Redbooks
ibm.com/redbooks

IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System
Host Attachment and Interoperability

Markus Oscheka

Bert Dufrasne

Roger Eriksson

Detlef Helmbrecht

Petar Kalachev

Stephen Solewin

Bruce Spell

Storage

IBM

Redbooks

International Technical Support Organization

IBM FlashSystem A9000, A9000R, and IBM XIV Storage System: Host Attachment and Interoperability

July 2019

SG24-8368-01

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

Second Edition (July 2019)

This edition applies to Version 12.2.1 of the IBM FlashSystem A9000 and IBM FlashSystem A9000R software, and Version 11.6.2 of the XIV Storage System software.

This document was created or updated on July 25, 2019.

Contents

Notices

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

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

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

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

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

Trademarks

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

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

Active Memory™
AIX®
FICON®
IBM®
IBM FlashSystem®
IBM Spectrum™
IBM Spectrum Accelerate™
IBM Spectrum Storage™
IBM Spectrum Virtualize™
IBM z Systems®
Interconnect®
Micro-Partitioning®
POWER®
Power Architecture®
Power Systems™
POWER6®
PowerVM®
Redbooks®
Redbooks (logo) ®
System Storage®
System z®
XIV®
z Systems®
z/VM®
z10™

The following terms are trademarks of other companies:

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

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

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

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

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

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

Preface

This IBM® Redbooks® publication provides information for attaching the IBM FlashSystem® A9000, IBM FlashSystem A9000R, and IBM XIV® Storage System to various host operating system platforms, such as IBM AIX® and Microsoft Windows.

This publication was last updated in May 2019 to cover the VLAN tagging and port trunking support available with software version 12.3.2 (see in particular section 2.4, “VLAN tagging” on page 67.

The goal is to give an overview of the versatility and compatibility of the IBM Spectrum™ Accelerate family of storage systems with various platforms and environments.

The information that is presented here is not meant as a replacement or substitute for the IBM Storage Host Attachment Kit publications or other product publications. It is meant as a complement and to provide usage guidance and practical illustrations.

You can download the IBM Storage Host Attachment Kit from IBM Fix Central:

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

This publication does not address attachments to a secondary system used for Remote Mirroring or data migration. These topics are covered in *IBM FlashSystem A9000 and IBM FlashSystem A9000 and A9000R Business Continuity Solutions*, REDP-5401.

Authors

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

Markus Oscheka is an IT Specialist for Proof of Concepts and Benchmarks with the Disk Solution Europe team in Germany. He has worked at IBM for fourteen years. He has performed many proof of concepts with Copy Services on IBM Spectrum Virtualize™ and IBM Spectrum Storage™, and also Performance-Benchmarks with IBM Spectrum Virtualize and IBM Spectrum Storage. He has written extensively, and acted as Project Lead for various Redbooks publications. He has spoken at several System Technical Universities. He holds a degree in Electrical Engineering from the Technical University in Darmstadt.

Bert Dufrasne is an IBM Certified Consulting IT Specialist and Project Leader for IBM System Storage® disk products at the International Technical Support Organization (ITSO), San Jose Center. He has worked at IBM in various IT areas. He has authored many Redbooks publications, and he has also developed and taught technical workshops. Before Bert joined the ITSO, he worked for IBM Global Services as an Application Architect. He holds a Master’s degree in Electrical Engineering.

Roger Eriksson works at IBM Systems Lab Services Nordic, based in Stockholm, Sweden. He is a Senior Accredited IBM Product Service Professional. Roger has over 20 years of experience working on IBM servers and storage. He has done consulting, proof of concepts, and education, mainly with the IBM Spectrum Accelerate™ product line, since December 2008. Working with both clients and various IBM teams worldwide is a normal day’s work. He holds a Technical College Graduation in Mechanical Engineering.

Detlef Helmbrecht is an Advanced Technical Skills (ATS) IT Specialist working for IBM Systems at the European Storage Competence Center (ESCC), Germany. Detlef has over 30 years of experience in IT, performing numerous roles, including software design, sales, and solution architecture. He joined IBM in 2013. His current areas of expertise include high-performance computing (HPC), application and database tuning, and IBM products, such as IBM Spectrum Virtualize, IBM Spectrum Accelerate, and IBM FlashSystem family. Detlef holds a degree in Mathematics from the Ruhr-Universität Bochum in Germany.

Petar Kalachev is a Product Field Engineer (PFE) working for IBM Systems at the Client Innovation Centre (CIC) in Sofia, Bulgaria. He has 7 years of experience in the IT industry spanning across various solutions combining servers, storage, SAN infrastructure, and virtualization. Since joining IBM in 2014 his focus has been on midrange and high-end storage systems, with current areas of expertise including the IBM Spectrum Accelerate and IBM FlashSystem family.

Stephen Solewin is an IBM Corporate Solutions Architect based in Tucson, Arizona. He has over 20 years of experience in working on IBM storage, including enterprise and midrange disk, Linear Tape-Open (LTO) drives, and libraries, SAN, storage virtualization, and storage software. Steve is a global resource, working with customers, IBM Business Partners, and fellow IBMers worldwide. He has been working on the XIV product line since 2008 and FlashSystem since 2013. He holds a BS degree in electrical engineering from the University of Arizona, where he graduated with honors.

Bruce Spell is a Storage Technical Advisor, located in Tucson, Arizona. As a Technical Advisor he delivers proactive and comprehensive consulting services for IBM storage products to ensure the highest availability for vital IT solutions. He is a customer advocate, providing focus for problem management, change management, crisis management, and best practices. Bruce has over 32 years of experience and a broad background spanning across development, test, field support, and program management. He has been a Technical Advisor since 2007 and recognized as a Notable Advocate over the last several years. Bruce holds a BS degree in Electrical Engineering from UCLA.

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

1

XIV host connectivity

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

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

For the latest information, see the IBM Storage Host Attachment Kit publications at Fix Central:

`https://ibm.biz/BdsnzX`

This chapter includes the following topics:

- 1.1, “Overview” on page 2
- 1.2, “Fibre Channel connectivity” on page 9
- 1.3, “iSCSI connectivity” on page 22
- 1.4, “Logical configuration for host connectivity” on page 31
- 1.5, “Performance tuning” on page 39
- 1.6, “Troubleshooting” on page 40

Note: Starting with version 2.6.0, IBM XIV Host Attachment Kit was renamed to *IBM Storage Host Attachment Kit.*

1.1 Overview

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

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

XIV is perfectly suited for integration into a new or existing FC storage area networks (SAN). After the host HBAs, cabling, and SAN zoning are in place, connecting an FC host to XIV is easy.

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

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

The XIV Storage System has up to 15 data modules, of which up to six are also interface modules. The number of interface modules and the activation status of the interfaces on those modules is dependent on the rack configuration. Table 1-1 lists the number of active interface modules and the FC and iSCSI ports for different rack configurations. As shown in Table 1-1, a six module XIV physically has three interface modules, but only two of them have active ports. An 11 module XIV physically has six interface modules, five of which have active ports.

Table 1-1 XIV host ports as capacity grows

Module	**6**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Module 9 host ports	Not present	Inactive ports	Inactive ports	Active	Active	Active	Active	Active
Module 8 host ports	Not present	Active	Active	Active	Active	Active	Active	Active
Module 7 host ports	Not present	Active	Active	Active	Active	Active	Active	Active
Module 6 host ports	Inactive ports	Inactive ports	Inactive ports	Inactive ports	Inactive ports	Active	Active	Active
Module 5 host ports	Active	Active	Active	Active	Active	Active	Active	Active
Module 4 host ports	Active	Active	Active	Active	Active	Active	Active	Active
FC Ports	8	16	16	20	20	24	24	24
iSCSI ports	6	14	14	18	18	22	22	22

Each active interface module (modules 4 - 9, if enabled) has four Fibre Channel ports. Each active interface module, except module 4 has four iSCSI ports. Module 4 has only two iSCSi ports. The maximum is therefore 22 ports.

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

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

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

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

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

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

1.1.1 Module, patch panel, and host connectivity

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

http://www.redbooks.ibm.com/abstracts/sg247659.html

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

- FC hosts connect to the XIV patch panel FC ports 1 and 3 on interface modules.
- Use XIV patch panel FC ports 2 and 4 for remote mirroring. They can also be used for data migration from another storage system.
- iSCSI hosts connect to at least one port on each active interface module.
- Connect hosts to multiple separate Interface Modules to avoid a single point of failure.

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

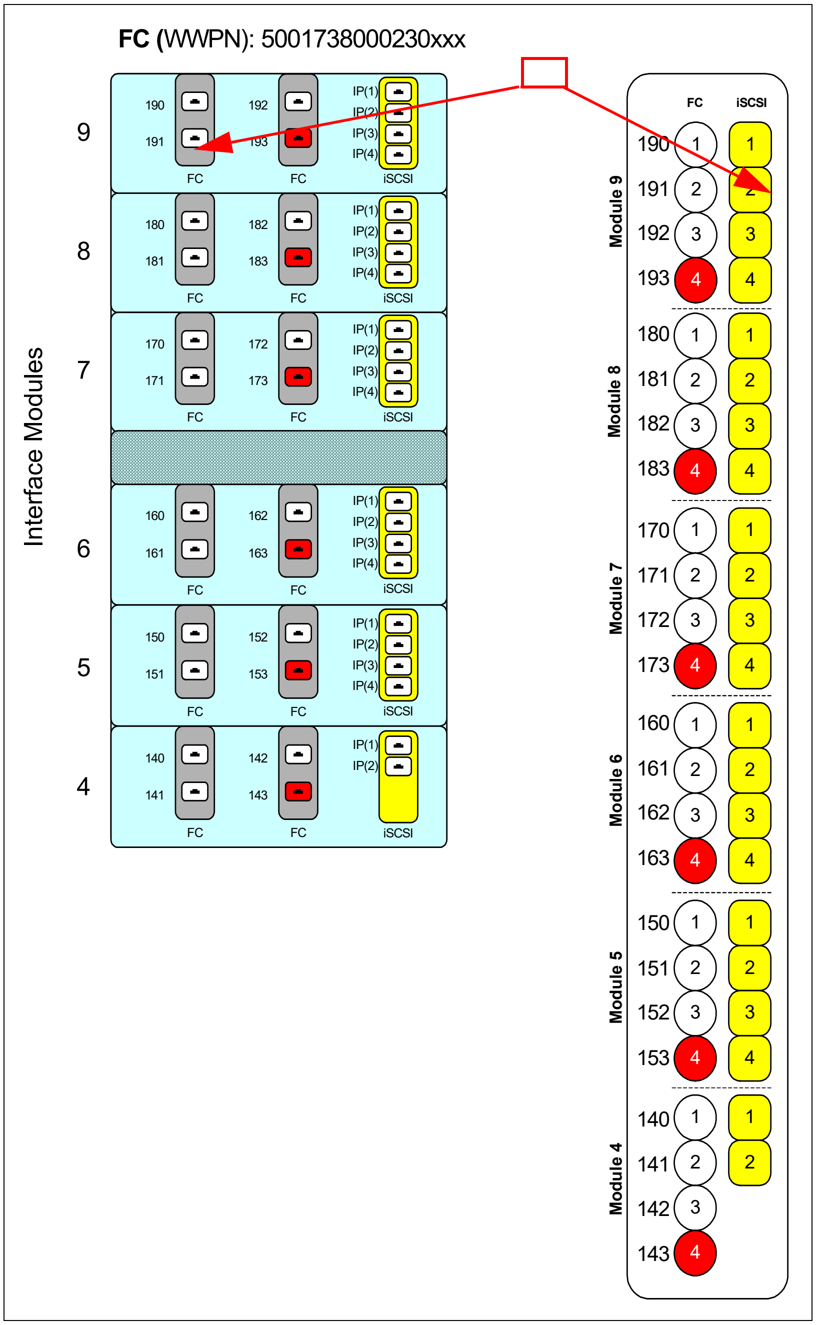


Figure 1-1 XIV Gen 3 Patch panel to FC and iSCSI port mappings

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

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

For more information about host connectivity and configuration options, see 1.2, “Fibre Channel connectivity” on page 9 and 1.3, “iSCSI connectivity” on page 22.

1.1.2 Host operating system support

The XIV Storage System supports many operating systems, and the list is constantly growing.

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

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

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

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

1.1.3 Host Attachment Kit

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

The Host Attachment Kit includes the following features:

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

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

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

You can download a Host Attachment Kit from Fix Central:

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

Commands provided by the Host Attachment Kit

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

The xiv_attach command

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

> **Tip:** AIX needs extra consideration. For more information, see “Installing the Host Attachment Kit for AIX” on page 194.

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

The xiv_detach command

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

The xiv_devlist command

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

- Size of the volume
- Number of paths (working and detected)
- Name and ID of each volume on the XIV
- ID of the XIV itself
- Name of the host definition on the XIV

The `xiv_devlist` command is one of the most powerful tools in your toolkit. Make sure that you are familiar with this command and use it whenever performing system administration. For more information about useful parameters that can be run with `xiv_devlist`, see *Host Attachment Kit User Guide*, GA32-1060.

The following parameters are especially useful:

xiv_devlist -u GiB
Displays the volume size in binary gigabytes (gibibyte, GiB). The **-u** is for unit size.

xiv_devlist -V
Displays the Host Attachment Kit version number. The **-V** is for version.

xiv_devlist -f filename.csv -t csv
Directs the output of the command to a file.

xiv_devlist -h
Opens the help page, which displays other available parameters. The **-h** is for help.

The xiv_diag command

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

The xiv_fc_admin command

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

xiv_fc_admin -P
Displays the WWPNs of the host server HBAs. The **-P** is for print.

xiv_fc_admin -V
Lists the tasks that **xiv_attach** would perform if it were run. Knowing the tasks is vital if you are using the portable version of the Host Attachment Kit. You must know what tasks the Host Attachment Kit needs to perform on your system before the change window. The **-V** is for verify.

xiv_fc_admin -C
Performs all the tasks that the **xiv_fc_admin -V** command identified as being required for your operating system. The **-C** is for configure.

xiv_fc_admin -R
This command scans for and configures new volumes that are mapped to the server. For a new host that is not yet connected to an XIV, use **xiv_attach**. However, if more volumes are mapped to such a host later, use **xiv_fc_admin -R** to detect them. You can use native host methods but the Host Attachment Kit command is an easier way to detect volumes. The **-R** is for rescan.

xiv_fc_admin -R --clean
Use the **clean** option to remove devices from the multipath. You can clean unreachable devices (use only with the **-R**/**--rescan** option). Option **clean** was added in Host Attachment Kit version 2.6.

xiv_fc_admin -h
Opens the help page that displays other available parameters. The **-h** is for help.

The xiv_iscsi_admin command

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

Coexistence with other multipathing software

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

A mix of different multipathing solution software on the same server is not supported. Each product can have different requirements for important system settings, which can conflict.

These conflicts can cause issues that range from poor performance to unpredictable behaviors, and even data corruption.

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

1.1.4 Fibre Channel versus iSCSI access

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

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

Consider the following information:

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

Figure 1-2 shows the simultaneous access to two different XIV volumes from one host by using both protocols.

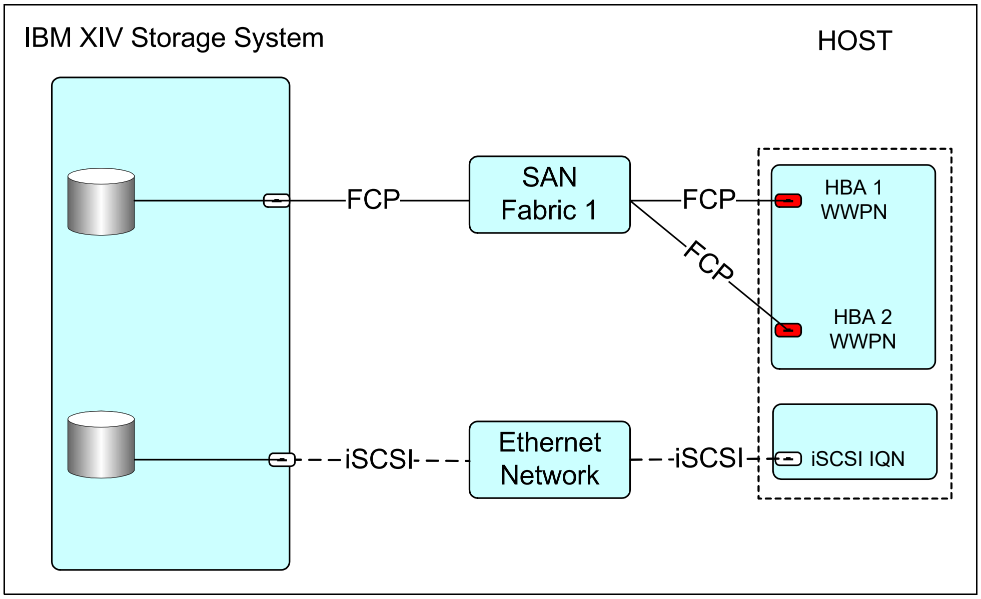


Figure 1-2 Connecting by using FCP and iSCSI simultaneously with separate host objects

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

1.2 Fibre Channel connectivity

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

1.2.1 Preparation steps

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

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

   `https://www.ibm.com/systems/support/storage/ssic/interoperability.wss`

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

   Also, review any documentation that comes from the HBA vendor and ensure that any additional conditions are met.
2. Check the LUN limitations for your host operating system and verify that there are enough adapters installed. You need enough adapters on the host server to manage the total number of LUNs that you want to attach.
3. Check the optimum number of paths that must be defined to help determine the zoning requirements.
4. Download and install the latest supported HBA firmware and driver, if needed.

HBA vendor resources

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

Platform and operating system vendor pages

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

1.2.2 Fibre Channel configurations

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

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

This section describes the following typical FC configurations that are supported and offer redundancy:

- Redundant configuration with twelve paths to each volume
- Redundant configuration with six paths to each volume
- Redundant configuration with minimal cabling

These configurations have no single point of failure. Consider the following points:

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

Redundant configuration with twelve paths to each volume

The fully redundant configuration is shown in Figure 1-3.

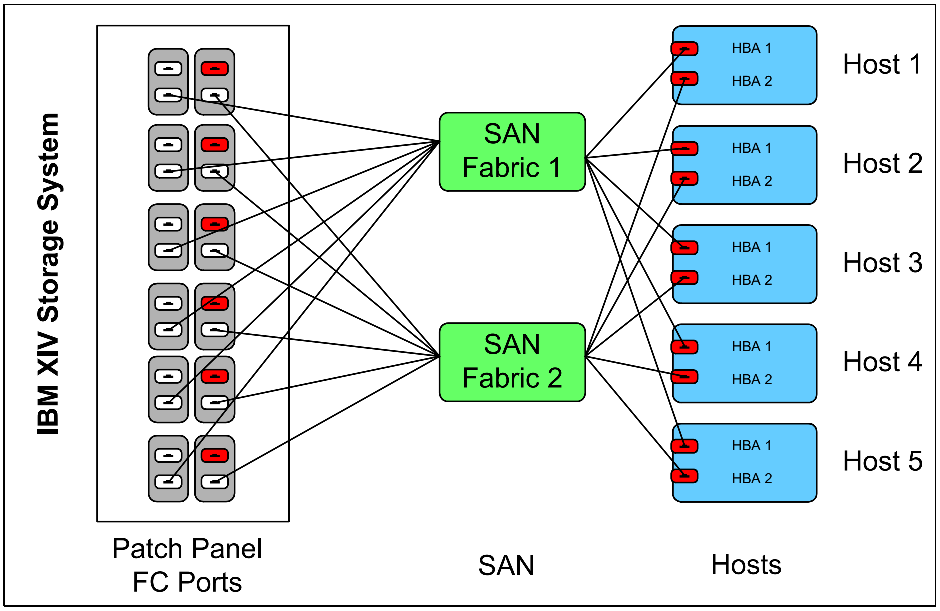


Figure 1-3 Fibre Channel fully redundant configuration

This configuration features the following characteristics:

- Each host is equipped with dual HBAs. Each HBA (or HBA port) is connected to one of two FC switches in different SAN fabrics.
- Each of the FC switches is connected to a separate FC port of each of the six interface modules.
- Each volume can be accessed through 12 paths. No benefit is realized in going beyond 12 paths because it can cause issues with host processor utilization and server reliability if a path failure occurs.

Redundant configuration with six paths to each volume

A redundant configuration that accesses all interface modules, but uses the ideal of six paths per LUN on the host, is shown in Figure 1-4.

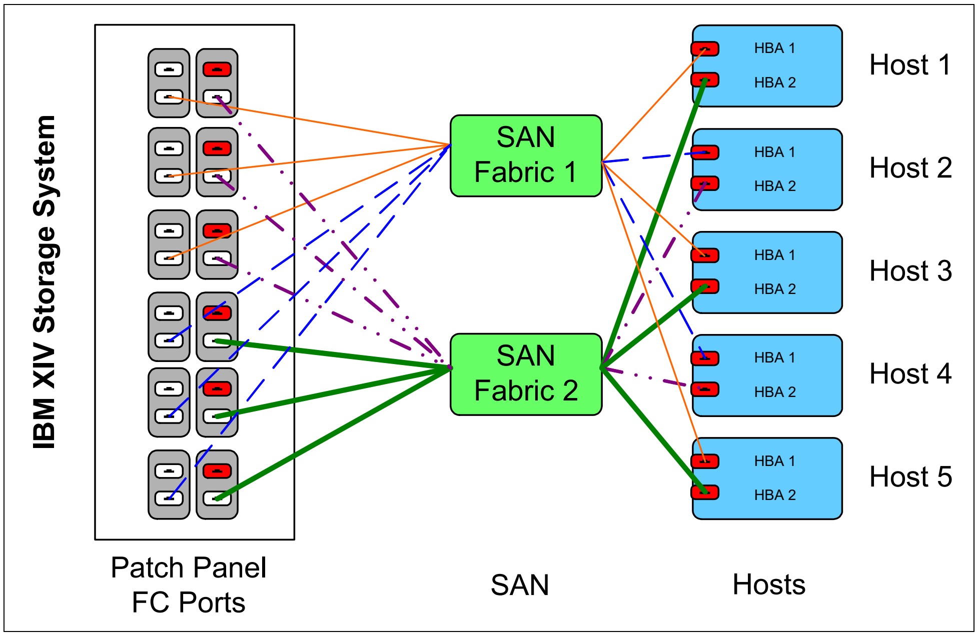


Figure 1-4 Fibre Channel redundant configuration

This configuration has the following characteristics:

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

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

Redundant configuration with minimal cabling

An even simpler redundant configuration is shown in Figure 1-5.

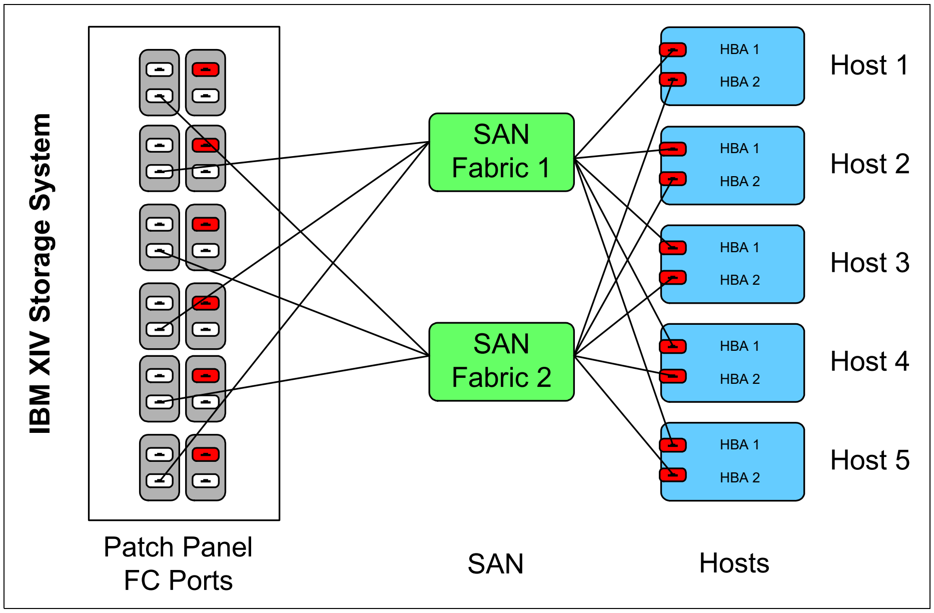


Figure 1-5 Fibre Channel simple redundant configuration

This configuration has the following characteristics:

- Each host is equipped with dual HBAs. Each HBA (or HBA port) is connected to one of two FC switches in different SAN fabrics.
- Each of the FC switches connects to three separate interface modules.
- Each volume has six paths.

Determining the ideal path count

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

> **Tip:** No performance or reliability benefit is realized by using too many paths. Going beyond 12 paths per volume has no benefit. More paths add more processor usage, which causes longer times for recovery. Going beyond six paths rarely has much benefit. Use four or six paths per volume as a standard.

Consider the configurations that are listed in Table 1-2 on page 13. The columns show the interface modules, and the rows show the number of installed modules. The table does not show how the system is cabled to each redundant SAN fabric, or how many cables are connected to the SAN fabric. You normally connect each module to each fabric and alternate which ports you use on each module.

- For a 6-module system, each host has four paths per volume: Two from module 4 and two from module 5. Port 1 on each module is connected to fabric A, whereas port 3 on each module is connected to fabric B. Each host is zoned to all four ports.
- For a 9- or 10-module system, each host has four paths per volume (one from each module). Port 1 on each module is connected to fabric A, whereas port 3 on each module is connected to fabric B.

Divide the hosts into two groups. Group 1 is zoned to port 1 on modules 4 and 8 in fabric A, and port 3 on modules 5 and 7 in fabric B. Group 2 is zoned to port 3 on modules 4 and 8 in fabric B, and port 1 on modules 5 and 7 in fabric A.

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

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

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

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

1.2.3 Zoning

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

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

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

- An error from an HBA that affects the zone or zone traffic is isolated to only the devices to which it is zoned.
- Any change in the SAN fabric triggers a *registered state change notification* (RSCN). Such changes can be caused by a server restarting or a new product being added to the SAN. An RSCN requires that any device that can “see” the affected or new device to acknowledge the change, interrupting its own traffic flow.

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

Zoning is affected by the following factors, among others:

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

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

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

For more information about SAN zoning, see *Introduction to Storage Area Networks*, SG24-5470:

http://www.redbooks.ibm.com/abstracts/sg245470.html

Soft zoning by using the *single initiator, multiple targets* method is shown in Figure 1-6.

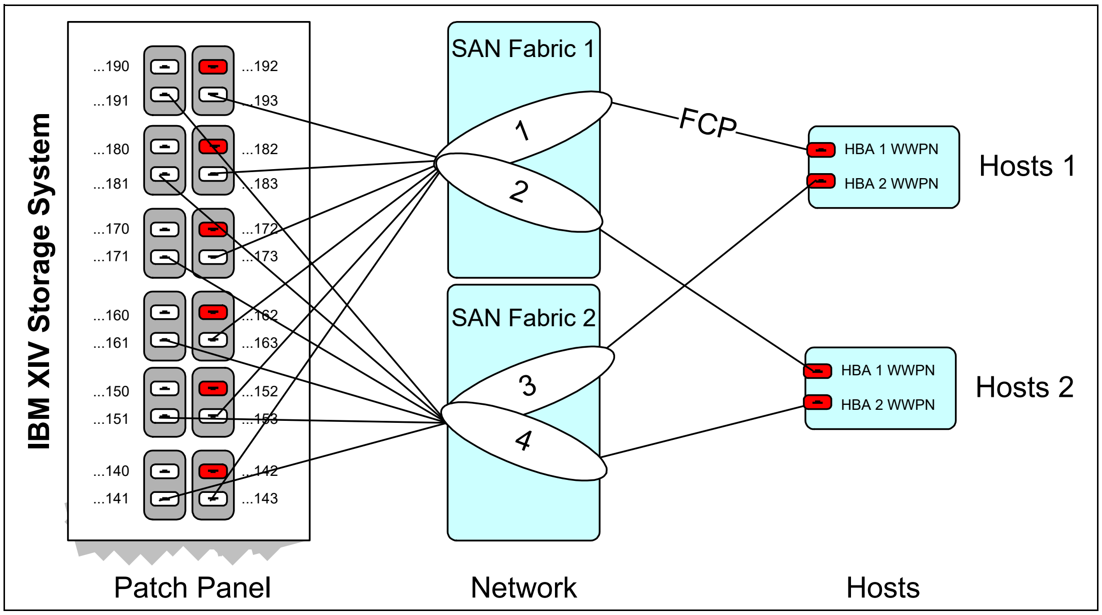


Figure 1-6 FC SAN zoning: The single initiator, multiple targets method

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

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

1.2.4 Identification of FC ports (initiator/target)

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

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

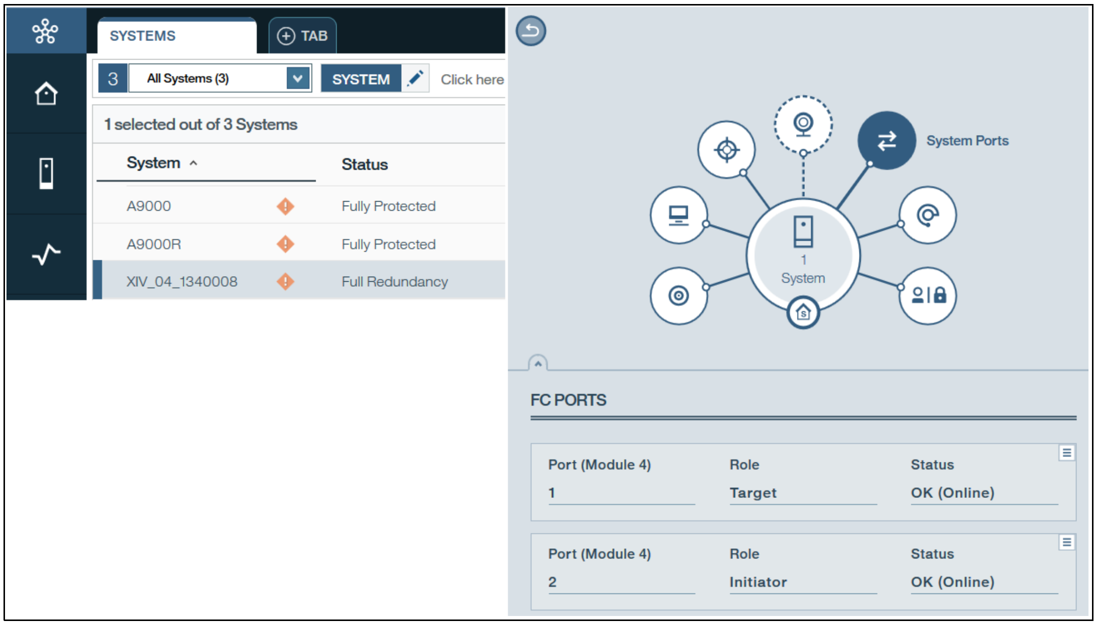
Example 1-1 Getting the WWPNs of an IBM XIV Storage System (XCLI)

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

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

To retrieve the same information from the Hyper-Scale Manager (GUI), see Figure 1-7 on page 16 and complete the following steps:

1. From the **Systems & Domains** view, select the system of interest.
2. From the Hub view of the System, select the **System Ports** spoke.
3. The ports and modules are listed at the bottom of the panel.



1 selected out of 3 Systems

System	Status
A9000	Fully Protected
A9000R	Fully Protected
XIV_04_1340008	Full Redundancy

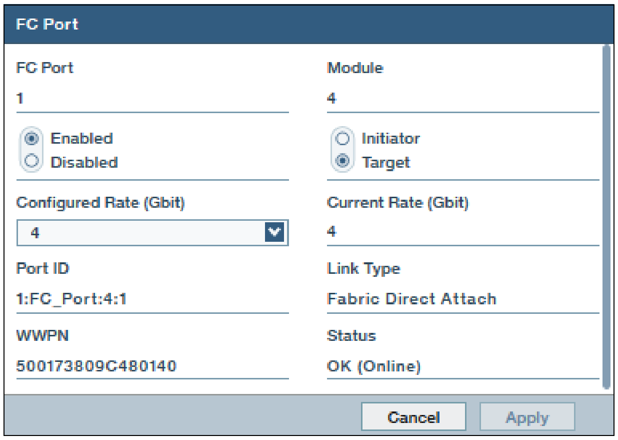
FC PORTS

Port (Module 4)	Role	Status
1	Target	OK (Online)

Port (Module 4)	Role	Status
2	Initiator	OK (Online)

Figure 1-7 Retrieving Fibre Channel port properties

4. Click the **Actions** icon for a particular port, and select **View/Edit FC Port**.
5. Scroll down if necessary to view the WWPN, as shown in Figure 1-8.



FC Port: 1	Module: 4
● Enabled ○ Disabled	○ Initiator ● Target
Configured Rate (Gbit): 4	Current Rate (Gbit): 4
Port ID: 1:FC_Port:4:1	Link Type: Fabric Direct Attach
WWPN: 500173809C480140	Status: OK (Online)

Figure 1-8 Viewing the FC port WWPN

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

As shown in Figure 1-8, the WWPN is `500173809C480140`, which means that the WWPN is from module 4, port 1.

The WWPNs for the port are numbered from 0 to 3, whereas the physical ports are numbered 1 - 4.

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

Example 1-2 Composition of the WWPN

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

5           NAA (Network Address Authority)
001738      IEEE Company ID from http://standards.ieee.org/regauth/oui/oui.txt
NNNNN       IBM XIV Serial Number in hexadecimal
RR          Rack ID   (01-FF, 00 for 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 or x64 based architecture

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

To boot from SAN, complete the following basic steps:

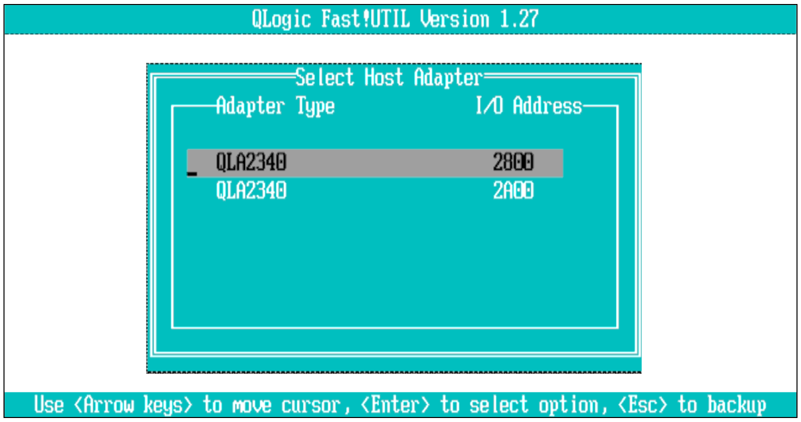
1. Go into the HBA configuration mode.
2. Set the HBA BIOS to `Enabled`.
3. Detect at least one XIV target port.
4. Select a LUN to boot from.

You often configure 2 to 4 XIV ports as targets. When using Hyper-Scale Mobility, make sure that in the adapter BIOS, targets (WWPNs) from both storage systems are defined. You might need to enable the BIOS on two HBAs, depending on the HBA, driver, and operating system. See the documentation that came with your HBA and operating systems.

For information about SAN boot for AIX, see Chapter 5, “AIX connectivity” on page 191.

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

1. Start your server. During the start process, press Ctrl+Q when prompted to load the configuration utility and display the **Select Host Adapter** menu (see Figure 1-9).



QLogic Fast!UTIL Version 1.27

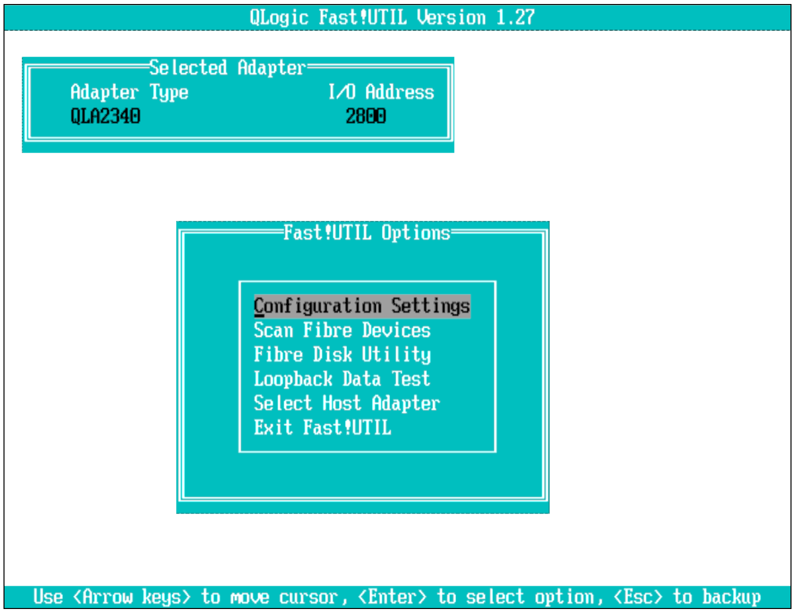
Select Host Adapter

Adapter Type	I/O Address
QLA2340	2800
QLA2340	2A00

Use <Arrow keys> to move cursor, <Enter> to select option, <Esc> to backup

Figure 1-9 Select Host Adapter menu

2. You normally see one or more ports. Select a port and press Enter to display the next panel (see Figure 1-10). If you are enabling the BIOS on only one port, make sure to select the correct port.

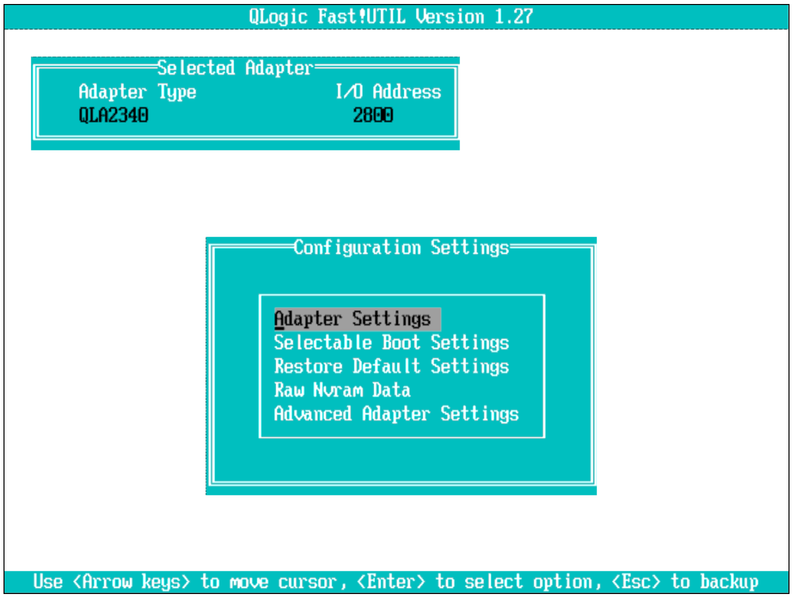


```
QLogic Fast!UTIL Version 1.27

Selected Adapter
Adapter Type          I/O Address
QLA2340                  2800

Fast!UTIL Options
Configuration Settings
Scan Fibre Devices
Fibre Disk Utility
Loopback Data Test
Select Host Adapter
Exit Fast!UTIL

Use <Arrow keys> to move cursor, <Enter> to select option, <Esc> to backup
```

Figure 1-10 Fast!UTIL Options menu

3. Select **Configuration Settings**.

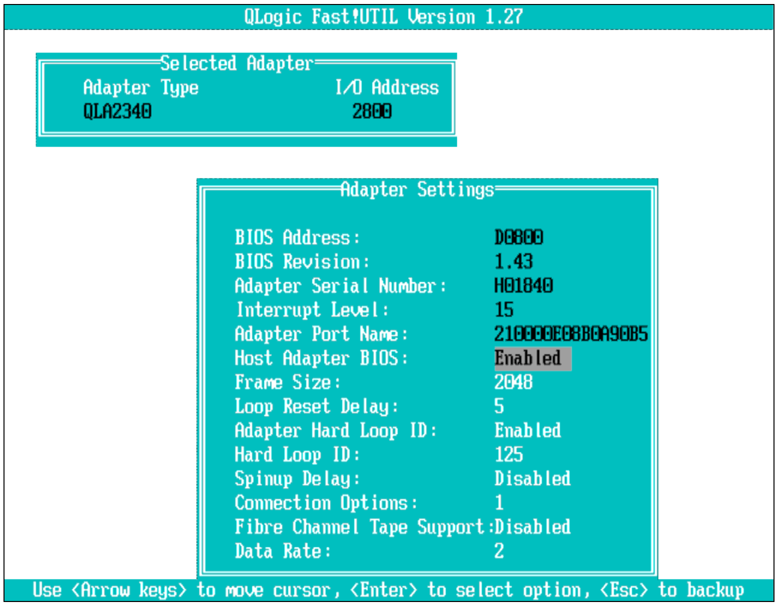
4. In the next panel (see Figure 1-11), select **Adapter Settings**.



```
QLogic Fast!UTIL Version 1.27

Selected Adapter
Adapter Type          I/O Address
QLA2340                  2800

Configuration Settings
Adapter Settings
Selectable Boot Settings
Restore Default Settings
Raw Nvram Data
Advanced Adapter Settings

Use <Arrow keys> to move cursor, <Enter> to select option, <Esc> to backup
```

Figure 1-11 Configuration Settings menu

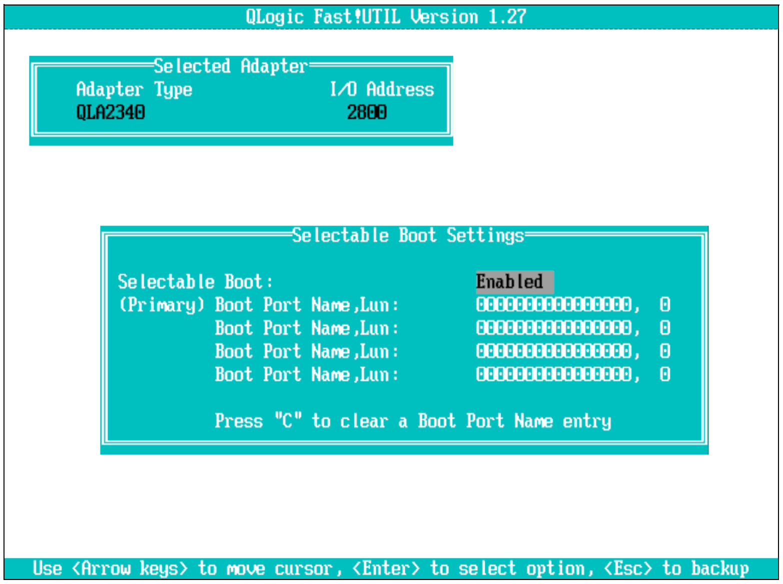
5. The Adapter Settings menu is displayed (see Figure 1-12). Change the Host Adapter BIOS setting to **Enabled**, and then press Esc to exit and go back to the Configuration Settings menu shown in Figure 1-11 on page 18.



```
                    QLogic Fast!UTIL Version 1.27

     Selected Adapter
     Adapter Type              I/O Address
     QLA2340                     2800

                 Adapter Settings

           BIOS Address:                 D0800
           BIOS Revision:                1.43
           Adapter Serial Number:        H01840
           Interrupt Level:              15
           Adapter Port Name:            210000E08B0A90B5
           Host Adapter BIOS:            Enabled
           Frame Size:                   2048
           Loop Reset Delay:             5
           Adapter Hard Loop ID:         Enabled
           Hard Loop ID:                 125
           Spinup Delay:                 Disabled
           Connection Options:           1
           Fibre Channel Tape Support:Disabled
           Data Rate:                    2

 Use <Arrow keys> to move cursor, <Enter> to select option, <Esc> to backup
```

Figure 1-12 Adapter Settings menu

6. From the Configuration Settings menu, select **Selectable Boot Settings** to get to the panel shown in Figure 1-13.



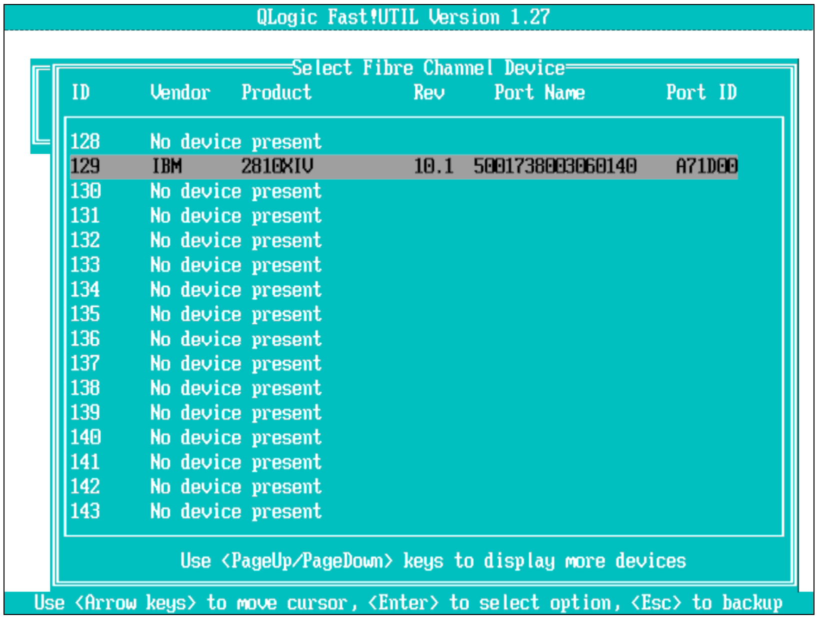
```
                    QLogic Fast!UTIL Version 1.27

     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

 Use <Arrow keys> to move cursor, <Enter> to select option, <Esc> to backup
```

Figure 1-13 Selectable Boot Settings menu

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

8. Select **Boot Port Name, Lun** and then press Enter.

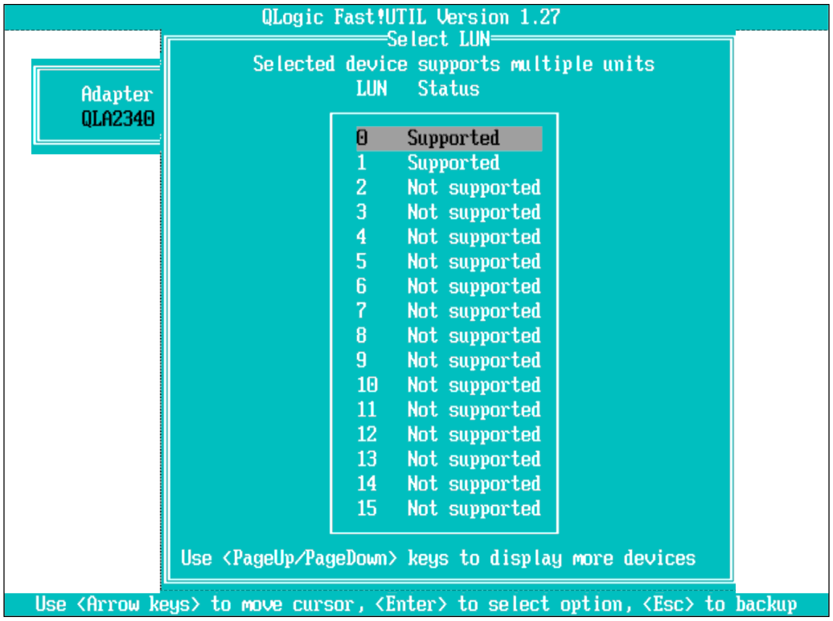
9. The Select Fibre Channel Device menu opens (see Figure 1-14). Select the **IBM 2810XIV** device, and press Enter.



```
                    QLogic Fast!UTIL Version 1.27

                    Select Fibre Channel Device
 ID      Vendor   Product          Rev      Port Name          Port ID

 128     No device present
 129     IBM      2810XIV          10.1  5001738003060140     A71D00
 130     No device present
 131     No device present
 132     No device present
 133     No device present
 134     No device present
 135     No device present
 136     No device present
 137     No device present
 138     No device present
 139     No device present
 140     No device present
 141     No device present
 142     No device present
 143     No device present

         Use <PageUp/PageDown> keys to display more devices
 Use <Arrow keys> to move cursor, <Enter> to select option, <Esc> to backup
```

Figure 1-14   Select Fibre Channel Device menu

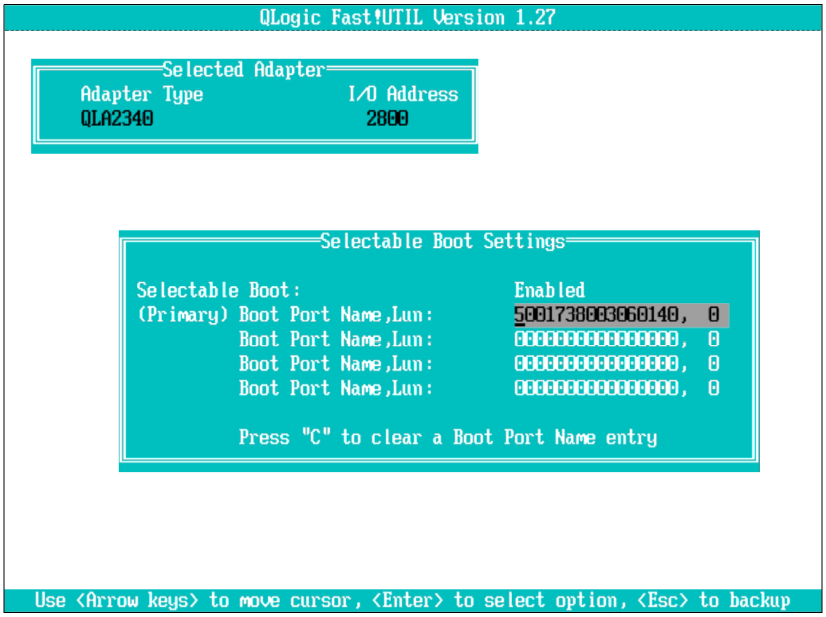
10. The Select LUN menu opens (see Figure 1-15). Select the boot LUN (in this example, LUN **0**).



```
                    QLogic Fast!UTIL Version 1.27
                             Select LUN
              Selected device supports multiple units
                            LUN   Status
 Adapter
 QLA2340
                         0    Supported
                         1    Supported
                         2    Not supported
                         3    Not supported
                         4    Not supported
                         5    Not supported
                         6    Not supported
                         7    Not supported
                         8    Not supported
                         9    Not supported
                         10   Not supported
                         11   Not supported
                         12   Not supported
                         13   Not supported
                         14   Not supported
                         15   Not supported

        Use <PageUp/PageDown> keys to display more devices
 Use <Arrow keys> to move cursor, <Enter> to select option, <Esc> to backup
```

Figure 1-15   Select LUN menu

You are returned to the Selectable Boot Setting menu, and the boot port with the boot LUN is displayed (see Figure 1-16).

