIVDSM 45
- XIVTop 14
- xpyv 47

Z

- zoning 12, 14, 32



XIV Storage System: Host Attachment and Interoperability

(0.5" spine)
0.475" <-> 0.873"
250 <-> 459 pages



Redbooks®

XIV Storage System

Host Attachment and Interoperability

**Integrate with DB2,
Oracle, VMware ESX,
and Citrix Xen Server**

**Get operating system
specifics for Host Side
tuning**

**Use with SVC, SONAS,
IBM i, N Series, and
 ProtecTier**

This IBM Redbooks publication provides information for attaching the XIV Storage System to various host operating system platforms, or in combination with databases and other storage oriented application software.

The book also presents and discusses solutions for combining the XIV Storage System with other storage platforms, host servers, or gateways.

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

The information presented here is not meant as a replacement or substitute for the Host Attachment kit publications available at:
<http://publib.boulder.ibm.com/infocenter/ibmxiv/r2/index.jsp>

The book is meant as a complement and to provide the readers with usage recommendations 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

SG24-7904-00

ISBN 0738435635