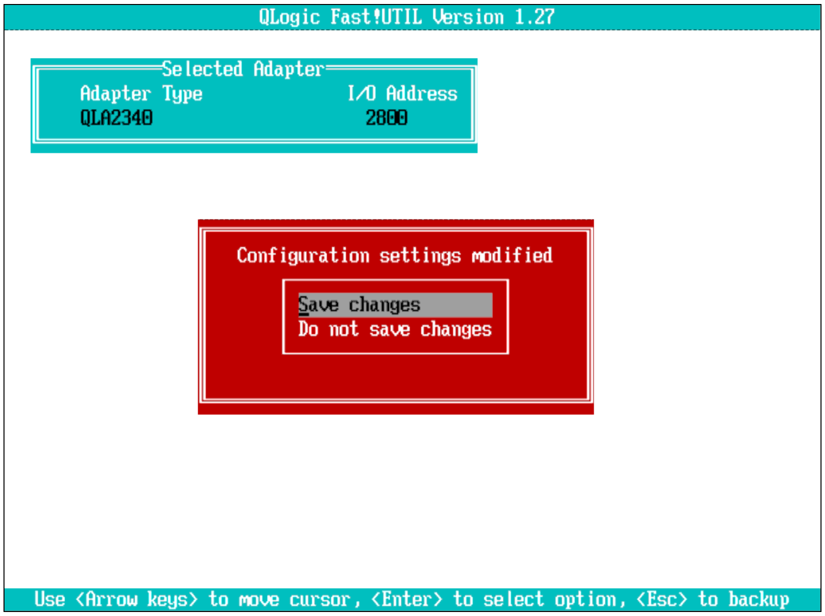


```
QLogic Fast!UTIL Version 1.27

Selected Adapter
Adapter Type                I/O Address
QLA2340                        2800

Selectable Boot Settings

Selectable Boot:                        Enabled
(Primary) Boot Port Name,Lun:           5001738003060140,  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

Use <Arrow keys> to move cursor, <Enter> to select option, <Esc> to backup
```

Figure 1-16 Boot port selected

11. Repeat the steps 8 on page 19 - 10 on page 20 to add controllers. Any extra controllers must be zoned so that they point to the same boot LUN.

12. After all the controllers are added, press Esc to exit the Configuration Setting panel. Press Esc again to get the **Save changes** option (see Figure 1-17).

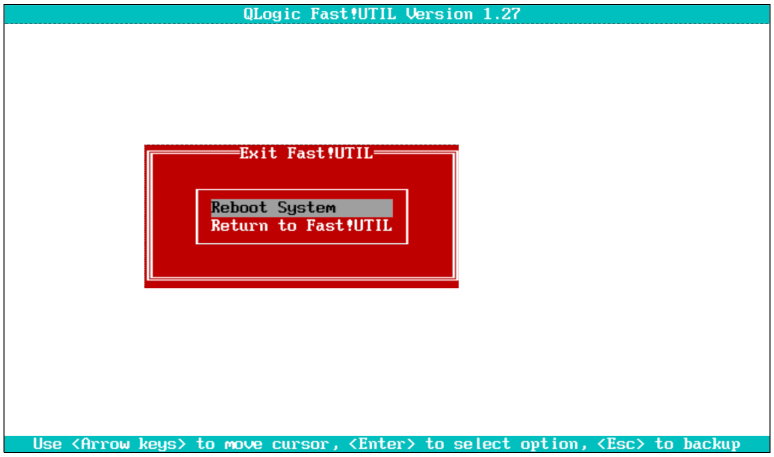


```
QLogic Fast!UTIL Version 1.27

Selected Adapter
Adapter Type                I/O Address
QLA2340                        2800

Configuration settings modified

Save changes
Do not save changes

Use <Arrow keys> to move cursor, <Enter> to select option, <Esc> to backup
```

Figure 1-17 Save changes

13. Select **Save changes** to return to the Fast!UTIL option panel. From there, select **Exit Fast!UTIL**.

14. The **Exit Fast!UTIL** menu is displayed (see Figure 1-18). Select **Reboot System** to reboot from the newly configured SAN drive.



Use <Arrow keys> to move cursor, <Enter> to select option, <Esc> to backup

Figure 1-18 Exit Fast!UTIL

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

1.3 iSCSI connectivity

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

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

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

1.3.1 Preparation steps

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

1. Connect the host to the XIV over iSCSI using a standard Ethernet port on the host server. Dedicate the port that you choose to iSCSI storage traffic only. This port must also be capable of a minimum of 1 Gbps. This port requires an IP address, subnet mask, and gateway.

   Also, review any documentation that came with your operating system about iSCSI to ensure that any additional conditions are met.

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

3. Check the optimum number of paths that must be defined, which helps determine the number of physical connections that must be made.

4. Install the latest supported adapter firmware and driver. If the latest version was not included with your operating system, download it.
5. Maximum transmission unit (MTU) configuration is required if your network supports an MTU that is larger than the default (1500 bytes). Anything larger is known as a *jumbo frame*. Specify the largest possible MTU.
6. Any device that uses iSCSI requires an iSCSI qualified name (IQN) and an attached host. The IQN uniquely identifies iSCSI devices. The IQN for the XIV Storage System is configured when the system is delivered and must not be changed. Contact IBM technical support if a change is required.

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

1.3.2 iSCSI configurations

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

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

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

Redundant configurations

A redundant configuration is shown in Figure 1-19 on page 24.

This configuration has the following characteristics:

- Each host is equipped with dual Ethernet interfaces. Each interface (or interface port) is connected to one of two Ethernet switches.
- Each of the Ethernet switches has a connection to a separate iSCSI port. The connection is to modules 4 - 9 on an XIV Gen 3.

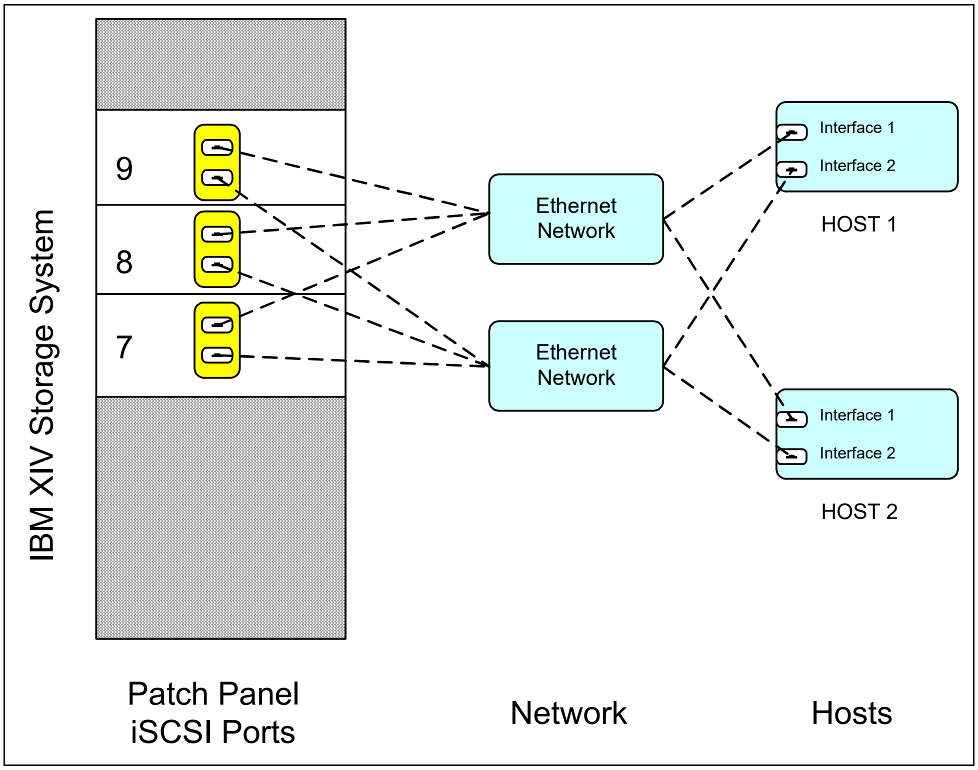


Figure 1-19 iSCSI redundant configuration using 2nd Generation XIV model A14 hardware

This configuration has no single point of failure. Consider the following points:

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

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

Non-redundant configurations

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

A non-redundant configuration is shown in Figure 1-20.

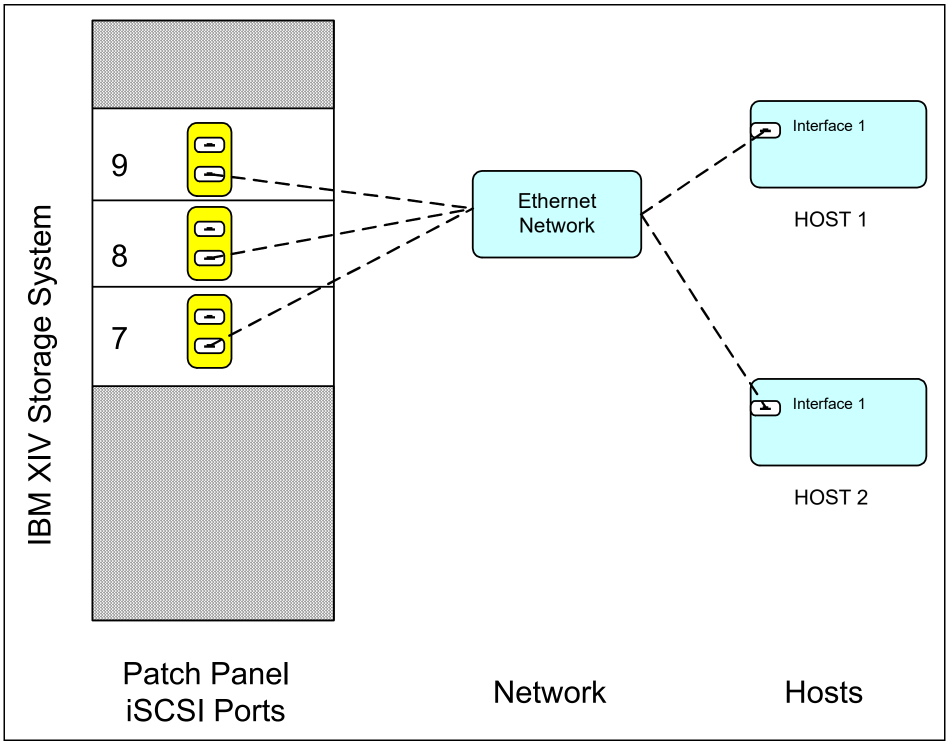


Figure 1-20 iSCSI single network switch configuration

> **Consideration:** Figure 1-19 on page 24 and Figure 1-20 show a second generation XIV (model A14). An XIV Gen 3 has more iSCSI ports on more modules.

1.3.3 Network configuration

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

1.3.4 IBM XIV Storage System iSCSI setup

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

Getting the XIV iSCSI Qualified Name

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 you need to change the IQN, you must engage IBM support.

To display the IQN of the storage system, complete the following steps:

1. From the Hyper-Scale Manager GUI, click the **Systems & Domains Views** icon and select **All Systems** from the menu, as shown in Figure 1-21.

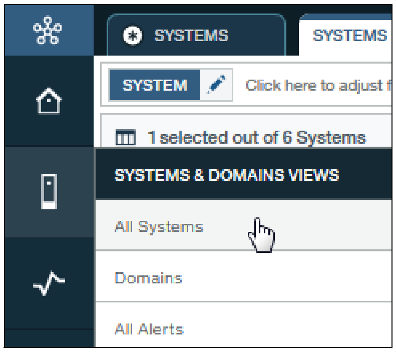
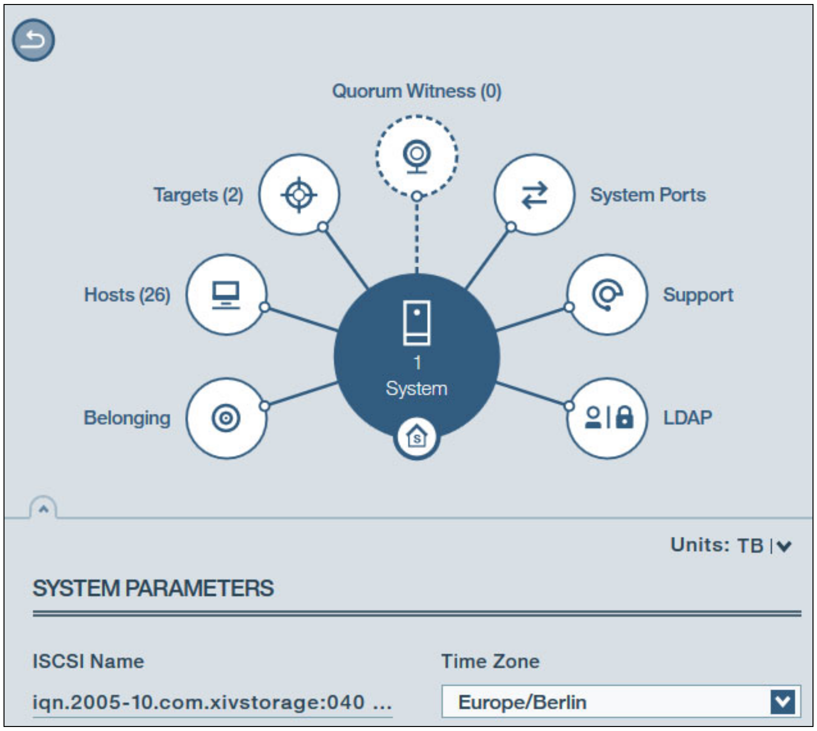


Figure 1-21 All Systems menu

2. The list of systems that are in the inventory is displayed in the Systems & Domains view. Select the system for which you need to retrieve the IQN.
3. Click in the circle that represents the system (see Figure 1-22) to display the System Properties. Scroll down in the System Properties to find the iSCSI name under System Parameters.



SYSTEM PARAMETERS

ISCSI Name: iqn.2005-10.com.xivstorage:040 ...

Time Zone: Europe/Berlin

Figure 1-22 Retrieving the iSCSI name (IQN)

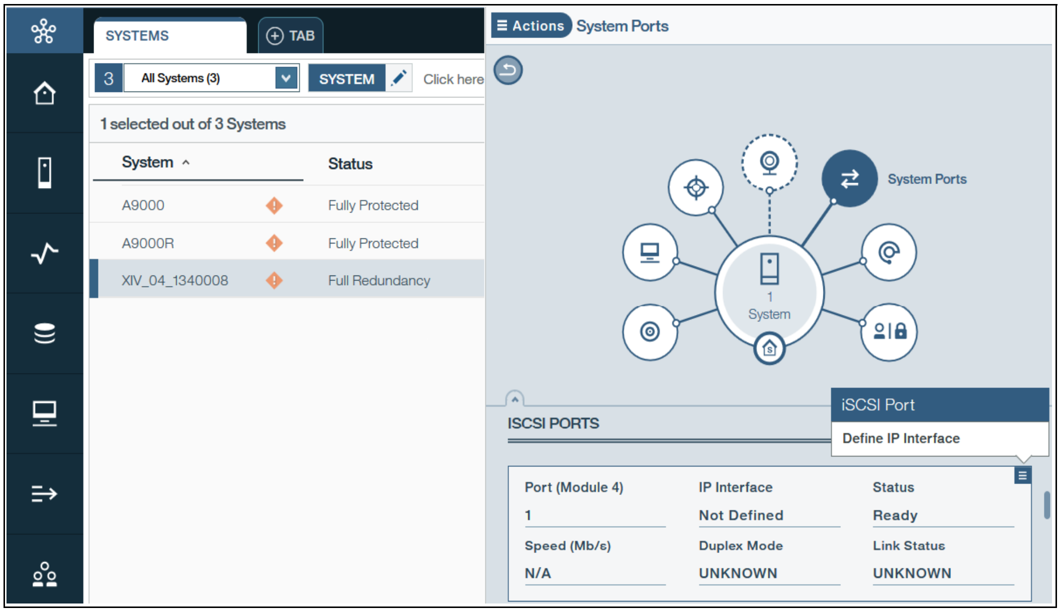
To show the same information in the XCLI, run **config_get** command (see Example 1-3).

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

```
>> config_get
Name                                    Value
dns_primary                             9.155.50.85
dns_secondary
system_name                             XIV_04_1340008
snmp_location                           Unknown
snmp_contact                            Unknown
snmp_community                          XIV
snmp_trap_community                     XIV
snmp_type                               V2C
snmpv3_user
snmpv3_encryption_type                  AES
snmpv3_encryption_passphrase            ****
snmpv3_authentication_type              SHA
snmpv3_authentication_passphrase        ****
system_id                               40008
machine_type                            2810
machine_model                           214
machine_serial_number                   1340008
email_sender_address                    admin@ibm.com
email_reply_to_address
email_subject_format                    {severity}: {description}
iscsi_name                              iqn.2005-10.com.xivstorage:040008
ntp_server                              9.155.112.20
support_center_port_type                Management
isns_server
ipv6_state                              enabled
ipsec_state                             disabled
ipsec_track_tunnels                     no
impending_power_loss_detection_method   UPS
```

Configuring the iSCSI port by using the GUI

To set up the iSCSI port by using the GUI, see Figure 1-23 and complete the following steps:



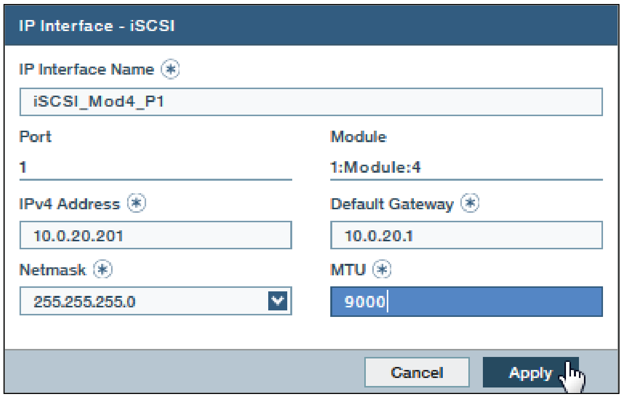
1 selected out of 3 Systems

System	Status
A9000	Fully Protected
A9000R	Fully Protected
XIV_04_1340008	Full Redundancy

Port (Module 4)	IP Interface	Status
1	Not Defined	Ready
Speed (Mb/s)	**Duplex Mode**	**Link Status**
N/A	UNKNOWN	UNKNOWN

Figure 1-23 Retrieving the IP port properties

1. From the **Systems & Domains** view, select the system of interest.
2. From the Hub view of the system, select the **System Ports** spoke.
3. The ports and modules are listed at the bottom of the panel. Scroll down to reach the iSCSI ports.
4. Click the **Actions** icon for a particular port, and select **Define IP Interface**.
5. The IP Interface window opens where you can define iSCSI ports settings (see Figure 1-24).



IP Interface - iSCSI

IP Interface Name: iSCSI_Mod4_P1

Port: 1

Module: 1:Module:4

IPv4 Address: 10.0.20.201

Default Gateway: 10.0.20.1

Netmask: 255.255.255.0

MTU: 9000

Figure 1-24 Setting the iSCSI port properties

Enter the following information in the IP Interface window:

– Name: Define the name for this interface.

– Address, netmask, and gateway: Enter the standard IP address details.

   – MTU: All devices in a network must use the same MTU. If in doubt, set MTU to `1500` because 1500 is the default value for Gigabit Ethernet. Performance might be affected if the MTU is set incorrectly.

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

> **Tip:** If the MTU that is used by the XIV is higher than the network can transmit, the frames are discarded. The frames are discarded because the do-not-fragment bit is normally set to `on`. Use the **`ping -l`** command to test and specify packet payload size from a Windows workstation in the same subnet. A **`ping`** command normally contains 28 bytes of IP and ICMP headers plus payload. Add the **`-f`** parameter to prevent packet fragmentation.
>
> For example, the **`ping -f -l 1472 10.1.1.1`** command sends a 1500-byte frame to the 10.1.1.1 IP address (1472 bytes of payload and 28 bytes of headers). If this command succeeds, you can use an MTU of 1500 in the XIV GUI or XCLI.

iSCSI XIV port configuration by using the XCLI

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

Example 1-4   iSCSI setup (XCLI)

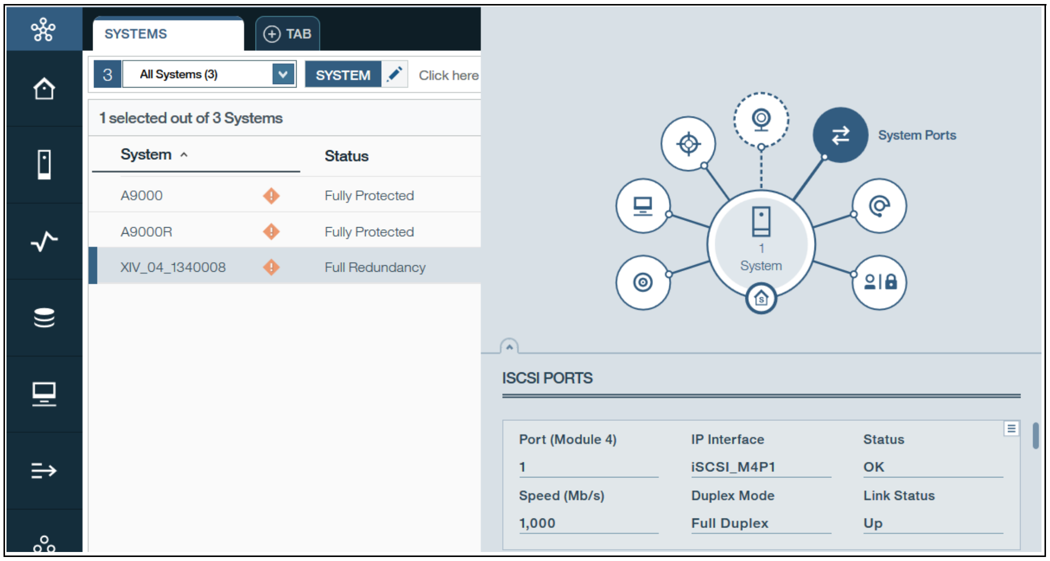
```
>> ipinterface_create ipinterface="Test" address=10.0.0.10 netmask=255.255.255.0
module=1:Module:5 ports="1" gateway=10.0.0.1 mtu=9000
```

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 the iSCSI configuration by using the GUI

To view the iSCSI port by using the GUI, see Figure 1-25 and complete the following steps:



1 selected out of 3 Systems

System		Status
A9000	◆	Fully Protected
A9000R	◆	Fully Protected
XIV_04_1340008	◆	Full Redundancy

ISCSI PORTS

Port (Module 4)	IP Interface	Status
1	iSCSI_M4P1	OK
Speed (Mb/s)	**Duplex Mode**	**Link Status**
1,000	Full Duplex	Up

Figure 1-25   Retrieving the IP port properties

1. From the **Systems & Domains** view, select the system of interest.
2. From the Hub view of the system, select the **System Ports** spoke.
3. The ports and modules are listed at the bottom of the panel. Scroll down to reach the iSCSI ports.

Viewing the iSCSI configuration by using the XCLI

The **ipinterface_list** command that is shown in Example 1-5 can be used to display configured network ports only. The output is truncated to show only the iSCSI connections that are of interest here. The command also displays all other Ethernet connections and settings. For clarity, several columns are removed from the example. This example shows a XIV Gen3 (model 114).

Example 1-5 Listing iSCSI ports with the ipinterface_list command

```
>> ipinterface_list
Name         Type         IP Address       Network Mask     Default Gateway     MTU    Module         Ports
iSCSI_M5P1   iSCSI        9.155.115.191    255.255.240.0    9.155.112.1         9000   1:Module:5     1
iSCSI_M6P1   iSCSI        9.155.115.192    255.255.240.0    9.155.112.1         9000   1:Module:6     1
iSCSI_M7P1   iSCSI        9.155.115.193    255.255.240.0    9.155.112.1         9000   1:Module:7     1
iSCSI_M8P1   iSCSI        9.155.115.194    255.255.240.0    9.155.112.1         9000   1:Module:8     1
```

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

1.3.6 iSCSI and CHAP authentication

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

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

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

For the iSCSI initiator to log in with CHAP, both the `iscsi_chap_name` and `iscsi_chap_secret` parameters must be set. After both of these parameters are set, the host can run an iSCSI login to the IBM XIV Storage System only if the login information is correct.

CHAP name and secret parameter guidelines

The following guidelines apply to the CHAP name and secret parameters:

- Both the `iscsi_chap_name` and `iscsi_chap_secret` parameters must be either specified or not specified. You cannot specify just one of them.
- The `iscsi_chap_name` and `iscsi_chap_secret` parameters must be unique. If they are not unique, an error message is displayed, although the command does not fail.
- The secret must be 96 - 128 bits. You can use one of the following methods to enter the secret:
  - Base64 requires that `0b` is used as a prefix for the entry. Each subsequent character that is entered is treated as a 6-bit equivalent length.

  - Hex requires that 0x is used as a prefix for the entry. Each subsequent character that is entered is treated as a 4-bit equivalent length.
  - String requires that a prefix is not used (it cannot be prefixed with 0b or 0x). Each character that is entered is treated as an 8-bit equivalent length.
- If the iscsi_chap_secret parameter does not conform to the required secret length (96 - 128 bits), the command fails.
- If you change the iscsi_chap_name or iscsi_chap_secret parameters, a warning message is displayed. The message says that the changes will apply the next time that the host is connected.

Configuring CHAP

CHAP can be configured by using XCLI commands:

- 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 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 this writing, you cannot start through iSCSI, even if an iSCSI HBA is used. For up-to-date information, see the SSIC web page:

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

1.4 Logical configuration for host connectivity

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

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 boot from SAN, install the operating system.

11. Install multipath drivers, if required. For more information about installing multipath drivers, see the appropriate section from the host-specific chapters of this book.
12. Restart the host server or scan for new disks.

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

1.4.1 Host configuration preparation

The environment that is shown in Figure 1-26 on page 33 is used to show the configuration tasks.

The example uses a second generation XIV. The following hosts are available:

- One host uses FC connectivity
- One host uses iSCSI

Figure 1-26 on page 33 also shows the unique names of components that are used in the configuration steps.

The following assumptions are made for the scenario that is shown in Figure 1-26 on page 33:

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

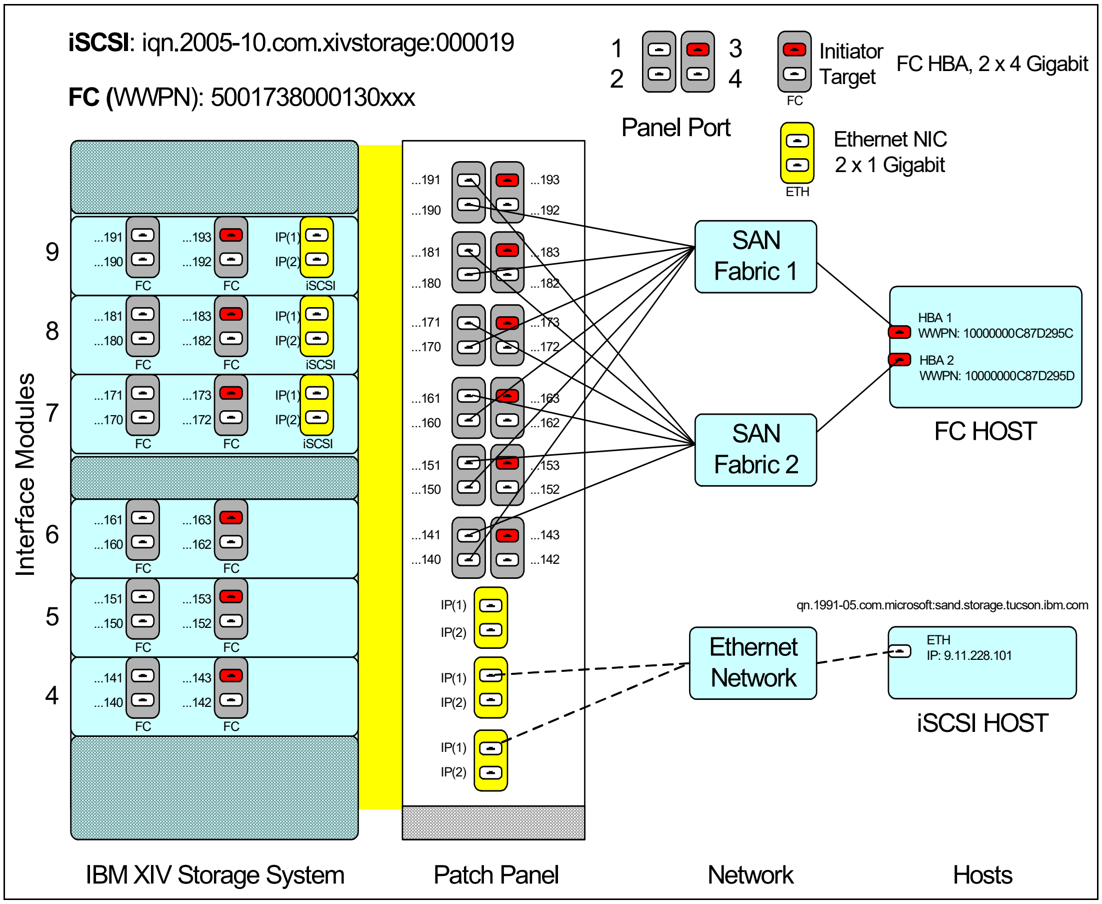


Figure 1-26 Overview of base host connectivity setup

Hardware information

Write down the component names and IDs because doing so saves time during the implementation. An example is listed in Table 1-3 for the sample scenario.

Table 1-3 Required component information

Component	FC environment	iSCSI environment
IBM XIV FC HBAs	WWPN: 5001738000130*nnn* ▸ *nnn* for Fabric1: 140, 150, 160, 170, 180, and 190 ▸ *nnn* for Fabric2: 142, 152, 162, 172, 182, and 192	N/A
Host HBAs	▸ HBA1 WWPN: 21000024FF24A426 ▸ HBA2 WWPN: 21000024FF24A427	N/A
IBM XIV iSCSI IPs	N/A	▸ Module7 Port1: 9.11.237.155 ▸ Module8 Port1: 9.11.237.156
IBM XIV iSCSI IQN *(do not change)*	N/A	iqn.2005-10.com.xivstorage:000019
Host IPs	N/A	9.11.228.101

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

> **Remember:** The OS Type is *default* for all hosts, except Windows 2008, HP-UX, and IBM z/VM®.

FC host-specific tasks

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

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

Fabric 1 HBA 1 zone

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

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

Fabric 2 HBA 2 zone

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

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

In the previous examples, the following aliases are used:

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

iSCSI host-specific tasks

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

1.4.2 Assigning LUNs to a host by using the GUI

Several steps are necessary to define a new host and assign LUNs to it. You must create the volumes in a storage system in advance.

Defining a host

To define a host, complete the following steps:

1. From the main GUI dashboard, click the **New** icon and select **Host** from the menu (see Figure 1-27).

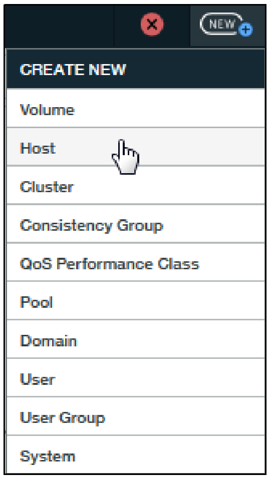
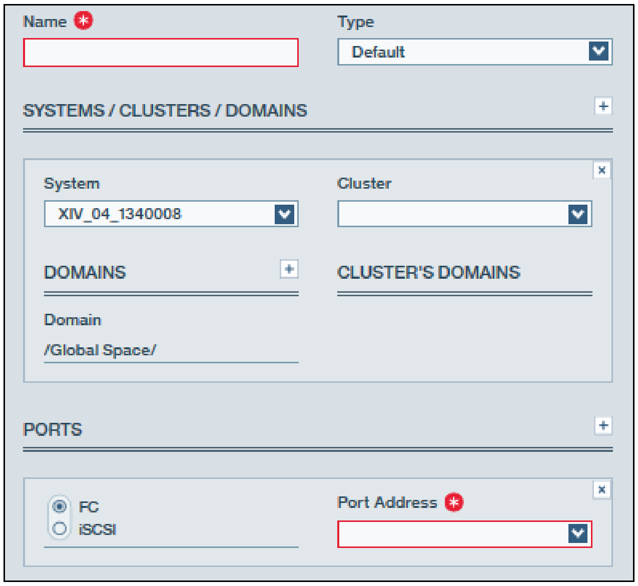


Figure 1-27 The Create New (hosts and clusters) menu

2. The new Host view is displayed. The Add Host settings window opens (see Figure 1-28).



Type: Default

System: XIV_04_1340008

Domain: /Global Space/

● FC
○ iSCSI

Figure 1-28 Add Host details

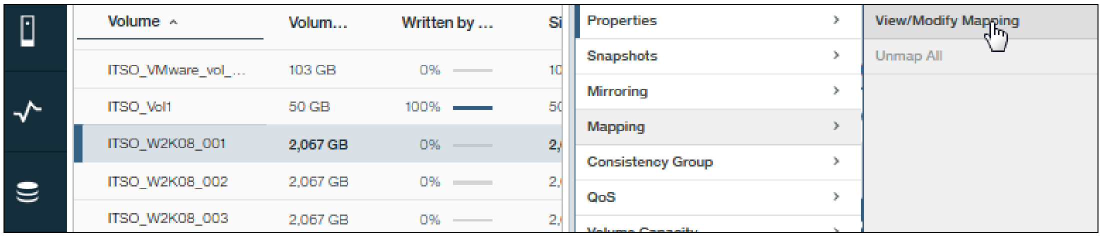
3. Enter the required information:
   - Specify a host name (required).
   - Select the type. In this example, **Default** is selected. If you use an HP-UX or a z/VM host, you must change the type to match your host type. For all other hosts, such as AIX, Linux, Solaris, VMware, and Windows, the Default option is correct. Starting with HAK version 2.7.0, HP-UX and Solaris are no longer supported.
   - Select whether you want an iSCSI or FC connection. Host access to LUNs is granted depending on the host adapter ID. For an FC connection, the host adapter ID is the FC HBA WWPN. For an iSCSI connection, the host adapter ID is the host IQN. To add a WWPN or an IQN to a host definition, click the plus sign (**+**) icon to the right of the PORTS heading (see Figure 1-28 on page 35).

     Ports can be added in any order.
   - Specify to which system in your inventory you want to attach the host. Alternatively, you can also add the host to a cluster. (A cluster can be created from this view.)
   - Optional: Add the host to a domain. (A domain can also be created from this view.)

Mapping LUNs to a host

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

1. From the Pools and Volumes Workspace view, select the volume to map, and select **Actions** →**Mapping** →**View/Modify Mapping** (see Figure 1-29).



Volume ^	Volum...	Written by ...	Si
ITSO_VMware_vol_...	103 GB	0%	10
ITSO_Vol1	50 GB	100%	50
ITSO_W2K08_001	2,067 GB	0%	2,
ITSO_W2K08_002	2,067 GB	0%	2,
ITSO_W2K08_003	2,067 GB	0%	2,

Figure 1-29 Mapping a LUN to a host

2. The Volume Mapping panel opens (see Figure 1-30). If no mapping exists yet, click the plus sign (**+**) to add a mapping.

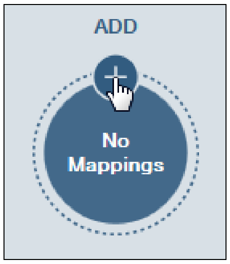
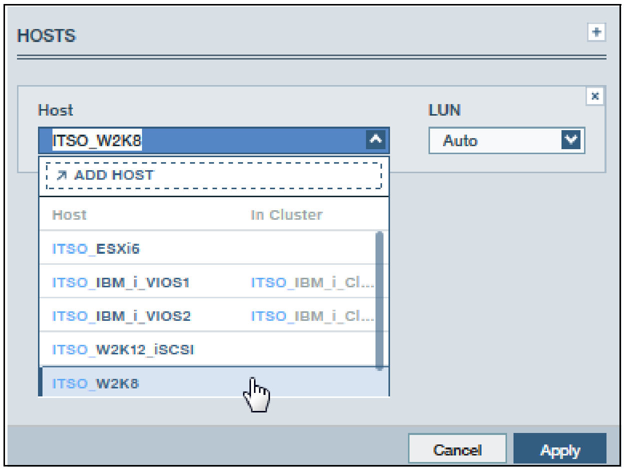


Figure 1-30 Add a mapping

3. Select an available host from the pull-down list. The GUI suggests a default (Auto) LUN ID to which to map the volume. However, this ID can be changed to meet your requirements. Click **Apply** and the volume is assigned immediately (see Figure 1-31 on page 37).



HOSTS

Host: ITSO_W2K8

LUN: Auto

Host	In Cluster
ITSO_ESXi6	
ITSO_IBM_i_VIOS1	ITSO_IBM_i_Cl...
ITSO_IBM_i_VIOS2	ITSO_IBM_i_Cl...
ITSO_W2K12_iSCSI	
ITSO_W2K8	

Figure 1-31 Mapping a volume to a host

No differences exist between mapping a volume to an FC host or an iSCSI host in the GUI.

1.4.3 Assigning LUNs to a host by using the XCLI

Several steps are necessary to define a new host and assign LUNs to it. Volumes must already be created in a storage pool.

Defining a new host

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

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

   Example 1-6 Creating host definition (XCLI)

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

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

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

   As Example 1-7 shows, the WWPN of the FC host for HBA1 and HBA2 is added with the **host_add_port** command by specifying an **fcaddress** value.

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

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

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

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

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

```
>> host_add_port host=itso_win2008_iscsi
iscsi_name=iqn.1991-05.com.microsoft:sand.storage.tucson.ibm.com
Command executed successfully
```

Mapping LUNs to a host

To map the LUNs, complete these steps:

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

   Example 1-9 Mapping volumes to hosts (XCLI)

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

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

   >> map_vol host=itso_win2008_iscsi vol=itso_win2008_vol3 lun=1
   Command executed successfully.
   ```

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

   Example 1-10 Checking host connectivity (XCLI)

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

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

   This example shows two paths per host FC HBA and two paths for the single Ethernet port that was configured.

At this stage, you might need to do operating system-dependent steps; these steps are described in the operating system chapters.

1.5 Performance tuning

This section provides some performance considerations to help you adjust your operating system to best fit your environment. The following are performance considerations for the operating system:

- Use multiple threads and asynchronous I/O to maximize performance on XIV.
- Use **iostat**, **perfmon**, or **esxtop** to check on a per path basis for the LUNs to make sure that the load is balanced across all paths.
- Verify the HBA queue depth and per device queue depth for the host are sufficient to prevent queue waits. However, make sure that they are not so large that they overrun the Storage System queues. For XIV queue limit is 1400 per XIV port and 256 per LUN per WWPN (host) per port. Do not submit more I/O per XIV port than the 1400 maximum it can handle.

Check the device queue depth. Consider the following points:

- To check the device queue depth on AIX, periodically run **iostat -D 5**. If avgwqsz (average wait queue size) or sqfull is consistently greater than zero, increase the device queue depth.

  For more information about AIX device queue depth tuning, see the following web page:

  http://www-01.ibm.com/support/docview.wss?uid=tss1td105745

- To check the device queue depth on Linux, periodically run **iostat -dx 5**. If the await-svctm values are consistently greater than zero or %util (disk utilization) is constantly close to 100%, increase the device queue depth.

  For more information about Linux disk tuning, see the following web page:

  http://cromwell-intl.com/linux/performance-tuning/disks.html

- To check the device queue depth on VMware ESXi, run **esxtop**, press the U key. The queue depth is listed under LQLEN. If the %USD (percentage of queue depth used) is constantly close to 100%, increase the device queue depth.

  For more information about VMware ESXi queues, see the following web page:

  http://blogs.vmware.com/apps/2015/07/queues-queues-queues-2.html

- To check the device queue depth on Windows, run **Administrative Tools →Performance Monitor** and select **Physical Disk →Current Disk Queue Length**.

  For more information about Windows disk performance monitoring, see the following web page:

  https://blogs.technet.microsoft.com/askcore/2012/03/16/windows-performance-monitor-disk-counters-explained/

Important: Do not start at the maximum and work down (except for Windows which starts by default with maximum) because you might flood the storage system with commands and waste memory on the host.

1.6 Troubleshooting

Troubleshooting connectivity problems can be difficult. However, the XIV Storage System includes built-in troubleshooting tools, including the tool that are listed in Table 1-4.

Table 1-4 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.

For more information, see *IBM XIV Storage System Command-Line Interface (CLI) Reference Guide*:

https://www.ibm.com/support/knowledgecenter/en/STJTAG/com.ibm.help.xivgen3.doc/xiv_pubsrelatedinfoic.dita

2

IBM FlashSystem A9000 and A9000R host connectivity

This chapter describes host connectivity for IBM FlashSystem A9000 and A9000R. It highlights key aspects of host connectivity. It also reviews concepts and requirements for the Fibre Channel (FC) and internet Small Computer System Interface (iSCSI) protocols.

This chapter covers common tasks that relate to most hosts. For more information, see the IBM Storage Host Attachment Kit publications at Fix Central:

https://ibm.co/2FAgkEp

This chapter includes the following topics:

2.1 Overview

FlashSystem A9000 and A9000R can be attached to various host systems by using the following methods:

- FC adapters that use Fibre Channel Protocol (FCP)

  Newer systems feature grid controllers that are equipped with FC-NVMe adapters. In these new controllers, the FC ports are dual-purpose and NVMe ready. A future software upgrade will enable these ports to connect to servers by using FC, FC-NVMe, or both.
- iSCSI software initiator or iSCSI host bus adapter (HBA) that uses the iSCSI protocol

FlashSystem A9000 and A9000R are perfectly suited for integration into a new or existing FC storage area network (SAN). After the HBAs, cabling, and SAN zoning are in place, connecting an FC host to FlashSystem A9000 or A9000R is easy. By using the IBM Hyper-Scale Manager graphical user interface (GUI) or the extended command-line interface (XCLI), the storage administrator defines hosts and ports, and then maps volumes to them.

You can also connect hosts to FlashSystem A9000 or A9000R through iSCSI by using an existing 10 Gb Ethernet network infrastructure. However, your workload might require a dedicated network to sustain the additional data traffic.

Grid controllers provide the interface node functionality to connect FlashSystem A9000 or A9000R to a host through FC and iSCSI (Intermix Protocol feature) or iSCSI only.

Each grid controller can be equipped with four FC ports and two iSCSI ports if you order the intermix protocol option, or each grid controller can be ordered with four iSCSI ports. The possible configurations are shown in Figure 2-1.

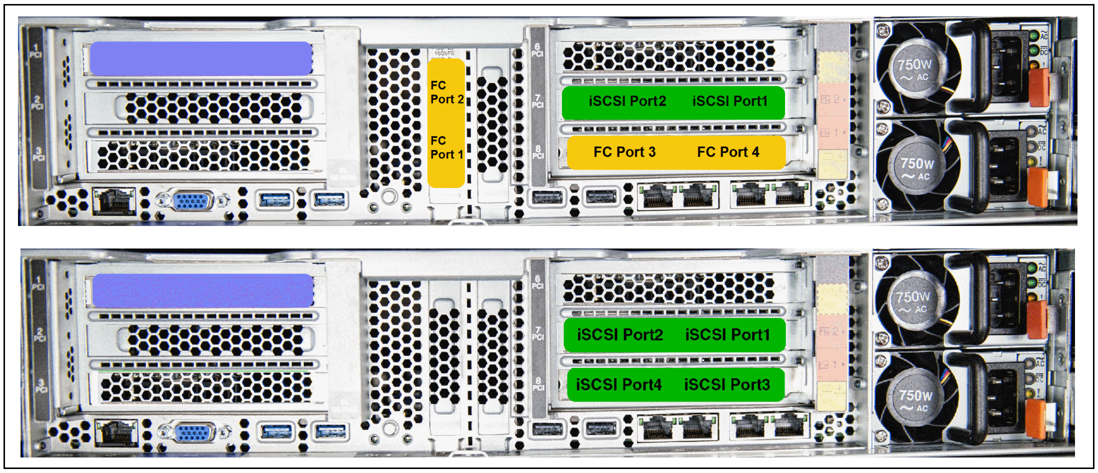


Figure 2-1 Grid controller configuration options

Important: All of the grid controllers in a single FlashSystem A9000 or A9000R must be identical.

Table 2-1 lists the number of FC ports and iSCSI ports for FlashSystem A9000 according to the system configuration.

Note: FlashSystem A9000 always includes grid controllers.

Table 2-1 FlashSystem A9000 attachment ports

FlashSystem A9000	Three grid controllers
FC ports with iSCSI ports	12 FC 6 iSCSI
iSCSI only	12 iSCSI

FlashSystem A9000R has up to 8 grid controllers for Model 425 and up to 12 grid controllers for Model 415. The Model 425 also offers a grid starter configuration with only 3 grid controllers. Table 2-2 lists the number of grid controllers for FlashSystem A9000R and the FC and iSCSI ports for different rack configurations.

Table 2-2 FlashSystem A9000R attachment ports

Number of grid controllers	Three (Model 425 Grid Starter)	Four	Six	Eight (max for Model 425)	Ten	Twelve (max for Model 415)
FC ports and iSCSI ports	12 FC 6 iSCSI	16 FC 8 iSCSI	24 FC 12 iSCSI	32 FC 16 iSCSI	40 FC 20 iSCSI	48 FC 24 iSCSI
iSCSI ports only	12	16	24	32	40	48

Hosts attach to the FC ports through an FC switch and to the iSCSI ports through a 10-Gigabit Ethernet switch. The FC ports for both FlashSystem A9000 and A9000R can negotiate FC speeds of 4 Gbps, 8 Gbps, or 16 Gbps. You must consider the switch port speed capability when you design the connectivity architecture. The negotiated FC transmission speeds can affect the overall I/O operations per second (IOPS) and bandwidth that is achievable per port. Whenever possible, use a 16 Gbps fabric and HBAs, for best performance.

Important: Direct attachment between hosts and FlashSystem A9000 or A9000R is not supported.

Host traffic can be directed to any of the grid controllers. The storage administrator must ensure that host connections avoid single points of failure and consider that A9000 and A9000R grid starter can tolerate a single grid controller failure, while A9000r with four grid controllers and above can tolerate two grid controller failures.

The server administrator also must ensure that the host workload is adequately balanced across the connections and host interface ports on the grid controllers. This balancing can be performed by installing the relevant Host Attachment Kit. Review the balancing periodically and when traffic patterns change.

With FlashSystem A9000 and A9000R, all grid controllers and all ports can be used concurrently to access any logical volume in the storage system. The only affinity is the mapping of logical volumes to host, which simplifies storage management. Balancing traffic and zoning for adequate performance and redundancy is a critical task and objective. You can connect through a SAN or a 10 Gb Ethernet network.

Figure 2-2 shows a logical connectivity example through an FC SAN to FlashSystem A9000R, which shows only ports 1 and 3, that are used for host connectivity.

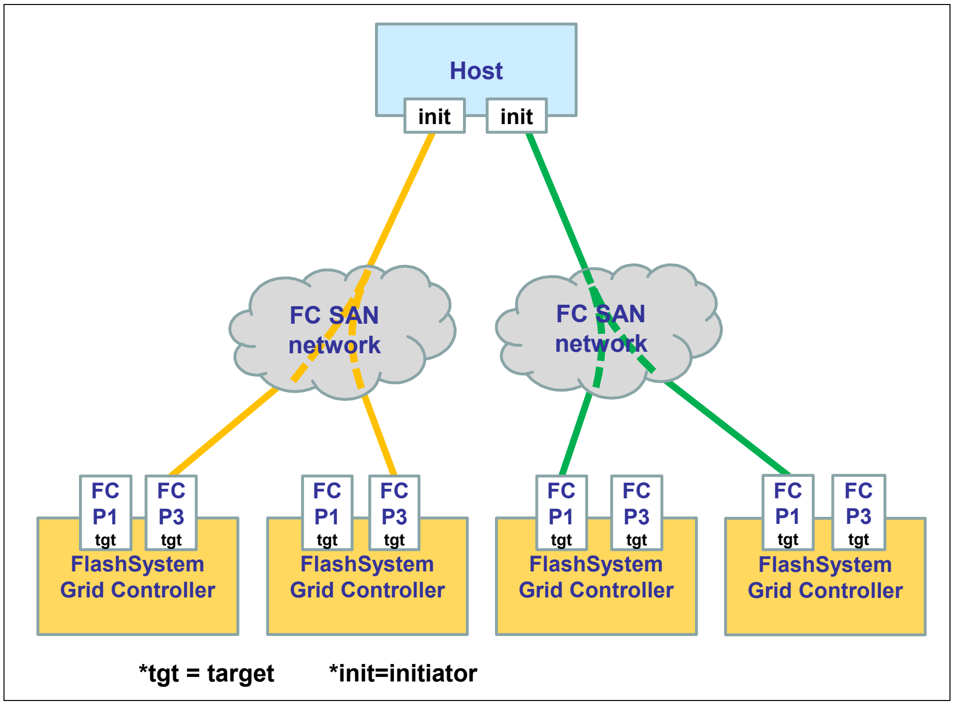


Figure 2-2 Host FC SAN connectivity overview example

Figure 2-3 shows a logical connectivity example through a 10-Gigabit Ethernet network, which shows only ports 1 and 3 that are used for host connectivity.

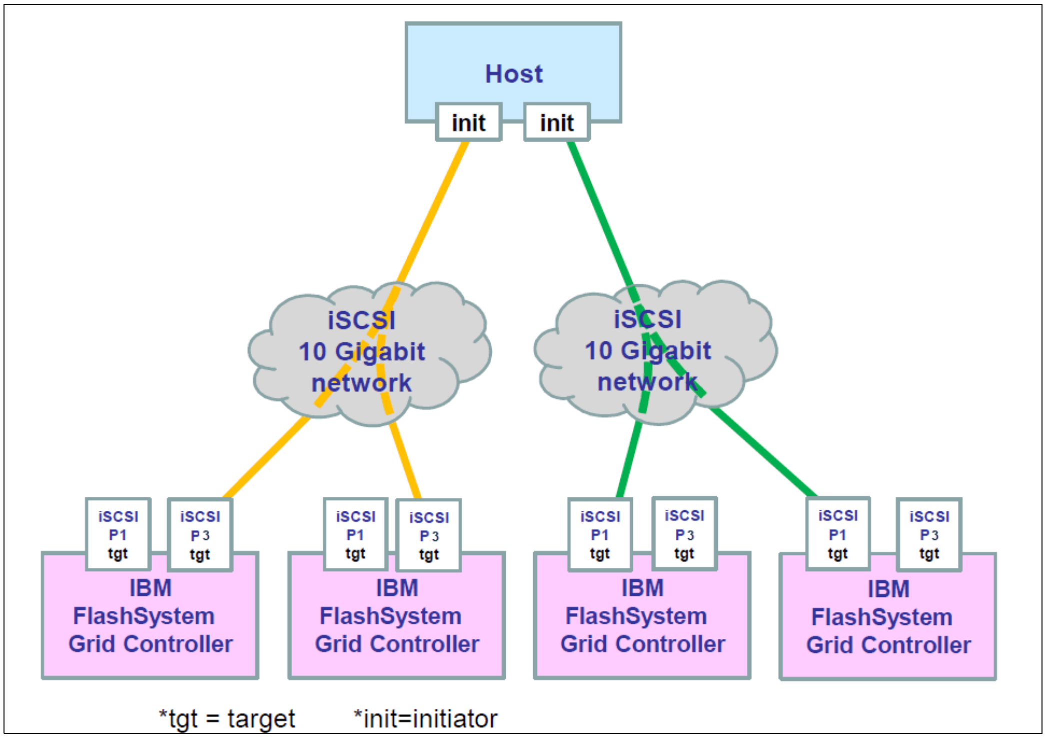


Figure 2-3 Host 10-Gigabit Ethernet connectivity overview example

2.1.1 Grid controller interface ports and host connectivity

This section presents a simplified view of the host connectivity to explain the relationship between individual system components and how they affect host connectivity.

When you connect hosts to FlashSystem A9000 or A9000R, no “one size fits all” solution can be applied because every environment differs. However, follow these guidelines to avoid single points of failure and ensure that hosts are connected to the correct ports:

- FC hosts connect to FlashSystem A9000 and A9000R FC ports 1 and 3 on grid controllers. This configuration provides increased reliability because FC ports 1 and 3 are on separate FC adapter cards (each grid controller has 2 FC adapters).
- FlashSystem A9000 and A9000R FC ports 2 and 4 can be used for mirroring to another FlashSystem A9000, A9000R, or XIV Gen3. They can also be used for data migration from another storage system.
- FC ports 2 and 4 can also be used as target ports for host attachment if they are required in a high connectivity scenario.
- FC port 4, by default, is set to the initiator role for use with the replication and data migration functions of FlashSystem A9000 and A9000R.
- FC port 4 can also be used as a target port. The port role must be changed through the XCLI or the GUI.

> **Tip:** Most figures in this book show that ports 1 and 3 are allocated for host connectivity. Ports 2 and 4 are reserved for extra host connectivity or remote mirroring and data migration connectivity. This configuration provides more resiliency because ports 1 and 3 are on separate FC adapters. This configuration provides increased availability. It might also increase performance because each adapter has its own IBM Peripheral Component Interconnect® (PCI) bus.
>
> Contact your IBM service support representative (SSR) or IBM Technical Advisor (TA) to determine the best port allocation for your environment.

- iSCSI hosts connect to at least one port on each active grid controller. The preferred practice is to use the same methodology as the FC connectivity recommendations when you connect through iSCSI with a FlashSystem A9000 or A9000R that is physically configured with the “all iSCSI adapters” option. That is, use iSCSI ports 1 and 3 for host connectivity and iSCSI ports 2 and 4 for replication. The iSCSI port numbering is shown in Figure 2-1 on page 42.
- Use every grid controller and spread connections evenly.

Figure 2-4 shows an overview of FC and iSCSI connectivity for a rack configuration that is configured with both FC and iSCSI ports.

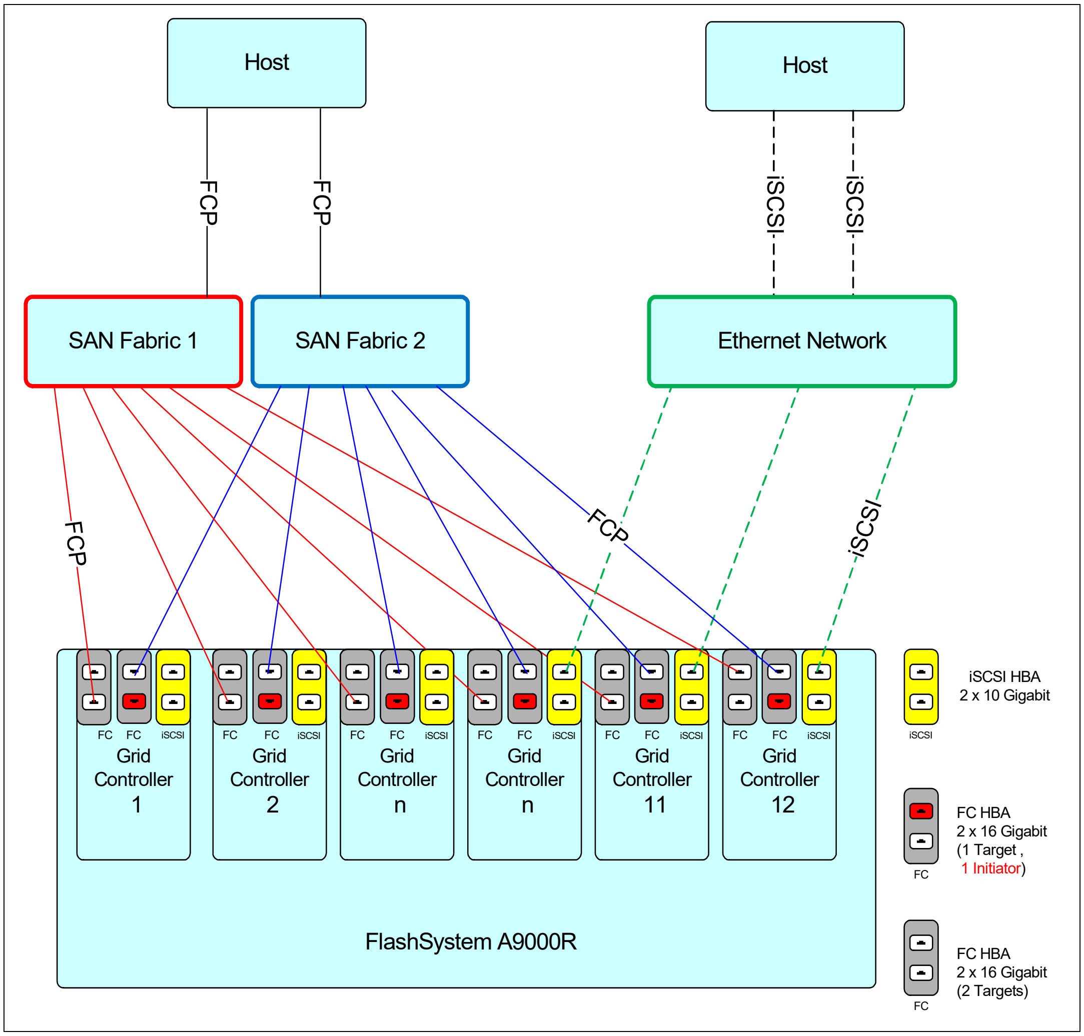


Figure 2-4 Host connectivity end-to-end view for mixed protocol configuration

2.1.2 Host operating system support

FlashSystem A9000 and A9000R support many operating systems. For more information about the current list, see the IBM System Storage Interoperation Center (SSIC):

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

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

If you cannot locate your current (or planned) combination of product versions, talk to your IBM Business Partner, IBM marketing representative, or IBM Pre-Sales Technical Support Representative.

2.1.3 Host Attachment Kits

For high availability, every host that is attached to FlashSystem A9000 or A9000R must have multiple paths to the connected system. Most operating systems, such as AIX, Windows, VMware, and Linux, can provide native multipathing. IBM has a Host Attachment Kit for most of these supported operating systems. These kits customize the host multipathing. The Host Attachment Kit also supplies powerful tools to assist the storage administrator in daily tasks.

The Host Attachment Kit includes the following features:

- Validates the host server patch and driver versions for the correct levels.
- Sets up multipathing on the host by using native multipathing.
- Adjusts select host system tunable parameters (if required) for performance.

> **Important:** One important host tunable parameter, LUN Queue Depth, is not modified. Review this parameter because you might need to increase it based on the type of workloads that are intended for the server and FlashSystem A9000 or A9000R. More about queue depth is in 2.7, “Performance tuning” on page 90.

- Provides an installation wizard (which might not be needed if you use the portable version).
- Provides management utilities, such as the **`xiv_devlist`** and **`xiv_attach`** commands.
- Provides support and troubleshooting utilities, such as the **`xiv_diag`** command.

At the time of this writing, the Host Attachment Kit (HAK) version that supports IBM A9000 and A9000R is Host Attachment Kit 2.9.x. The latest version might differ for the specific operating system.

To search for and download a Host Attachment Kit from Fix Central, see this web page:

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

Commands that are provided by the Host Attachment Kit

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

The xiv_attach command

This command locally configures the operating system and defines the host on FlashSystem A9000 or A9000R.

Sometimes, after you run the **`xiv_attach`** command, you might be prompted to reboot the host. This reboot might be needed because the command can perform system modifications that force a reboot based on the normal behavior of the operating system. For example, a reboot is required when you install a Windows hotfix. You need to run this command only one time, when you perform the initial host configuration. After the first time, use **`xiv_fc_admin -R`** or **`xiv_iscsi_admin -R`** to detect newly mapped volumes. The **`--clean`** option removes unavailable paths that were left in the multipathing configuration for the operation system.

The xiv_detach command

This command is used on a Microsoft Windows Server to remove all multipathing settings from the host. For other operating systems, use the uninstallation option.

The xiv_devlist command

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

- Size of the volume
- Number of paths (working and detected)
- Name and ID of each volume on FlashSystem A9000 or A9000R
- ID of the FlashSystem A9000 or A9000R system
- Name of the host definition on the FlashSystem A9000 or A9000R system

The **`xiv_devlist`** command is one of the most powerful tools in your toolkit. Ensure that you are familiar with this command and use it whenever you perform system administration. The *Host Attachment Kit User Guide*, GA32-1060 lists many useful parameters that can be run with the **`xiv_devlist`** command.

The following parameters are especially useful:

`xiv_devlist -u GiB` Displays the volume size in binary GB. The **`-u`** is for unit size.

`xiv_devlist -V` Displays the Host Attachment Kit version number. The **`-V`** is for version.

`xiv_devlist -f filename.csv -t csv`
Directs the output of the command to a file.

`xiv_devlist -h` Opens the help page that displays other available parameters. The **`-h`** is for help.

The xiv_diag command

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

The xiv_fc_admin command

This command is similar to **`xiv_attach`**. However, unlike **`xiv_attach`**, you can use the **`xiv_fc_admin`** command to perform individual steps and tasks.

The following parameters are useful:

`xiv_fc_admin -P` Displays the worldwide port names (WWPNs) of the host server HBAs. The **`-P`** is for print.

`xiv_fc_admin -V` Lists the tasks that **`xiv_attach`** will perform if it is run. Knowing the tasks is vital if you are using the portable version of the Host Attachment Kit. The **`-V`** is for verify.

`xiv_fc_admin -C` Performs all of the tasks that the **`xiv_fc_admin -V`** command identified as required for your operating system. The **`-C`** is for configure.

`xiv_fc_admin -R` This command scans for and configures new volumes that are mapped to the server. For a new host that is not yet connected to an XIV, use **`xiv_attach`**. However, if more volumes are mapped to this host later, use **`xiv_fc_admin -R`** to detect them. You can use native host methods but the Host Attachment Kit command is an easier way to detect volumes. The **`-R`** is for rescan.

`xiv_fc_admin -R --clean`
Use the **`clean`** option to remove devices from the multipath. You can clean unreachable devices (use only with the **`-R`**/**`--rescan`** option).

`xiv_fc_admin -h` Opens the help page that displays other available parameters. The **`-h`** is for help.

The xiv_iscsi_admin command

This command is similar to **`xiv_fc_admin`**, but this command is used on hosts with iSCSI interfaces rather than FC.

Coexistence with other multipathing software

The Host Attachment Kit is not a multipath driver. It enables and configures native multipathing rather than providing it.

> **Important:** An IBM requirement is that you install the correct Host Attachment Kit for each operating system (OS) type.
>
> A mix of different multipathing solution software on the same server is not supported. Each product can have different requirements for important system settings, which can conflict. These conflicts can cause issues that range from poor performance to unpredictable behaviors, and even data corruption.
>
> If you need coexistence and a support statement does not exist, apply for a support statement from IBM. This statement is known as a Solution for Compliance in a Regulated Environment (SCORE), or sometimes a request for price quotation (RPQ).

2.1.4 Fibre Channel versus iSCSI access

Hosts can attach to FlashSystem A9000 or A9000R over an FC or Ethernet network (by using iSCSI). The version of FlashSystem A9000 and A9000R system software at the time of writing this book supports iSCSI by using the software initiator and converged network adapters (CNAs). To obtain the current list of supported operating systems and adapters for iSCSI, see the IBM System Storage Interoperation Center (SSIC):

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

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

Consider the following information:

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

Figure 2-5 shows the simultaneous access to two separate FlashSystem A9000 and A9000R volumes from one host by using both protocols.

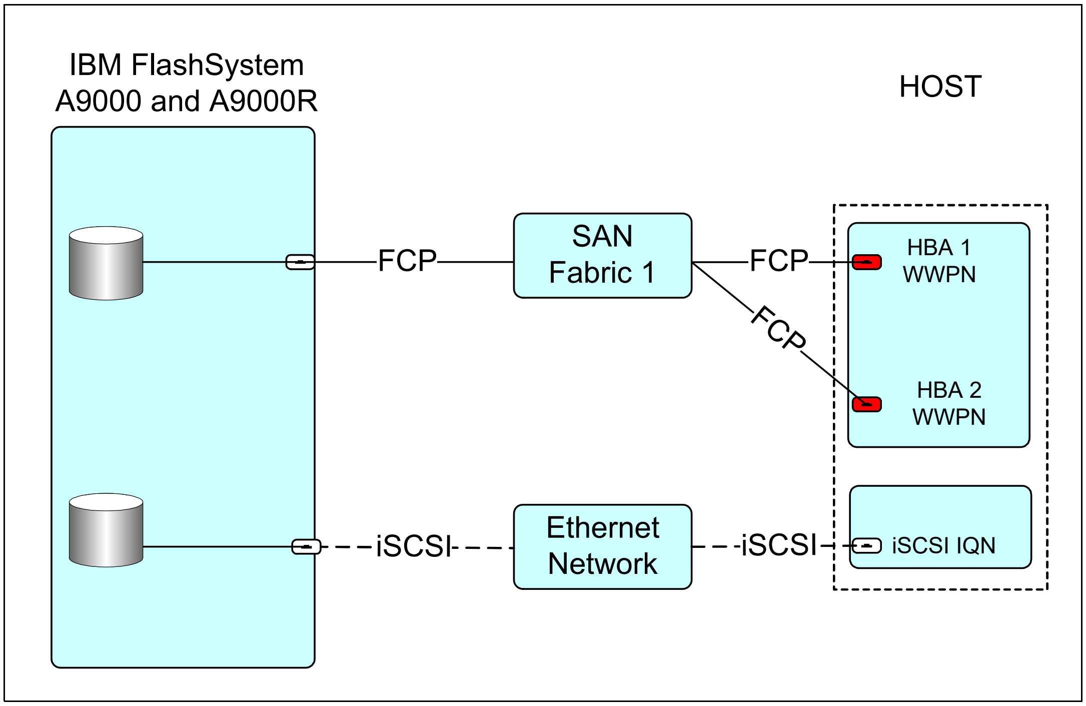


Figure 2-5 Connecting by using FCP and iSCSI simultaneously with separate host objects

A host can connect through FC and iSCSI simultaneously. However, you cannot access the same logical unit number (LUN) through both protocols.

2.2 Fibre Channel connectivity

This section highlights information about FC connectivity that applies to FlashSystem A9000 and A9000R in general.

2.2.1 Preparation steps

Before you can attach an FC host to FlashSystem A9000 or A9000R, you must complete several procedures. The following general procedures pertain to all hosts. However, you also must review any procedures that pertain to your specific hardware and operating system.

Complete the following steps:

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

   `https://www.ibm.com/systems/support/storage/ssic/interoperability.wss`

   For each query, select FlashSystem A9000 or A9000R, a host server model, an operating system, and an HBA vendor. 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 enough adapters are installed. You need enough adapters on the host server to manage the total number of LUNs that you want to attach.
3. Check the optimum number of paths that must be defined to help determine the zoning requirements.
4. Download and install the latest supported HBA firmware and driver, if needed.

HBA vendor resources

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

2.2.2 Fibre Channel configurations

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

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

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

- If a grid controller fails, each host remains connected to all other grid controllers.
- If an FC switch fails, each host remains connected to multiple grid controllers.
- If an HBA fails, each host remains connected to multiple grid controllers.
- If a host cable fails, each host remains connected to multiple grid controllers.

Redundant configuration with multiple paths to each volume

A redundant configuration, which assumes six grid controllers but uses only six paths per LUN on the host, is shown in Figure 2-6.

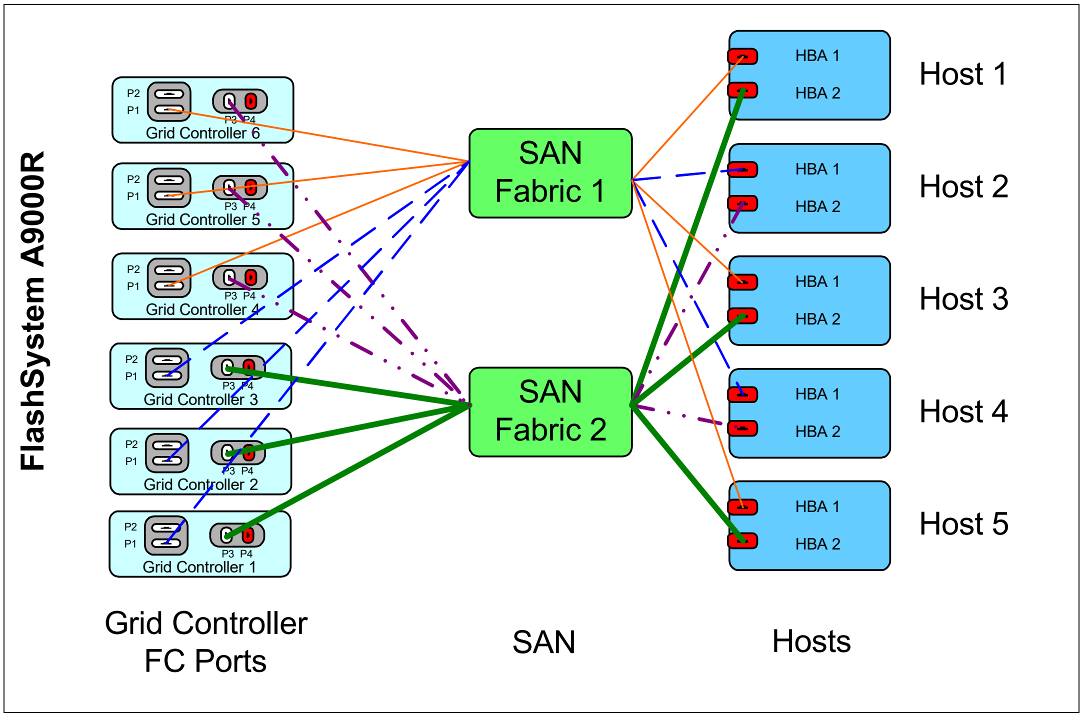


Figure 2-6 Fibre Channel redundant configuration

This configuration has the following characteristics:

- Each host is equipped with dual HBAs. Each HBA (or HBA port) is connected to one of two FC switches in separate fabrics.
- Each of the FC switches has a connection to a separate FC port of each of the six grid controllers (in FlashSystem A9000R).
- Each fabric has two zones (noted by the different colors of the connections), with three paths per fabric, per zone for each host, giving a total of six paths per volume. If a fabric fails, all grid controllers that are connected are still used.

Important: The configuration that is shown in Figure 2-6 is a good overall multipathing configuration that consists of six paths per LUN for a system with six grid controllers.

If the system had eight grid controllers such as in a fully configured FlashSystem A9000R Model 425, each of the FC switches would have two zones, each zone having a connection to four separate grid controllers.

An even simpler redundant configuration, which still assumes six grid controllers, is shown in Figure 2-7.

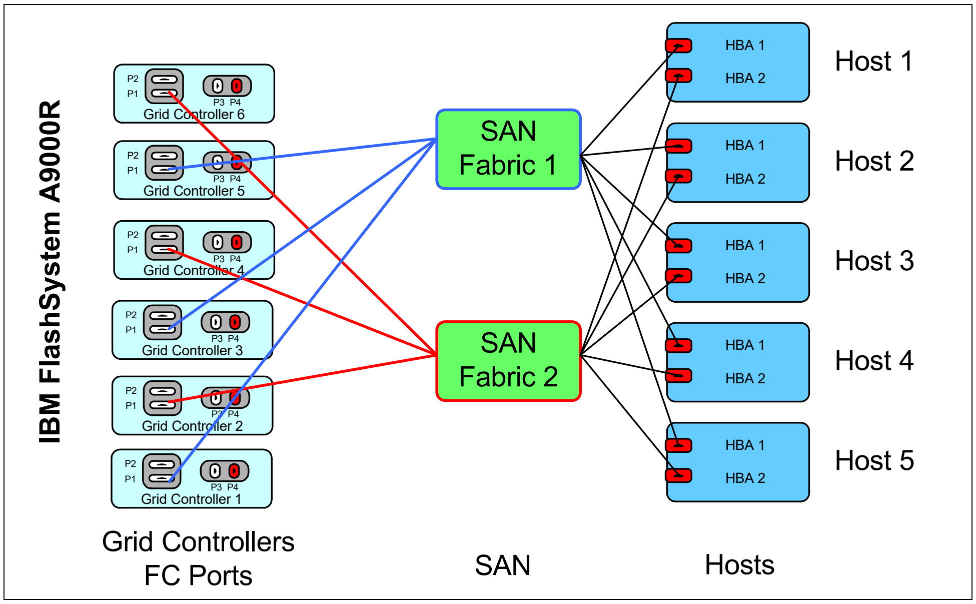


Figure 2-7 Fibre Channel simple redundant configuration

This configuration has the following characteristics:

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

A fully redundant configuration, which assumes six grid controllers, is shown in Figure 2-8.

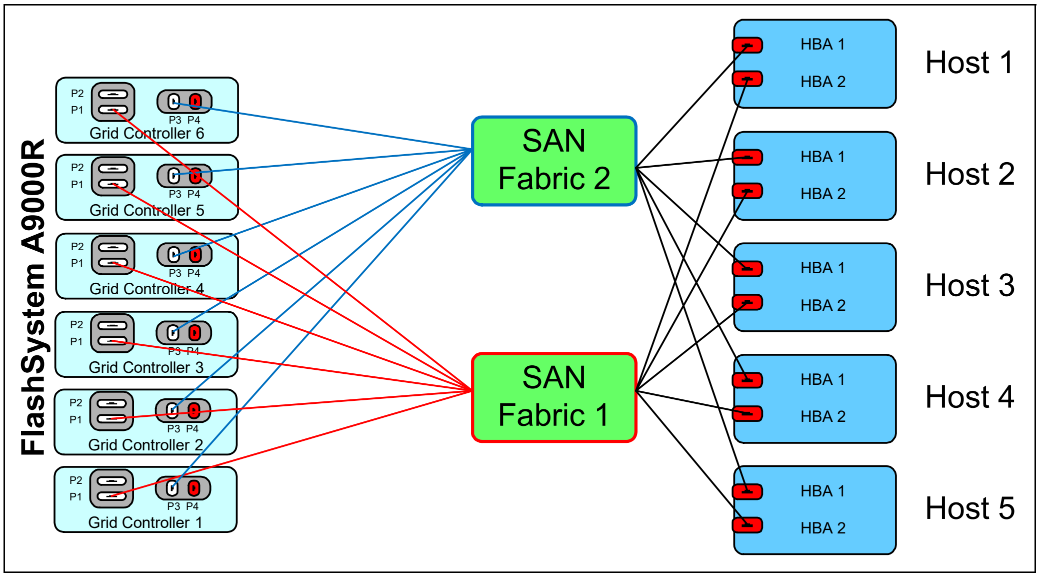


Figure 2-8 Fibre Channel fully redundant configuration

This configuration has the following characteristics:

- Each host is equipped with dual HBAs. Each HBA (or HBA port) is connected to one of two FC switches in separate fabrics.
- Each volume can be accessed through 12 paths. No benefit occurs from exceeding 12 paths (even with more grid controllers) because it can cause issues with host processor utilization and server reliability if a path failure occurs.

For FlashSystem A9000, with only three grid controllers, a redundant configuration that still gives six paths per host is shown in Figure 2-9.

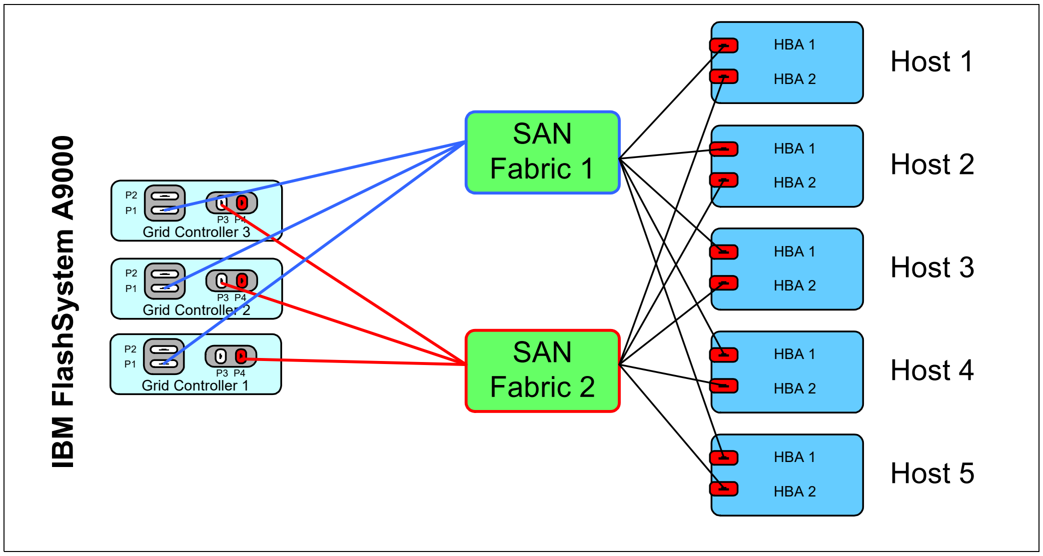


Figure 2-9 FlashSystem A9000 redundant configuration

In a fully configured FlashSystem A9000R Model 415, 12 grid controllers are present. This type of configuration can benefit from attaching each physical connection to each grid controller's port 1 and port 3, as shown in Figure 2-10. It is important to equally distribute the workload over all grid controllers.

The diagram shows that:

- Host 1 has HBA1 connected to SAN Fabric 1 and HBA2 connected to SAN Fabric 2
- Host 2 has HBA1 connected to SAN Fabric 1 and HBA2 connected to SAN Fabric 2

SAN Fabric1 has:

- Six paths to Port1 (FC adapter 1) in grid controllers 7 - 12 (Zone 1)
- Six paths to Port1 (FC adapter 1) in grid controllers 1 - 6 (Zone 3)

SAN Fabric2 has:

- Six paths to Port3 (FC adapter 2) in grid controllers 1 - 6 (Zone 2)
- Six paths to Port3 (FC adapter 2) in grid controllers 7 - 12 (Zone4)

Host 1 is zoned to each Grid Controller with 12 paths in total, via:

- HBA 1 zoned to port 1 (FC adapter 1) in grid controllers 7 - 12 via SAN fabric 1
- HBA 2 zoned to port 3 (FC adapter 2) in grid controllers 1 - 6 via SAN fabric 2

Host 2 is zoned to each Grid Controller with 12 paths in total, via:

- HBA 1 zoned to port 1 (FC adapter 1) in grid controllers 1 - 6 via SAN fabric 1
- HBA 2 zoned to port 3 (FC adapter 2) in grid controllers 7 - 12 via SAN fabric 2

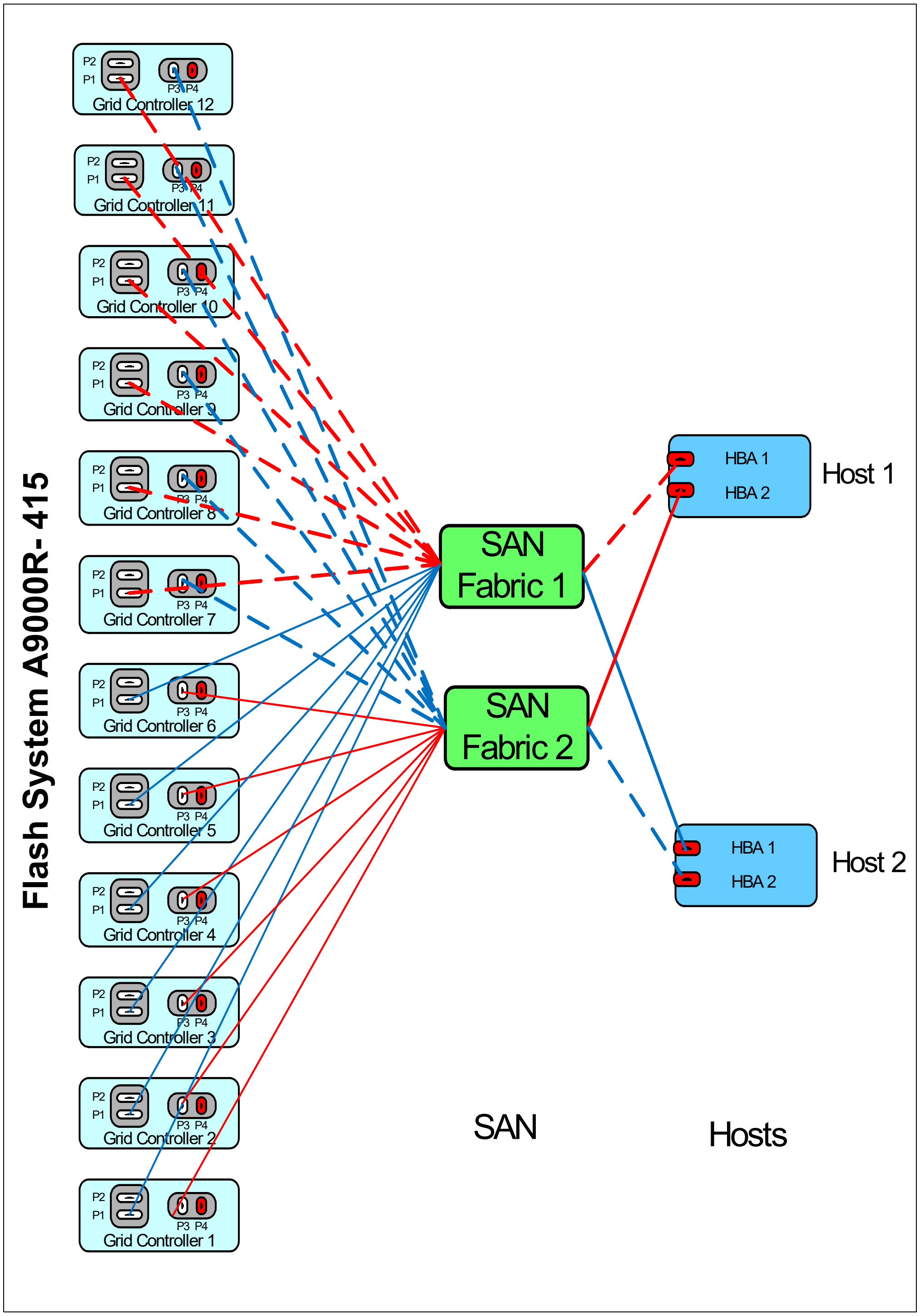


Figure 2-10 Full FlashSystem A9000R Model 415 configuration for maximum I/O connectivity

An alternative is to attach all SAN 1 connections to port 1 of grid controllers 1 - 6 and all SAN 2 connections to port 3 of grid controllers 7 - 12. This configuration allows the operations staff to help keep a connectivity scheme that splits the connections across SAN fabrics. Either choice allows the host I/Os to be spread to all 12 grid controllers instead of using both FC port 1 and port 3 attachment recommendations and restricting the number of physical grid controller attachments.

2.2.3 Zoning

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

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

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

- An error from an HBA that affects the zone or zone traffic is isolated to only the devices to which it is zoned.
- Any change in the SAN fabric triggers a *registered state change notification* (RSCN). These changes can be caused by a server restart or by adding a new product to the SAN. An RSCN requires that any device that can "see" the affected or new device acknowledge the change, interrupting its own traffic flow.

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

Zoning is affected by the following factors, among others:

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

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

- Each zone (excluding those zones for the IBM SAN Volume Controller) has one initiator HBA (the host) and multiple target HBA ports from a single FlashSystem A9000 or A9000R.
- Zone each host to ports from at least two grid controllers.
- Do not mix disk and tape traffic in a single zone. Also, avoid having disk and tape traffic on the same HBA.

More information about SAN zoning is in *Introduction to Storage Area Networks*, SG24-5470:

http://www.redbooks.ibm.com/abstracts/sg245470.html

2.2.4 Identification of FC ports (initiator/target)

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

The easiest way to get a record of all of the WWPNs on FlashSystem A9000 or A9000R is to use the CLI. However, this information is also available from the GUI.

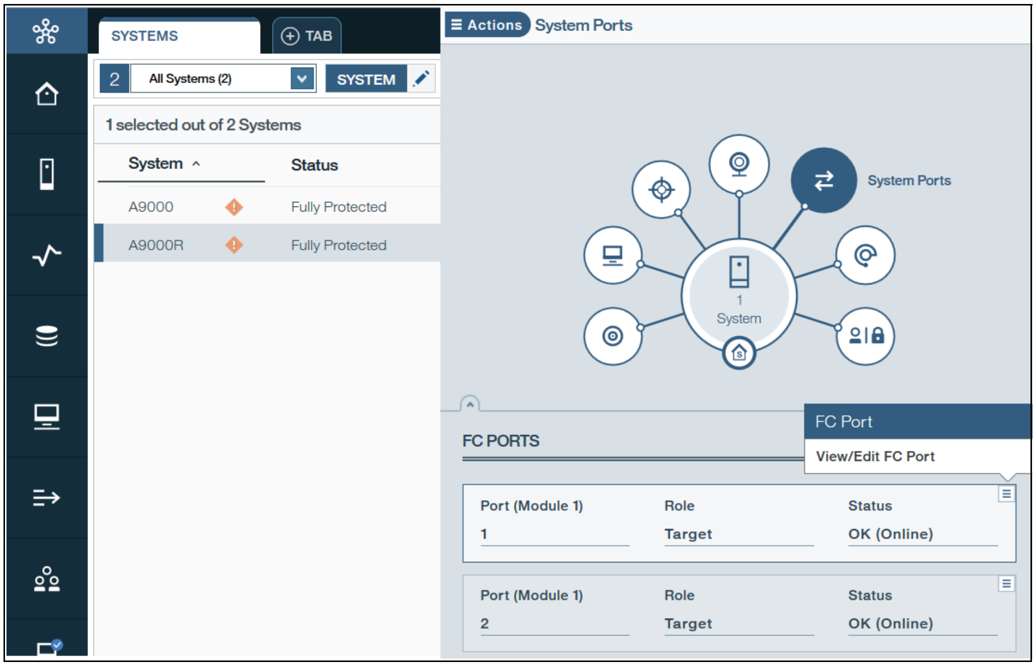
Example 2-1 shows all WWPNs for one FlashSystem A9000R system that was used in the preparation of this book. It also shows the CLI command that was used to list them. For clarity, several columns are removed from the example.

Example 2-1 Getting the WWPN on FlashSystem A9000R

```
>> fc_port_list
Component ID    Status   Currently Functioning   WWPN             Port ID    Role
1:FC_Port:1:1    OK      yes                   5001738035CF0110  FFFFFFFF   Target
1:FC_Port:1:2    OK      yes                   5001738035CF0111  00010200   Target
1:FC_Port:1:3    OK      yes                   5001738035CF0112  00020300   Target
1:FC_Port:1:4    OK      yes                   5001738035CF0113  FFFFFFFF   Target
1:FC_Port:2:1    OK      yes                   5001738035CF0120  FFFFFFFF   Target
1:FC_Port:2:2    OK      yes                   5001738035CF0121  00010100   Target
....
1:FC_Port:4:4   OK        yes                  5001738035CF0143   004BA400   Initiator
```

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

To retrieve the same information from the GUI, see Figure 2-11 and complete the following steps:



1 selected out of 2 Systems

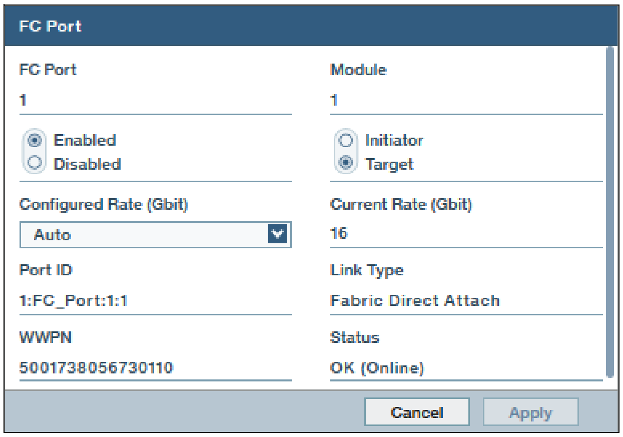
System	Status
A9000	Fully Protected
A9000R	Fully Protected

Port (Module 1)	Role	Status
1	Target	OK (Online)

Port (Module 1)	Role	Status
2	Target	OK (Online)

Figure 2-11 Retrieving Fibre Channel port properties

1. From the **Systems & Domains** view, select the system of interest.
2. From the Hub view of the System, select the **System Ports** spoke.
3. The ports and modules are listed at the bottom of the window.
4. Click the **Actions** icon for a particular port, and select **View/Edit FC Port**.
5. Scroll down if necessary to view the WWPN (see Figure 2-12).



FC Port: 1

Module: 1

● Enabled
○ Disabled

○ Initiator
● Target

Configured Rate (Gbit): Auto

Current Rate (Gbit): 16

Port ID: 1:FC_Port:1:1

Link Type: Fabric Direct Attach

WWPN: 5001738056730110

Status: OK (Online)

Figure 2-12 Viewing the FC port WWPN

2.2.5 Start from SAN

For more information, see 1.2.5, “Boot from SAN on x86 or x64 based architecture” on page 17.

2.3 iSCSI connectivity

This section focuses on iSCSI connectivity as it applies to FlashSystem A9000 and A9000R in general.

2.3.1 IP configuration

This section summarizes the various IP connectivity interfaces in the storage system in relation to the different Ethernet physical ports and their specific usage and configuration.

Starting with software version 12.3.2, FlashSystem A9000 and FlashSystem A9000R support Virtual LAN (VLAN) tagging. For more information, see 2.4, “VLAN tagging” on page 67.

iSCSI host attachment ports

Remember that two grid controller hardware configurations are available. Depending on the configuration that is ordered, grid controllers can be equipped with one dual port iSCSI adapter or two dual port iSCSI adapters, as described in 2.1, “Overview” on page 42.

Each iSCSI port is defined with its own IP interface and address. If you use VLAN tagging, multiple IP interfaces with their respective VLAN tag can be defined on ports that are used for host connectivity. For more information, see 2.4, “VLAN tagging” on page 67.

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

For high availability and performance, ensure that the following requirements are met:

- 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. If your grid controllers have two iSCSI adapters, use one port in each adapter. In any grid controller, use iSCSI adapter ports 1 and 3 for host connectivity, use ports 2 and 4 for replication.
- Spread connections to all grid controllers. Remember that the A9000 or A9000R minimum rack configuration (three grid controllers and one flash enclosure) can support only one grid controller failure.
- Avoid switch port oversubscription and disable TCP delayed.
- Use jumbo frames, thus with more than 1500 bytes of payload and up 9000 bytes. Use 9000 bytes if possible and specify an MTU of 9000.

Management ports

In addition to the Ethernet ports used for host connectivity, the storage system includes management ports. These ports are dedicated for CLI and GUI communications, outgoing SNMP and SMTP connections, and connecting with NTP, Encryption Key servers, DNS, and LDAP.

To ensure management redundancy if a module fails, the storage system management function is accessible from two different IP addresses in IBM FlashSystem A9000, and three in IBM FlashSystem A9000R. Each of the three IP addresses is handled by a different hardware module.

The various IP addresses are not apparent to the user. Management functions can be performed by using any of the IP addresses. These addresses can be accessed simultaneously by multiple users.

Important: All management IP interfaces must be connected to the same subnet and use the same network mask, gateway, and MTU. Use caution because changing the MTU without supporting it on the adjacent switches can break management connectivity.
CLI and GUI management is run over TCP port 7778, with all traffic encrypted through the Secure Sockets Layer (SSL) protocol

Remote support ports

Two remote support ports are used to establish a VPN for remote support by IBM personnel. They are also referred to as the *VPN ports*.

Technician port

The system also includes one or two Ethernet ports that are reserved for use by the technician to connect their notebook for initial setup or for servicing the system.

The iSCSI ports can be easily identified. Use the GUI or XCLI to display the current settings.

The **ipinterface_list** command (see Example 2-5) can be used to display only all configured network ports on the various grid elements (modules). The command also displays all other Ethernet connections and settings. For clarity, several columns are removed from the Example 2-5.

Example 2-2 Listing iSCSI ports with the ipinterface_list command

```
>> ipinterface_list
Name         Type         IP Address     Network Mask    Default Gateway ..... MTU    Module
management   Management   9.11.237.109   255.255.254.0   9.11.236.1            1500   1:Module:1
management   Management   9.11.237.107   255.255.254.0   9.11.236.1            1500   1:Module:2
management   Management   9.11.237.108   255.255.254.0   9.11.236.1            1500   1:Module:3
VPN          VPN                                                               1500   1:Module:3
VPN          VPN                                                               1500   1:Module:4
port1        iSCSI        9.11.230.76    255.255.254.0   9.11.230.1            1500   1:Module:1
port2        iSCSI        9.11.230.79    255.255.254.0   9.11.230.1            1500   1:Module:2
```

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

2.3.2 Preparation steps

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

1. Connect the host to FlashSystem A9000 or A9000R over iSCSI by using a standard Ethernet port or a Converged Network Adapter (CNA) on the host server. Dedicate the port that you choose to iSCSI storage traffic only. This port must also support 10 Gbps. This port requires the definition of an IP interface, with an IP address, subnet mask, and gateway.

   Also, review any documentation that came with your operating system about iSCSI to ensure that any other conditions are met.

2. Check the LUN limitations for your host operating system. Verify that enough adapters are installed on the host server to manage the total number of LUNs that you want to attach.
3. Check the optimum number of paths that must be defined, which helps determine the number of physical connections that must be made.

4. Install the latest supported adapter firmware and driver on the host. Download the latest version if it was not included with your operating system.
5. Maximum transmission unit (MTU) configuration is required if your network supports an MTU that is larger than the default (1500 bytes). Anything larger is known as a *jumbo frame*. Specify the largest possible MTU (9000 for host ports and 1500 for management and VPN ports).

   > **Note:** Before changing an Ethernet port MTU, make sure that the adjacent IP switch is properly configured. If the switch uses a smaller MTU size than specified for the Ethernet port, it breaks existing application connectivity. For example, if the system sends 9000 byte packets while the switch can receive packets up to 1500 bytes only, the switch drops extra packets.

6. Any device that uses iSCSI requires an iSCSI qualified name (IQN) and an attached host. The IQN uniquely identifies iSCSI devices. The IQN for FlashSystem A9000 or A9000R is configured when the system is delivered, and the IQN must not be changed. Contact IBM technical support if a change is required.
7. Starting with software version 12.3.2, consider whether you will use VLAN tagging and plan accordingly. For more information, see 2.4, “VLAN tagging” on page 67.

2.3.3 Network configuration for iSCSI host connectivity

Disk access is susceptible to network latency. Latency can cause timeouts, delayed writes, and data loss. To get the best performance from iSCSI, place all iSCSI IP traffic on a dedicated network. Physical switches or VLANs can be used to provide a dedicated network. This network must support 10 Gbps, and the hosts need interfaces that are dedicated to iSCSI only. You might need to purchase more host Ethernet ports.

To achieve high performance, it is important to spread the host connections to each grid controller evenly.

Also, use the CPU in each grid controller as much as possible. Assuming you have a system that is equipped with two iSCSI controllers (four iSCSI ports on each grid controller), it is better to use ports 1 and 3 for host connectivity and then ports 2 and 4 for mirroring. For host connectivity, use half the ports in each grid controller (ports 1 and 3) and create a single subnet (an identical subnet in each switch).

Figure 2-13 on page 62 shows an example of how the host connections are divided.

Consider the following guidelines when configuring the iSCSI connectivity. Most apply whether you use VLAN tagging:

- Disable Spanning Tree (STP) on switch ports that are connected to a FlashSystem A9000/R system.

  Using spanning tree on the switches can delay ports recovery after system hot upgrade due to spanning tree discovery process.

  A standard “port-fast” (also called *edge*) disable configuration is not sufficient for trunk ports. You must specifically disable spanning-tree on the relevant ports, even if the ports were originally marked as connected to edge devices and not to other switches.
- Disable the TCP delayed acknowledgment feature on host side.
- Disable Nagle’s algorithm on host side.
- Avoid switch port oversubscription.

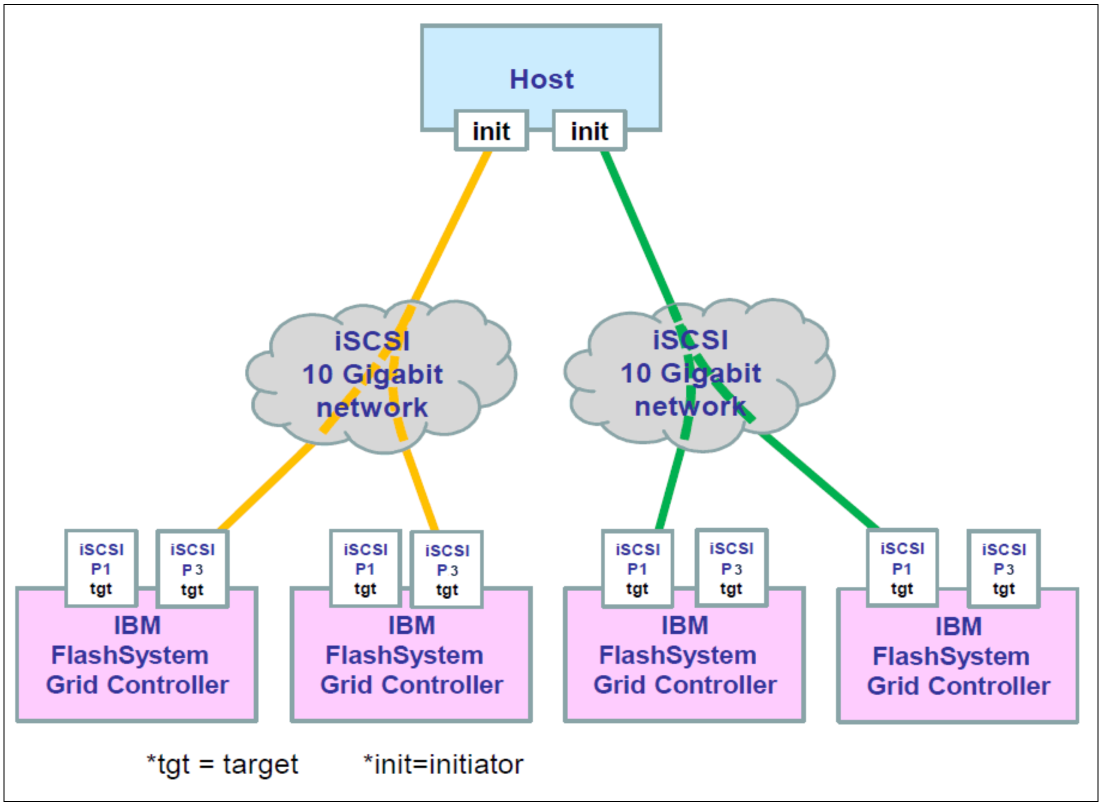


Figure 2-13 Host IP connectivity

2.3.4 iSCSI start from LUN

At the time of this writing, you cannot start through iSCSI, even if an iSCSI HBA is used. For more information, see the SSIC web page:

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

2.3.5 iSCSI setup

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

Getting the iSCSI qualified name

Every FlashSystem A9000 or A9000R has a unique IQN. The format of the IQN is simple, and it includes a fixed text string, which is followed by the last digits of the system serial number.

To display the IQN of the storage system, complete the following steps:

1. From the Hyper-Scale Manager GUI, click the **Systems & Domains Views** icon and select **Systems** from the menu, as shown in Figure 2-14 on page 63.

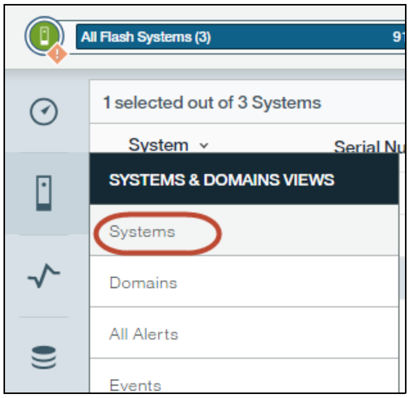
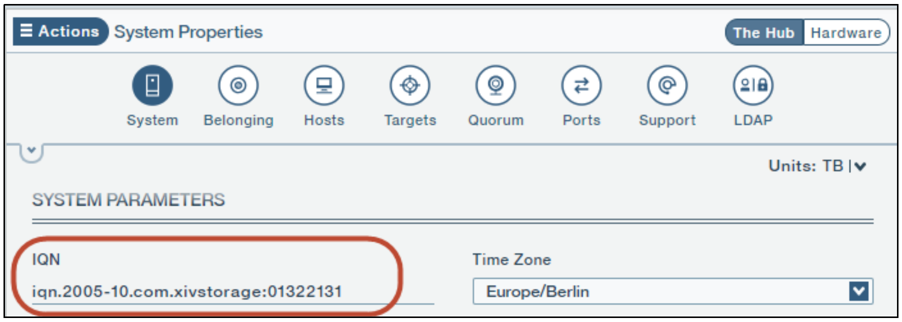


Figure 2-14 Systems menu

2. The list of systems that are in the inventory is displayed in the Systems & Domains view. Select the system for which you need to retrieve the IQN.
3. Click in the circle that represents the system from (see Figure 2-15) to display the System Properties. Scroll down in the System Properties to find the iSCSI name under System Parameters.



SYSTEM PARAMETERS

IQN: iqn.2005-10.com.xivstorage:01322131

Time Zone: Europe/Berlin

Figure 2-15 Retrieving the iSCSI name (IQN)

> **Important:** Do not attempt to change the IQN. If you need to change the IQN, you must engage IBM Support.

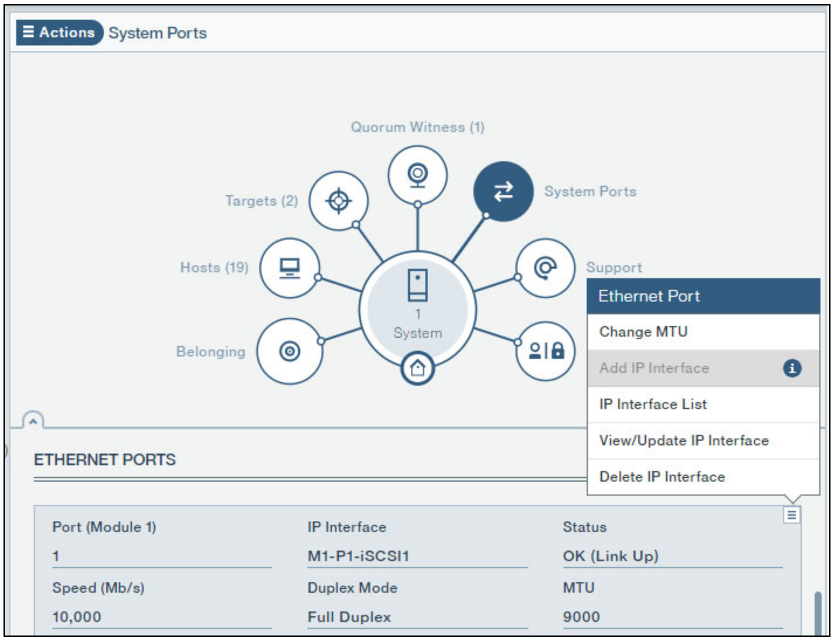
To get the information in the XCLI, run the **config_get** command as shown in Example 2-3. The output is truncated for clarity.

Example 2-3 Use the CLI to get the iSCSI name (IQN)

```
>> config_get
Name                                    Value
system_name                             ITSO_2_A9000R
...
system_id                               13775
machine_type                            9835
machine_model                           415
machine_serial_number                   6013000
...
iscsi_name                              iqn.2005-10.com.xivstorage:01322131
...
```

Configuring the Ethernet ports by using the GUI

To set up the iSCSI port by using the GUI, see Figure 2-16 and complete the following steps:

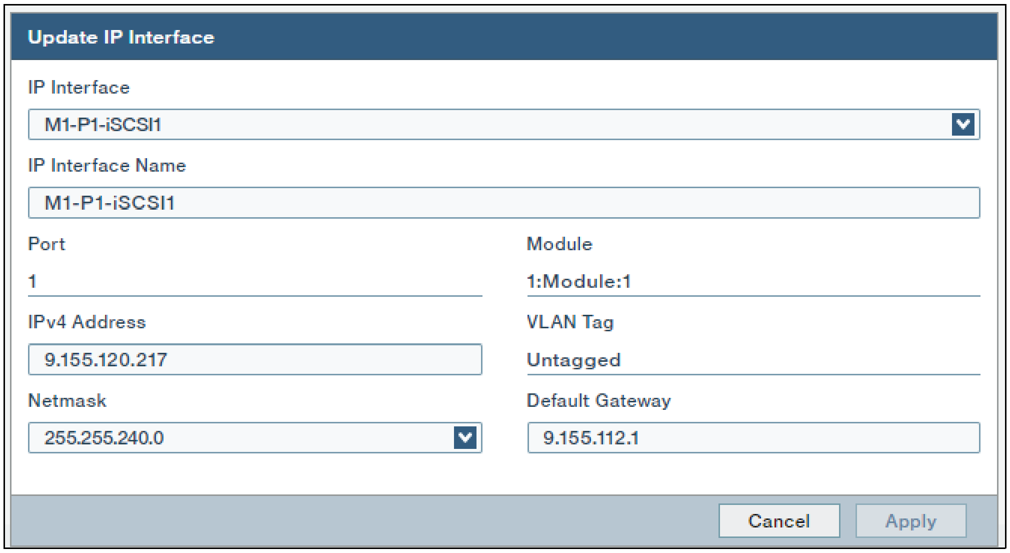


ETHERNET PORTS

Port (Module 1)	IP Interface	Status
1	M1-P1-iSCSI1	OK (Link Up)
Speed (Mb/s)	Duplex Mode	MTU
10,000	Full Duplex	9000

Figure 2-16 Retrieving the IP port properties

1. From the **Systems & Domains** view, select the system of interest.
2. From the Hub view of the system, select the **System Ports** spoke.
3. The ports and modules are listed at the bottom of the window. Scroll down to reach the section for Ethernet ports.
4. Click the **Actions** icon for a particular port, and if the port is configured, select **View/Update IP Interface**; otherwise, select **Add IP Interface** (see Figure 2-16).
5. You can view and change the IP interface settings as shown in Figure 2-17.



Update IP Interface

IP Interface: M1-P1-iSCSI1

IP Interface Name: M1-P1-iSCSI1

Port	Module
1	1:Module:1
IPv4 Address	VLAN Tag
9.155.120.217	Untagged
Netmask	Default Gateway
255.255.240.0	9.155.112.1

Figure 2-17 Setting the iSCSI port properties

Enter the following information (see Figure 2-17 on page 64):

- Name: Define the name for this interface.
- Address, netmask, and gateway: Enter the standard IP address information.
- VLAN Tag: by default, it is untagged. Only assign a VLAN ID to the interface if you use VLAN tagging. For more information, see 2.4, “VLAN tagging” on page 67.

6. Click **Apply** to complete the IP interface and iSCSI setup.
7. From the Actions menu that is shown in Figure 2-16 on page 64, you can also change the MTU value.

   All devices in a network must use the same MTU. If in doubt, set MTU to `1500` because 1500 is the default value for Gigabit Ethernet. Performance might be affected if the MTU is set incorrectly.

> **Tip:** If the MTU that is used by FlashSystem A9000 or A9000R is greater than the network can transmit, the frames are discarded. The frames are discarded because the do-not-fragment bit is normally set to `on`.
>
> Use the **`ping -l`** command to test to specify packet payload size from a Windows workstation in the same subnet. A **`ping`** command normally contains 28 bytes of IP and Internet Control Message Protocol (ICMP) headers plus payload. Add the **`-f`** parameter to prevent packet fragmentation.
>
> For example, the **`ping -f -l 1472 10.1.1.1`** command sends a 1500-byte frame to the 10.1.1.1 IP address (1472 bytes of payload and 28 bytes of headers). If this command succeeds, you can use an MTU of 1500.

Configuring the iSCSI port by using the CLI

To configure iSCSI ports by using the CLI session tool, issue the **`ipinterface_create`** command, as shown in Example 2-4. If VLAN tagging is used, see 2.4.5, “Managing ports and interfaces with VLANs definitions” on page 72 for more information.

Example 2-4 iSCSI setup by using the CLI

```
>> ipinterface_create ipinterface="Test" address=10.0.0.10 netmask=255.255.255.0
module=1:Module:1 ports="1" gateway=10.0.0.1 mtu=9000
```

Identifying iSCSI ports

The iSCSI ports can be easily identified. Use the GUI or XCLI to display the current settings.

The **`ipinterface_list`** command (see Example 2-5) can be used to display only all of the configured network ports. The use of the command also displays all other Ethernet connections and settings. For clarity, several columns are removed from the example.

Example 2-5 Listing iSCSI ports with the ipinterface_list command

```
>> ipinterface_list
Name         Type         IP Address     Network Mask    Default Gateway ..... MTU    Module
management   Management   9.11.237.109   255.255.254.0   9.11.236.1            1500   1:Module:1
management   Management   9.11.237.107   255.255.254.0   9.11.236.1            1500   1:Module:2
management   Management   9.11.237.108   255.255.254.0   9.11.236.1            1500   1:Module:3
VPN          VPN                                                               1500   1:Module:3
VPN          VPN                                                               1500   1:Module:4
port1        iSCSI        9.11.230.76    255.255.254.0   9.11.230.1            1500   1:Module:1
port2        iSCSI        9.11.230.79    255.255.254.0   9.11.230.1            1500   1:Module:2
```

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

2.3.6 iSCSI and CHAP authentication

IBM FlashSystem A9000 and A9000R support industry-standard unidirectional iSCSI Challenge Handshake Authentication Protocol (CHAP). The iSCSI target of FlashSystem A9000 or A9000R can validate the identity of the iSCSI Initiator that attempts to log on to the system.

The CHAP configuration in the IBM FlashSystem A9000 or A9000R is defined on a per-host basis. No global configurations for CHAP affect all the hosts that are connected to the system.

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

For the iSCSI initiator to log in with CHAP, the `iscsi_chap_name` and `iscsi_chap_secret` parameters must be set. After both of these parameters are set, the host can run an iSCSI login to the Storage System only if the login information is correct.

CHAP name and secret parameter guidelines

The following guidelines apply to the CHAP name and secret parameters:

- Both the `iscsi_chap_name` and `iscsi_chap_secret` parameters must be specified or not specified. You cannot specify only one parameter.
- The `iscsi_chap_name` and `iscsi_chap_secret` parameters must be unique. If they are not unique, an error message is displayed, although the command does not fail.
- The secret must be 96 - 128 bits. You can use one of the following methods to enter the secret:
  - Base64 requires that `0b` is used as a prefix for the entry. Each subsequent character that is entered is treated as a 6-bit equivalent length.
  - Hex requires that `0x` is used as a prefix for the entry. Each subsequent character that is entered is treated as a 4-bit equivalent length.
  - String requires that a prefix is not used (it cannot be prefixed with `0b` or `0x`). Each character that is entered is treated as an 8-bit equivalent length.
- If the `iscsi_chap_secret` parameter does not conform to the required secret length (96 - 128 bits), the command fails.
- If you change the `iscsi_chap_name` or `iscsi_chap_secret` parameters, a warning message indicates that the changes will apply the next time that the host is connected.

Configuring CHAP

CHAP can be configured by using XCLI commands. Consider the following points:

- 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 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=
```

2.4 VLAN tagging

IBM FlashSystem A9000 and A9000R software version 12.3.2 introduces VLAN tagging and port trunking support for iSCSI environments. The feature allows organizations to operate multiple private virtual networks, and share FlashSystem A9000 and A9000R systems among the virtual networks.

Dynamically pooling a high-density FlashSystem A9000 or A9000R among multiple virtual networks maximizes its capacity and performance utilization. For example, private cloud and Managed Service Provider (MSP) environments can use VLANs to provision a private virtual network per tenant, yet securely share FlashSystem A9000/R systems among the tenants.

2.4.1 What is a VLAN?

A virtual local area network (VLAN) is a logical grouping of devices within one or more local area networks. A VLAN is a logically isolated network. VLANs allow you to logically isolate networks without physically separating them through various switches. That is, VLANs make it possible to have multiple isolated networks over a single port by creating different broadcast domains. Each VLAN features its own broadcast domain or subnet.

The most practical scenario for VLANs is to virtualize multiple networks on the same physical infrastructure. As shown in Figure 2-28, this technique is typically used by service providers to segregate and secure mulitenant network traffic on a shared physical network.

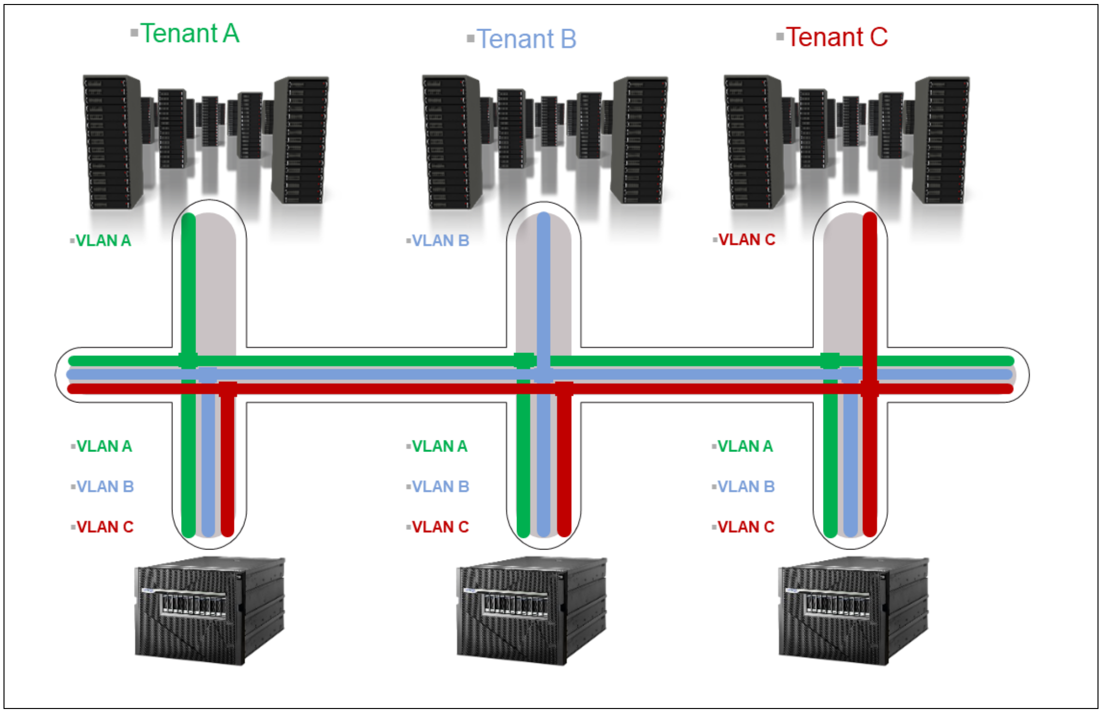


Figure 2-18 Using VLANs to segregate network traffic in a multi-tenant environment

2.4.2 Tagging and trunking

VLANs are typically associated with a particular IP subnet. Devices in the same VLAN cannot communicate with devices in other VLANs, and traffic between VLANs must be routed.

Ethernet interfaces on an IP switch can be defined as access ports or trunk ports. Consider the following points:

- An access port has only one VLAN configured and can carry traffic for only one VLAN.
- A trunk port can have two or more VLANs configured and can carry traffic for multiple VLANs.

A high-level overview of VLAN tagging and port trunking is shown in Figure 2-19.

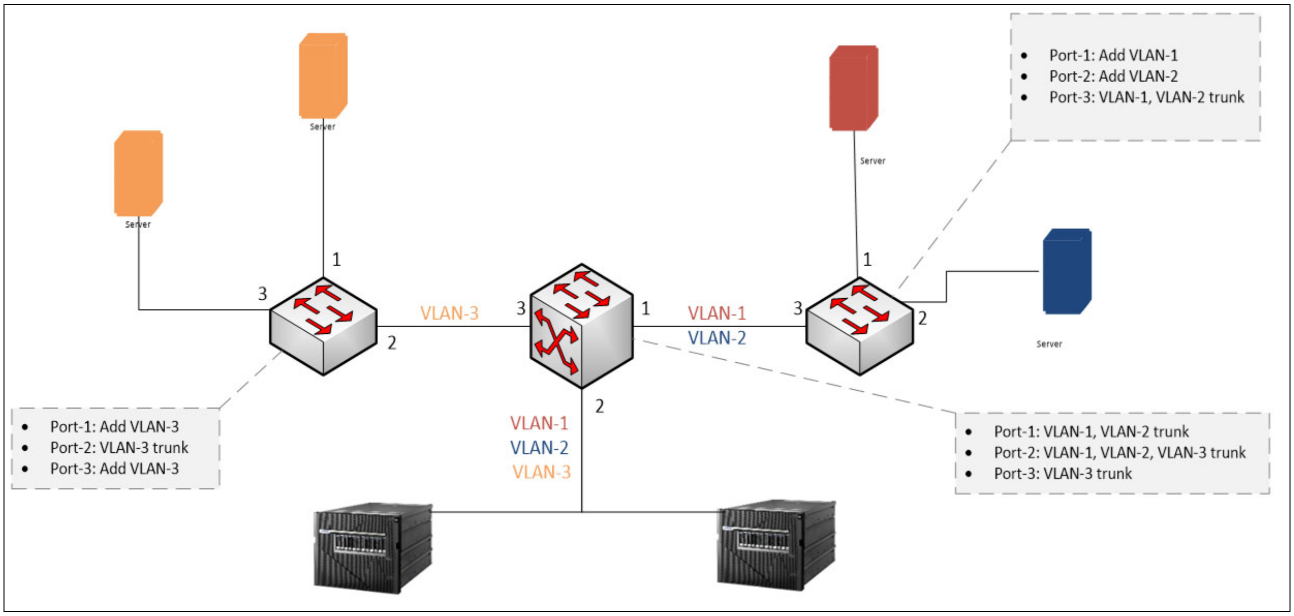
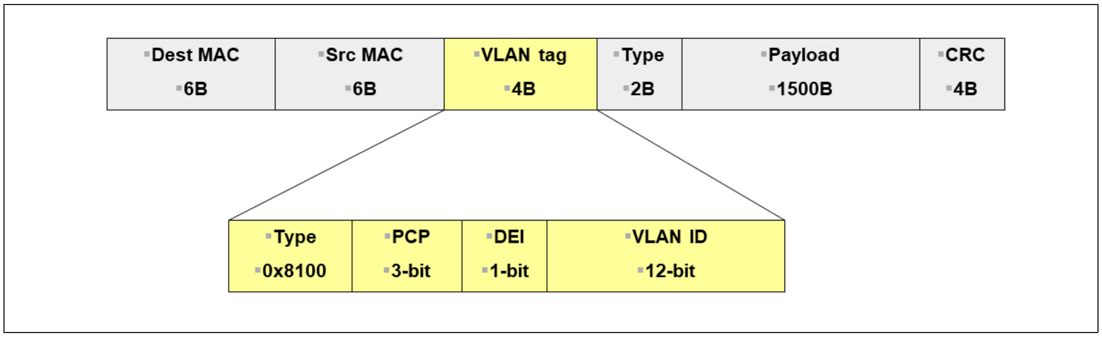


Figure 2-19 VLAN tagging and port trunking concept

To route the traffic on a trunk port to a specific VLAN, the switch uses an encapsulation method by inserting a VLAN tag in the network frames header, as defined by the IEEE 802.1Q standard. In IEEE 802 1Q conformant networks, when a frame enters the VLAN-aware portion of the network, a tag is added to represent the VLAN membership.

This method (see Figure 2-20), which is known as *VLAN tagging*, was developed by Cisco. The standard also contains provisions for a quality of service prioritization scheme, which is indicated by the PCP field in the VLAN tag. The value of the priority ranges 0 - 7.



Dest MAC	Src MAC	VLAN tag	Type	Payload	CRC
6B	6B	4B	2B	1500B	4B

Type	PCP	DEI	VLAN ID
0x8100	3-bit	1-bit	12-bit

Figure 2-20 VLAN frame tagging

2.4.3 VLAN tagging and port trunking support in FlashSystem A9000/R

Starting with software version 12.3.2, FlashSystem A9000/R supports VLAN tagging and port trunking. Before this software version, iSCSI data ports on the storage device only allowed the definition of a single IP interface on each physical port. Any port can be an access port only and therefore, did not support VLAN tagging.

The feature that was introduced in software V12.3.2 supports the configuration of more than one IP interface on any single port, which allows the port to be set up as a trunk port, as shown in Figure 2-21.

Restriction: Be aware that A9000/R ports that you want to use for replication (target connectivity) can have only one IP interface defined. That is, the A9000/R system can support only one VLAN definition when connected to another A9000/R for replication. It is better to use a separate VLAN for the replication.

Also, the HSM GUI does not support target autoconfiguration for iSCSI connectivity. If VLANs are configured, the iSCSI target connectivity must be configured manually.

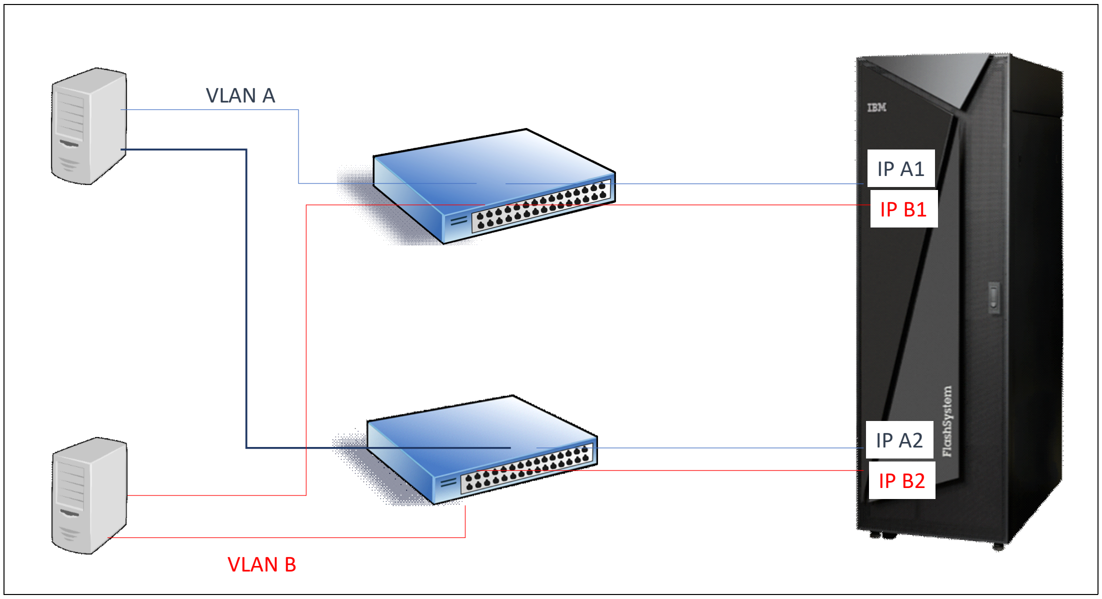


Figure 2-21 VLAN tagging support for host traffic n FlashSystem A9000/R

Note: Host systems do *not* need to be aware or understand VLAN tagging. That is, no specific configuration must be done on the host to account for VLAN tagging.

The host connects to an access port. If the switch is configured for VLAN, it adds the configured VLAN ID to the frames that are received from the host by using this port and sends them towards the network. Likewise, the switch removes the VLAN ID from the frames that are transmitted to the host by using this port.

2.4.4 FlashSystem A9000/R port trunking reference model

A typical use case for VLAN tagging and port trunking is in multi-tenant environments where network traffic must be segregated.

In Figure 2-22, a reference model example is shown that includes three Infrastructure as a Service (IaaS) providers, each assigned specific VLAN IDs, with their own layer 2 broadcast domains and IP subnets.

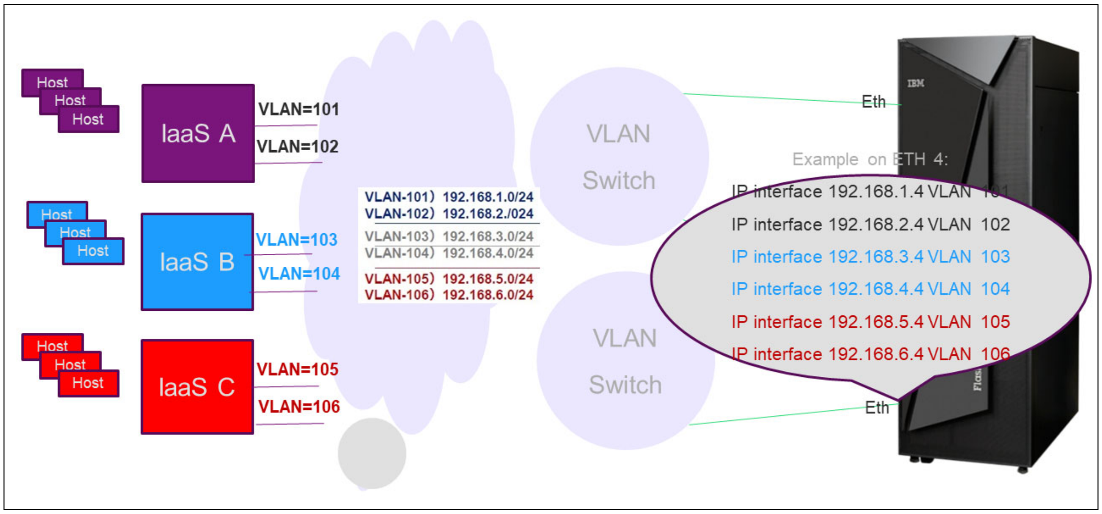


Figure 2-22 FlashSystem A9000/R trunking reference model

With A9000/R software version 12.3.2 or later, the multiple interfaces can be defined on any iSCSI port that is used for host connectivity, as shown for port ETH 4 in Figure 2-22.

FlashSystem A9000/R VLAN tagging implementation

From a design and implementation perspective, the A9000/R IP interface model is composed of the following layers, as shown in Figure 2-23 on page 71:

- The Ethernet ports

  This physical layer is where the iSCSI ports definition is stored. The ports are automatically defined by the storage system software upon discovery of the physical iSCSI adapters in the storage system.

- IP interface and VLAN

  This layer is where the storage administrator can define IP interfaces on the physical ports and might decide to optionally assign VLANs.

- iSCSI/Data

  This layer is where the storage administrator makes the host attachment and multipathing definitions.

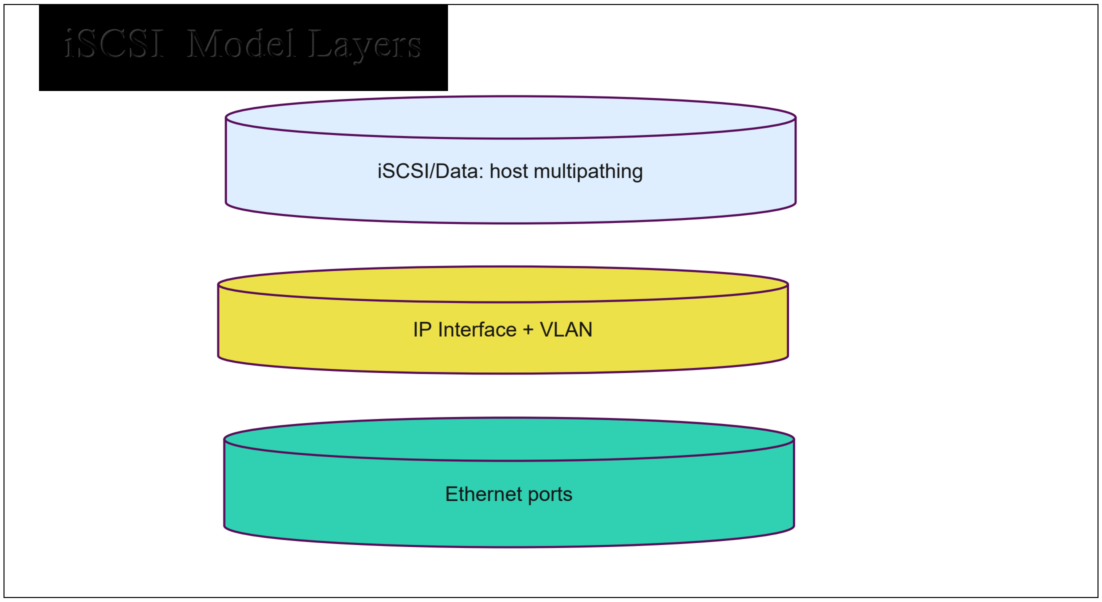


Figure 2-23 FlashSystem A900/R iSCSI interface model

If a port does not have a VLAN ID defined, it is known as Untagged.

System configuration capabilities

Regarding IP interfaces and support for VLAN, the system has the following capabilities:

- A maximum of 100 VLANs per FlashSystem A9000/R system.
- Up to 512 IP addresses per FlashSystem A9000/R system. It is on top of IP addresses that are assigned to management ports or VPN ports.
- Up to 100 IP addresses per physical Ethernet port.
- A total of 12 iSCSI paths per host, per VLAN.
- Permitted VLAN ID range is 1 - 4094 (for more information, see the section about the `ip_interface_create` command in “Managing IP interfaces with XCLI” on page 73).

Restrictions: When VLAN tagging is enabled, all management or VPN ports must belong to the same VLAN ID one per port type and with the same subnet gateway and MTU).

Multiple IPs cannot be assigned to the same VLAN and Ethernet port. For example, if `192.0.2.1` is assigned to Ethernet port 3 and VLAN 100, `192.0.2.2` cannot be assigned to Ethernet port 3 and VLAN 100, but can be assigned to a different Ethernet port on VLAN 100 or a different VLAN on Ethernet port 3.

Ports that are used for replication and other non-data ports, such as VPN and management ports, can have only one IP interface that is defined.

System Proprietary MIB

The FlashSystem proprietary MIB was updated with VLAN support to include OIDs of Ethernet ports and IP interfaces. Each subtree includes a table of the objects, where each entry has the following attributes:

- For Ethernet ports: Name, role, status, and counters
- For IP Interfaces: Address, subnet mask, default gateway, VLAN ID, Ethernet port, and counters

2.4.5 Managing ports and interfaces with VLANs definitions

Defining or monitoring resources in each layer is achieved by using specific XCLI commands or in part by using the Hyper-Scale Manager GUI. To support VLAN tagging, FlashSystem A9000/A9000R software V12.3.2 or later offers several new or modified XCLI commands and similar actions in HSM GUI.

In this section, we focus on commands or GUI elements that were created or modified in support of VLANs. Those commands are related to ports and interfaces mainly. Commands at the host data layer did not change as a result of VLAN support.

For more information about the commands that are related to ports, interfaces, and hosts configuration, see *Command-Line Reference Guide for IBM FlashSystem A9000 or A9000R*, SC27-8559.

Managing ports with XCLI

The following XCLI commands are available at the Ethernet port layer (assuming FlashSystem A9000/R software level 12.3.2 or later):

- `ethernet_port_list`

  This command lists all physical Ethernet ports and includes the following information, which is not an exhaustive list:

  - Component ID: Module number for iSCSI or switch number for management/field technician port
  - Port number on module/switch
  - Role: Data, management, VPN, and so on
  - IP interface that contains the ports (or none, if port is not configured as part of IP interface)
  - Status up/down
  - MTU
  - Auto-negotiation: Half-full duplex and 1000/100/10

- `ethernet_port_counter_list`

  Use this command to display Ethernet (physical) port statistics in the system.

- `ethernet_port_update`

  This command is used to update the configuration of the Ethernet port in changing the Maximum Transmission Unit (MTU) value.

  The command includes the following syntax:

  ```
  ethernet_port_update ethernet_port_name=PortName mtu=MTU[force_mtu_change=<yes|no>]
  ```

  Where `m` is the `module_id` and `p` is the `port`.

  MTU is an integer 1500 - 9000. For host traffic, the default and recommendation is 9000.

Important: Be careful when changing Ethernet port MTU. If management connectivity is on, or a VPN is being used, this process might break application connectivity, unless the adjacent switch is properly configured.

This break is the result of the switch using a fixed MTU size that is lower than the new size that is being defined on the system. For example, if the system sends 9000-byte packets but the switch can receive packets up to 1500 bytes only, the switch drops the extra packets.

Managing IP interfaces with XCLI

To manage IP interfaces and configure VLANs on those interfaces, the following XCLI commands are available:

- `ipinterface_create`

  The following syntax is used in the command:

  ```
  ipinterface_create ipinterface=IPInterfaceName address=Address netmask=NetworkMask
  [gateway=DefaultGateway ] < < ethernet_port_name=EthernetPortName
  [ vlan_id=VlanID ]> | < module=ModuleNumber port=PortNumber [ mtu=MTU ]
  [speed=<auto|10mb|100mb|1000mb|1gb|2500mb|2.5gb|10000mb|10gb> ] > >
  ```

  The parameter’s values are listed in Table 2-3.

Table 2-3 Parameter values

Name	Description	Mandatory
`ipinterface`	The name of the IP interface to be created. Do not use the names Management or VPN.	Y
`address`	IP address of the interface.	Y
`netmask`	Network mask of the interface.	Y
`gateway`	IP address of the default gateway for this interface.	N
`ethernet_port_name`	Name of the Ethernet port. The name format is `ethernet_m_p`, where `m` is the `module_id` and `p` is the `port`.	N
`vlan_id`	Must be an integer. Valid values are 1 - 4094. The VLAN value is optional. If not specified, the interface port is marked as untagged; that is, no VLAN. The VLAN ID is not an arbitrary number; it must match the VLAN configuration as set up by your Network Administrator.	N
`module`	Component identifier (rack and module) of the module that contains Ethernet ports.	N
`port`	An integer that designates the port number.	N
`mtu`	Maximum Transmission Unit: The supported packet size by the connecting Ethernet switch.	N
`speed`	Interface’s speed. A specific speed turns off auto-negotiation. Valid values: 'auto' or 10mb, 100mb, 1000mb, 1gb, 2500mb, 2.5gb, 10000mb or, 10gb	N

> **Note:** The system prevents IP interface creation if it detects IP addresses collision among defined IP interfaces. In particular, the system does not permit to use same IP address on different IP interfaces that share the VLAN

- `ip_interface_delete`

  `ipinterface_delete ipinterface=IPInterfaceName`

  Only the interfaces that are defined for iSCSI traffic can be deleted. Management and VPN interfaces cannot be deleted.

  The command is primarily used if you need to update a VLAN defined for that interface. To update the VLAN, you must first delete and then re-create the IP interface.

- `ipinterface_run_traceroute`

  ```
  ipinterface_run_traceroute localipaddress=IPaddress remote=remoteHost [ vlan_id=VlanID ]
  ```

  The command is used to perform a route trace of a specified remote host through the specified IP interface, and for a specified VLAN.

- `ip_interface_list`

  ```
  ipinterface_list [ipinterface=IPInterfaceName | address=Address | address6=IPv6address]
  ```

  Use the command to list the configuration of a specific IP interface or all IP interfaces.

- **`ipinterface_vlan_update`**

  Use this command to update the priority code points of all IP interfaces that have a VLAN ID defined.

  ```
  ipinterface_vlan_update vlan_pcp=VlanPcp
  ```

  The VLAN priority code point parameter (PCP) VlanPCP must be an integer 0 - 7. Valid values are 0 - 7.

> **Note:** The system supports a single priority value (PCP field) for all ports. The default value is set to 6. All packets that are leaving the system (outbound communication) feature the same VLAN priority.

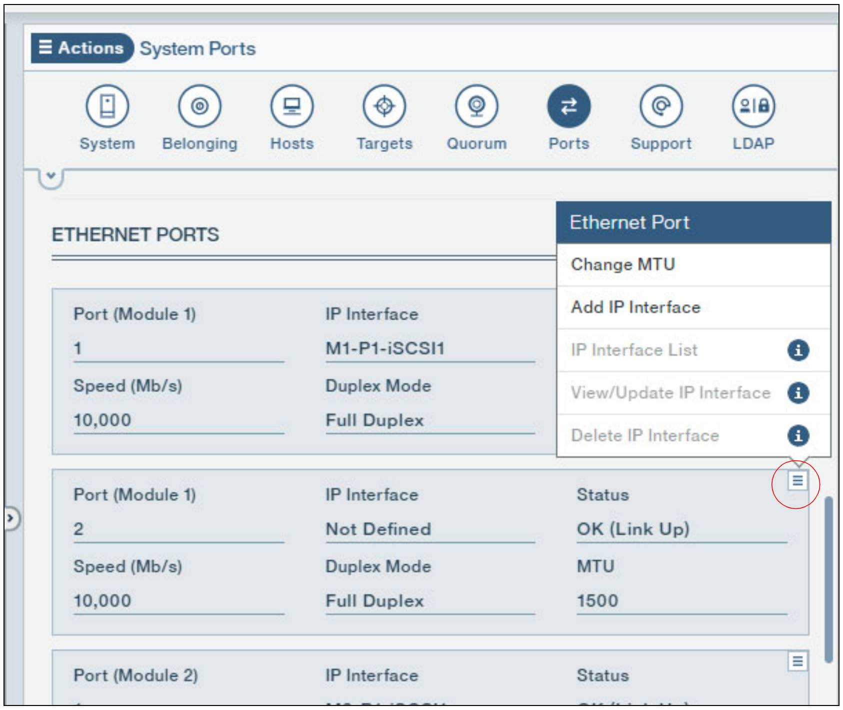
For more information about the commands that are related to interfaces, see *Command-Line Reference Guide for IBM FlashSystem A9000 or A9000R*, SC27-8559.

Managing ports and interfaces with HSM GUI

Several HSM GUI windows were changed in HSM version 5.6 to accommodate VLAN definitions.

From the systems list, select the system that you want to configure. From the Hub View, select the **Ports** spoke or icon.

Scroll down the port list to display the Ethernet ports section and then, select the port to configure and click the **action** icon (which is highlighted by the red circle in Figure 2-24 on page 75) for that specific port to display the Ethernet Port pop-up menu.



ETHERNET PORTS

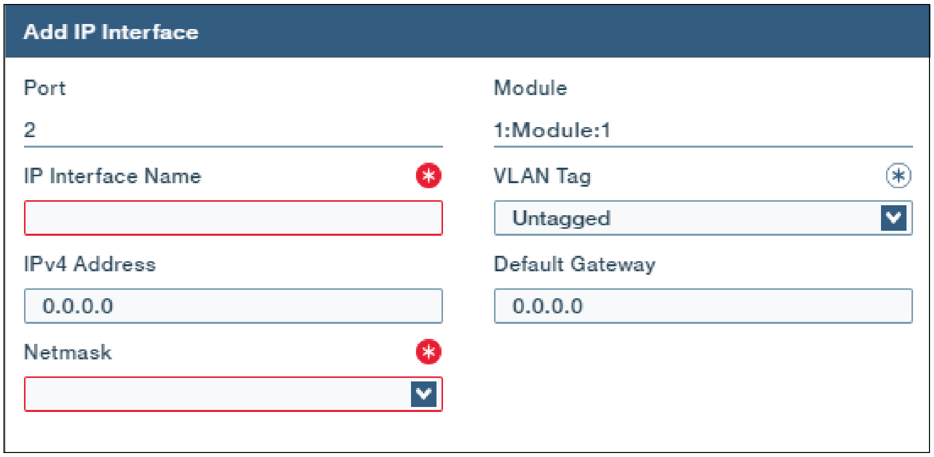
Port (Module 1)	IP Interface	
1	M1-P1-iSCSI1	
Speed (Mb/s)	Duplex Mode	
10,000	Full Duplex	

Port (Module 1)	IP Interface	Status
2	Not Defined	OK (Link Up)
Speed (Mb/s)	Duplex Mode	MTU
10,000	Full Duplex	1500

Port (Module 2)	IP Interface	Status

Figure 2-24 Ethernet Ports list

To define or add a new IP interface, select **Add IP interface**. This selection is the only valid choice if no IP interfaces were defined for that port. This action opens the Add IP Interface window (see Figure 2-25).



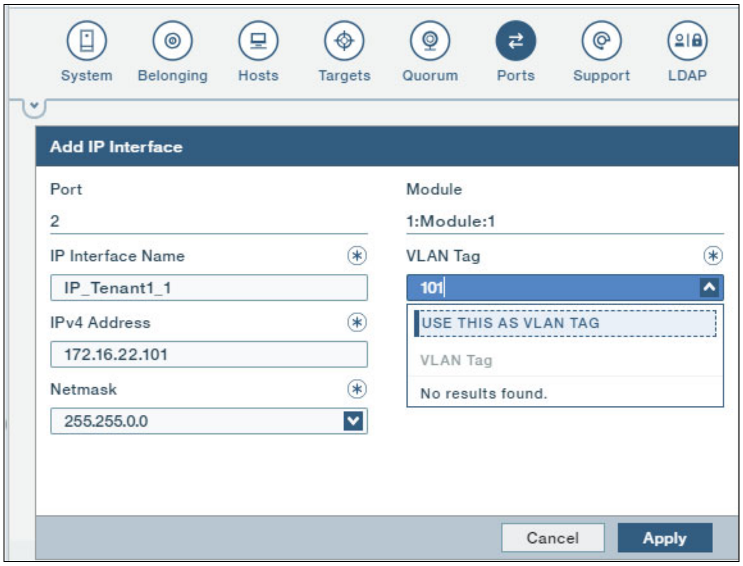
Add IP Interface

Port	Module
2	1:Module:1
IP Interface Name	VLAN Tag
	Untagged
IPv4 Address	Default Gateway
0.0.0.0	0.0.0.0
Netmask	

Figure 2-25 Add IP Interface window

Enter the IP interface name, IP address, and Default Gateway for the IP interface that is being defined.

If you set up VLANs and define more than one IP interface for this particular port (which is the case of our example), assign a VLAN ID. You must enter a number for the VLAN ID, and then select **USE THIS AS VLAN TAG** from the drop-down menu, as shown in Figure 2-26 on page 76.



Add IP Interface

Port: 2

Module: 1:Module:1

IP Interface Name: IP_Tenant1_1

VLAN Tag: 101

IPv4 Address: 172.16.22.101

Netmask: 255.255.0.0

Figure 2-26   Assign VLAN ID

> **Note:** The VLAN Tag is not an arbitrary number; it must match the VLAN configuration as set up on the IP switch by your Network Administrator.

Click **Apply**. You are returned to the Ethernet Ports list. To add another IP interface to the same port, repeat the actions that are shown in Figure 2-25 and Figure 2-26, by specifying a new, unique IP-Interface Name, IP address, and VLAN ID.

After you added multiple IP interfaces, you can review and list them by selecting the IP interlace list option from the port action menu, as shown in Figure 2-27.

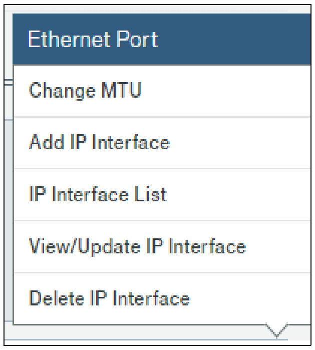
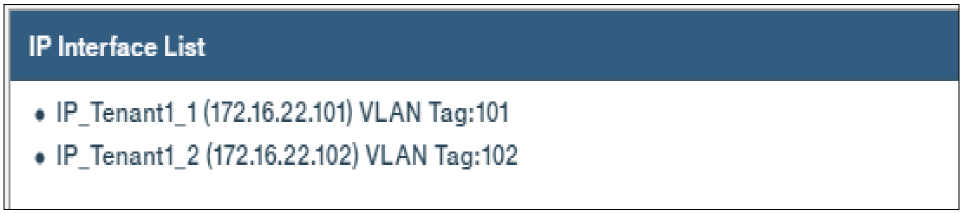


Figure 2-27   Ethernet Port menu

The list of IP interfaces is displayed, as shown inFigure 2-28.

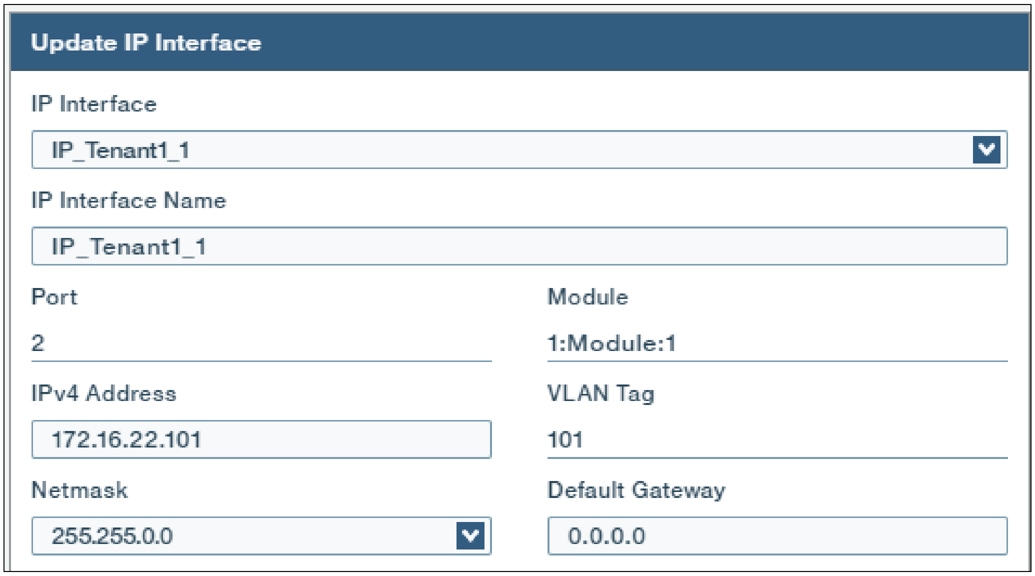


IP Interface List

- IP_Tenant1_1 (172.16.22.101) VLAN Tag:101
- IP_Tenant1_2 (172.16.22.102) VLAN Tag:102

Figure 2-28   IP Interfaces list

If you select View/Update IP Interfaces from the menu that is shown in Figure 2-27 on page 76, the Update IP Interfaces window opens (see Figure 2-29).



Update IP Interface

IP Interface: IP_Tenant1_1

IP Interface Name: IP_Tenant1_1

Port: 2

Module: 1:Module:1

IPv4 Address: 172.16.22.101

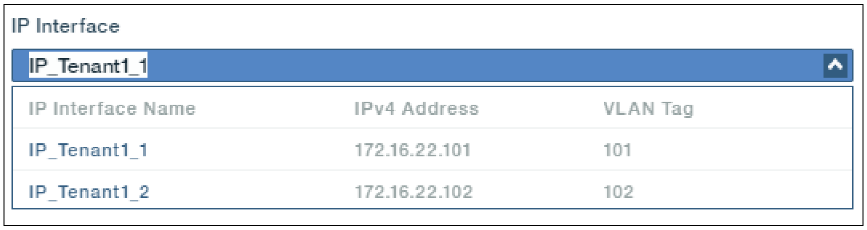
VLAN Tag: 101

Netmask: 255.255.0.0

Default Gateway: 0.0.0.0

Figure 2-29 Update IP Interfaces window

From the IP_Interface pull-down menu that is shown in Figure 2-30, you can select which IP interface to update. You can then change the IP Interface name, which must remain unique on the storage system and the IP address, Netmask, and Gateway.



IP Interface

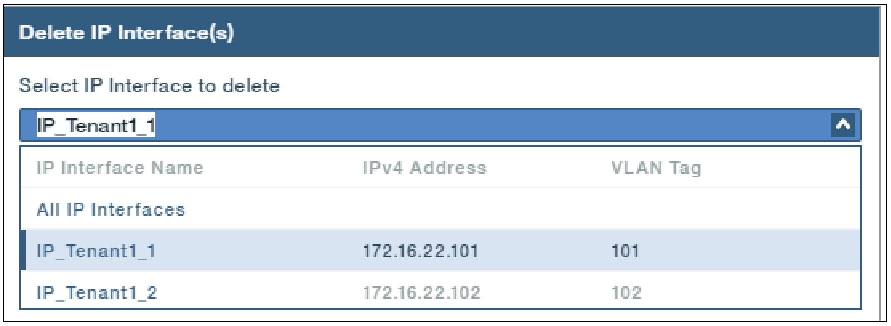
IP_Tenant1_1

IP Interface Name	IPv4 Address	VLAN Tag
IP_Tenant1_1	172.16.22.101	101
IP_Tenant1_2	172.16.22.102	102

Figure 2-30 Select IP Interface to update

The VLAN Tag cannot be updated. If you must change the VLAN Tag, you must first delete the IP interface and start over.

Figure 2-31 shows the Delete IP Interface window. You can select a specific IP interface to delete, or delete them all.



Delete IP Interface(s)

Select IP Interface to delete

IP_Tenant1_1

IP Interface Name	IPv4 Address	VLAN Tag
All IP Interfaces		
IP_Tenant1_1	172.16.22.101	101
IP_Tenant1_2	172.16.22.102	102

Figure 2-31 Delete IP Interface(s) menu

Host connectivity

When VLAN is used, you must assign the host to a port for which you defined IP interfaces with a VLAN tag that matches the VLAN where the host is stored (see Figure 2-21 on page 69).

In the following scenario, we describe the connectivity of a RedHat Linux host that is on a VLAN with a VLAN tag of 50 and connects to three storage systems (A9000 or A9000R) by using an Ethernet switch. The configuration for the scenario is shown in Figure 2-32.

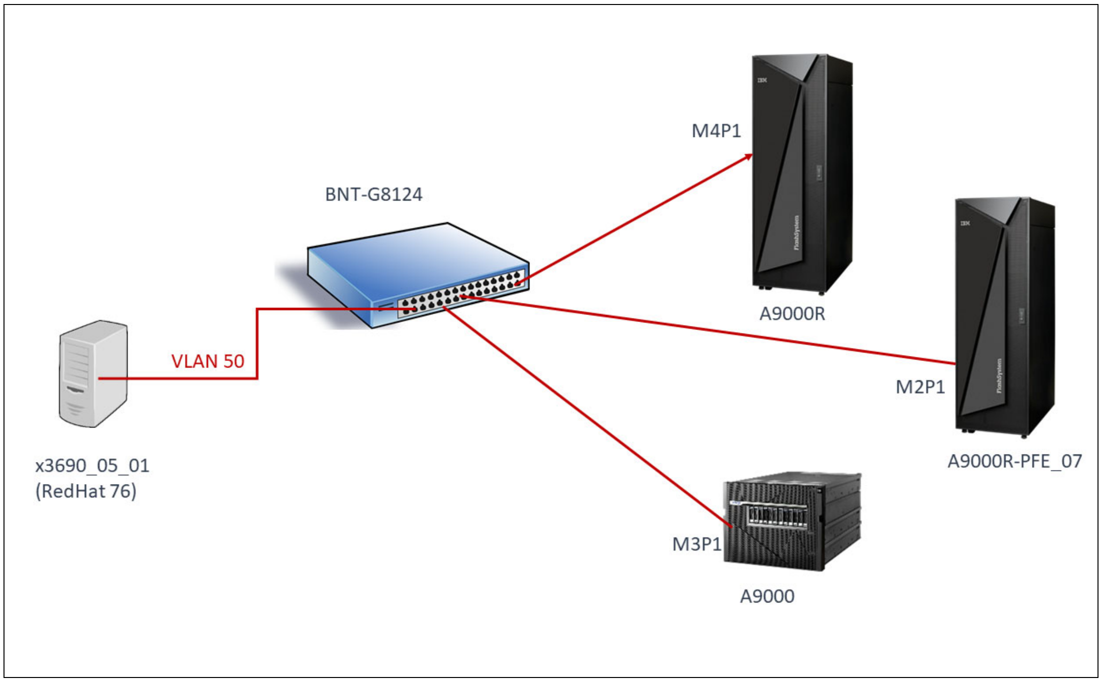


Figure 2-32 Host Connectivity and VLAN

The configuration includes the following components:

- The Redhat76 host is connected to Ethernet switch port 1.
- The A9000R system connects from its Ethernet Port 1 Module 4 to Ethernet switch port 2.
- The A9000R-PFE-07 system connects from its Ethernet Port 1 Module 2 to Ethernet switch port 3.
- The A9000 system connects from its Ethernet Port 1 Module 3 to Ethernet switch port 6.

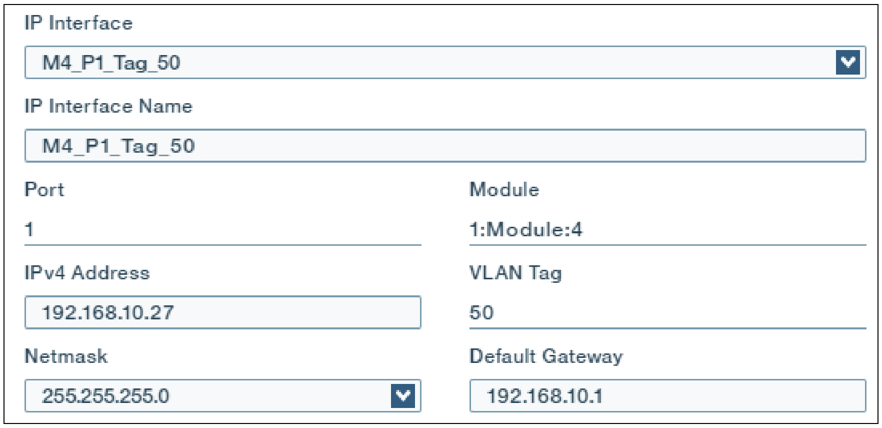
The Ethernet switch configuration for the relevant ports is shown in Example 2-6 on page 79. You can see that the ports to which the RedHAt76 host is attached (port 1) has a port VLAN ID (pvid) of 50 (pvid 50). The ports to which the A9000 systems are attached also have a pvid of 50, and are set for tagging for that pvid (tagging and tag-pvid). We disabled the spanning tree, as recommended.

This example configuration is done by the Network administrator and looks different depending on the brand and model number of a particular switch. In our case, the Ethernet switch is a Blade Network Technologies (BNT) Rackswitch G8124.

Example 2-6 BNT G8124 switch configuration for VLAN 50

```
Current configuration:
!
version "6.3.2"
switch-type "Blade Network Technologies RackSwitch G8124"
!
......
!
interface port 1
        name "RedHat76"
        pvid 50
!
interface port 2
        name "A9000R"
        tagging
        tag-pvid
        pvid 50
!
interface port 3
        name "A9000R-PFE_07"
        tagging
        tag-pvid
        pvid 50
!
interface port 6
        name "A9000"
        tagging
        tag-pvid
        pvid 50
!
vlan 50
        enable
        name "VLAN 50"
        member 1
        member 2
        member 3
        member 6
!
spanning-tree stp 1 vlan 50
--More--
```

Figure 2-33, Figure 2-34, and Figure 2-35 show the Ethernet port and interface definitions for the connected ports in each of the storage Systems. They all feature a VLAN tag of 50.



IP Interface

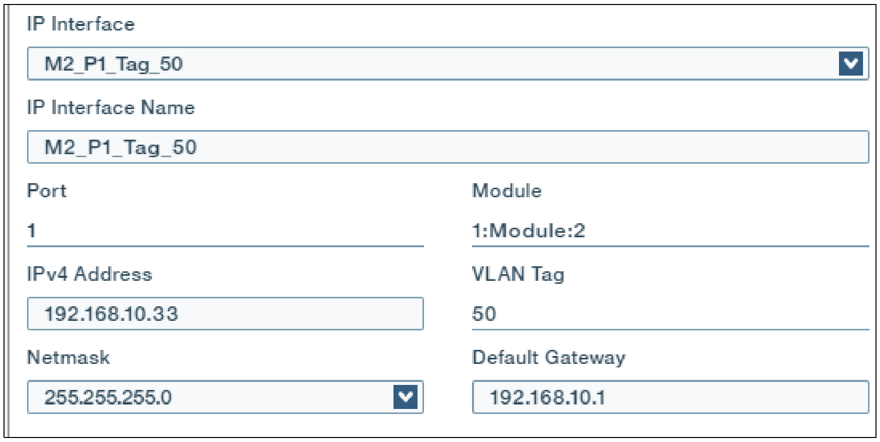
M4_P1_Tag_50

IP Interface Name

M4_P1_Tag_50

Port	Module
1	1:Module:4
IPv4 Address	VLAN Tag
192.168.10.27	50
Netmask	Default Gateway
255.255.255.0	192.168.10.1

Figure 2-33 IP interface definition for Port 1 Module 4 in A9000R



IP Interface

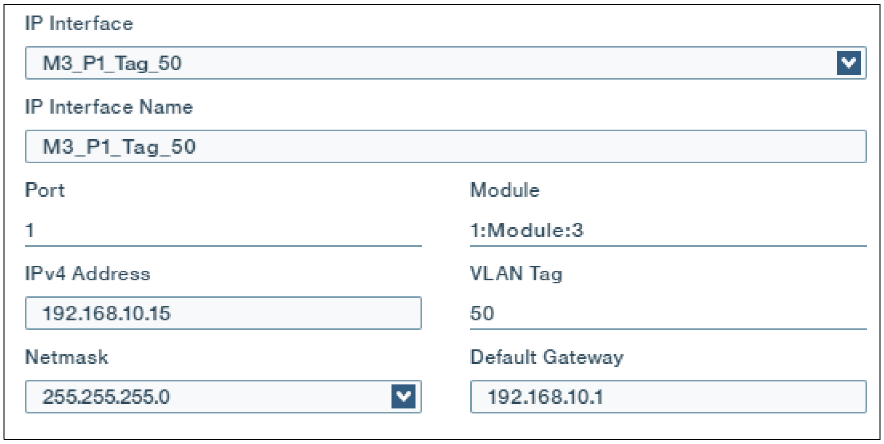
M2_P1_Tag_50

IP Interface Name

M2_P1_Tag_50

Port	Module
1	1:Module:2
IPv4 Address	VLAN Tag
192.168.10.33	50
Netmask	Default Gateway
255.255.255.0	192.168.10.1

Figure 2-34 IP interface definition for Port 1 Module 2 in A9000R_PFE_07



IP Interface

M3_P1_Tag_50

IP Interface Name

M3_P1_Tag_50

Port	Module
1	1:Module:3
IPv4 Address	VLAN Tag
192.168.10.15	50
Netmask	Default Gateway
255.255.255.0	192.168.10.1

Figure 2-35 P interface definition for Port 1 Module 3 in A9000

The RedHat76 host is added to each of the Storage Systems, with connection to the corresponding ports. Figure 2-36 shows the resulting connectivity in the HSM GUI.

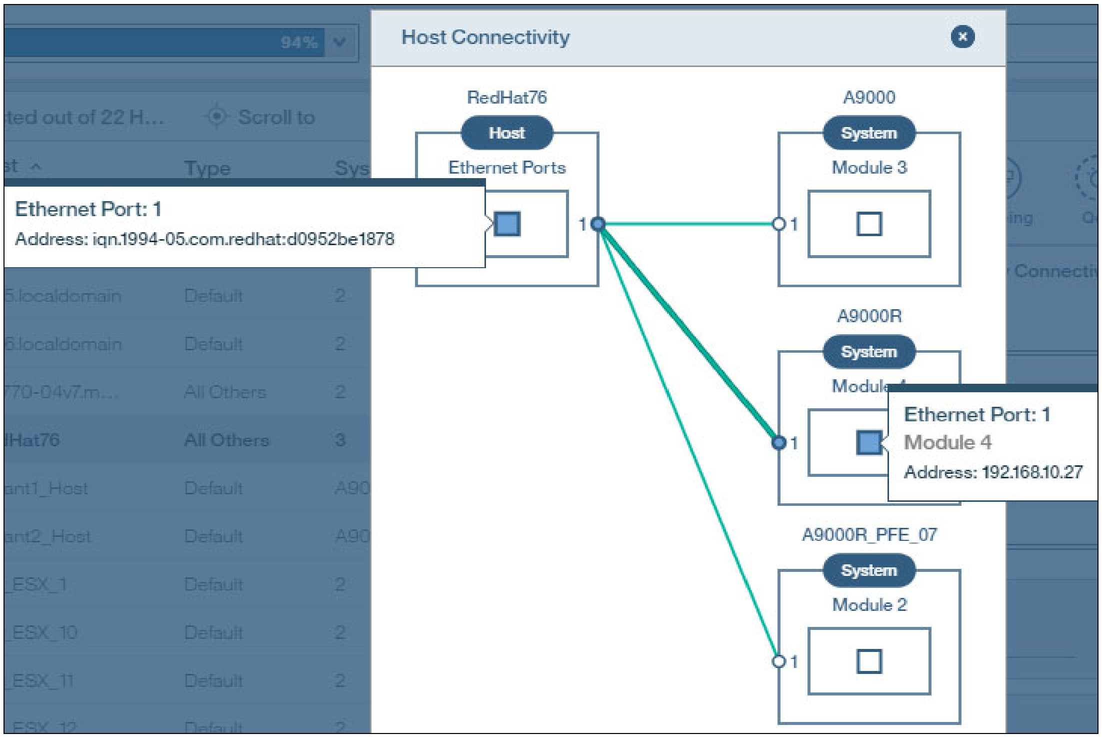


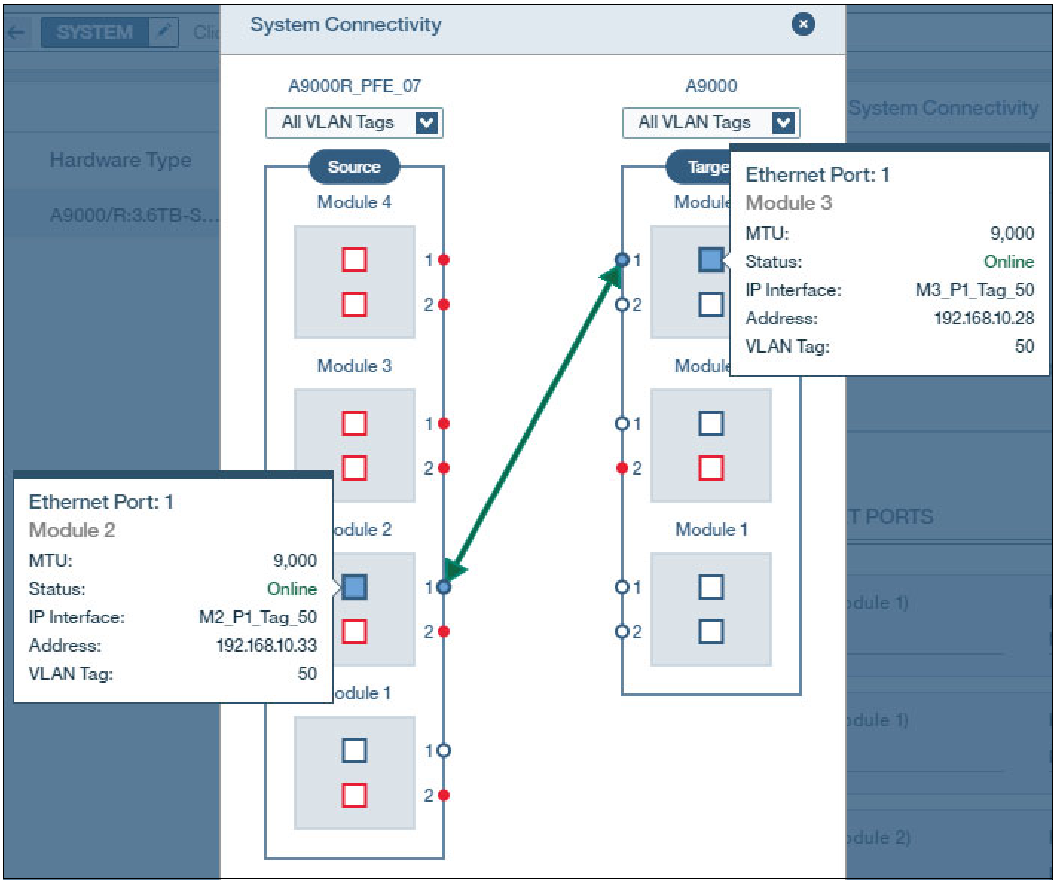
Figure 2-36 Host connectivity

Mirroring connectivity

If you use ports that have an IP interface with a specified VLAN tag for replication, you can have only one interface that is defined on that port; otherwise, the system rejects the connection. the same restriction applies to all non-data ports, such as VPN and management ports.

Figure 2-37 on page 82 shows an example of iSCSI mirroring connectivity between storage systems over Ethernet ports. Each connected port has only one IP interface with a VLAN Tag defined.

> **Note:** The HSM GUI does not support target autoconfiguration for iSCSI connectivity. If VLANs are configured, the iSCSI target connectivity must be configured manually.



Ethernet Port: 1
Module 3

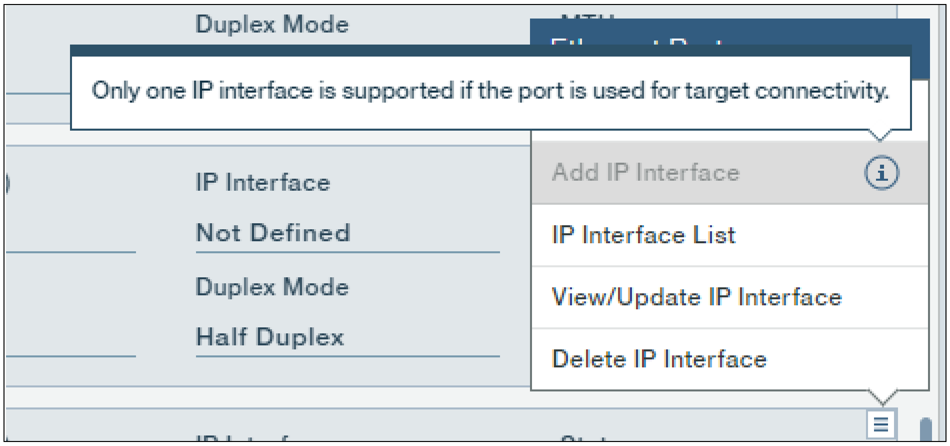
MTU:	9,000
Status:	Online
IP Interface:	M3_P1_Tag_50
Address:	192.168.10.28
VLAN Tag:	50

Ethernet Port: 1
Module 2

MTU:	9,000
Status:	Online
IP Interface:	M2_P1_Tag_50
Address:	192.168.10.33
VLAN Tag:	50

Figure 2-37 iSCSI mirroring connectivity with VLAN

If you go to the ports definition, the Add interface option is grayed out in the Action menu for that port. The information message indicates that only one IP interface is supported when the port is used for target connectivity (see Figure 2-38).

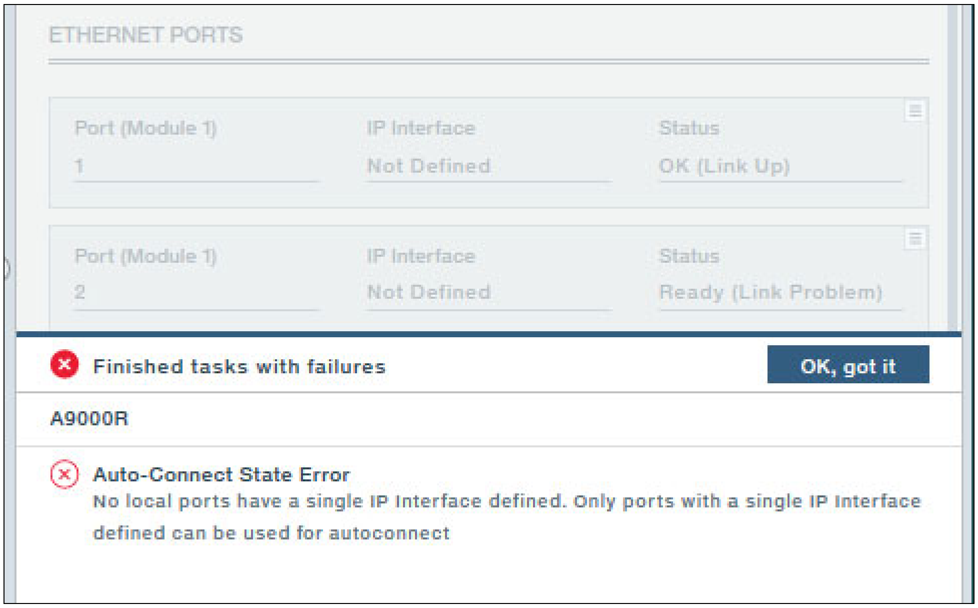


Only one IP interface is supported if the port is used for target connectivity.

IP Interface	Not Defined
Duplex Mode	Half Duplex

Figure 2-38 Only one IP interface for target port

However, if you define multiple interfaces on a port that you then try to use as a target connectivity port, you cannot establish the connectivity, as shown in Figure 2-39.



Finished tasks with failures

A9000R

Auto-Connect State Error
No local ports have a single IP Interface defined. Only ports with a single IP Interface defined can be used for autoconnect

Figure 2-39 Target connectivity rejected

2.5 VLAN usage scenario

VLANs are typically used to segregate traffic on the same physical network. The ability to limit broadcast network traffic and secure portions of the network is particularly appealing to service providers who want to share host, network, and storage resources among multiple tenants.

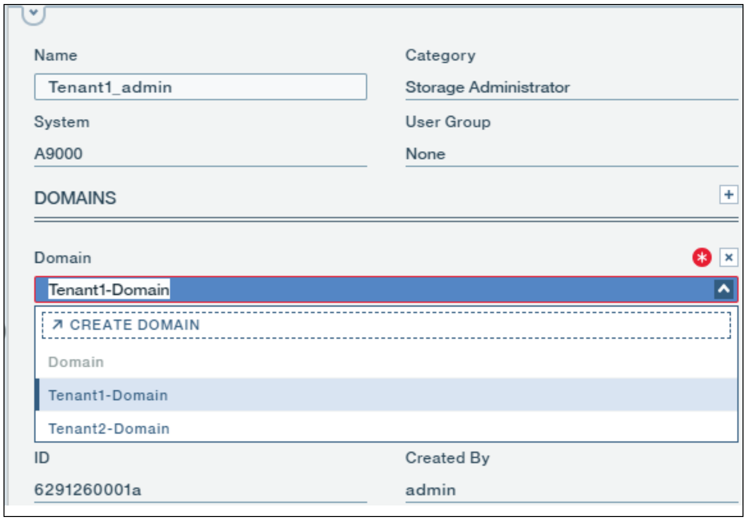
FlashSystem A9000 or A9000R also support the concept of multitenancy through the definition of independent administrative domains. A domain is a logical partitioning of the system (more than one can exist in a system) with its own administrators and storage resources. Domains are defined by the global storage administrator.

When defining a domain, a global administrator also assigns another storage administrator user as the domain administrator. After assigned, the designated domain administrator user loses their global privileges. User Domain administrators cannot modify physical system resources and they cannot access resources that belong to another domain, including the “global domain” resources of the system.

A domain perfectly addresses the needs of a multitenant environment with its inherent capability of isolating some system resources within secure domains. With software release V12.3.2, the definition of domains can now also be restricted to a specific VLAN by including hosts in the domain that are assigned to specific VLANs only.

Consider the following scenario:

1. The global storage administrator creates a domain for each tenant (called `Tenant1-Domain` and `Tenant2-Domain`), creates user administrator `Tenant1_admin` and `Tenant2_admin`, and assign them as domain administrator for `Tenant1-Domain` and `Tenant2_Domain`, as shown in Figure 2-40 on page 84 for `Tenant1-admin`.



Name: Tenant1_admin

Category: Storage Administrator

System: A9000

User Group: None

DOMAINS

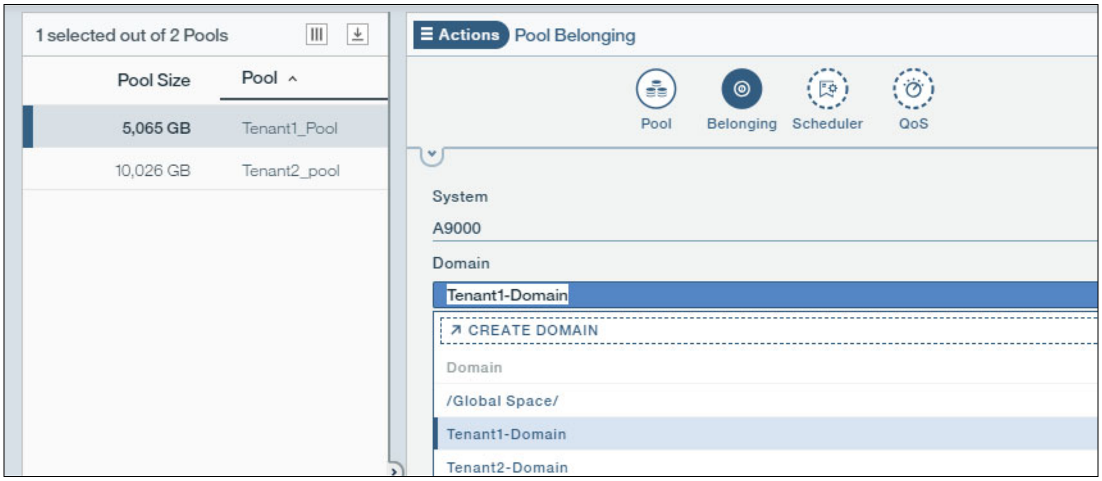
Domain: Tenant1-Domain

ID: 6291260001a

Created By: admin

Figure 2-40 Assigning a domain administrator

2. The global administrator now creates and assigns a pool `Tenant1_pool` and `Tenant2_Pool` for each domain (see Figure 2-41).



1 selected out of 2 Pools

Pool Size	Pool ^
5,065 GB	Tenant1_Pool
10,026 GB	Tenant2_pool

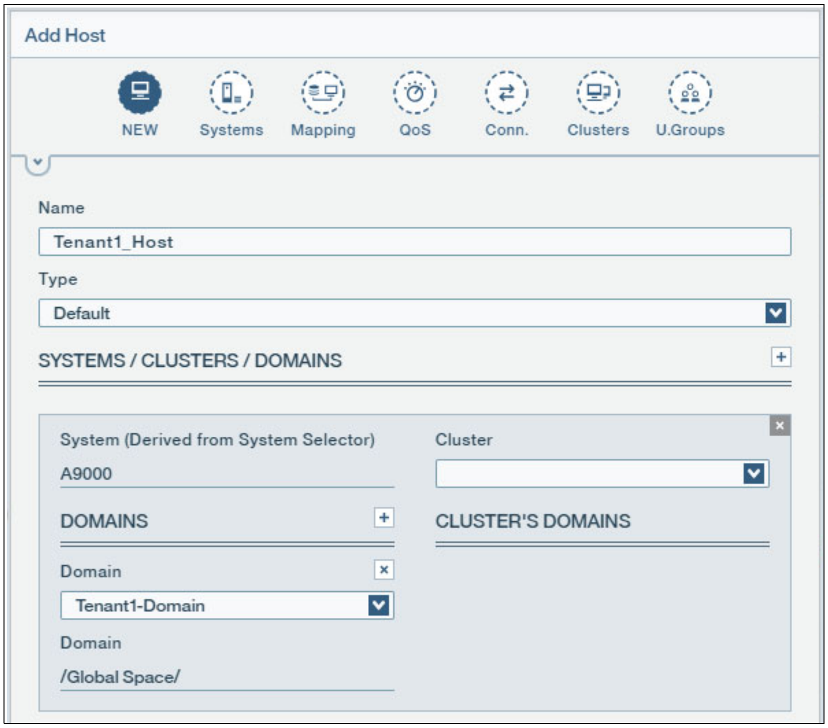
System
A9000

Domain
Tenant1-Domain

Figure 2-41 Assign pools to domains

3. The Global Admin defines interfaces and assigns VLAN IDs. Ethernet ports configuration and interface set up must be done by the global administrator because a domain administrator cannot configure a global physical resource, such as Ethernet ports.

4. The Global Admin creates the hosts for the respective tenants and assigns hosts to respective domains, as shown for host `Tenant1_Host` in Figure 2-42.



Name: Tenant1_Host

Type: Default

SYSTEMS / CLUSTERS / DOMAINS

System (Derived from System Selector): A9000

Cluster:

DOMAINS

Domain: Tenant1-Domain

Domain: /Global Space/

CLUSTER'S DOMAINS

Figure 2-42 Create Tenant1_host and assign to Tenant1-Domain

5. Define connectivity by attaching the host to a port. For an iSCSI connection, the host adapter ID is the host IQN. Knowing which VLAN the host belong to, attach it to the A9000 port on which you defined an interface for the corresponding VLAN, as described in “Host connectivity” on page 78.

2.6 Logical configuration for host connectivity

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

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 FlashSystem A9000 or A9000R system.
9. Complete any operating system-specific tasks.
10. If the server will SAN boot, install the operating system.
11. Install multipath drivers, if required. For information about installing multipath drivers, see the appropriate section from the host-specific chapters of this book.
12. Reboot the host server or scan new disks.

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

2.6.1 Host configuration preparation

Write down the component names and IDs to save time during the implementation.

FC host-specific tasks

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

For more information about configuring zoning, see your FC switch documentation.

iSCSI host-specific tasks

For iSCSI connectivity, ensure that any configurations, such as VLAN membership or port configuration, are completed so that the hosts and the FlashSystem A9000/R can communicate over IP.

2.6.2 Assigning LUNs to a host by using the GUI

Many steps are involved in defining a new host and assigning LUNs to it. You must create the volumes in a storage system in advance.

Defining a host

To define a host, complete the following steps:

1. From the main GUI Dashboard, click **New**, and select **Host** from the menu (see Figure 2-43).

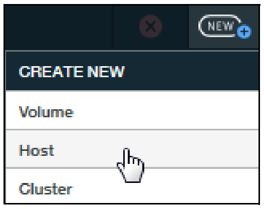
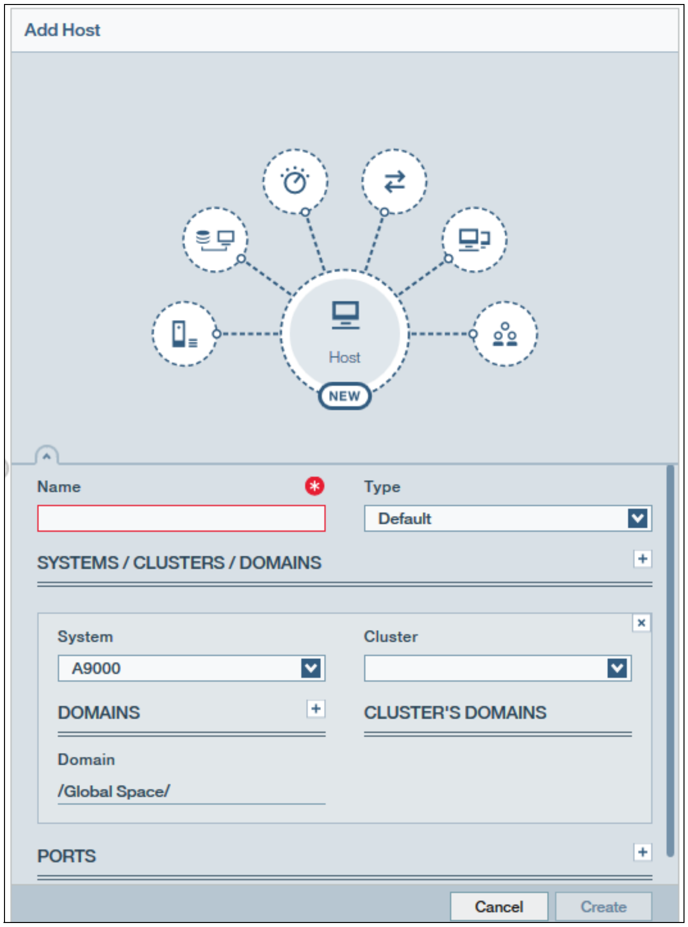


Figure 2-43 The Create New (hosts and clusters) menu

2. The new Host view is displayed. The Add Host setting window opens (see Figure 2-44) where you provide information.



Name

Type

Default

System

A9000

Cluster

Domain

/Global Space/

Figure 2-44 Add Host details

Enter the following required information:

- Specify a host name (required).
- Select the type. In this example, **Default** is selected. If you use Windows 2008, HP-UX or z/VM host, you must change the type to match your host type.

For all other hosts, such as AIX, Linux, Solaris, VMware, and Windows (except Windows 2008), the Default (or All Others) option is correct. Note that starting with HAK version 2.7.0, HP-UX and Solaris are no longer supported.

- Select whether you want an iSCSI or FC connection. Host access to LUNs is granted depending on the host adapter ID. For an FC connection, the host adapter ID is the FC HBA WWPN. For an iSCSI connection, the host adapter ID is the host IQN. To add a WWPN or an IQN to a host definition, click the plus sign (**+**) icon to the right of the PORTS heading (see Figure 2-44 on page 87).

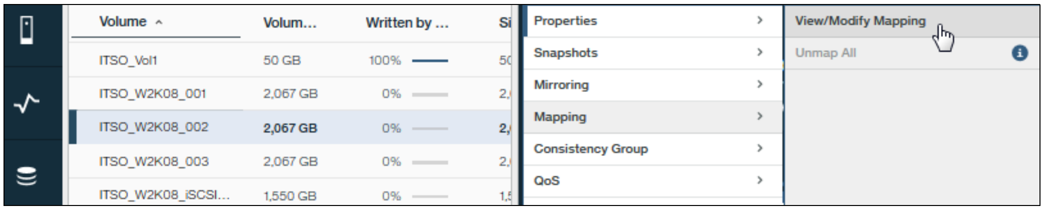
  Ports can be added in any order.

- Specify to which FlashSystem A9000 or A9000R in your inventory you want to attach the host. Alternatively, you can also add the host to a cluster. (A cluster can be created from this view.)
- Optional: Add the host to a domain. (A domain can also be created from this view.)

Mapping LUNs to a host

The final configuration steps are to map LUNs to the host as follows:

1. From the **Pools and Volumes** Workspace →**Volumes** view, select the volume to map, and select **Actions** →**Mapping** →**View/Modify Mapping** (see Figure 2-45).



Volume ^	Volum...	Written by ...	Si
ITSO_Vol1	50 GB	100%	50
ITSO_W2K08_001	2,067 GB	0%	2,
ITSO_W2K08_002	2,067 GB	0%	2,
ITSO_W2K08_003	2,067 GB	0%	2,
ITSO_W2K08_iSCSI...	1,550 GB	0%	1,5

Figure 2-45 Mapping a LUN to a host

2. The Volume Mapping window is displayed (see Figure 2-46). If no mapping exists yet, click the plus sign (**+**) to add a mapping.

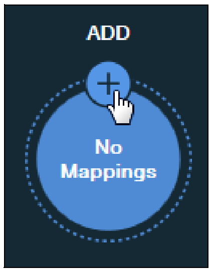
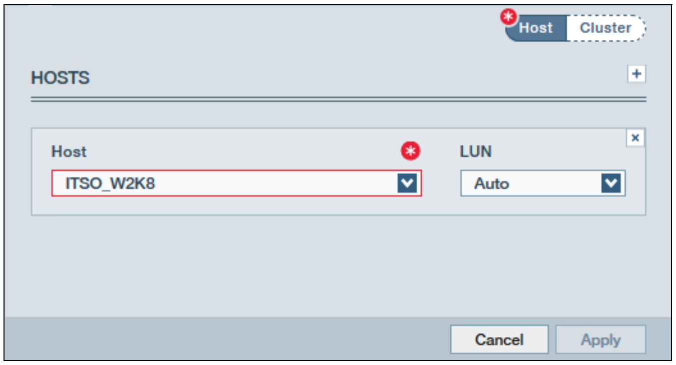


Figure 2-46 Add a mapping

3. In the Hosts window (see Figure 2-47 on page 89), select an available host from the pull-down list. The GUI suggests a default (Auto) LUN ID to which to map the volume. However, this ID can be changed to meet your requirements. Click **Apply** and the volume is assigned immediately.



Host: ITSO_W2K8

LUN: Auto

Figure 2-47 Mapping a volume to a host

No differences exist between mapping a volume to an FC host or an iSCSI host in the GUI.

2.6.3 Assigning LUNs to a host by using the XCLI

Several steps are involved to define a new host and assign LUNs to it. You must create the volumes in a storage pool in advance.

Defining a new host

To use the CLI to prepare for a new host, complete the following steps:

1. Create a host definition for your FC and iSCSI hosts by using the **host_define** command, as shown in Example 2-7.

   Example 2-7 Creating a host definition

   ```
   >> host_define host=itso_win2012
   Command executed successfully.

   >> host_define host=itso_win2012_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.

   As Example 2-8 shows, the WWPNs of the FC host for HBA1 and HBA2 are added with the **host_add_port** command by specifying an **fcaddress**.

   Example 2-8 Creating the FC port and adding it to the host definition

   ```
   >> host_add_port host=itso_win2012 fcaddress=21000024FF24A426
   Command executed successfully.

   >> host_add_port host=itso_win2012 fcaddress=21000024FF24A427
   Command executed successfully.
   ```

   As Example 2-9 shows, the IQN of the iSCSI host is added. This command is the same **host_add_port** command, but it uses the **iscsi_name** parameter.

   Example 2-9 Creating an iSCSI port and adding it to the host definition

   ```
   >> host_add_port host=itso_win2012_iscsi
   iscsi_name=iqn.1991-05.com.microsoft:sand.storage.tucson.ibm.com
   Command executed successfully.
   ```

Mapping LUNs to a host

To map the LUNs, complete these steps:

1. Map the LUNs to the host definition. For a cluster, the volumes are mapped to the cluster host definition. No differences exist between the FC mapping to a host and the iSCSI mapping to a host. Both commands are shown in Example 2-10.

   Example 2-10 Mapping 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. Power on the server and check the host connectivity status from the storage system point of view by using the **`host_connectivity_list host=host_name`** command. Example 2-11shows the output for both hosts.

   Example 2-11 Checking host connectivity (XCLI)

   ```
   >> host_connectivity_list host=itso_win2008
   Host           Host Port          Module       Local FC port    Local iSCSI port
   itso_win2008   21000024FF24A427   1:Module:1   1:FC_Port:1:1
   itso_win2008   21000024FF24A427   1:Module:2   1:FC_Port:2:1
   itso_win2008   21000024FF24A427   1:Module:3   1:FC_Port:3:1
   itso_win2008   21000024FF24A426   1:Module:4   1:FC_Port:4:3
   itso_win2008   21000024FF24A426   1:Module:5   1:FC_Port:5:3
   itso_win2008   21000024FF24A426   1:Module:6   1:FC_Port:6:3

   >> 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:1                    iSCSI
   itso_win2008_iscsi   iqn.1991-05.com.microsoft:sand.storage.tucson.ibm.com   1:Module:2                    iSCSI
   ```

At this stage, operating system-dependent steps might need to be performed. These steps are described in the operating system chapters.

2.7 Performance tuning

This section provides some performance considerations to help you adjust your operating system to best fit your environment. The following performance considerations are for the operating system:

- Use multiple threads and asynchronous I/O to maximize performance on the FlashSystem A9000 and A9000R.
- Check with **`iostat`**, **`perfmon`**, or **`esxtop`** on a per path basis for the LUNs to make sure that the load is balanced across all paths.
- Verify the HBA queue depth and per device queue depth for the host are sufficient to prevent queue waits. However, make sure that they are not so large that they overrun the storage system queues. For FlashSystem A9000 and A9000R queue limit is 2048 per storage system port and 256 per LUN per WWPN (host) per port. Do not submit more I/O per storage system port than the 2048 maximum it can handle.

Check the device queue depth:

- To check the device queue depth on AIX, periodically run **iostat -D 5**. If the avgwqsz (average wait queue size) or sqfull is consistently greater zero, increase the device queue depth.

  For more information about AIX device queue depth tuning, see the following web page:

  http://www-01.ibm.com/support/docview.wss?uid=tss1td105745

- To check the device queue depth on Linux, periodically run **iostat -dx 5**. If the await-svctm values are consistently greater than zero or %util (disk utilization) is constantly close to 100%, increase the device queue depth.

  For more information about Linux disk tuning, see the following web page:

  http://cromwell-intl.com/linux/performance-tuning/disks.html

- To check the device queue depth on VMware ESXi, run **esxtop**, press the U key. The queue depth is listed under LQLEN. If the %USD (percentage of queue depth used) is constantly close to 100%, increase the device queue depth.

  For more information about VMware ESXi queues, see the following web page:

  http://blogs.vmware.com/apps/2015/07/queues-queues-queues-2.html

- To check the device queue depth on Windows, run **Administrative Tools** → **Performance Monitor** and select **Physical Disk** →**Current Disk Queue Length**.

  For more information about Windows disk performance monitoring, see the following web page:

  https://blogs.technet.microsoft.com/askcore/2012/03/16/windows-performance-monitor-disk-counters-explained/

> **Important:** Do not start at the maximum and work down (except for Windows, which starts by default at maximum) because you might flood the storage system with commands and waste memory on the host.

2.8 Troubleshooting

Troubleshooting connectivity problems can be difficult. However, FlashSystem A9000 and A9000R have built-in troubleshooting tools. Some of the built-in tools are listed in Table 2-4. For more information, see the CLI manual that is available at IBM Knowledge Center:

https://www.ibm.com/support/knowledgecenter/en/STJKMM_12.1.0/fs9k_kc_welcome.html

Table 2-4 Connectivity troubleshooting commands

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 Address Resolution Protocol (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.

3

Windows connectivity

This chapter addresses the specific considerations for attaching various Microsoft Windows hosts to IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System.

> **Important:** The procedures and instructions that are provided here are based on code that was available at the time of writing. For the latest support information and instructions, see IBM System Storage Interoperation Center (SSIC):
>
> https://www.ibm.com/systems/support/storage/ssic/interoperability.wss
>
> In addition, you can download the Host Attachment Kit and related publications from Fix Central:
>
> http://www.ibm.com/support/fixcentral/

This chapter includes the following topics:

3.1 Prerequisites

To successfully attach a Windows host to IBM FlashSystem A9000 or A9000R, or IBM XIV and access storage, a number of prerequisites must be met. Generic prerequisites are indicated in the following list. Your environment might have extra requirements. Further prerequisites are also be outlined in specific sections for specific Windows versions.

- Complete the cabling.
- Complete the zoning.
- Install any required service packs and fixes.
- Create volumes to be assigned to the host.

3.1.1 Supported versions of Windows

At the time of this writing, the supported versions of Windows (including cluster configurations) are listed in Table 3-1.

Table 3-1 Supported Windows operating versions

Operating system	Storage	Supported with HAK version[a]
Microsoft Windows Server 2016	XIV, A9000/R	2.6.0 - 2.8.2
Microsoft Windows Server 2016 Hyper-V	XIV, A9000/R	2.6.0 - 2.8.2
Microsoft Windows Server 2012 R2	XIV, A9000/R	2.3.0 - 2.8.2
Microsoft Windows Server 2012 R2 Hyper-V	XIV, A9000/R	2.3.0 - 2.8.2
Microsoft Windows Server 2012	XIV, A9000/R	2.3.0 - 2.8.2
Microsoft Windows Server 2012 Hyper-V	XIV, A9000/R	2.3.0 - 2.8.2
Microsoft Windows Server 2008 R2	XIV, A9000/R	2.3.0 - 2.8.2
Microsoft Windows Server 2008 R2 Hyper-V	XIV, A9000/R	2.3.0 - 2.8.2
Microsoft Windows Server 2008	XIV, A9000/R	2.3.0 - 2.8.2
Microsoft Windows Server 2008 Hyper-V	XIV, A9000/R	2.3.0 - 2.8.2
Microsoft Windows Server 2003 R2 - EOL[b]	XIV	2.2.0 (end of support 12/2016)[c]
Microsoft Windows Server 2003 - EOL[b]	XIV	2.2.0 (end of support 12/2016)[c]

[a] Windows host attachment for IBM FlashSystem A9000 and A9000R is only supported with Host Attachment Kit (HAK) version 2.6.0 or later. Starting with version 2.6.0, IBM XIV Host Attachment Kit was renamed to *IBM Storage Host Attachment Kit.*

[b] End of life (EOL). This is where interoperability items are no longer supported by the vendor either generally or by extended service contract. IBM will continue to support the environment where possible. Where issues occur that are deemed by IBM support to be directly related to items that are no longer generally supported by the vendor, IBM might direct customers to upgrade a component to a recommended level.

[c] The Host Attachment Kit (HAK) lifecycle and support matrix. See details of the HAK lifecycle with compatible storage system microcode versions and supported OS releases: https://www.ibm.com/support/knowledgecenter/en/SSEPRF/landing/css_lifecycle_support_matrix_hak.html

3.1.2 Supported FC HBAs

Supported FC HBAs are available from Brocade, Emulex, IBM, and QLogic. For more information about driver versions, see this web page:

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

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

3.1.3 Multipath support and Clustering Options

FlashSystem A9000, A9000R, and XIV Storage System support the multipath solutions and clustering options as listed in Table 3-2.

Table 3-2 MultiPathing support and clustering options

Operating system	XIV		A9000/R	
	Multipathing	**Clustering**	**Multipathing**	**Clustering**
Windows Server 2003	IBM Device Specific Module (XIV DSM)[c], Microsoft MPIO[a], Symantec Veritas Volume Manager with DMP[b], Symantec Veritas Volume Manager with SDDDSM[c]	Microsoft Cluster Service, Microsoft MSCS GeoCluster, Symantec Veritas Cluster Server	N/A	N/A
Windows Server 2008 Windows Server 2008 Hyper-V	IBM Device Specific Module (XIV DSM)[c], Microsoft MPIO[a], Symantec Veritas Volume Manager with DMP[b], Symantec Veritas Volume Manager with SDDDSM[c]	Microsoft Cluster Service, Microsoft MSCS GeoCluster, Microsoft Windows Failover Clustering, Symantec Veritas Cluster Server	Microsoft MPIO[a]	Microsoft Cluster Service, Microsoft Windows Failover Clustering
Windows Server 2008 R2 Windows Server 2008 R2 Hyper-V	IBM Device Specific Module (XIV DSM)[c], IBM SDD DSM[c], Microsoft MPIO[a], Symantec Veritas Volume Manager with DMP[b], Symantec Veritas Volume Manager with SDDDSM[c]	Microsoft Cluster Service, Microsoft MSCS GeoCluster, Microsoft Windows Failover Clustering, Symantec Veritas Cluster Server	Microsoft MPIO[a]	Microsoft Cluster Service, Microsoft Windows Failover Clustering

Operating system	XIV		A9000/R	
	Multipathing	**Clustering**	**Multipathing**	**Clustering**
Windows Server 2012 Windows Server 2012 Hyper-V	IBM Device Specific Module (XIV DSM)[c], Microsoft MPIO[a], Symantec Veritas Volume Manager with DMP[b], Symantec Veritas Volume Manager with SDDDSM[c]	Microsoft Cluster Service, Microsoft MSCS GeoCluster, Microsoft Windows Failover Clustering, Symantec Veritas Cluster Server	Microsoft MPIO[a]	Microsoft Cluster Service, Microsoft Windows Failover Clustering
Windows Server 2012 R2 Windows Server 2012 R2 Hyper-V	IBM Device Specific Module (XIV DSM)[c], Microsoft MPIO[a], Symantec Veritas Volume Manager with DMP[b], Symantec Veritas Volume Manager with SDDDSM[c]	Microsoft Cluster Service, Microsoft MSCS GeoCluster, Microsoft Windows Failover Clustering, Symantec Veritas Cluster Server	Microsoft MPIO[a]	Microsoft Cluster Service, Microsoft Windows Failover Clustering
Windows Server 2016 Windows Server 2016 Hyper-V	Microsoft MPIO[a]	Microsoft Cluster Service, Microsoft Windows Failover Clustering, Microsoft MSCS GeoCluster (Windows 2016 only)	Microsoft MPIO[a]	Microsoft Cluster Service, Microsoft Windows Failover Clustering

a. Native multipath I/O (MPIO) installed from the Server Manager. MPIO allows the host HBAs to establish multiple sessions with the same target LUN, but present them to Windows as a single LUN. The Windows MPIO driver enables a true active/active path policy, allowing I/O over multiple paths simultaneously. Starting with Microsoft Windows 2008, the MPIO device driver is part of the operating system.

b. Veritas Dynamic Multipathing (DMP) 5.1.

c. The driver development kit allows storage vendors to create DSMs for MPIO. You can use DSMs to build interoperable multi-path solutions that integrate tightly with the Microsoft Windows family of products.

For more information about Microsoft MPIO, see the online guide at this web page:

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

Note: Using more than one multipath I/O framework on the same host is not supported.

3.1.4 Required software on the host

Prior to installing the IBM Storage Host Attachment Kit for Windows and depending on the installed Windows Server version, specific operating system updates must be installed on the host. These updates are required for error-free functionality.

Important: This book has information that is known at the time of the General Availability (GA) date. Newer or additional fixes might be required in distinct cases or in particular production environments. Contact IBM Support if you encounter any errors or difficulties.

Although the Host Attachment Kit automatically installs certain Microsoft hotfixes for Windows Server, you should always ensure that the most updated and relevant hotfixes are installed manually in addition to the bundled hotfixes. You can find any relevant fix and download it from the following Microsoft Support website:

http://support.microsoft.com

Always see the Microsoft documentation (KB article) for required hotfixes.

3.1.5 Boot from SAN support

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

- Windows Server 2016 with MSDSM
- Windows Server 2012 R2 with MSDSM
- Windows Server 2012 with MSDSM
- Windows Server 2008 R2 with MSDSM
- Windows Server 2008 with MSDSM
- Windows Server 2003 with XIVDSM (IBM XIV Storage System only)

For more information about SAN boot see 1.2.5, “Boot from SAN on x86 or x64 based architecture” on page 17.

3.2 Attaching a Microsoft Windows Server 2016, 2012 R2, or 2008 R2 host

This section highlights specific instructions for Fibre Channel (FC) and internet Small Computer System Interface (iSCSI) connections. All the information here relates only to Windows Server 2016, 2012 R2 and 2008 R2 unless otherwise specified.

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

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

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

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

3.2.1 Windows Server 2012 R2 and 2008 R2 Prerequisites

For more information about general prerequisites to successfully attach a Windows host to FlashSystem A9000, A9000R, or XIV and access storage, see 3.1, “Prerequisites” on page 94.

Table 3-3 on page 98 lists the prerequisites in addition to those prerequisites that were described and are specific to Windows Server 2012 R2 and 2008 R2.

Table 3-3 Windows 2008 R2 and 2012 R2 required software

Operating system	Service pack	Required components that must be installed manually	Automatically installed by the Host Attachment Kit
Windows Server 2008 R2	Service Pack 1	► Microsoft Hotfix set* KB2468345 (http://support.microsoft.com/kb/2468345) ► Microsoft Hotfix set* KB2545685 (http://support.microsoft.com/kb/2545685) if Windows Clustering is used.	► Microsoft Hotfix KB2460971 (http://support.microsoft.com/kb/2460971) ► Microsoft Hotfix KB2522766 (http://support.microsoft.com/kb/2522766)
	None	► Microsoft Multipath I/O Framework (enable from the Server Manager applet). ► Microsoft Hotfix set* KB980054 (http://support.microsoft.com/kb/980054) if Windows Clustering is used.	► Microsoft Hotfix KB979711 (http://support.microsoft.com/kb/979711) ► Microsoft Hotfix KB981208 (http://support.microsoft.com/kb/981208) ► Microsoft Hotfix KB2460971 (http://support.microsoft.com/kb/2460971) ► Microsoft Hotfix KB2522766 (http://support.microsoft.com/kb/2522766)
Windows Server 2012 R2	None	None	► Microsoft Hotfix set* KB4019217 (http://support.microsoft.com/kb/4019217)
* Includes more than one hotfix. Refer to the Microsoft KB article on the provided link			

3.2.2 Windows host FC configuration

This section describes attaching to FlashSystem A9000, A9000R, or XIV over FC and provides detailed descriptions and installation instructions for the various software components required.

Installing HBA drivers

Windows Server 2016, 2012 R2 and 2008 R2 include drivers for many HBAs. However, they probably are not the latest versions. Install the latest available driver that is supported. HBA drivers are available from the supported HBA vendor website, and include instructions.

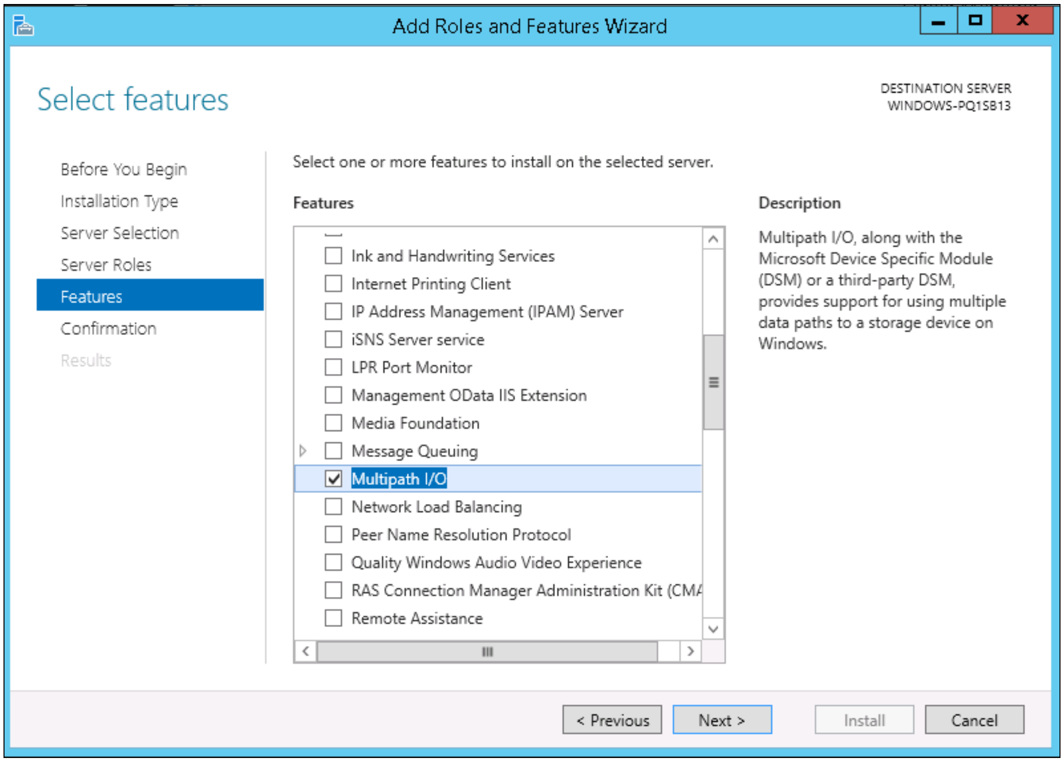
With Windows operating systems, the queue depth settings are specified as part of the host adapter configuration. These settings can be specified through the BIOS settings or by using software that is provided by the HBA vendor.

Optimize the storage environment by evenly spreading the I/O load across all available ports. Account for the load on a particular server, its queue depth, and the number of volumes.

Installing Multipath I/O (MPIO) feature

MPIO is provided as a built-in feature in Windows Server 2016, 2012 R2 and 2008 R2. It can be installed as part of the automated host attachment procedure using the IBM Storage Host Attachment Kit for Windows, as shown in Figure 3-3. If a manual installation is preferred, complete the following steps, which pertain to Windows Server 2012 R2, but also are valid for Windows Server 2016 and 2008 R2:

1. Open Server Manager and select **Add roles and features** from the Dashboard for Windows Server 2016 and 2012 R2. For Windows Server 2008 R2 open Server Manager and select **Features Summary**, select **Add Features**, and then select **Multipath I/O**.
2. For Windows Server 2016 and 2012 R2 click **Next** at the bottom of the dialog until Select Features is displayed and then select **Multipath I/O** in the Features window, as shown in Figure 3-1.



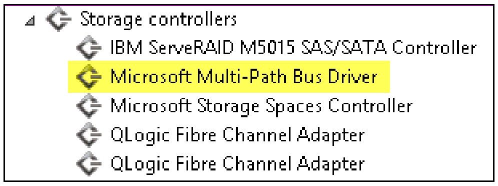
Select one or more features to install on the selected server.

- ☐ Ink and Handwriting Services
- ☐ Internet Printing Client
- ☐ IP Address Management (IPAM) Server
- ☐ iSNS Server service
- ☐ LPR Port Monitor
- ☐ Management OData IIS Extension
- ☐ Media Foundation
- ☐ Message Queuing
- ☑ Multipath I/O
- ☐ Network Load Balancing
- ☐ Peer Name Resolution Protocol
- ☐ Quality Windows Audio Video Experience
- ☐ RAS Connection Manager Administration Kit (CMA
- ☐ Remote Assistance

Multipath I/O, along with the Microsoft Device Specific Module (DSM) or a third-party DSM, provides support for using multiple data paths to a storage device on Windows.

Figure 3-1 Selecting the Multipath I/O feature

3. Follow the instructions on the panel to complete the installation. This process might require a reboot.
4. Check that the driver is installed correctly by loading **Device Manager**. Verify that it now includes `Microsoft Multi-Path Bus Driver`, as shown in Figure 3-2.



Storage controllers
- IBM ServeRAID M5015 SAS/SATA Controller
- Microsoft Multi-Path Bus Driver
- Microsoft Storage Spaces Controller
- QLogic Fibre Channel Adapter
- QLogic Fibre Channel Adapter

Figure 3-2 Microsoft Multi-Path Bus Driver

Windows Host Attachment Kit installation

The IBM Storage Host Attachment Kit (HAK) for Windows is a software pack that simplifies the tasks of connecting a Microsoft Windows Server host to IBM XIV, IBM Spectrum Accelerate, IBM FlashSystem A9000 and A9000R storage systems.

> **Note:** Previous versions of HAK were identified as the IBM XIV Host Attachment Kit, which was renamed beginning with version 2.6.0.

The HAK provides a set of command-line interface (CLI) tools that help host administrators perform different host-side tasks, such as: detect any physically connected storage system (single system or multiple systems), detect storage volumes, define the host on the storage system, run diagnostics, and apply best practice native multipath connectivity configuration on the host.

The IBM Storage Host Attachment Kit for Windows version 2.6.0, or later, must be installed to gain access to FlashSystem A9000 and A9000R storage. Previous versions may be used for connecting Windows host servers to IBM XIV or IBM Spectrum Accelerate.

For Windows Server 2016, 2012 R2 and 2008 R2, the 64-bit Windows version is required. A Host Attachment Kit can be downloaded from Fix Central:

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

Portable Storage Host Attachment Kit installation and usage

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

Performing a local installation

The following instructions are based on the installation performed at the time of writing. For more information, see the instructions in the *Windows Host Attachment Kit User Guide* (which is available through Fix Central with the Host Attachment Kit). These instructions show the GUI installation, and can change over time. For information about command-line instructions, see the *Windows Host Attachment Kit User Guide*.

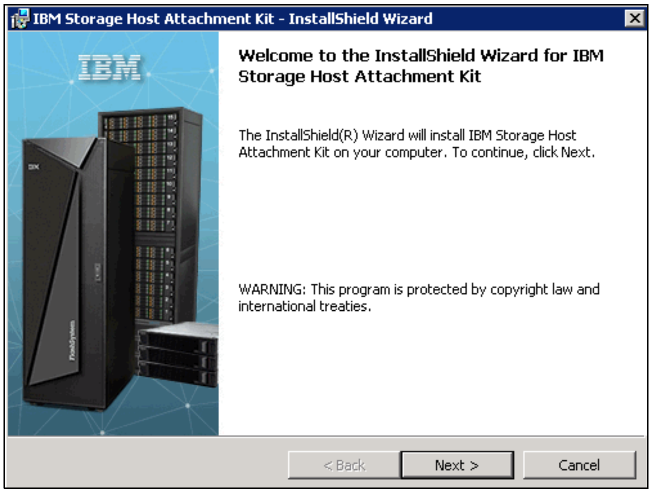
> **Attention:** Before installing the Host Attachment Kit, remove any non-supported multipathing software that was previously installed. Failure to do so can lead to unpredictable behavior or even loss of data.

Install the IBM Storage Host Attachment Kit (it is a mandatory prerequisite for support by completing the following steps:

1. Run the following installation setup executable file (at time of writing, this is the file name):

   ```
   IBM_Storage_Host_Attachment_Kit_2.8.2-b2877_Windows-x64.exe
   ```

   The process starts the Python engine (*xpyv)*. When you are prompted, select a language, and then proceed with the installation wizard instructions (see Figure 3-3 on page 101).


IBM Storage Host Attachment Kit - InstallShield Wizard

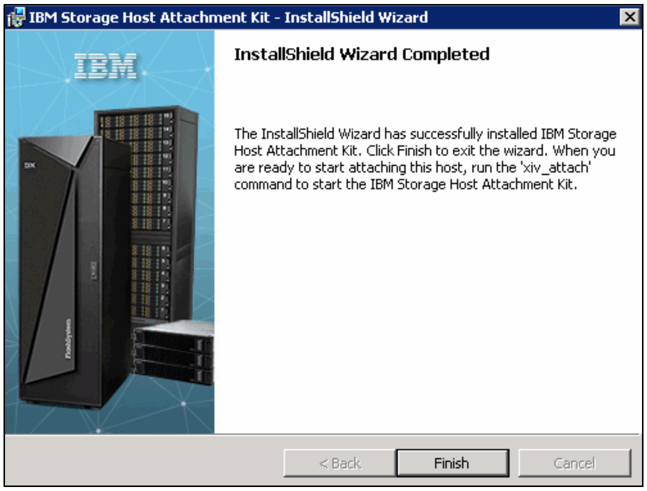
Welcome to the InstallShield Wizard for IBM Storage Host Attachment Kit

The InstallShield(R) Wizard will install IBM Storage Host Attachment Kit on your computer. To continue, click Next.

WARNING: This program is protected by copyright law and international treaties.

Figure 3-3 Welcome to IBM Storage Host Attachment Kit installation wizard

2. When the installation completes, click **Finish** (see Figure 3-4). The IBM Storage Host Attachment Kit is added to the list of installed Windows programs.


IBM Storage Host Attachment Kit - InstallShield Wizard

InstallShield Wizard Completed

The InstallShield Wizard has successfully installed IBM Storage Host Attachment Kit. Click Finish to exit the wizard. When you are ready to start attaching this host, run the 'xiv_attach' command to start the IBM Storage Host Attachment Kit.

Figure 3-4 IBM Storage Host Attachment Kit installation wizard completed

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

Running the xiv_attach program

Complete the procedure that is shown in Example 3-1 for a FC connection.

Example 3-1 Using the IBM Storage Host Attachment Wizard tor attachment over Fibre Channel

```
PS C:\> xiv_attach
-------------------------------------------------------------------------------
Welcome to the IBM Storage Host Attachment wizard, version 2.8.2.
This wizard will help you attach this host to one or more IBM storage systems

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

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

```
-------------------------------------------------------------------------------
Please wait while the wizard validates your existing configuration...
Verifying Previous HAK versions                                              OK
Verifying Disk timeout setting                                               OK
Verifying Built-In MPIO feature                                              OK
Verifying Multipath I/O feature compatibility with IBM storage devices   NOT OK
Verifying IBM storage system MPIO Load Balancing (service)                   OK
Verifying IBM storage system MPIO Load Balancing (agent)                     OK
Verifying Windows Hotfix 2460971                                         NOT OK
Verifying Windows Hotfix 2522766                                         NOT OK
Verifying LUN0 device driver                                             NOT OK
-------------------------------------------------------------------------------
The wizard needs to configure this host for the IBM storage system.
Do you want to proceed? [default: yes ]:
Please wait while the host is being configured...
-------------------------------------------------------------------------------
Configuring Previous HAK versions                                            OK
Configuring Disk timeout setting                                             OK
Configuring Built-In MPIO feature                                            OK
Configuring Multipath I/O feature compatibility with IBM storage devices REBOOT
Configuring IBM storage system MPIO Load Balancing (service)                 OK
Configuring IBM storage system MPIO Load Balancing (agent)                   OK
Configuring Windows Hotfix 2460971                                       REBOOT
Configuring Windows Hotfix 2522766                                       REBOOT
Configuring LUN0 device driver                                               OK
-------------------------------------------------------------------------------
This host requires a reboot in an orderly manner without interruptions.
Please reboot this host in an orderly manner and then run the HAK utility again.


Press [ENTER] to exit.

After reboot:
PS C:\> xiv_attach
-------------------------------------------------------------------------------
Welcome to the IBM Storage Host Attachment wizard, version 2.8.2.
This wizard will help you attach this host to one or more IBM storage systems

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

-------------------------------------------------------------------------------
Please specify the connectivity type: [f]c / [i]scsi : f
-------------------------------------------------------------------------------
Please wait while the wizard validates your existing configuration...
Verifying Previous HAK versions                                              OK
Verifying Disk timeout setting                                               OK
Verifying Built-In MPIO feature                                              OK
Verifying Multipath I/O feature compatibility with IBM storage devices       OK
Verifying LUN0 device driver                                                 OK
This host is already configured for the IBM storage system.
-------------------------------------------------------------------------------
Please define zoning for this host and add its World Wide Port Names (WWPNs) to the IBM
storage system:
21:00:00:24:ff:35:c6:6c: [QLogic QLE2562 Fibre Channel Adapter]: QLE2562
21:00:00:24:ff:35:c6:6d: [QLogic QLE2562 Fibre Channel Adapter]: QLE2562
Press [ENTER] to proceed.

Would you like to rescan for new storage devices? [default: yes ]:
Please wait while rescanning for IBM storage devices...
```

```
-------------------------------------------------------------------------------
This host is connected to the following IBM storage arrays:
Storage Type        Serial   System Version  Host Defined  Ports Defined     Protocol  Host
Name(s)
XIV                 1340010  11.6.2          No            No ports defined  FC        N/A
XIV                 1340008  11.6.2.a        No            No ports defined  FC        N/A
FlashSystem A9000   1322131  12.1.0.b        Yes           All               FC,iSCSI
ITSO_W2K12,ITSO_W2K12_iSCSI
FlashSystem A9000R  1320902  12.1.0.b        No            No ports defined  FC,iSCSI  N/A
This host is not defined on some FC-attached IBM storage systems.
Do you want to define this host on these IBM storage systems now? [default: yes ]:
Please enter a name for this host [default: WINDOWS-PQ1SB13 ]: ITSO_W2K12
Please enter a username for system 1340010 [default: admin ]:
Please enter the password of user admin for system 1340010:


Please enter a username for system 1340008 [default: admin ]:
Please enter the password of user admin for system 1340008:


Please enter a username for system 1320902 [default: admin ]:
Please enter the password of user admin for system 1320902:

Press [ENTER] to proceed.

-------------------------------------------------------------------------------
The IBM Storage Host Attachment wizard has successfully configured this host.

Press [ENTER] to exit.
```

Scanning for new LUNs

Before scanning for new LUNs in Windows, the host must be created, configured, and have LUNs assigned. The **xiv_attach** command that is shown in Example 3-1 on page 101 automated the host creation and configuration for the storage systems already attached to this host.

For more information about assigning LUNs to a host for XIV, see 1.4.2, “Assigning LUNs to a host by using the GUI” on page 35 and 1.4.3, “Assigning LUNs to a host by using the XCLI” on page 37.

For more information about assigning LUNs to a host for FlashSystem A9000 and A9000R, see 2.6.2, “Assigning LUNs to a host by using the GUI” on page 86, and 2.6.3, “Assigning LUNs to a host by using the XCLI” on page 89.

The instructions that are described next, it is assumed that these operations are complete.

To scan for LUNs, complete the following steps:

1. Open Administrative Tools from the Windows desktop (alternately choose **Control Panel** →**System and Security** →**Administrative Tools**). Next, select **Computer Management** to open its window and select **Device Manager** (see Figure 3-5).

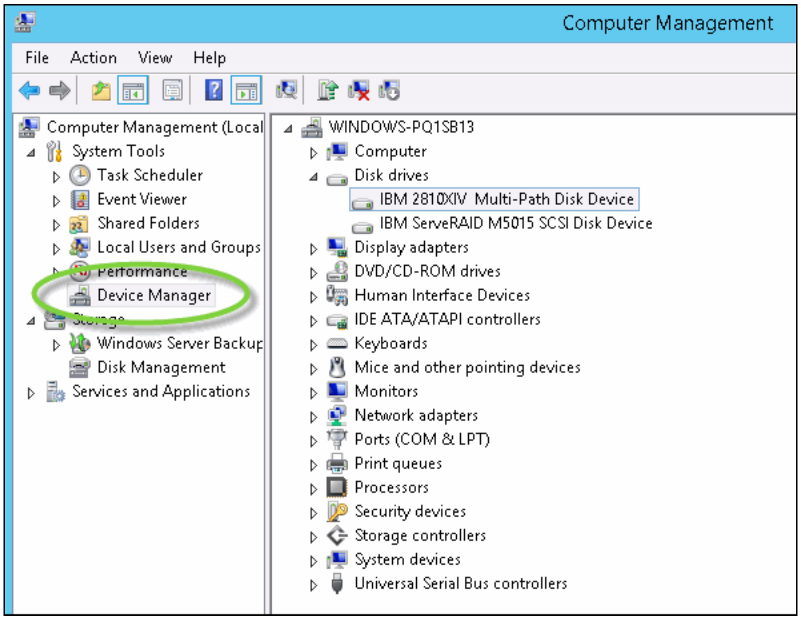
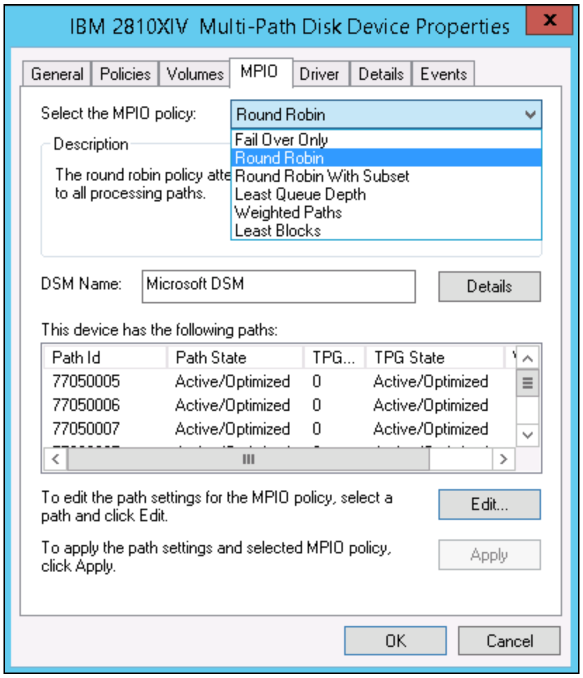


Figure 3-5 Device Manager

2. Right-click **Disk drives** in the center panel and select **Scan for hardware changes**. FlashSystem A9000, A9000R and XIV LUNs are displayed as `IBM 2810XIV Multi-Path Disk Device` in the Device Manager tree under `Disk drives` (see Figure 3-5). The number of objects that are named `IBM 2810XIV Multi-Path Disk Device` depends on the number of LUNs mapped to the host.

3. Right-click an **IBM 2810 Multi-Path Disk Device** object and select **Properties**.

4. Click the MPIO tab to set the load balancing, as shown in Figure 3-6.



IBM 2810XIV Multi-Path Disk Device Properties

Select the MPIO policy: Round Robin

Description

The round robin policy atte
to all processing paths.

DSM Name: Microsoft DSM

This device has the following paths:

Path Id	Path State	TPG...	TPG State
77050005	Active/Optimized	0	Active/Optimized
77050006	Active/Optimized	0	Active/Optimized
77050007	Active/Optimized	0	Active/Optimized

To edit the path settings for the MPIO policy, select a path and click Edit.

To apply the path settings and selected MPIO policy, click Apply.

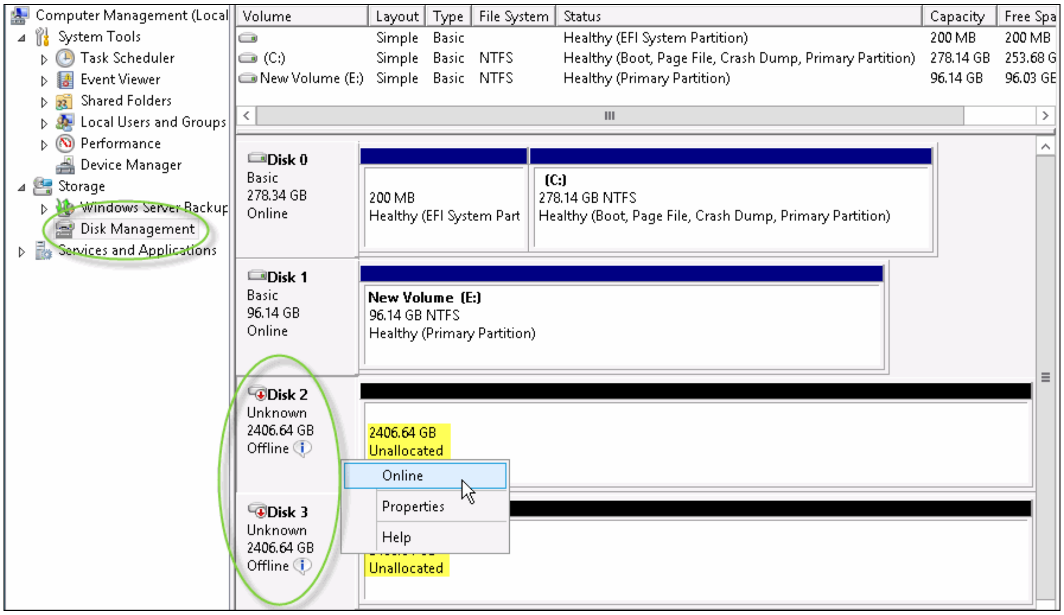
Figure 3-6 MPIO load balancing

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

Load balancing has the following possible options:

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

5. New volumes show under Disk Management as `Offline` and `Unallocated`. Right-click on the disk name and select **Online** to bring the volume to online status (see Figure 3-7).



Volume	Layout	Type	File System	Status	Capacity	Free Spa
	Simple	Basic		Healthy (EFI System Partition)	200 MB	200 MB
(C:)	Simple	Basic	NTFS	Healthy (Boot, Page File, Crash Dump, Primary Partition)	278.14 GB	253.68 G
New Volume (E:)	Simple	Basic	NTFS	Healthy (Primary Partition)	96.14 GB	96.03 GE

Disk 0
Basic
278.34 GB
Online

200 MB
Healthy (EFI System Part

(C:)
278.14 GB NTFS
Healthy (Boot, Page File, Crash Dump, Primary Partition)

Disk 1
Basic
96.14 GB
Online

New Volume (E:)
96.14 GB NTFS
Healthy (Primary Partition)

Disk 2
Unknown
2406.64 GB
Offline

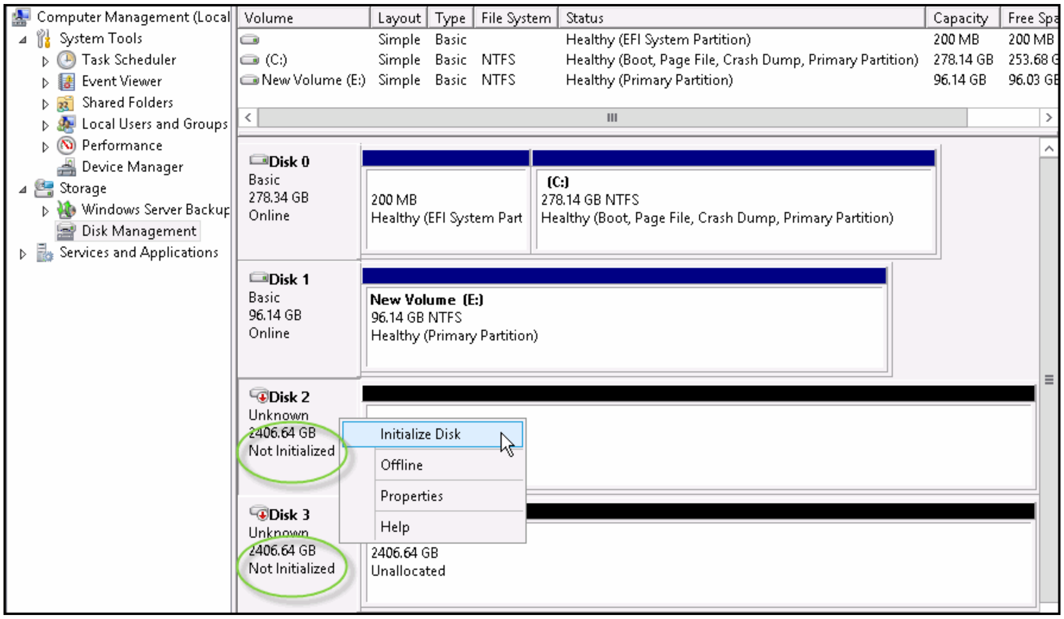
2406.64 GB
Unallocated

Disk 3
Unknown
2406.64 GB
Offline

Unallocated

Figure 3-7 New volumes presented from FlashSystem A9000 storage device

6. After the volume (or volumes) are set to `Online`, they will show as `Not Initialized`. Right-click on the volume name and select **Initialize Disk** to prepare it for use with Logical Disk Manager (see Figure 3-8).



Volume	Layout	Type	File System	Status	Capacity	Free Spa
	Simple	Basic		Healthy (EFI System Partition)	200 MB	200 MB
(C:)	Simple	Basic	NTFS	Healthy (Boot, Page File, Crash Dump, Primary Partition)	278.14 GB	253.68 G
New Volume (E:)	Simple	Basic	NTFS	Healthy (Primary Partition)	96.14 GB	96.03 GE

Disk 0
Basic
278.34 GB
Online

200 MB
Healthy (EFI System Part

(C:)
278.14 GB NTFS
Healthy (Boot, Page File, Crash Dump, Primary Partition)

Disk 1
Basic
96.14 GB
Online

New Volume (E:)
96.14 GB NTFS
Healthy (Primary Partition)

Disk 2
Unknown
2406.64 GB
Not Initialized

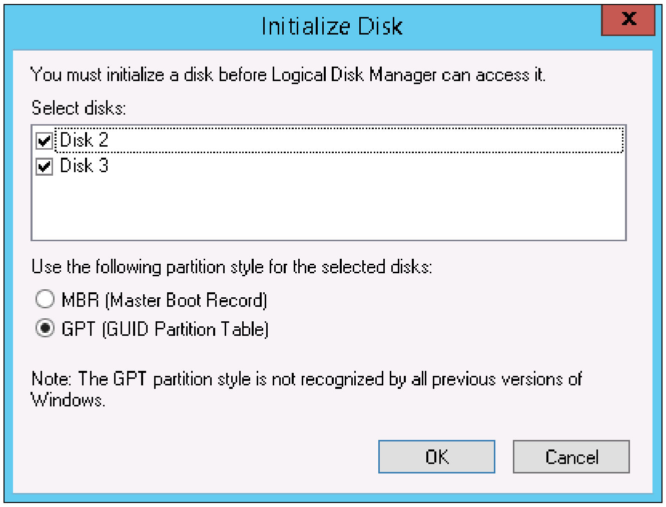
Disk 3
Unknown
2406.64 GB
Not Initialized

2406.64 GB
Unallocated

Figure 3-8 New volumes need to be Initialized before Logical Disk Manager can access them

Note that multiple disks can be initialized at the same time when selecting the **Initialize Disk** option.

The resulting window is shown in Figure 3-9.



Initialize Disk

You must initialize a disk before Logical Disk Manager can access it.

Select disks:

☑ Disk 2
☑ Disk 3

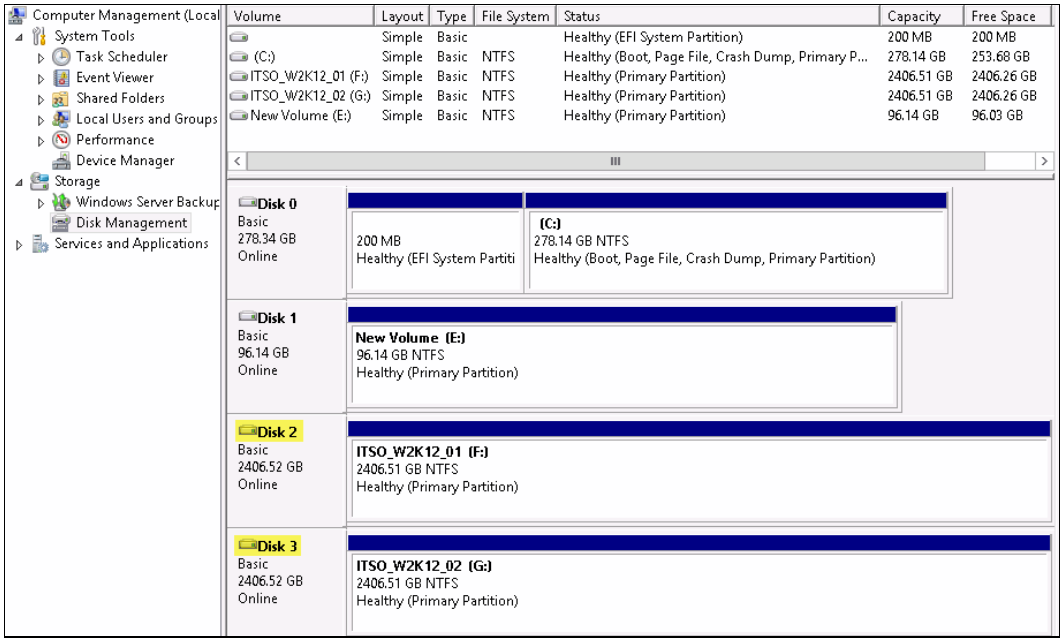
Use the following partition style for the selected disks:

○ MBR (Master Boot Record)
● GPT (GUID Partition Table)

Note: The GPT partition style is not recognized by all previous versions of Windows.

Figure 3-9 Initialize Disk dialog box

7. The mapped LUNs on the host are listed under Disk Management as shown in Figure 3-10 (shown after creating a Simple Volume, formatting, and naming).



Volume	Layout	Type	File System	Status	Capacity	Free Space
	Simple	Basic		Healthy (EFI System Partition)	200 MB	200 MB
(C:)	Simple	Basic	NTFS	Healthy (Boot, Page File, Crash Dump, Primary P...	278.14 GB	253.68 GB
ITSO_W2K12_01 (F:)	Simple	Basic	NTFS	Healthy (Primary Partition)	2406.51 GB	2406.26 GB
ITSO_W2K12_02 (G:)	Simple	Basic	NTFS	Healthy (Primary Partition)	2406.51 GB	2406.26 GB
New Volume (E:)	Simple	Basic	NTFS	Healthy (Primary Partition)	96.14 GB	96.03 GB

Disk 0
Basic
278.34 GB
Online

200 MB
Healthy (EFI System Partiti

(C:)
278.14 GB NTFS
Healthy (Boot, Page File, Crash Dump, Primary Partition)

Disk 1
Basic
96.14 GB
Online

New Volume (E:)
96.14 GB NTFS
Healthy (Primary Partition)

Disk 2
Basic
2406.52 GB
Online

ITSO_W2K12_01 (F:)
2406.51 GB NTFS
Healthy (Primary Partition)

Disk 3
Basic
2406.52 GB
Online

ITSO_W2K12_02 (G:)
2406.51 GB NTFS
Healthy (Primary Partition)

Figure 3-10 Mapped LUNs as displayed in Disk Management

3.2.3 Windows host iSCSI configuration

In Windows Server 2016, 2012 R2 and 2008 R2, the iSCSI Software Initiator is part of the operating system. Before configuring the host, however, establish the physical iSCSI connection to the XIV (see 1.3, “iSCSI connectivity” on page 22) or FlashSystem A9000 or A9000R (see 2.3, “iSCSI connectivity” on page 59).

IBM XIV, FlashSystem A9000, and A9000R support the iSCSi Challenge Handshake Authentication Protocol (CHAP). These examples assume that CHAP is not required. If it is, specify the settings for the required CHAP parameters on both the host and Storage system sides.

Supported CNAs

For Windows, IBM XIV, FlashSystem A9000, and A9000R all support various CNAs. More details about supported CNAs are available from the SSIC:

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

Ethernet interface adapters that use an iSCSI software initiator can be used also.

Windows multipathing feature and Host Attachment Kit installation

To install the Windows multipathing feature, follow the procedure given in “Installing Multipath I/O (MPIO) feature” on page 99.

To install the Windows Host Attachment Kit, use the procedure explained in “Windows Host Attachment Kit installation” on page 100.

Running the xiv_attach program

Run the **xiv_attach** program, as shown in Example 3-2.

Example 3-2 Using the IBM Storage Host Attachment Wizard for attachment over iSCSI

```
--------------------------------------------------------------------------------
Welcome to the IBM Storage Host Attachment wizard, version 2.8.2.
This wizard will help you attach this host to one or more IBM storage systems

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

--------------------------------------------------------------------------------
Please specify the connectivity type: [f]c / [i]scsi : i
--------------------------------------------------------------------------------
Please wait while the wizard validates your existing configuration...
Verifying Previous HAK versions                                               OK
Verifying Disk timeout setting                                                OK
Verifying iSCSI service                                                   NOT OK
Verifying Built-In MPIO feature                                               OK
Verifying Multipath I/O feature compatibility with IBM storage devices        OK
Verifying LUN0 device driver                                                  OK
--------------------------------------------------------------------------------
The wizard needs to configure this host for the IBM storage system.
Do you want to proceed? [default: yes ]:
Please wait while the host is being configured...
--------------------------------------------------------------------------------
Configuring Previous HAK versions                                             OK
Configuring Disk timeout setting                                              OK
Configuring iSCSI service                                                     OK
Configuring Built-In MPIO feature                                            OK
Configuring Multipath I/O feature compatibility with IBM storage devices     OK
Configuring LUN0 device driver                                               OK
The host is now configured for the IBM storage system
-------------------------------------------------------------------------------
Discovering new iSCSI targets ...
New iSCSI target for IBM storage system 01320902: 9.155.116.206 was discovered and configured successfully.
New iSCSI target for IBM storage system 01322131: 9.155.120.21 was discovered and configured successfully.
-------------------------------------------------------------------------------
Would you like to discover a new iSCSI target? [default: yes ]: no
Would you like to rescan for new storage devices? [default: yes ]:
-------------------------------------------------------------------------------
This host is connected to the following IBM storage arrays:
Storage Type        Serial   System Version  Host Defined  Ports Defined  Protocol  Host Name(s)
XIV                 1340010  11.6.2          Yes           All            FC        ITSO_W2K12
XIV                 1340008  11.6.2.a        Yes           All            FC        ITSO_W2K12
FlashSystem A9000   1322131  12.1.0.b        Yes           All            FC,iSCSI
ITSO_W2K12,ITSO_W2K12_iSCSI
FlashSystem A9000R  1320902  12.1.0.b        Yes           Not all        FC,iSCSI  ITSO_W2K12
This host is not defined on some iSCSI-attached IBM storage systems.
Do you want to define this host on these IBM storage systems now? [default: yes ]:
Please enter a name for this host [default: WINDOWS-PQ1SB13 ]: ITSO_W2K12_iSCSI
This host is already defined as "ITSO_W2K12" at IBM storage system 1320902. Defining the missing
ports.
Please enter a username for system 1320902 [default: admin ]:
Please enter the password of user admin for system 1320902:

Press [ENTER] to proceed.
-------------------------------------------------------------------------------
The IBM Storage Host Attachment wizard has successfully configured this host.
Press [ENTER] to exit.
```

Now assign the storage volumes to the defined Windows host as described in “Scanning for new LUNs” on page 103.

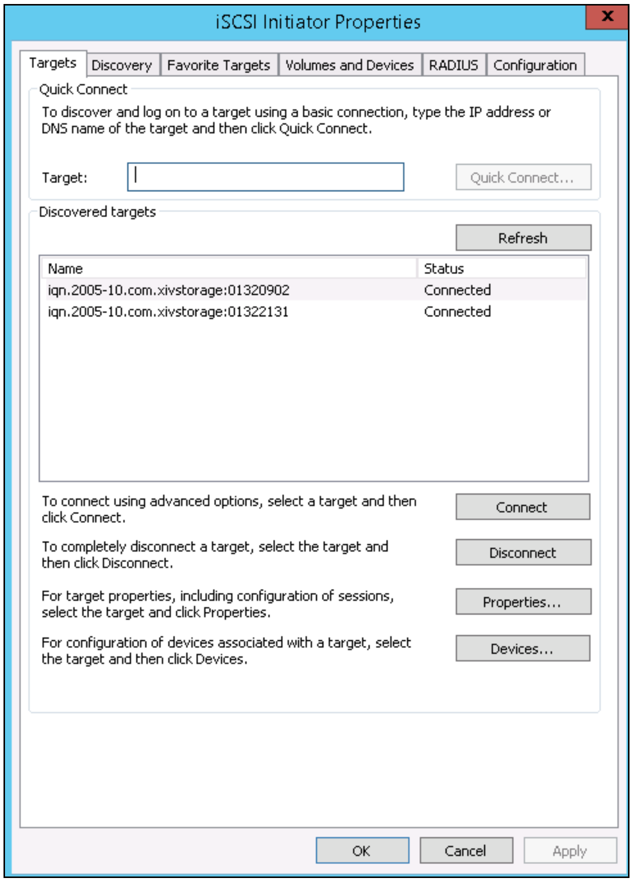
Configuring Microsoft iSCSI software initiator

The iSCSI connection must be configured on both the Windows host and the IBM Storage. By using the IBM Storage Host Attachment Kit for Windows, both the Windows host and the IBM Storage are configured automatically. Skip to step 13 on page 118.

To configure the Windows host and the IBM Storage manually, follow these instructions to complete the iSCSI configuration:

> **Note:** These steps use a Windows Server 2012 R2 host and a FlashSystem A9000 storage device with Hyper-Scale Manager.

1. Open Administrative Tools from the Windows desktop (alternately, choose **Control Panel →System and Security →Administrative Tools**).
2. Select **iSCSI Initiator**. The iSCSI Initiator Properties window opens (see Figure 3-11).



To discover and log on to a target using a basic connection, type the IP address or DNS name of the target and then click Quick Connect.

Name	Status
iqn.2005-10.com.xivstorage:01320902	Connected
iqn.2005-10.com.xivstorage:01322131	Connected

To connect using advanced options, select a target and then click Connect.

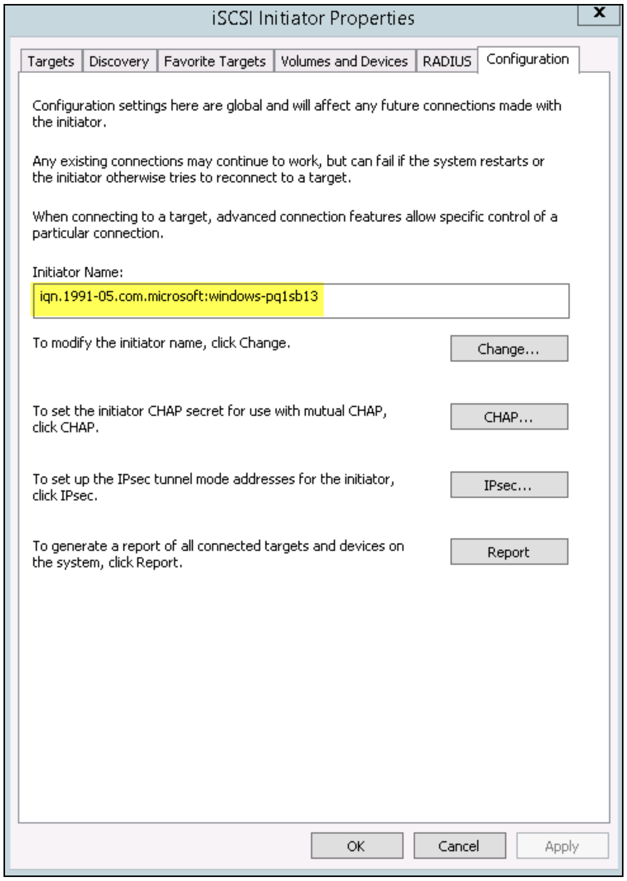
To completely disconnect a target, select the target and then click Disconnect.

For target properties, including configuration of sessions, select the target and click Properties.

For configuration of devices associated with a target, select the target and then click Devices.

Figure 3-11 iSCSI Initiator Properties window

3. Get the iSCSI Qualified Name (IQN) of the host from the Configuration tab (see Figure 3-12). In this example, the IQN is `iqn.1991-05.com.microsoft:windows-pq1sb13`. Copy this IQN to the clipboard and use this IQN to define this host on FlashSystem A9000, A9000R, or XIV in the next step (step 4 on page 112).



Configuration settings here are global and will affect any future connections made with the initiator.

Any existing connections may continue to work, but can fail if the system restarts or the initiator otherwise tries to reconnect to a target.

When connecting to a target, advanced connection features allow specific control of a particular connection.

Initiator Name:
iqn.1991-05.com.microsoft:windows-pq1sb13

To modify the initiator name, click Change.

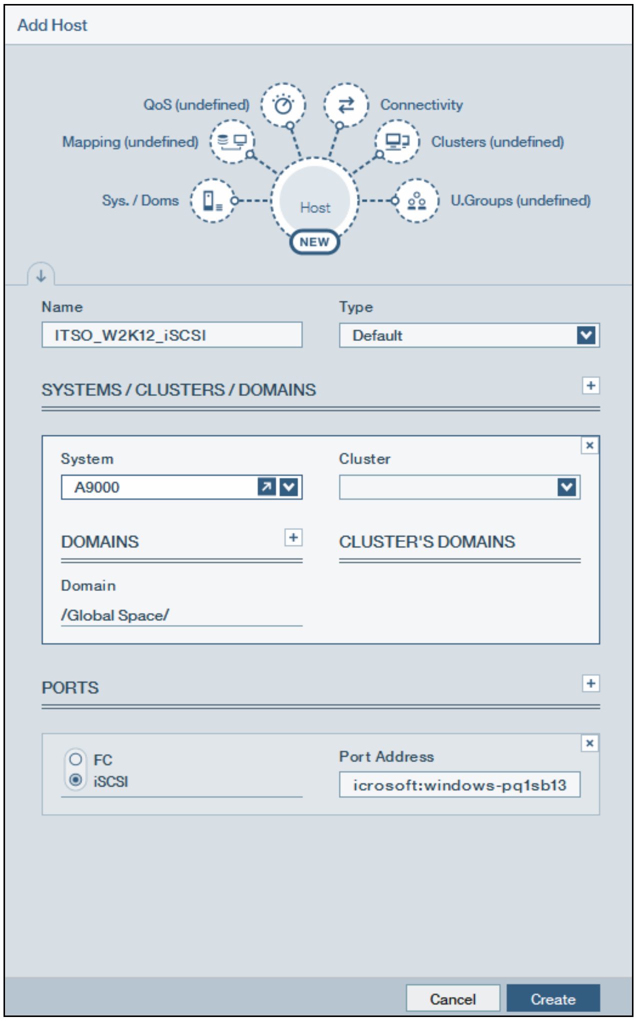
To set the initiator CHAP secret for use with mutual CHAP, click CHAP.

To set up the IPsec tunnel mode addresses for the initiator, click IPsec.

To generate a report of all connected targets and devices on the system, click Report.

Figure 3-12 iSCSI Configuration tab

4. Define the host on FlashSystem A9000, A9000R, or XIV, as shown in Figure 3-13 (Hyper-Scale Manager used).



Name: ITSO_W2K12_iSCSI

Type: Default

System: A9000

Cluster:

Domain: /Global Space/

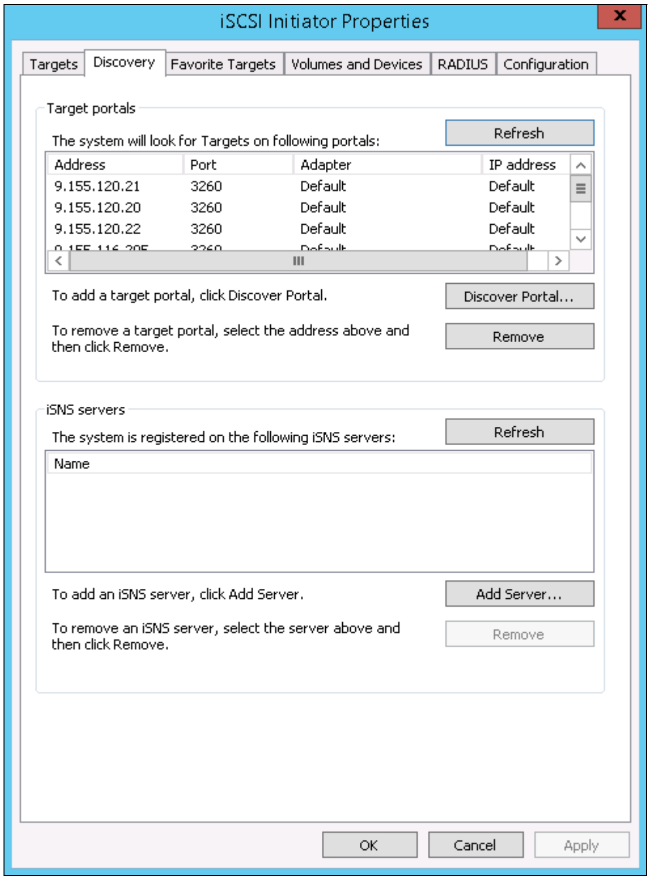
○ FC
● iSCSI

Port Address: icrosoft:windows-pq1sb13

Figure 3-13   Defining the host

5. Back on the host, click the **Discovery** tab in the iSCSI Initiator Properties dialog (see Figure 3-14).

   Click **Discover Portal** and use one of the iSCSI IP addresses from the storage system. Repeat this step for more target portals. Figure 3-14 shows the results.



The system will look for Targets on following portals:

Address	Port	Adapter	IP address
9.155.120.21	3260	Default	Default
9.155.120.20	3260	Default	Default
9.155.120.22	3260	Default	Default
9.155.116.205	3260	Default	Default

To add a target portal, click Discover Portal.

To remove a target portal, select the address above and then click Remove.

The system is registered on the following iSNS servers:

Name

To add an iSNS server, click Add Server.

To remove an iSNS server, select the server above and then click Remove.

Figure 3-14 iSCSI targets portals defined

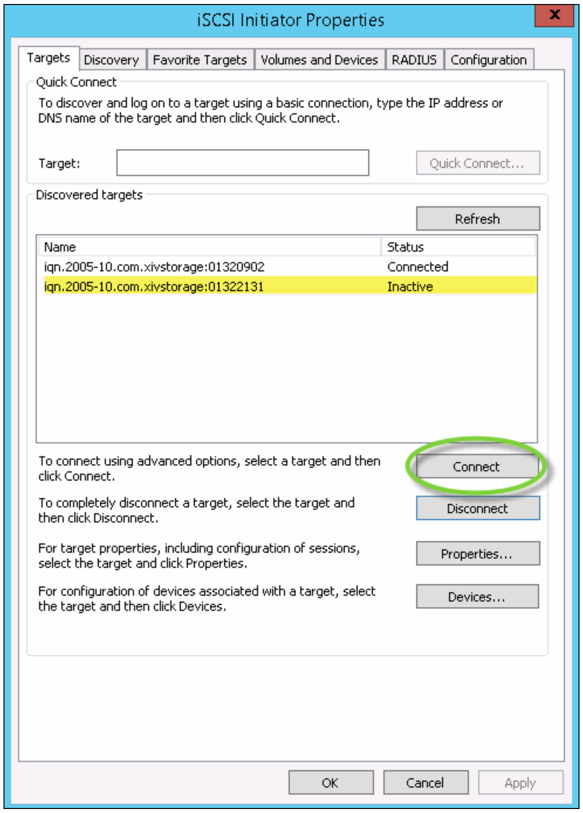
To improve performance, you can increase the MTU size if your network supports it, as shown in Example 3-3 (MTU of 9000 used).

Example 3-3 Increase MTU size

```
A9000>>ipinterface_list
Name         Type         IP Address       Network Mask     Default Gateway    MTU     Module         Port
management   Management   10.0.20.108      255.255.255.0    10.0.20.1          1500    1:Module:1
VPN          VPN          9.155.120.218    255.255.240.0    9.155.112.1        1500    1:Module:2
management   Management   10.0.20.109      255.255.255.0    10.0.20.1          1500    1:Module:3
VPN          VPN          9.155.120.219    255.255.240.0    9.155.112.1        1500    1:Module:3
M2-iSCSI1    iSCSI        9.155.120.21     255.255.240.0    9.155.112.1        9000    1:Module:2     1
M3-iSCSI1    iSCSI        9.155.120.22     255.255.240.0    9.155.112.1        9000    1:Module:3     1
M1-iSCSI1    iSCSI        9.155.120.20     255.255.240.0    9.155.112.1        9000    1:Module:1     1
```

The iSCSI IP addresses that were used in the test environment are `9.155.120.21`, `9.155.120.22`, and `9.155.120.20`. If you want to change the MTU size, use the **`ipinterface_update`** command.

6. The storage system is discovered by the initiator and displayed in the Targets tab (see Figure 3-15). At this stage, the Target shows as `Inactive`.



To discover and log on to a target using a basic connection, type the IP address or DNS name of the target and then click Quick Connect.

Name	Status
iqn.2005-10.com.xivstorage:01320902	Connected
iqn.2005-10.com.xivstorage:01322131	Inactive

To connect using advanced options, select a target and then click Connect.

To completely disconnect a target, select the target and then click Disconnect.

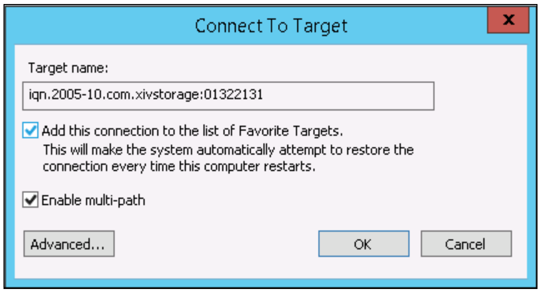
For target properties, including configuration of sessions, select the target and click Properties.

For configuration of devices associated with a target, select the target and then click Devices.

Figure 3-15 A discovered XIV Storage with Inactive status

7. To activate the connection, click **Connect.**

8. In the Connect To Target window, select **Enable multi-path**, and **Add this connection to the list of Favorite Targets** (see Figure 3-16). These settings automatically restore this connection when the system boots.



Target name:

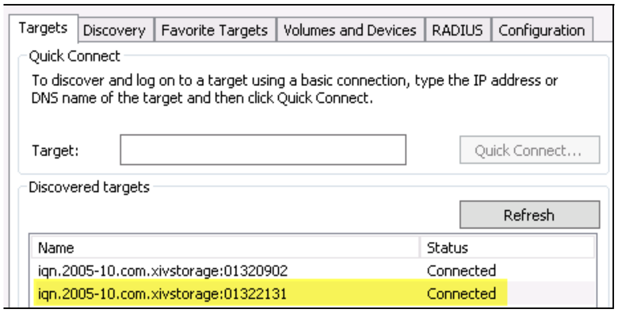
iqn.2005-10.com.xivstorage:01322131

☑ Add this connection to the list of Favorite Targets.
This will make the system automatically attempt to restore the connection every time this computer restarts.

☑ Enable multi-path

Figure 3-16 Connect To Target window

The iSCSI Target connection status now shows as `Connected` (see Figure 3-17).



To discover and log on to a target using a basic connection, type the IP address or DNS name of the target and then click Quick Connect.

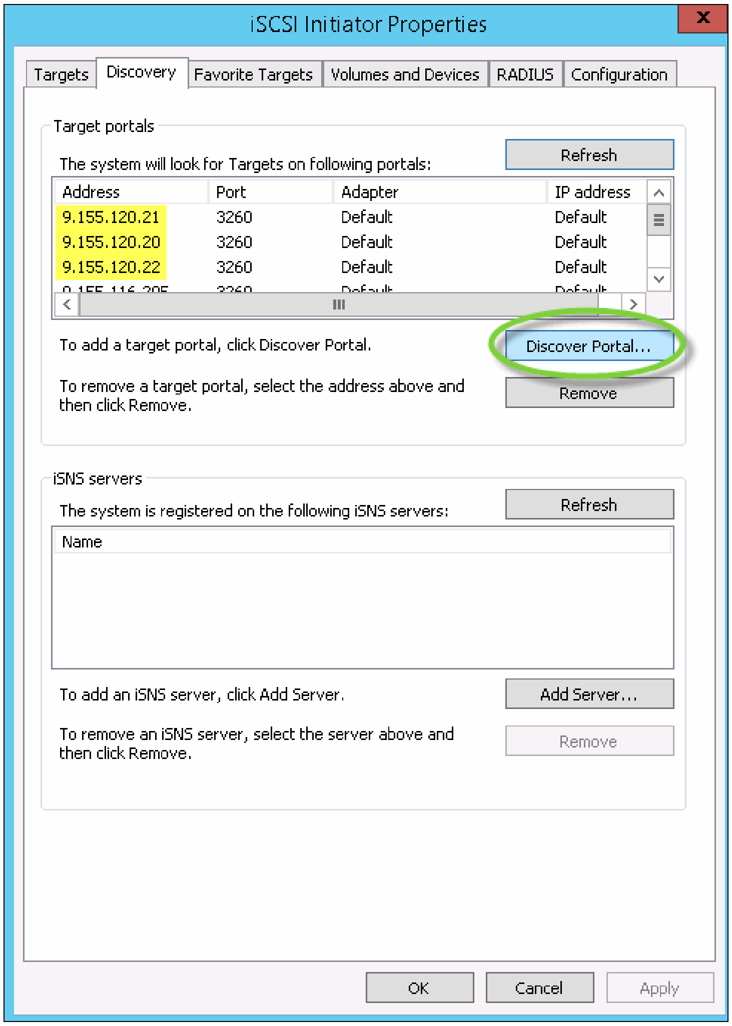
Target:

Name	Status
iqn.2005-10.com.xivstorage:01320902	Connected
iqn.2005-10.com.xivstorage:01322131	Connected

Figure 3-17 Connect to Target is active

9. Click the **Discovery** tab.

10. Select **Discover Portal** and enter the IP address of the FlashSystem A9000, A9000R, or XIV system in the resulting window (see Figure 3-18 and Figure 3-19 on page 116).



The system will look for Targets on following portals:

Address	Port	Adapter	IP address
9.155.120.21	3260	Default	Default
9.155.120.20	3260	Default	Default
9.155.120.22	3260	Default	Default
9.155.116.205	3260	Default	Default

To add a target portal, click Discover Portal.

To remove a target portal, select the address above and then click Remove.

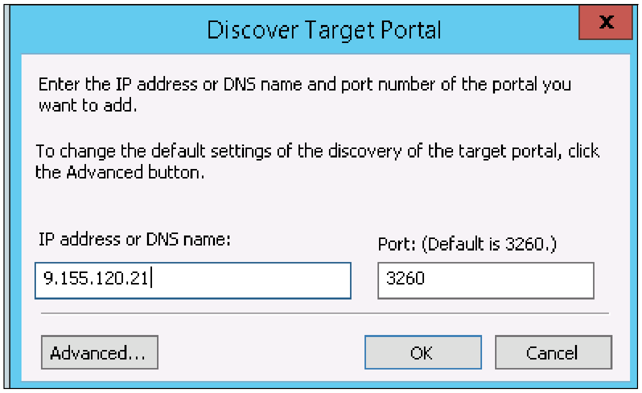
The system is registered on the following iSNS servers:

Name

To add an iSNS server, click Add Server.

To remove an iSNS server, select the server above and then click Remove.

Figure 3-18 Discovering the iSCSI connections



Discover Target Portal

Enter the IP address or DNS name and port number of the portal you want to add.

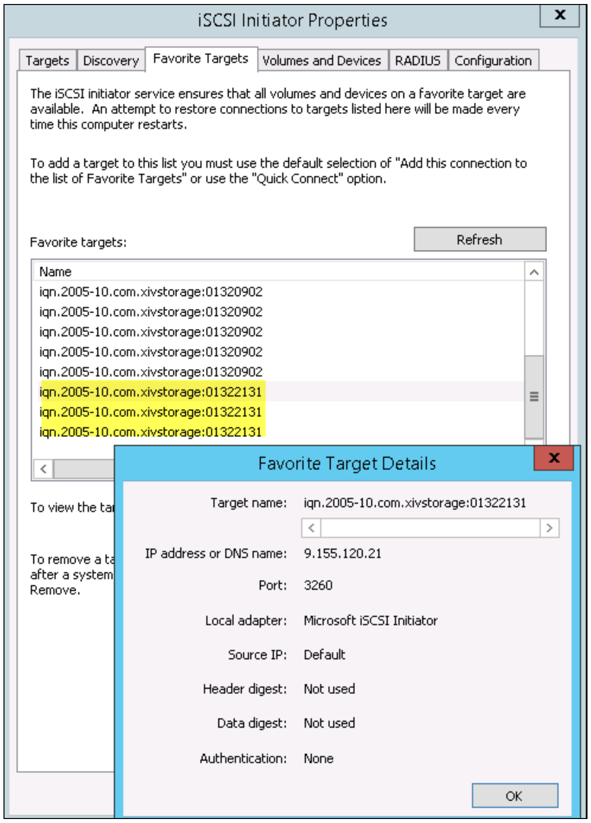
To change the default settings of the discovery of the target portal, click the Advanced button.

IP address or DNS name: 9.155.120.21

Port: (Default is 3260.) 3260

Figure 3-19 Discovering the FlashSystem A9000 iSCSI IP addresses

The Favorite Targets tab shows the connected IP addresses (see Figure 3-20).



iSCSI Initiator Properties

The iSCSI initiator service ensures that all volumes and devices on a favorite target are available. An attempt to restore connections to targets listed here will be made every time this computer restarts.

To add a target to this list you must use the default selection of "Add this connection to the list of Favorite Targets" or use the "Quick Connect" option.

Favorite targets:

Name
iqn.2005-10.com.xivstorage:01320902
iqn.2005-10.com.xivstorage:01320902
iqn.2005-10.com.xivstorage:01320902
iqn.2005-10.com.xivstorage:01320902
iqn.2005-10.com.xivstorage:01320902
iqn.2005-10.com.xivstorage:01322131
iqn.2005-10.com.xivstorage:01322131
iqn.2005-10.com.xivstorage:01322131

To view the ta

To remove a ta after a system Remove.

Favorite Target Details

Target name: iqn.2005-10.com.xivstorage:01322131

IP address or DNS name: 9.155.120.21

Port: 3260

Local adapter: Microsoft iSCSI Initiator

Source IP: Default

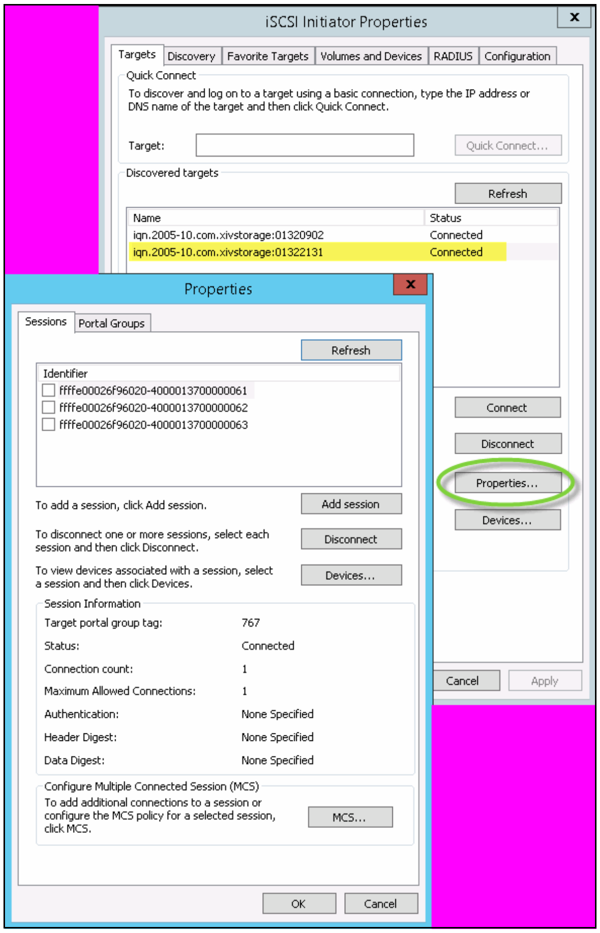
Header digest: Not used

Data digest: Not used

Authentication: None

Figure 3-20 A discovered FlashSystem A9000 with Connected status

11. View the iSCSI sessions by clicking the **Targets** tab, highlighting the target, and clicking **Properties**. Verify the sessions of the connection, as shown in Figure 3-21.



To discover and log on to a target using a basic connection, type the IP address or DNS name of the target and then click Quick Connect.

Name	Status
iqn.2005-10.com.xivstorage:01320902	Connected
iqn.2005-10.com.xivstorage:01322131	Connected

Identifier

- □ ffffe00026f96020-4000013700000061
- □ ffffe00026f96020-4000013700000062
- □ ffffe00026f96020-4000013700000063

To add a session, click Add session.

To disconnect one or more sessions, select each session and then click Disconnect.

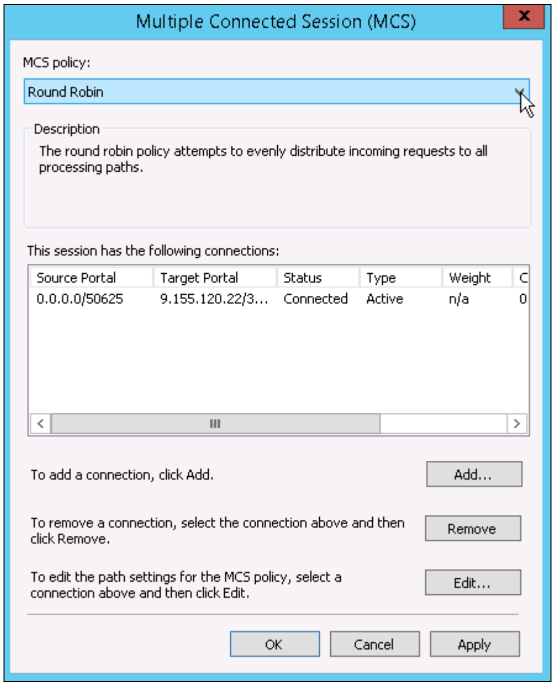
To view devices associated with a session, select a session and then click Devices.

Target portal group tag:	767
Status:	Connected
Connection count:	1
Maximum Allowed Connections:	1
Authentication:	None Specified
Header Digest:	None Specified
Data Digest:	None Specified

To add additional connections to a session or configure the MCS policy for a selected session, click MCS.

Figure 3-21 Target connection details

12. To see further details or change the load balancing policy, click **MCS** (Multiple Connected Session), as shown in Figure 3-22.


Multiple Connected Session (MCS)
MCS policy:
Round Robin
Description
The round robin policy attempts to evenly distribute incoming requests to all processing paths.
This session has the following connections:

Source Portal	Target Portal	Status	Type	Weight	C
0.0.0.0/50625	9.155.120.22/3...	Connected	Active	n/a	0

To add a connection, click Add.
Add...
To remove a connection, select the connection above and then click Remove.
Remove
To edit the path settings for the MCS policy, select a connection above and then click Edit.
Edit...
OK
Cancel
Apply

Figure 3-22 Connected sessions

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

The following available options are available:

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

13. If no volumes are mapped to this host yet, assign them now.

For more information about assigning LUNs to a host for XIV, see 1.4.2, “Assigning LUNs to a host by using the GUI” on page 35, and 1.4.3, “Assigning LUNs to a host by using the XCLI” on page 37.

For more information about assigning LUNs to a host for FlashSystem A9000 and A9000R, see 2.6.2, “Assigning LUNs to a host by using the GUI” on page 86, and 2.6.3, “Assigning LUNs to a host by using the XCLI” on page 89.

The instructions that are described next assume that these operations are complete.

14. To verify an assigned disk, open the Windows Device Manager, which shows all XIV, FlashSystem A9000, and A9000R disks connected via iSCSI under `Storage controllers` (see Figure 3-23).

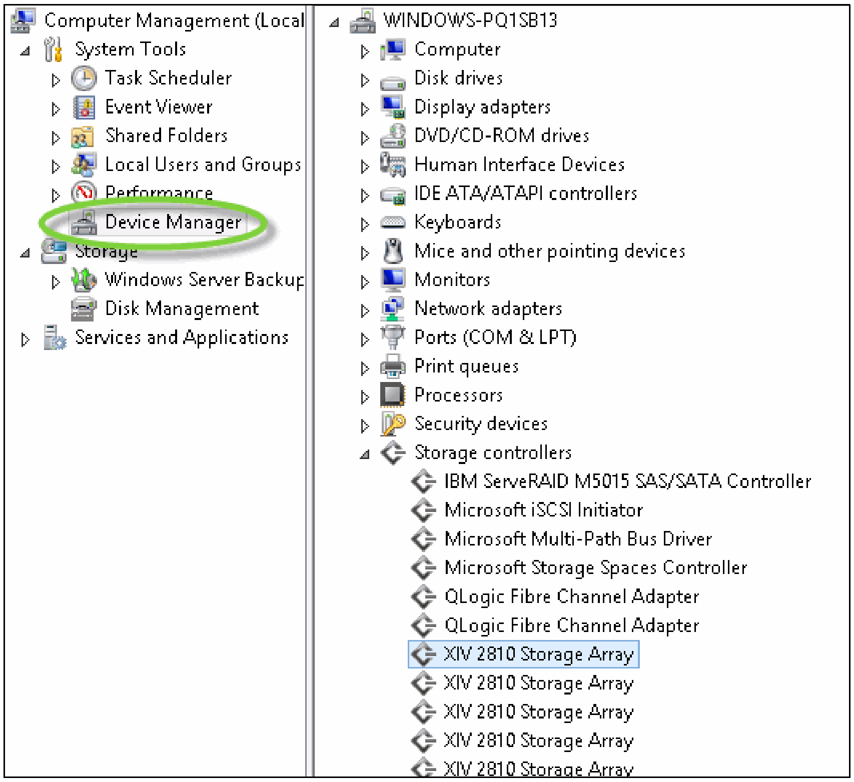
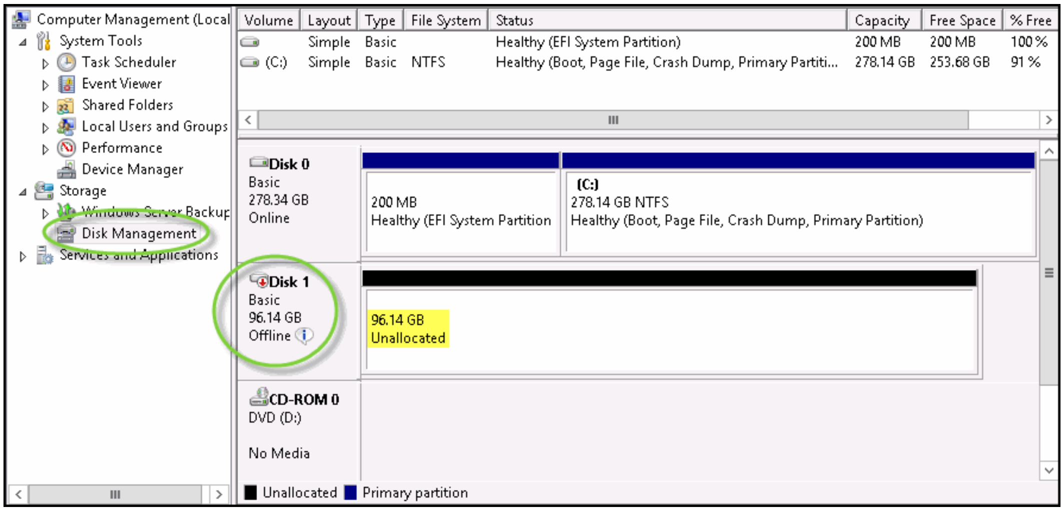


Figure 3-23 Windows Device Manager with IBM Storage disks connected through iSCSI

New disks are shown in the Disk Management window as `Offline` and `Unallocated` (see Figure 3-24).



Volume	Layout	Type	File System	Status	Capacity	Free Space	% Free
	Simple	Basic		Healthy (EFI System Partition)	200 MB	200 MB	100 %
(C:)	Simple	Basic	NTFS	Healthy (Boot, Page File, Crash Dump, Primary Partiti...	278.14 GB	253.68 GB	91 %

Figure 3-24 FlashSystem A9000 volume newly mapped to Windows 2012 host server

15. Right-click the disk and select **Online** (see Figure 3-25) to bring the volume online. Then, initialize the disks.

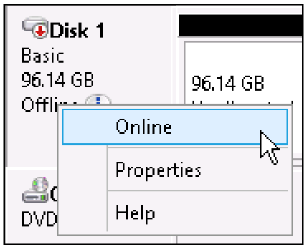


Figure 3-25 Changing the Disk to Online

At this stage, the volume can be administered in Windows as needed. For this example, right-click the **Unallocated** volume and select **New Simple Volume** (see Figure 3-26), and follow the New Simple Volume Wizard.

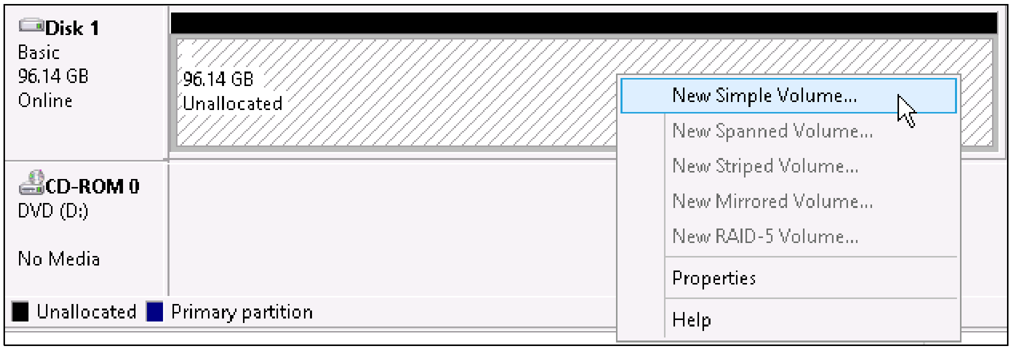
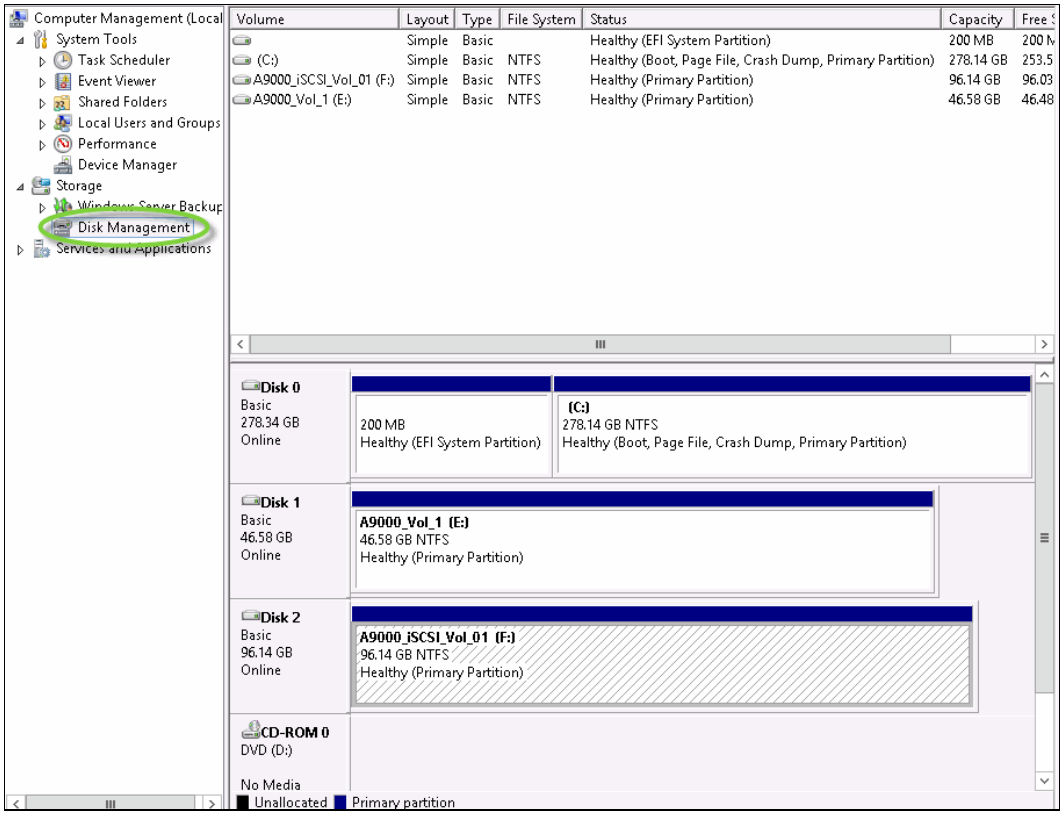


Figure 3-26 New Simple Volume

The end result after naming and formatting the volume is shown in the Disk Management window, as shown in Figure 3-27.



Volume	Layout	Type	File System	Status	Capacity	Free S
	Simple	Basic		Healthy (EFI System Partition)	200 MB	200 M
(C:)	Simple	Basic	NTFS	Healthy (Boot, Page File, Crash Dump, Primary Partition)	278.14 GB	253.5
A9000_iSCSI_Vol_01 (F:)	Simple	Basic	NTFS	Healthy (Primary Partition)	96.14 GB	96.03
A9000_Vol_1 (E:)	Simple	Basic	NTFS	Healthy (Primary Partition)	46.58 GB	46.48

Disk 0
Basic
278.34 GB
Online

200 MB
Healthy (EFI System Partition)

(C:)
278.14 GB NTFS
Healthy (Boot, Page File, Crash Dump, Primary Partition)

Disk 1
Basic
46.58 GB
Online

A9000_Vol_1 (E:)
46.58 GB NTFS
Healthy (Primary Partition)

Disk 2
Basic
96.14 GB
Online

A9000_iSCSI_Vol_01 (F:)
96.14 GB NTFS
Healthy (Primary Partition)

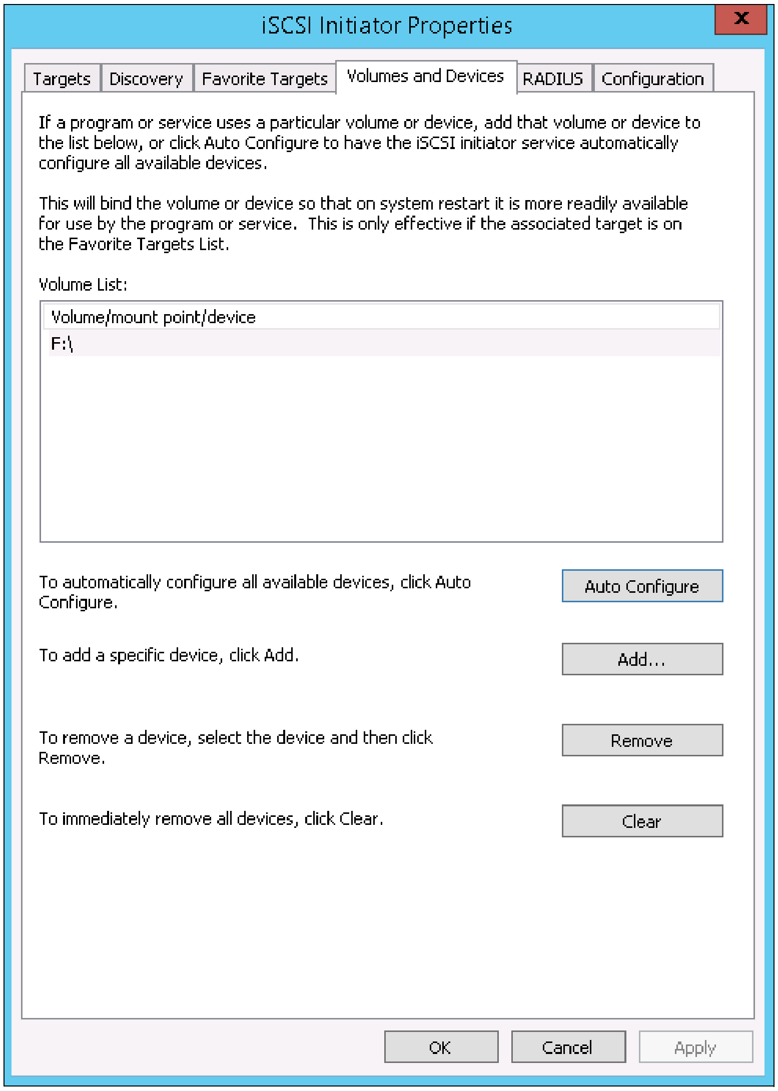
CD-ROM 0
DVD (D:)

No Media

■ Unallocated ■ Primary partition

Figure 3-27 Mapped LUNs are displayed in Disk Management

16. Select **iSCSI Initiator** from the Administrative Tools Window to display the iSCSI Initiator Properties window. Click the **Volumes and Devices** tab to show any newly created volumes (see Figure 3-28).



If a program or service uses a particular volume or device, add that volume or device to the list below, or click Auto Configure to have the iSCSI initiator service automatically configure all available devices.

This will bind the volume or device so that on system restart it is more readily available for use by the program or service. This is only effective if the associated target is on the Favorite Targets List.

Volume List:

Volume/mount point/device
F:\

To automatically configure all available devices, click Auto Configure.

To add a specific device, click Add.

To remove a device, select the device and then click Remove.

To immediately remove all devices, click Clear.

Figure 3-28 Connected Volumes list

3.2.4 Host Attachment Kit utilities

The Host Attachment Kit includes several utilities including the following examples:

- The xiv_devlist utility
- The xiv_diag utility

The xiv_devlist utility

This utility requires Administrator privileges. The utility lists the XIV, FlashSystem A9000, and A9000R volumes available to the host. Other disks are also listed separately. To run this utility, enter **xiv_devlist** at a command prompt (see Example 3-4).

Example 3-4 The xiv_devlist command results

```
PS C:\> xiv_devlist
IBM storage devices
------------------------------------------------------------------------------------------------------------------------

Device              Size (GB)  Paths  Vol Name               Vol ID  Storage ID  Storage Type       Hyper-Scale Mobility

------------------------------------------------------------------------------------------------------------------------

\\.\PHYSICALDRIVE1  50.1       6/6    ITSO_Vol1              14401   1322131     FlashSystem A9000  Idle

------------------------------------------------------------------------------------------------------------------------

\\.\PHYSICALDRIVE2  103.4      3/3    ITSO_Win2K12_iSCSI_01  14531   1322131     FlashSystem A9000  Idle

------------------------------------------------------------------------------------------------------------------------



Non-IBM storage devices
------------------------------------
Device              Size (GB)  Paths
------------------------------------
\\.\PHYSICALDRIVE0  299.0      N/A
------------------------------------
```

The xiv_diag utility

This utility requires Administrator privileges. It gathers diagnostic information from the operating system. The resulting compressed file can then be sent to IBM support teams for review and analysis. To run this utility, enter **xiv_diag** at a command prompt (see Example 3-5).

Example 3-5 The xiv_diag command results

```
PS C:\> xiv_diag
Welcome to the IBM storage host diagnostics tool, version 2.8.2.
This tool will gather essential support information from this host and save the information
to a file.
Specify the directory into which the xiv_diag file should be saved [default:
c:\users\admini~1\appdata\local\temp\2]:
Creating archive xiv_diag-results_2017-10-27_17-48-54
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 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 (1/4)...                               SKIPPED
INFO: Gathering Windows Setup API (2/4)...                                  DONE
INFO: Gathering Windows Setup API (3/4)...                                  DONE
```

```
INFO: Gathering Windows Setup API (4/4)...                                  DONE
INFO: Gathering Hardware Registry Subtree...                                DONE
INFO: Gathering xiv_devlist...                                              DONE
INFO: Gathering xiv_syslist...                                              DONE
INFO: Gathering xiv_syslist -L...                                           DONE
INFO: Gathering xiv_host_profiler --create --local --debug...               DONE
INFO: Gathering xhop results...                                             DONE
INFO: Gathering HAK version...                                              DONE
INFO: Gathering xiv_fc_admin -V...                                          DONE
INFO: Gathering xiv_fc_admin -P...                                          DONE
INFO: Gathering xiv_iscsi_admin -V...                                       DONE
INFO: Gathering xiv_iscsi_admin -P...                                       DONE
INFO: Gathering SCSI Inquiries...                                           DONE
INFO: Gathering Driver Versions...                                          DONE
INFO: Gathering WMI Disk And MPIO Objects...                                DONE
INFO: Gathering mpio_dump.py...                                             DONE
INFO: Gathering xiv_mscs_admin --report...                               SKIPPED
INFO: Gathering xiv_mscs_admin --verify...                               SKIPPED
INFO: Gathering xiv_mscs_admin --version...                              SKIPPED
INFO: Gathering build-revision file...                                      DONE
INFO: Gathering host_attach logs...                                         DONE
INFO: Gathering xiv logs...                                                 DONE
INFO: Gathering ibm products logs...                                        DONE
INFO: Gathering vss provider logs...                                     SKIPPED

INFO: Closing xiv_diag archive file                                         DONE
Deleting temporary directory...                                             DONE
INFO: Information gathering has been completed.
INFO: You can now send
c:\users\admini~1\appdata\local\temp\2\xiv_diag-results_2017-10-27_17-48-54.tar.gz to IBM
Support
 for further review.
INFO: Exiting.
```

3.3 Attaching a Microsoft Windows cluster

This section addresses the attachment of Microsoft Windows Server cluster nodes to IBM Storage. The procedure outlined is described with IBM XIV Storage although the process is similar for FlashSystem A9000 and A9000R.

> **Important:** The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, see the System Storage Interoperability Center (SSIC):
>
> https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

For more information, see the *IBM Storage Host Attachment Kit, Version 2.6.0, User Guide*, which is available at IBM Fix Central:

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

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

3.3.1 Prerequisites

To successfully attach a Windows Server cluster node to XIV and access storage, a number of prerequisites must be met. Generic prerequisites are indicated in the following list. Your environment might have extra requirements:

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

Supported FC HBAs

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

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

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

Multipath support

Microsoft provides a multipath framework and development kit that is called Multipath I/O (MPIO). The driver development kit allows storage vendors to create DSMs for MPIO. You can use DSMs to build interoperable multipath solutions that integrate tightly with Microsoft Windows.

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

Further information about Microsoft MPIO is available at this web page:

https://technet.microsoft.com/en-us/library/ee619778%28v=ws.10%29.aspx

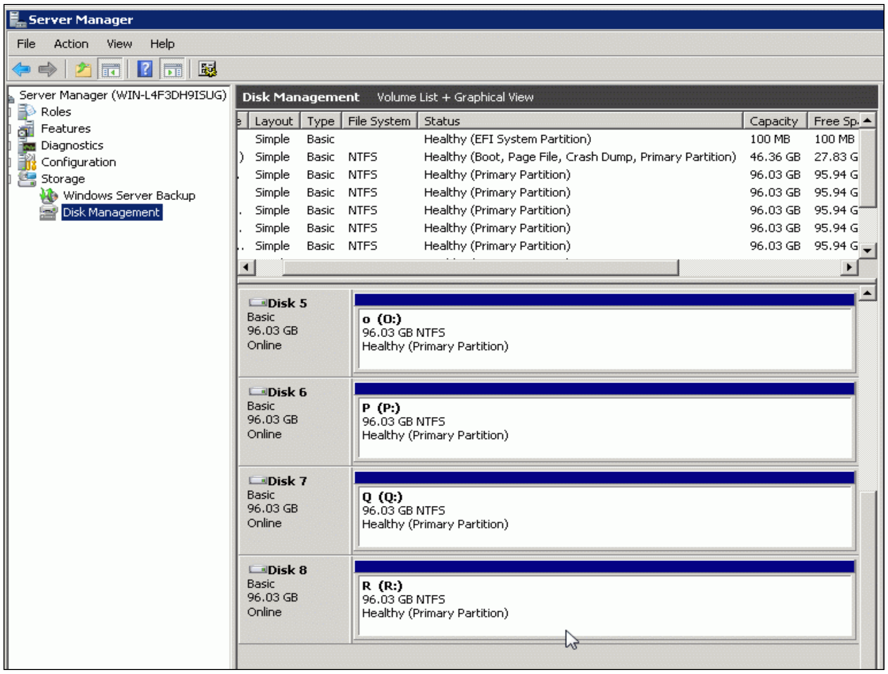
3.3.2 Installing cluster services

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

To install the cluster, complete the following steps:

1. In the Hyper-Scale Manager GUI, select **Hosts and Clusters Views →Cluster**. Create a cluster and create two hosts in the cluster as described in “Defining a host” on page 35. In this example, an XIV cluster named *clu_solman* was created and both nodes were placed in it.
2. Map all the LUNs to the cluster as described “Mapping LUNs to a host” on page 36. All LUNs are mapped to the XIV cluster, but not to the individual hosts.
3. Set up a cluster-specific configuration that includes the following characteristics:
   - All nodes are in the same domain.
   - Has network connectivity.
   - Has private (heartbeat) network connectivity.
   - Node 2 must not do any disk I/O.
   - Run the cluster configuration check.
4. On node 1, scan for new disks. Then, initialize, partition, and format them with NTFS. The following requirements are for shared cluster disks:
   - These disks must be basic disks.
   - For Windows Server 2008, you must decide whether they are Master Boot Record (MBR) disks or GUID Partition Table (GPT) disks.

   Figure 3-29 shows what this configuration looks like on node 1.

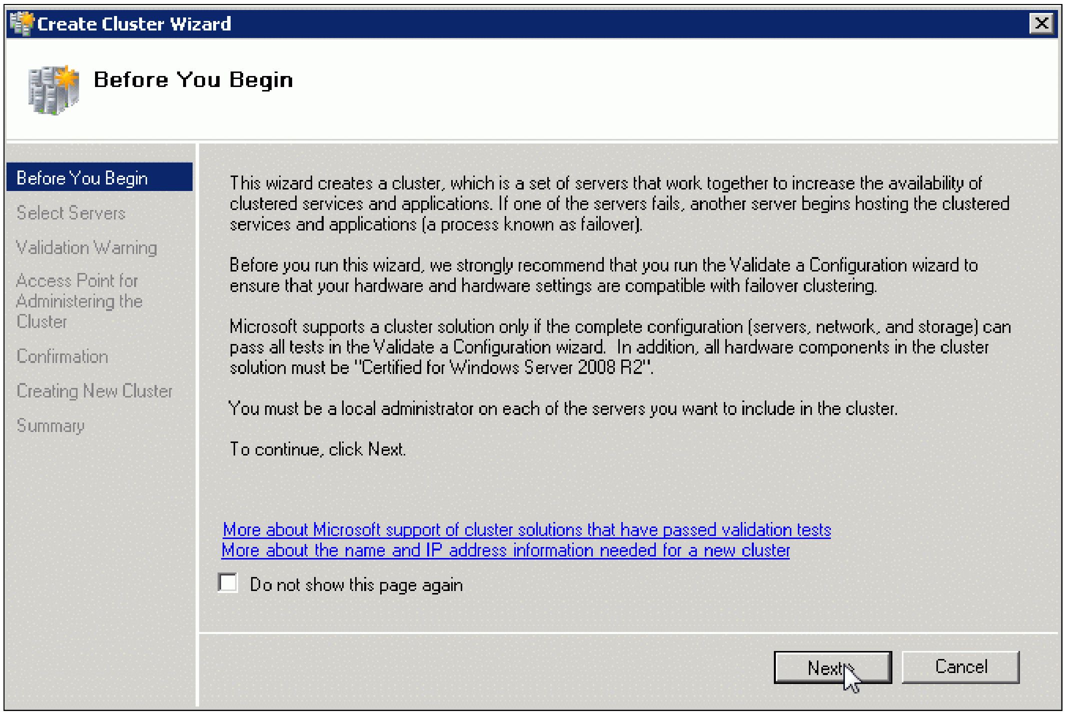


e	Layout	Type	File System	Status	Capacity	Free Sp.
	Simple	Basic		Healthy (EFI System Partition)	100 MB	100 MB
)	Simple	Basic	NTFS	Healthy (Boot, Page File, Crash Dump, Primary Partition)	46.36 GB	27.83 G
.	Simple	Basic	NTFS	Healthy (Primary Partition)	96.03 GB	95.94 G
	Simple	Basic	NTFS	Healthy (Primary Partition)	96.03 GB	95.94 G
.	Simple	Basic	NTFS	Healthy (Primary Partition)	96.03 GB	95.94 G
.	Simple	Basic	NTFS	Healthy (Primary Partition)	96.03 GB	95.94 G
..	Simple	Basic	NTFS	Healthy (Primary Partition)	96.03 GB	95.94 G

Disk 5 Basic 96.03 GB Online	**o (O:)** 96.03 GB NTFS Healthy (Primary Partition)
Disk 6 Basic 96.03 GB Online	**P (P:)** 96.03 GB NTFS Healthy (Primary Partition)
Disk 7 Basic 96.03 GB Online	**Q (Q:)** 96.03 GB NTFS Healthy (Primary Partition)
Disk 8 Basic 96.03 GB Online	**R (R:)** 96.03 GB NTFS Healthy (Primary Partition)

Figure 3-29 Initialized, partitioned, and formatted disks

5. Ensure that only one node accesses the shared disks until the cluster service is installed on all nodes. This restriction must be done before you continue to the Cluster wizard (see Figure 3-30). You no longer must turn off all nodes as you did with Windows Server 2003. You can bring all nodes into the cluster in a single step. However, no one is allowed to work on the other nodes.



This wizard creates a cluster, which is a set of servers that work together to increase the availability of clustered services and applications. If one of the servers fails, another server begins hosting the clustered services and applications (a process known as failover).

Before you run this wizard, we strongly recommend that you run the Validate a Configuration wizard to ensure that your hardware and hardware settings are compatible with failover clustering.

Microsoft supports a cluster solution only if the complete configuration (servers, network, and storage) can pass all tests in the Validate a Configuration wizard. In addition, all hardware components in the cluster solution must be "Certified for Windows Server 2008 R2".

You must be a local administrator on each of the servers you want to include in the cluster.

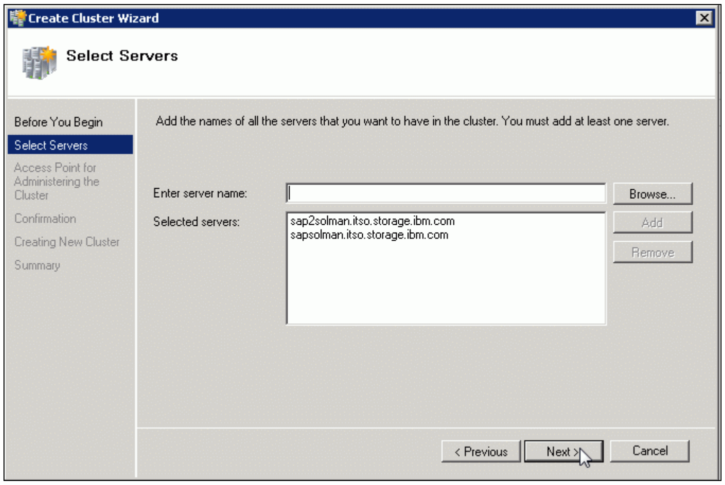
To continue, click Next.

More about Microsoft support of cluster solutions that have passed validation tests
More about the name and IP address information needed for a new cluster

□ Do not show this page again

Figure 3-30 Create Cluster Wizard welcome window

6. Select all nodes that belong to the cluster (see Figure 3-31).



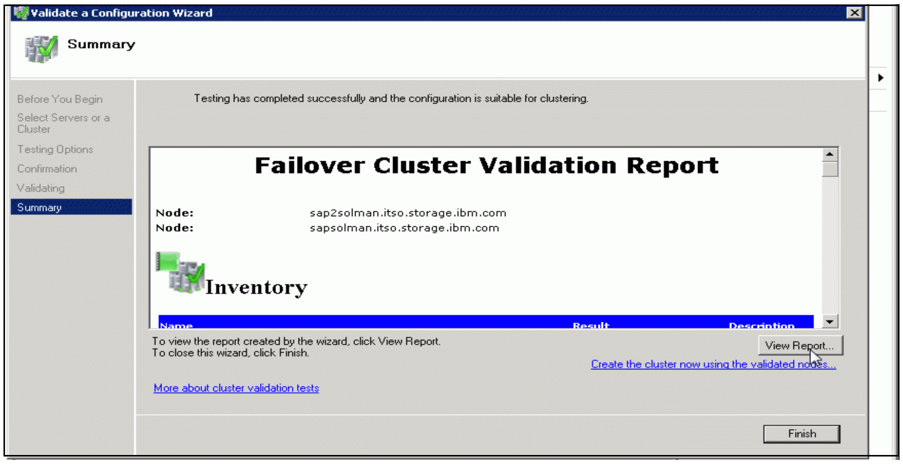
Add the names of all the servers that you want to have in the cluster. You must add at least one server.

Enter server name:

Selected servers:
sap2solman.itso.storage.ibm.com
sapsolman.itso.storage.ibm.com

Figure 3-31 Selecting your nodes

7. After the Create Cluster wizard completes, a summary panel shows that the cluster was created successfully (see Figure 3-32). Keep this report for documentation purposes.



Testing has completed successfully and the configuration is suitable for clustering.

Failover Cluster Validation Report

Node: sap2solman.itso.storage.ibm.com
Node: sapsolman.itso.storage.ibm.com

Inventory

To view the report created by the wizard, click View Report.
To close this wizard, click Finish.

Create the cluster now using the validated nodes...

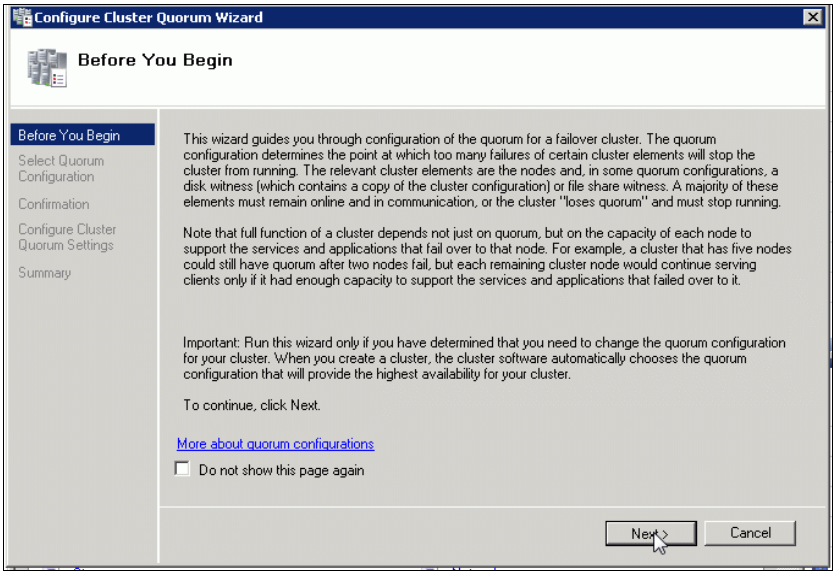
More about cluster validation tests

Figure 3-32 Failover Cluster Validation Report window

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

9. Check the access from node 2 to the shared disks and power node 1 on again.

10. Make sure that you have the correct cluster witness model, as shown in Figure 3-33. The old cluster model had a quorum as a single point of failure.


Configure Cluster Quorum Wizard
Before You Begin
Before You Begin
Select Quorum Configuration
Confirmation
Configure Cluster Quorum Settings
Summary
Next >
Cancel

This wizard guides you through configuration of the quorum for a failover cluster. The quorum configuration determines the point at which too many failures of certain cluster elements will stop the cluster from running. The relevant cluster elements are the nodes and, in some quorum configurations, a disk witness (which contains a copy of the cluster configuration) or file share witness. A majority of these elements must remain online and in communication, or the cluster "loses quorum" and must stop running.

Note that full function of a cluster depends not just on quorum, but on the capacity of each node to support the services and applications that fail over to that node. For example, a cluster that has five nodes could still have quorum after two nodes fail, but each remaining cluster node would continue serving clients only if it had enough capacity to support the services and applications that failed over to it.

Important: Run this wizard only if you have determined that you need to change the quorum configuration for your cluster. When you create a cluster, the cluster software automatically chooses the quorum configuration that will provide the highest availability for your cluster.

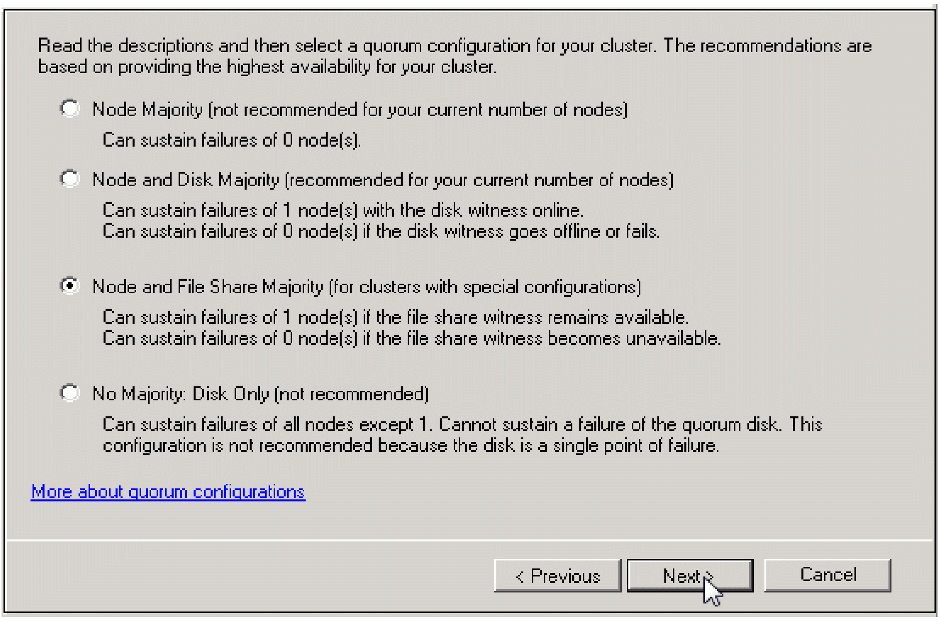
To continue, click Next.

More about quorum configurations

☐ Do not show this page again

Figure 3-33 Configure Cluster Quorum Wizard window

In this example, the cluster witness model is changed, which assumes that the witness share is in a third data center (see Figure 3-34).



Read the descriptions and then select a quorum configuration for your cluster. The recommendations are based on providing the highest availability for your cluster.

○ Node Majority (not recommended for your current number of nodes)

Can sustain failures of 0 node(s).

○ Node and Disk Majority (recommended for your current number of nodes)

Can sustain failures of 1 node(s) with the disk witness online.
Can sustain failures of 0 node(s) if the disk witness goes offline or fails.

● Node and File Share Majority (for clusters with special configurations)

Can sustain failures of 1 node(s) if the file share witness remains available.
Can sustain failures of 0 node(s) if the file share witness becomes unavailable.

○ No Majority: Disk Only (not recommended)

Can sustain failures of all nodes except 1. Cannot sustain a failure of the quorum disk. This configuration is not recommended because the disk is a single point of failure.

More about quorum configurations

Figure 3-34 Selecting the witness model

3.3.3 Configuring the IBM Storage Enabler for Windows Failover Clustering

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

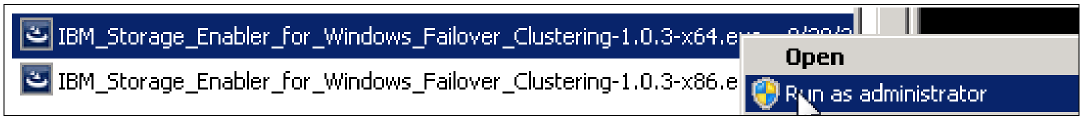
To find the software, release notes, and the user guide, go to IBM Fix Central:

https://ibm.biz/Bdse5N

Installing and configuring the Storage Enabler

The following instructions are based on the installation that was performed at the time of writing. For more information, see the instructions in the release notes and the user guide. These instructions are subject to change over time:

1. Start the installer as administrator (see Figure 3-35).



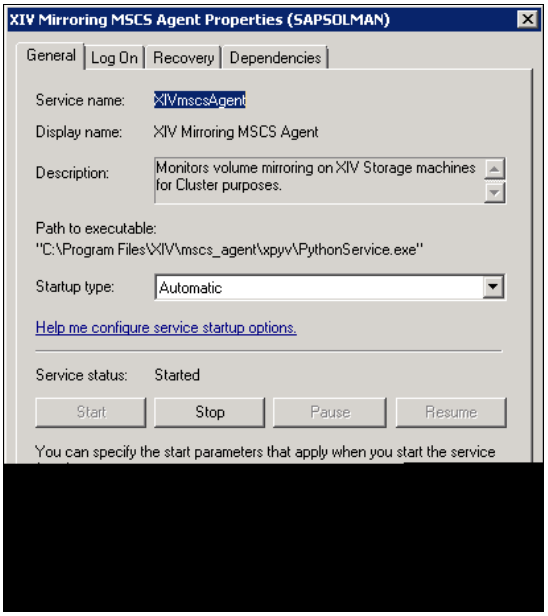
IBM_Storage_Enabler_for_Windows_Failover_Clustering-1.0.3-x64.e

IBM_Storage_Enabler_for_Windows_Failover_Clustering-1.0.3-x86.e

Figure 3-35 Starting the installation

2. Follow the wizard instructions.

   After the installation is complete, observe a new service named `XIVmscsAgent` (see Figure 3-36 on page 130).



XIV Mirroring MSCS Agent Properties (SAPSOLMAN)

Service name: XIVmscsAgent

Display name: XIV Mirroring MSCS Agent

Description: Monitors volume mirroring on XIV Storage machines for Cluster purposes.

Path to executable:
"C:\Program Files\XIV\mscs_agent\xpyv\PythonService.exe"

Startup type: Automatic

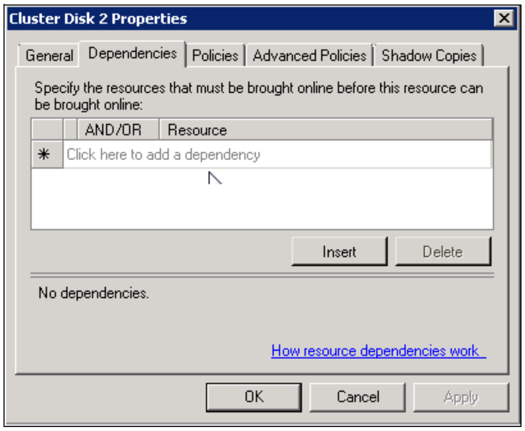
Help me configure service startup options.

Service status: Started

You can specify the start parameters that apply when you start the service

Figure 3-36 XIVmscsAgent as a service

No configuration took place until now. Therefore, the dependencies of the Storage LUNs did not change (see Figure 3-37).



Cluster Disk 2 Properties

Specify the resources that must be brought online before this resource can be brought online:

	AND/OR	Resource
*	Click here to add a dependency	

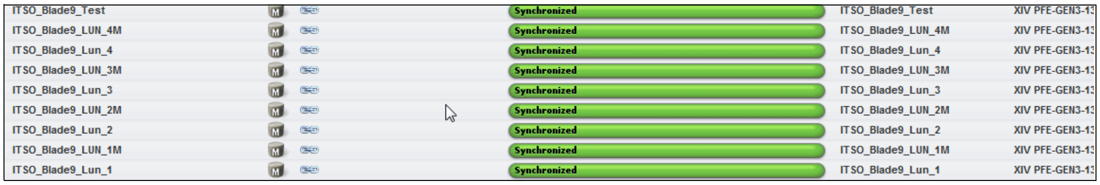
No dependencies.

How resource dependencies work

Figure 3-37 Dependencies of drive properties

3. Define the mirror connections for your LUNs between the two XIVs (see Figure 3-38). For more information about how to define the mirror pairs, see *IBM XIV Storage System Business Continuity Functions*, SG24-7759:

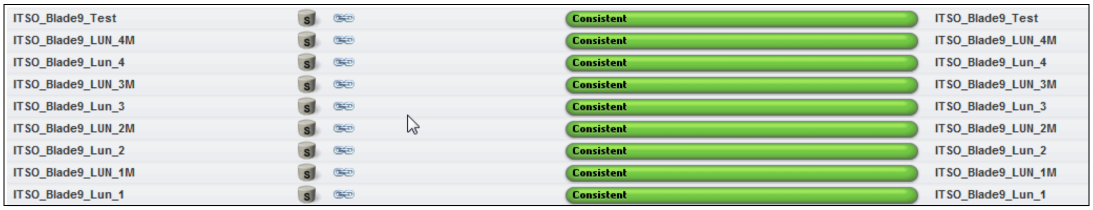
   http://www.redbooks.ibm.com/abstracts/sg247759.html



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

Figure 3-38 Mirror definitions at the master side and side of node 1

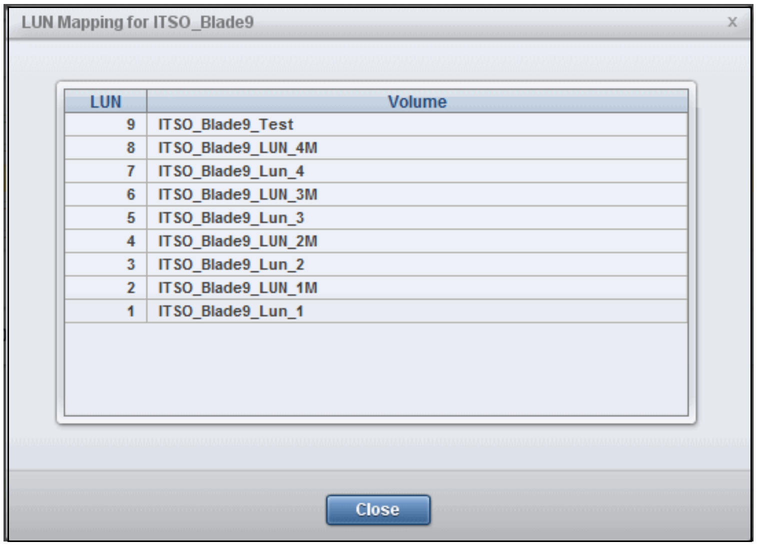
4. Also, define the connections on the subordinate side (see Figure 3-39).



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

Figure 3-39 Mirror definitions on the subordinate side

5. Redefine the host mapping of the LUNs on both XIVs. For a working cluster, both nodes and their HBAs must be defined in a cluster group. All of the LUNs that are provided to the cluster must be mapped to the cluster group itself, not to the nodes. When using the XIVmscsAgent, you must remap those LUNs to their specific XIV/node combination. Figure 3-40 shows the mapping for node 1 on the master side.

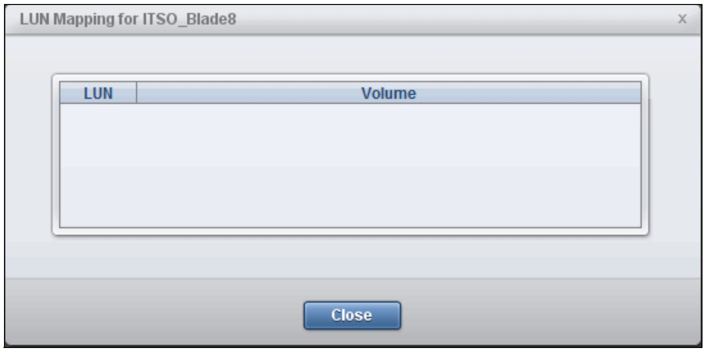


LUN Mapping for ITSO_Blade9

LUN	Volume
9	ITSO_Blade9_Test
8	ITSO_Blade9_LUN_4M
7	ITSO_Blade9_Lun_4
6	ITSO_Blade9_LUN_3M
5	ITSO_Blade9_Lun_3
4	ITSO_Blade9_LUN_2M
3	ITSO_Blade9_Lun_2
2	ITSO_Blade9_LUN_1M
1	ITSO_Blade9_Lun_1

Figure 3-40 Selecting the private mapping for node 1 on the master side

Figure 3-41 shows the mapping for node 2 on the master side.

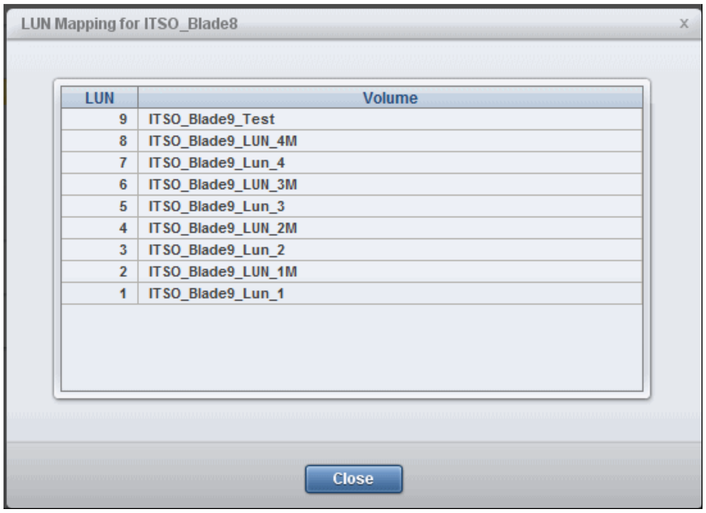


LUN Mapping for ITSO_Blade8

LUN	Volume

Figure 3-41 Changing the default mapping: Node 2 has no access to the master side

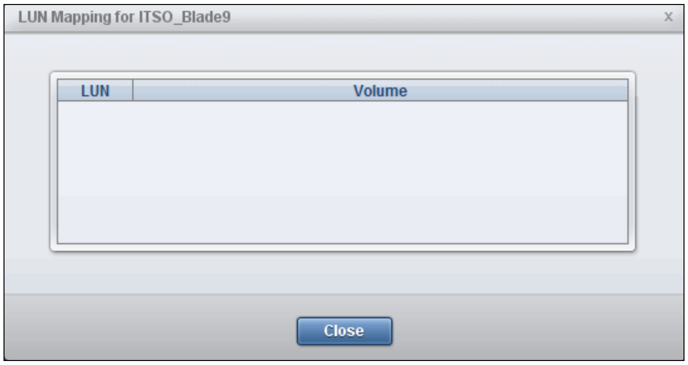
Figure 3-42 shows the mapping for node 2 on the subordinate side.


LUN Mapping for ITSO_Blade8

LUN	Volume
9	ITSO_Blade9_Test
8	ITSO_Blade9_LUN_4M
7	ITSO_Blade9_Lun_4
6	ITSO_Blade9_LUN_3M
5	ITSO_Blade9_Lun_3
4	ITSO_Blade9_LUN_2M
3	ITSO_Blade9_Lun_2
2	ITSO_Blade9_LUN_1M
1	ITSO_Blade9_Lun_1

Figure 3-42 Selecting the private mapping on the subordinate side for node 2

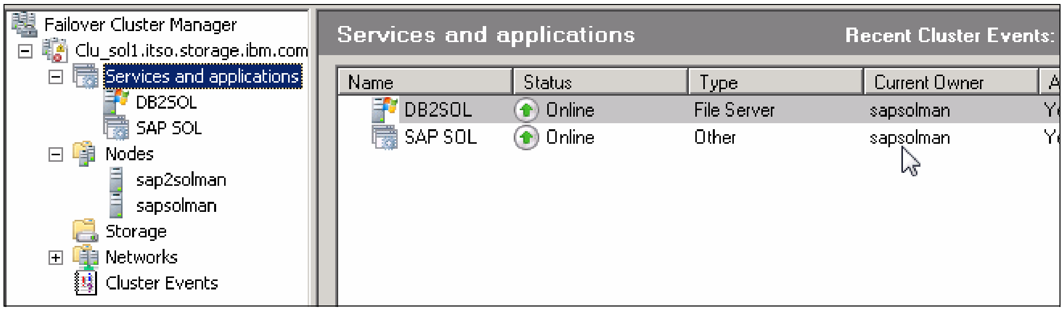
Figure 3-43 shows the mapping for node 1 on the subordinate side.


LUN Mapping for ITSO_Blade9

LUN	Volume

Figure 3-43 Node 1 has no access to XIV 2

6. Check that all resources are on node 1, where the Mirror Master side is defined, as shown in Figure 3-44.



Name	Status	Type	Current Owner	A
DB2SOL	Online	File Server	sapsolman	Y
SAP SOL	Online	Other	sapsolman	Y

Figure 3-44 All resources are on node 1

7. To configure the XIVmscsAgent, run the **admin** tool with the **-install** option (see Example 3-6).

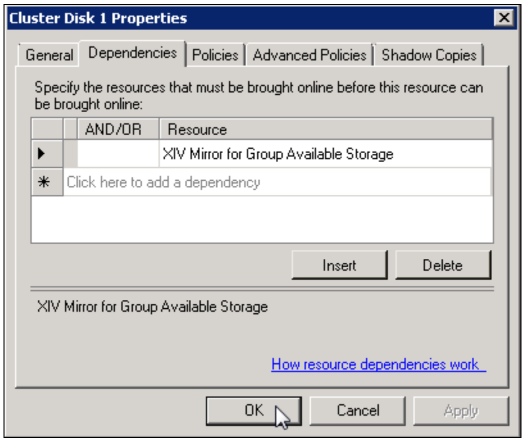
Example 3-6 How to use mcsc_agent

```
C:\Users\Administrator.ITSO>cd C:\Program Files\XIV\mscs_agent\bin
C:\Program Files\XIV\mscs_agent\bin>dir
 Volume in drive C has no label.
 Volume Serial Number is CA6C-8122
Directory of C:\Program Files\XIV\mscs_agent\bin
09/29/2011  11:13 AM    <DIR>          .
09/29/2011  11:13 AM    <DIR>          ..
09/14/2011  03:37 PM             1,795 project_specific_pyrunner.py
09/13/2011  07:20 PM             2,709 pyrunner.py
09/14/2011  11:48 AM           134,072 xiv_mscs_admin.exe
09/14/2011  11:48 AM           134,072 xiv_mscs_service.exe
               4 File(s)        272,648 bytes
               2 Dir(s)  14,025,502,720 bytes free
C:\Program Files\XIV\mscs_agent\bin>xiv_mscs_admin.exe
Usage: xiv_mscs_admin [options]
Options:
  --version              show program's version number and exit
  -h, --help             show this help message and exit
  --install              installs XIV MSCS Agent components on this node and
cluster Resource Type
  --upgrade              upgrades XIV MSCS Agent components on this node
  --report               generates a report on the cluster
  --verify               verifies XIV MSCS Agent deployment
  --fix-dependencies     fixes dependencies between Physical Disks and XIV Mirror
resources
  --deploy-resources     deploys XIV Mirror resources in groups that contain
Physical Disk Resources
  --delete-resources     deletes all existing XIV Mirror resources from the
cluster
  --delete-resourcetype  deletes the XIV mirror resource type
  --uninstall            uninstalls all XIV MSCS Agent components from this node
  --change-credentials   change XIV credentials
  --debug                enables debug logging
  --verbose              enables verbose logging
  --yes                  confirms distruptive operations
XCLI Credentials Options:
    --xcli-username=USERNAME
    --xcli-password=PASSWORD
C:\Program Files\XIV\mscs_agent\bin>xiv_mscs_admin.exe --install --verbose
--xcli-username=itso  --xcli-password=<PASSWORD>
2011-09-29 11:19:12 INFO classes.py:76 checking if the resource DLL exists
2011-09-29 11:19:12 INFO classes.py:78 resource DLL doesn't exist, installing it
2011-09-29 11:19:12 INFO classes.py:501 The credentials MSCS Agent uses to connect
to the XIV Storage System have been change
d. Check the guide for more information about credentials.
Installing service XIVmscsAgent
Service installed
Changing service configuration
Service updated
2011-09-29 11:19:14 INFO classes.py:85 resource DLL exists
C:\Program Files\XIV\mscs_agent\bin>
```

8. To deploy the resources into the geo cluster, run the `xiv_mcs_admin.exe` utility:

```
C:\Program Files\XIV\mscs_agent\bin>xiv_mscs_admin.exe --deploy-resources
--verbose  --xcli-username=itso  --xcli-password=<PASSWORD> --yes
```

The cluster dependencies that result are shown in Figure 3-45.



Specify the resources that must be brought online before this resource can be brought online:

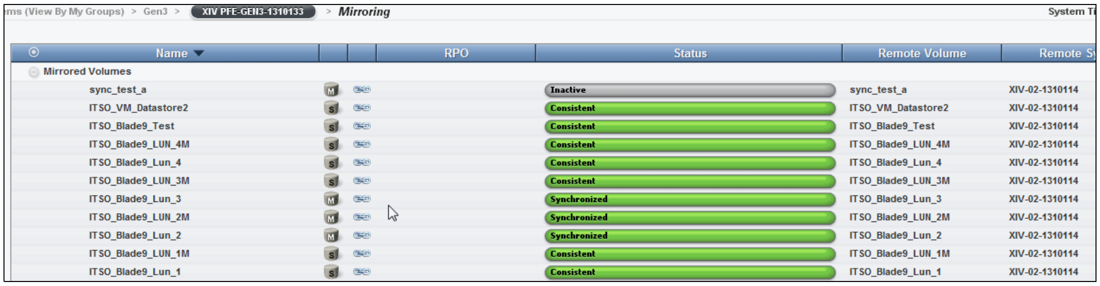
	AND/OR	Resource
▶		XIV Mirror for Group Available Storage
✳	Click here to add a dependency	

XIV Mirror for Group Available Storage

Figure 3-45   Dependencies after deploying the resources

9. Power down node 1. Then, on node 2, repeat step 5 on page 131 through step 8.

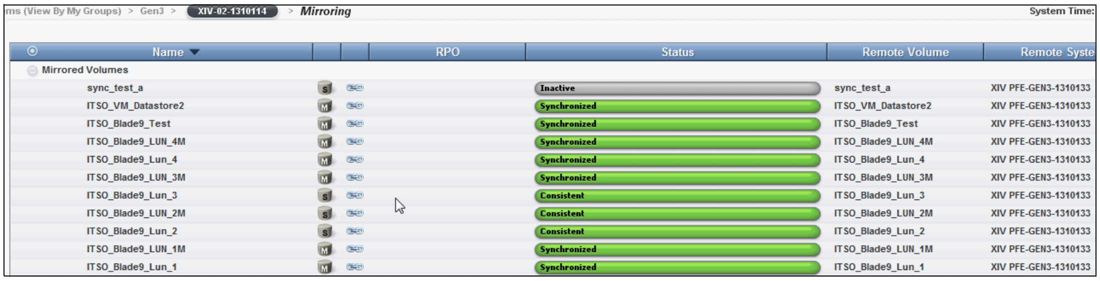
A switch of the cluster resource group from node 1 to node 2 leads to a change of the replication direction. XIV2 becomes the master and XIV1 becomes the subordinate, as shown in Figure 3-46 and Figure 3-47.


ms (View By My Groups) > Gen3 > XIV PFE-GEN3-1310133 > Mirroring
System Ti

Name			RPO	Status	Remote Volume	Remote Sy
Mirrored Volumes						
sync_test_a	M	🔗		Inactive	sync_test_a	XIV-02-1310114
ITSO_VM_Datastore2	S	🔗		Consistent	ITSO_VM_Datastore2	XIV-02-1310114
ITSO_Blade9_Test	S	🔗		Consistent	ITSO_Blade9_Test	XIV-02-1310114
ITSO_Blade9_LUN_4M	S	🔗		Consistent	ITSO_Blade9_LUN_4M	XIV-02-1310114
ITSO_Blade9_Lun_4	S	🔗		Consistent	ITSO_Blade9_Lun_4	XIV-02-1310114
ITSO_Blade9_LUN_3M	S	🔗		Consistent	ITSO_Blade9_LUN_3M	XIV-02-1310114
ITSO_Blade9_Lun_3	M	🔗		Synchronized	ITSO_Blade9_Lun_3	XIV-02-1310114
ITSO_Blade9_LUN_2M	M	🔗		Synchronized	ITSO_Blade9_LUN_2M	XIV-02-1310114
ITSO_Blade9_Lun_2	M	🔗		Synchronized	ITSO_Blade9_Lun_2	XIV-02-1310114
ITSO_Blade9_LUN_1M	S	🔗		Consistent	ITSO_Blade9_LUN_1M	XIV-02-1310114
ITSO_Blade9_Lun_1	S	🔗		Consistent	ITSO_Blade9_Lun_1	XIV-02-1310114

Figure 3-46   XIVmscsAgent changing the replication direction

The results are shown in Figure 3-47.



Name			RPO	Status	Remote Volume	Remote Syste
Mirrored Volumes						
sync_test_a	S	🔗		Inactive	sync_test_a	XIV PFE-GEN3-1310133
ITSO_VM_Datastore2	M	🔗		Synchronized	ITSO_VM_Datastore2	XIV PFE-GEN3-1310133
ITSO_Blade9_Test	M	🔗		Synchronized	ITSO_Blade9_Test	XIV PFE-GEN3-1310133
ITSO_Blade9_LUN_4M	M	🔗		Synchronized	ITSO_Blade9_LUN_4M	XIV PFE-GEN3-1310133
ITSO_Blade9_Lun_4	M	🔗		Synchronized	ITSO_Blade9_Lun_4	XIV PFE-GEN3-1310133
ITSO_Blade9_LUN_3M	M	🔗		Synchronized	ITSO_Blade9_LUN_3M	XIV PFE-GEN3-1310133
ITSO_Blade9_Lun_3	S	🔗		Consistent	ITSO_Blade9_Lun_3	XIV PFE-GEN3-1310133
ITSO_Blade9_LUN_2M	S	🔗		Consistent	ITSO_Blade9_LUN_2M	XIV PFE-GEN3-1310133
ITSO_Blade9_Lun_2	S	🔗		Consistent	ITSO_Blade9_Lun_2	XIV PFE-GEN3-1310133
ITSO_Blade9_LUN_1M	M	🔗		Synchronized	ITSO_Blade9_LUN_1M	XIV PFE-GEN3-1310133
ITSO_Blade9_Lun_1	M	🔗		Synchronized	ITSO_Blade9_Lun_1	XIV PFE-GEN3-1310133

Figure 3-47   A switch of the cluster resources leads to a “change role” on XIV

3.4 Attaching a Microsoft Hyper-V Server

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

System hardware requirements

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

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

  https://www.windowsservercatalog.com/

Installing Hyper-V in Windows Server 2016, 2012 R2, and 2008 R2

The Server Core option on Windows Server provides a subset of the features of Windows Server 2016, 2012 R2, and 2008 R2. This option runs these supported server roles without a full Windows installation:

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

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

Using Hyper-V on a Server Core installation reduces the *attack surface*. The attack surface is the scope of interfaces, services, APIs, and protocols that a hacker can use to attempt to gain entry into the software. As a result, a Server Core installation reduces management requirements and maintenance. Microsoft provides management tools to remotely manage the Hyper-V role and virtual machines (VMs).

For more information about using the XIV Storage System with Microsoft Hyper-V, see *Microsoft Hyper-V with IBM XIV Storage System Gen3*:

http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102456

Note: The cited paper describes Microsoft Hyper-V with IBM XIV Storage System Gen 3, however the information is applicable to FlashSystem A9000 and A9000R.

3.5 Microsoft System Center Virtual Machine Manager Storage Automation

With *Microsoft System Center Virtual Machine Manager (*SCVMM) alongside IBM FlashSystem A9000, A9000R, or XIV Gen3, administrators have more features to extend Microsoft Hyper-V virtualization for Cloud. They can now perform storage management tasks through an integrated administrative interface. For example, within the SCVMM's graphical user interface (GUI), administrators can discover, classify, allocate, provision, map, assign, and decommission storage. This storage can be associated with clustered and stand-alone virtualization hosts. Usually, these tasks require multiple steps that use different applications and skills.

IBM has complied with the Storage Networking Industry Association (SNIA) standards requirements for a long time. In addition, the Common Information Model (CIM) framework is embedded in FlashSystem A9000, A9000R, and XIV Gen3. Therefore, these IBM Storage devices are an SCVMM supported device, and are fully compliant with the SNIA Storage Management Initiative Specification (SMI-S) 1.4.

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

This section addresses storage automation concepts and provides some implementation guidelines for SCVMM with IBM Storage.

3.5.1 Open API overview

The XIV Open API complies to SMI-S, as specified by the SNIA to manage the XIV Storage System Gen. 3. This API can be used by any storage resource management application to configure and manage the XIV. Similarly, the FlashSystem A9000 and A9000R Open API is a non-proprietary storage-management client application. The Open API uses the Storage Management Initiative Specification (SMI-S), as defined by the Storage Networking Industry Association (SNIA) to view LUN information.

Microsoft SCVMM uses the Open API to communicate with the System's CIM Agent, as shown in Figure 3-48. The main components are the CIM object manager (CIMOM) and the SMI-S Provider (device provider). With IBM Storage's single tier design, the SMI-S provider and CIM agent are treated identically due to the CIM component collective functionality. The SMI-S provider is also referred to as the SCVMM storage provider to align with the SCVMM application interfaces and documentation.

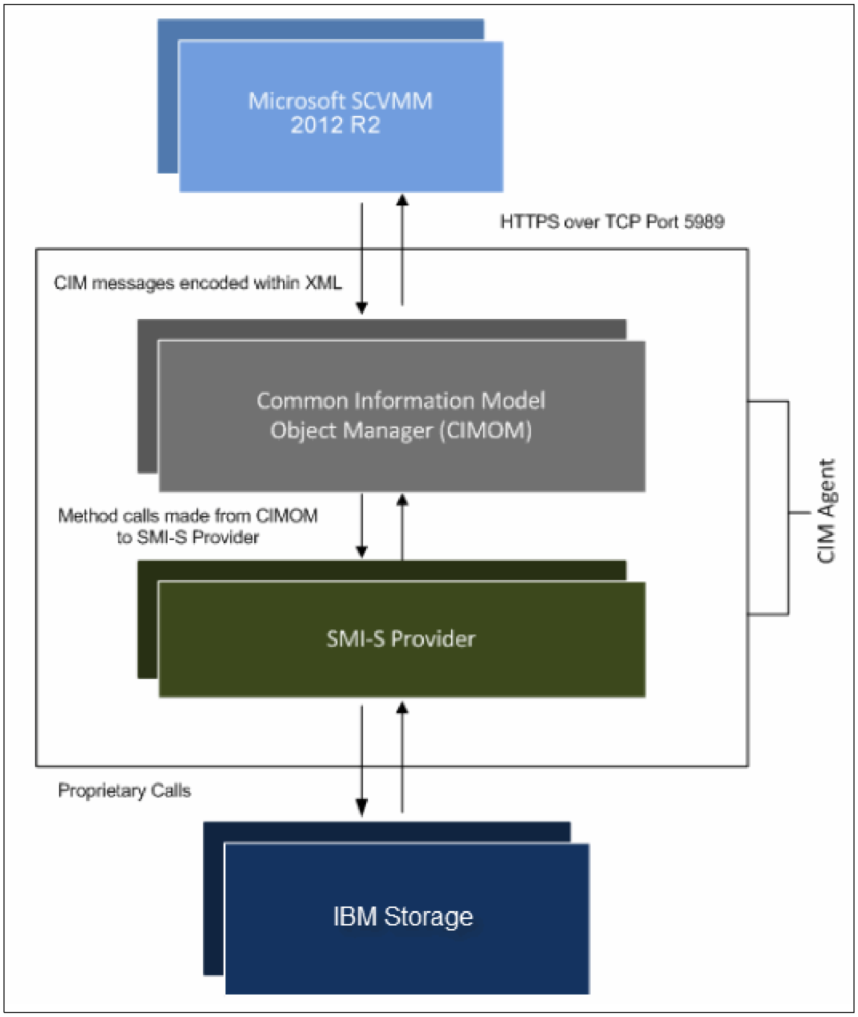


Figure 3-48 SCVMM 2012 R2 CIM communication with IBM Storage

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

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

The CIM Agent is embedded in the administrative module and does not require configuration. It is also enabled by default.

The CIM agent has the following limitations:

- The CIM agent is able to manage only the system on which the administrative module is installed.
- The secure mode must be used over port 5989.
- The CIM Agent uses the same account that is used to manage the IBM Storage with the Hyper-Scale Manager, XIV GUI, or the XCLI.

For more information about Open API and CIM framework, see the following guides:

- *IBM FlashSystem A9000 and IBM FlashSystem A9000R: Open API Reference Guide*, publication number SC27-8561:

  http://www.ibm.com/support/knowledgecenter/en/STJKMM_12.0.2/PDFs/IBM_FlashSystem_A9000x_12.0.x_Open_API.pdf?view=kc

- *IBM XIV Storage System: Application Programming Interface Reference*, publication number GC27-3916:

  http://www.ibm.com/support/knowledgecenter/en/STJTAG/com.ibm.help.xivgen3.doc/docs/XIV_11.6_API_RG.pdf?view=kc

3.5.2 System Center Virtual Machine Manager overview

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

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

Within a cloud, private or public, fabric is a key concept. *Fabric* is composed by hosts, host groups, library servers, networking, and storage configuration. In SCVMM 2012, this concept simplifies the operations that are required to use and define resource pools. From a user's perspective, deploying VMs and provisioning storage capacity without performing storage administrative tasks becomes possible.

The SCVMM fabric management window is shown in Figure 3-49.

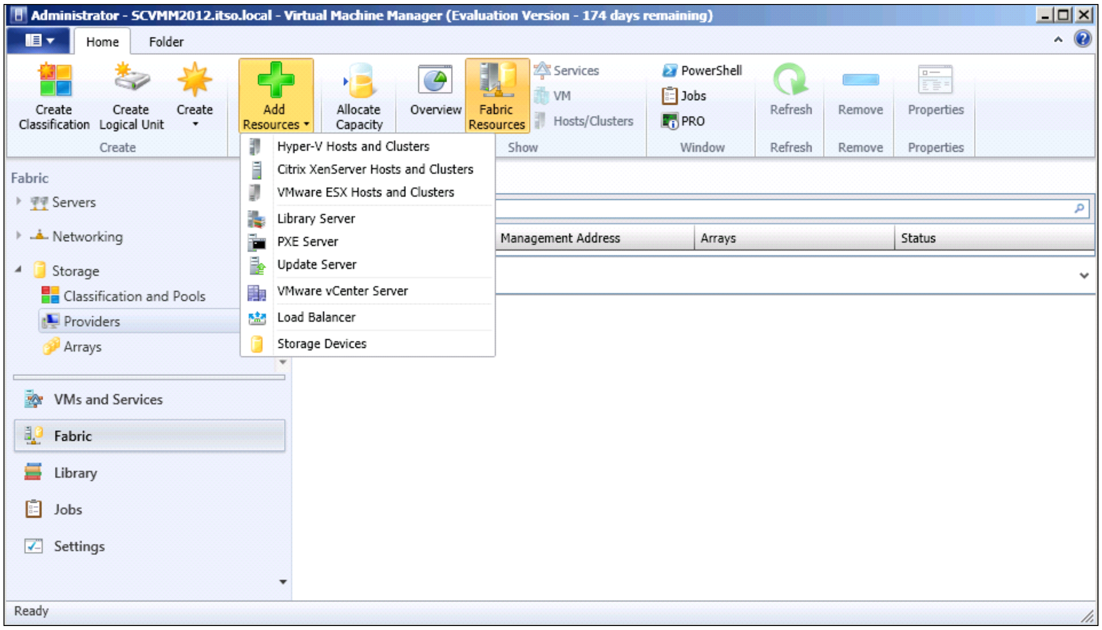


Figure 3-49 SCVMM fabric management

For more information about configuring and deploying fabric resources, see the IBM technical white paper *Microsoft System Center Virtual Machine Manager 2012 R2 storage automation with IBM XIV Storage System Gen3*:

http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102517

Note: This white paper details SCVMM 2012 R2 with IBM XIV Storage System Gen3, however the information and procedures described in the white paper are also applicable to FlashSystem A9000 and A9000R.

For more information about the procedures to define storage resources, see the topic about “SCVMM storage automation step-by-step processes” in the referenced white paper. These resources include storage classifications, logical units, and storage pools that are made available to Hyper-V hosts and host clusters. The following SCVMM 2012 R2 key features that use IBM XIV Storage System Gen3 are also documented:

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

For more information, see the “Configuring Storage in VMM” overview:

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

4

Linux connectivity

This chapter addresses the specifics for attaching FlashSystem A9000, A9000R, and XIV Storage System to host systems that are running Linux. Although it does not cover every aspect of connectivity, it addresses all of the basics. The examples usually use the Linux console commands because they are more generic than the GUIs that are provided by vendors.

This guide covers the following hardware architectures that are supported for the attachment:

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

Older Linux versions are supported to work with the storage systems. However, the scope of this chapter is limited to more recent enterprise level distributions of SUSE Linux Enterprise Server, and Red Hat Enterprise Linux.

> **Important:** The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, *always* see the System Storage Interoperation Center (SSIC):
>
> `https://www.ibm.com/systems/support/storage/ssic/interoperability.wss`
>
> You can retrieve the Host Attachment Kit publications from Fix Central:
>
> `http://www.ibm.com/support/fixcentral/`

This chapter includes the following topics:

4.1 IBM storage systems and Linux support overview

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

4.1.1 Issues that distinguish Linux from other operating systems

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

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

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

IBM supports mainly the following Linux distributions that are targeted at enterprise clients:

- Red Hat Enterprise Linux
- SUSE Linux Enterprise Server

Furthermore, CentOS is supported, which derives from Red Hat Enterprise Linux, but without a comparable support.

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

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

4.1.2 Reference material

A wealth of information is available to help you set up your Linux server and attach it to the storage system. The IBM Storage Host Attachment Kit for Linux release notes and user guide contain up-to-date materials. You can retrieve them from Fix Central for FlashSystem A9000, IBM A9000R, or XIV:

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

Primary references

Other useful references about Linux distributions are as follows:

- Red Hat Enterprise Linux 5: Online Storage Reconfiguration Guide:

  https://ibm.biz/BdseNA

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

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

- Red Hat Enterprise Linux 6: DM Multipath Configuration and Administration:

  https://ibm.biz/BdseN9

  The DM Multipath has the following changes in Red Hat Enterprise Linux 6:

  - The **mpathconf** utility can be used to change the configuration file, **mulitpathd** daemon, and **chkconfig**.
  - The new path selection algorithms **queue-length** and **service-time** provide benefits for certain workloads.
  - The location of the bindings file etc/multipath/bindings is different.
  - Output of **user_friendly_names=yes**, results in mpath*n* being an alphabetic character and not numeric as in Red Hat Enterprise Linux 5.

- Red Hat Enterprise Linux 7: Storage Administration Guide:

  https://ibm.biz/BdseNu

  This guide is part of the documentation that is provided for Red Hat Enterprise Linux 7.

- Red Hat Enterprise Linux 7: DM Multipath Configuration and Administration:

  https://ibm.biz/BdseNC

  Another part of the Red Hat Enterprise Linux 7 documentation. It contains useful information for anyone who works with Device Mapper Multipathing (DM-MP). Most of the information is valid for Linux in general. It includes:

  - Understanding how Device Mapper Multipathing works
  - Setting up and configuring DM-MP within Red Hat Enterprise Linux 7
  - Troubleshooting DM-MP

- SUSE Linux Enterprise Server 12: Storage Administration Guide:

  https://www.suse.com/documentation/sles-12/index.html

  This publication is part of the documentation for Novell SUSE Linux Enterprise Server 12, Service Pack 1. Although written specifically for SUSE Linux Enterprise Server, it contains useful information for any Linux user who is interested in storage-related subjects. The following are the most useful topics in the book:

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

- The Linux on Power Community Wiki:

  https://ibm.biz/BdDKbG

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

  - A discussion forum
  - An announcement section
  - Technical articles

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

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

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

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

Other sources of information

The Linux distributor documentation pages are good starting points for installation, configuration, and administration of Linux servers. The following documentation resources are especially useful for server-platform-specific issues:

- SUSE Linux Enterprise Server documentation:

  https://www.suse.com/documentation/

- Red Hat Enterprise Linux documentation:

  https://access.redhat.com/documentation/en/red-hat-enterprise-linux/

Other IBM resources:

- IBM z Systems has its own web page for storage attachment using FCP:

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

- *IBM z Systems Connectivity Handbook*, SG24-5444:

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

  This book describes connectivity options that available for use within and beyond the data center for IBM z Systems servers. It has a section for FC attachment, although it is outdated with regards to multipathing.

4.1.3 Storage-related improvements to Linux

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

Past issues

The following partial list of storage-related issues in older Linux versions are addressed in newer versions:

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

Dynamic generation of device nodes

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

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

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

Persistent device naming

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

Multipathing

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

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

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

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

Adding and removing volumes online

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

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

Dynamic LUN resizing

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

Write barriers availability for ext4 file system

Red Hat Enterprise Linux 6 by default uses the ext4 file system. This file system uses the *write barriers* feature to improve performance. A write barrier kernel mechanism ensures that file system metadata is correctly written and ordered on persistent storage. The write barrier continues to do so even when storage devices with volatile write caches lose power.

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

> **Important:** IBM has confirmed that write barriers have a negative impact on XIV performance. Ensure that all of your mounted disks use the following switch:
>
> ```
> mount -o nobarrier /fs
> ```

For more information, see the *Red Hat Linux Enterprise 6: Storage Administration Guide*:

https://ibm.biz/Bdse7n

4.2 Basic host attachment

This section addresses the steps to make FlashSystem A9000, A9000R, and XIV volumes available to your Linux host. It addresses attaching storage for the different hardware architectures. It also describes configuration of the Fibre Channel HBA driver, setting up multipathing, and any required special settings.

4.2.1 Platform-specific remarks

The most popular hardware system for Linux is the Intel x86 (32- or 64-bit) architecture. However, this architecture allows only direct mapping of volumes to hosts through Fibre Channel fabrics and HBAs, or IP networks with or without CNAs. IBM z Systems and IBM Power Systems provide extra mapping methods so you can use their much more advanced virtualization capabilities.

IBM Power Systems

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

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

Direct attachment is not described because it works the same way as with the other systems. VIOS attachment requires specific considerations. More information about how VIOS works and how it is configured is in Chapter 6, “Clients connecting through VIOS” on page 219.

For more information, see the following publications:

- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940:

  http://www.redbooks.ibm.com/abstracts/sg247940.html
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590:

  http://www.redbooks.ibm.com/abstracts/sg247590.html

Virtual VSCSI disks through VIOS

Linux on Power distributions contain a kernel module (driver) for a virtual SCSI HBA. This driver is called `ibmvscsi`, and attaches the virtual disks that are provided by the VIOS to the Linux system. The devices as seen by the Linux system are shown in Example 4-1.

Example 4-1 Virtual SCSI disks

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

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

Virtual Fibre Channel adapters through NPIV

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

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

Linux on Power distributions come with a kernel module for the virtual HBA called `ibmvfc`. This module presents the virtual HBA to the Linux operating system as though it were a real FC HBA. Volumes that are attached to the virtual HBA are displayed as though they are connected through a physical adapter (see Example 4-2).

Example 4-2 Volumes that are mapped through NPIV virtual HBAs

```
p6-570-lpar13:~ # lsscsi
[1:0:0:0]    disk    IBM      2810XIV          10.2  /dev/sdc
[1:0:0:1]    disk    IBM      2810XIV          10.2  /dev/sdd
[1:0:0:2]    disk    IBM      2810XIV          10.2  /dev/sde
[2:0:0:0]    disk    IBM      2810XIV          10.2  /dev/sdm
[2:0:0:1]    disk    IBM      2810XIV          10.2  /dev/sdn
[2:0:0:2]    disk    IBM      2810XIV          10.2  /dev/sdo
```

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

z Systems

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

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

Linux on z Systems running natively in a z Systems LPAR

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

> **Tip:** In IBM z Systems, the term *adapters* is better suited than the more common *channels* term that is often used in the z Systems environment.

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

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

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

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

- Fibre Channel (FCP) attached SCSI devices

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

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

- FCP attachment of SCSI devices through NPIV

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

Tip: Linux on z Systems can also use *count-key-data devices* (CKDs). CKDs are the traditional mainframe method to access disks. However, FlashSystem A9000, A9000R, and XIV Storage System do not support the CKD protocol, so it is not described in this book.

4.2.2 Configuring for Fibre Channel attachment

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

Loading the Linux Fibre Channel drivers

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

- QLogic: The most used HBAs for Linux on the Intel X86 system. The kernel module `qla2xxx` is a unified driver for all types of QLogic FC HBAs. It is included in the enterprise Linux distributions. The shipped version is supported for storage system attachment.
- Emulex: Sometimes used in Intel x86 servers and, rebranded by IBM, the standard HBA for the Power platform. The kernel module `lpfc` is a unified driver that works with all Emulex FC HBAs. A supported version is also included in the enterprise Linux distributions for both Intel x86 and Power Systems.
- Brocade: Provides *Converged Network Adapters* (CNAs) that operate as FC and Ethernet adapters. They are supported on Intel x86 for FC attachment. The kernel module version that is provided with the current enterprise Linux distributions is not supported. You must download the supported version from the Brocade website. The driver package comes with an installation script that compiles and installs the module. The script might cause support issues with your Linux distributor because it modifies the kernel. The FC kernel module for the CNAs is called `bfa`. The driver can be downloaded from Brocade:

  http://www.brocade.com/services-support/drivers-downloads/index.page
- IBM FICON Express: These are the HBAs for the z Systems system. They can either operate in FICON mode for traditional CKD devices, or FCP mode for FB devices. Linux deals with them directly only in FCP mode. The driver is part of the enterprise Linux distributions for z Systems, and is called `zfcp`.

Kernel modules (drivers) are loaded with the **`modprobe`** command. They can be removed if they are not in use as shown in Example 4-3.

Example 4-3 Loading and unloading a Linux Fibre Channel HBA Kernel module

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

After the driver is loaded, the FC HBA driver examines the FC fabric, detects attached volumes, and registers them in the operating system. To discover whether a driver is loaded, and what dependencies exist for it, use the **`lsmod`** command (Example 4-4 on page 150).

Example 4-4 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, ....
```

To get detailed information about the kernel module, such as the version number and what options it supports, use the **modinfo** command. You can see a partial output in Example 4-5.

Example 4-5 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 Linux on z Systems automatically scans and registers the attached volumes, but only in the most recent Linux distributions and only if NPIV is used. Otherwise, you must tell it explicitly which volumes to access. The reason is that the Linux virtual machine might not be intended to use all volumes that are attached to the HBA. For more information, see “Linux on z Systems running in a virtual machine under z/VM” on page 148, and “Adding volumes to a Linux on z Systems system” on page 162.

Using the FC HBA driver at installation time

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

When you have an unsupported driver version included with your Linux distribution, either replace it immediately after installation, or use a driver disk during the installation.

Considerations:

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

For more information, see 4.5, “Boot from SAN in Linux” on page 185.

Making the FC driver available early in the boot process

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

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

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

SUSE Linux Enterprise Server

Kernel modules that must be included in the initramfs are listed in the `/etc/sysconfig/kernel` file on the line that starts with `INITRD_MODULES`. The order that they show up on this line is the order that they are loaded at system startup (see Example 4-6).

Example 4-6 Telling SUSE Linux Enterprise Server to include a kernel module in the initramfs

```
x3650lab9:~ # cat /etc/sysconfig/kernel
...
# This variable contains the list of modules to be added to the initial
# ramdisk by calling the script "mkinitrd"
# (like drivers for scsi-controllers, for lvm or reiserfs)
#
INITRD_MODULES="thermal aacraid ata_piix ... processor fan jbd ext3 edd qla2xxx"
...
```

After you add the HBA driver module name to the configuration file, rebuild the initramfs with the **`mkinitrd`** command. This command creates and installs the image file with standard settings and to standard locations as shown in Example 4-7.

Example 4-7 Creating 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, use parameters for **`mkinitrd`**. For more information, see the **`mkinitrd`** man page on your Linux system.

Red Hat Enterprise Linux 5

Kernel modules that must be included in the initramfs are listed in the `/etc/modprobe.conf` file. The order that they show up in the file is the order that they are loaded at system startup, as shown in Example 4-8.

Example 4-8 Telling Red Hat Enterprise Linux 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 you add the HBA driver module to the configuration file, rebuild the initramfs with the **`mkinitrd`** command. The Red Hat version of **`mkinitrd`** requires the following information as parameters (see Example 4-9):

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

Example 4-9 Creating 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 exists, use the **`-f`** option to force **`mkinitrd`** to overwrite the existing one. The command shows more detailed output with the **`-v`** option.

You can discover the kernel version that is running on the system with the **`uname`** command (see Example 4-10).

Example 4-10 Determining the kernel version

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

Red Hat Enterprise Linux 6 and 7

The **`dracut`** utility is for Red Hat Enterprise Linux 6 and 7 and is important to the boot process. In previous versions of Red Hat Enterprise Linux, the initial RAM disk image preinstalled the block device modules, such as for SCSI or RAID. The root file system, on which those modules are normally located, can then be accessed and mounted.

With Red Hat Enterprise Linux 6 and 7 systems, the **`dracut`** utility is always called by the installation scripts to create an `initramfs`. This process occurs whenever a new kernel is installed by using the Yum, PackageKit, or Red Hat Package Manager (RPM).

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

Verify that an initramfs corresponding to your current kernel version exists and is specified correctly in the bootloader (GRUB by default for Red Hat Enterprise Linux 6 and GRUB2 for Red Hat Enterprise 7) configuration file by using the following procedure:

1. As root, list the contents in the `/boot/` directory.
2. Find the kernel (`vmlinuz-<kernel_version>`) and `initramfs-<kernel_version>` with the most recent version number, as shown in Figure 4-1.

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

Figure 4-1 Red Hat Enterprise Linux 6 display of matching initramfs and kernel

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

For more information about options and usage, see `man dracut` and `man dracut.conf`.

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

4.2.3 Determining the WWPN of the installed HBAs

To create a host port on the storage system that can map volumes to an HBA, you need the WWPN of the HBA. The WWPN is shown in *sysfs*, a Linux pseudo file system that reflects the installed hardware and its configuration. Example 4-11 shows how to discover which SCSI host instances are assigned to the installed FC HBAs. You can then determine their WWPNs.

Example 4-11 Finding 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
```

Map volumes to a Linux host as described for XIV in 1.4, “Logical configuration for host connectivity” on page 31 or for FlashSystem A9000 and A9000R, described in 2.6, “Logical configuration for host connectivity” on page 86.

Tip: For Intel host systems, the Host Attachment Kit can create the host object and host port objects for you automatically from the Linux operating system. For more information, see 4.2.4, “Attaching volumes to an Intel x86 host using the Host Attachment Kit” on page 154.

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

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

Installing the Host Attachment Kit

For multipathing with Linux, FlashSystem A9000, A9000R, and XIV, IBM provides a single, unified *Host Attachment Kit*. This section explains how to install the Host Attachment Kit on a Linux server.

Consideration: Although manually configuring Linux on Intel x86 servers for the attachment is possible, IBM strongly recommends using the Host Attachment Kit.The kit not only automates the host definition and LUN mappings but also provides valuable data collection and troubleshooting tools.

At the time of writing, Host Attachment Kit version 2.8 is the current version for Linux.

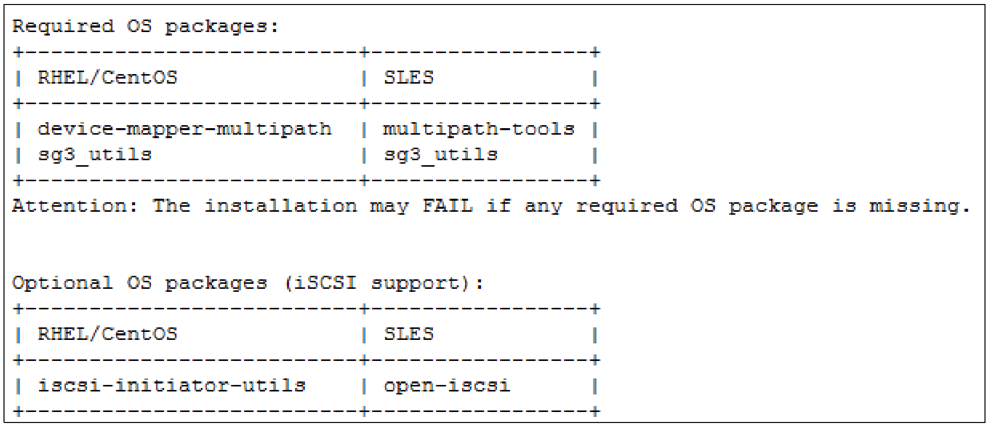
Some additional troubleshooting checklists and tips are available in 4.4, “Troubleshooting and monitoring” on page 181.

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

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

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

The required packages are shown in Figure 4-2.



```
Required OS packages:
+--------------------------+-----------------+
| RHEL/CentOS              | SLES            |
+--------------------------+-----------------+
| device-mapper-multipath  | multipath-tools |
| sg3_utils                | sg3_utils       |
+--------------------------+-----------------+
Attention: The installation may FAIL if any required OS package is missing.


Optional OS packages (iSCSI support):
+--------------------------+-----------------+
| RHEL/CentOS              | SLES            |
+--------------------------+-----------------+
| iscsi-initiator-utils    | open-iscsi      |
+--------------------------+-----------------+
```

Figure 4-2   Required packages

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

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

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

> **Consideration:** Illustrations and examples are based on version 2.6, a recent version of the Linux Host Attachment Kit.

Example 4-12   Installing the Host Attachment Kit package

```
# tar -xvzf IBM_Storage_Host_Attachment_Kit_2.8.0-b2850_Linux_x86-64.tar.gz
HAK_2.8.0/
HAK_2.8.0/scripts/
HAK_2.8.0/scripts/platform.linux.sh
HAK_2.8.0/scripts/motd.linux.txt
HAK_2.8.0/scripts/pkginst.linux.sh
HAK_2.8.0/scripts/is_installed.linux.sh
HAK_2.8.0/scripts/xpyv_upgrade.py
HAK_2.8.0/scripts/postinst.linux.sh
HAK_2.8.0/packages/
HAK_2.8.0/packages/pkg.list
HAK_2.8.0/packages/host_attach-2.8.0-b2850.x86_64.rpm
HAK_2.8.0/packages/tools.list
HAK_2.8.0/packages/platform.list
HAK_2.8.0/install.sh
# cd HAK_2.8.0/
# ./install.sh
Welcome to the IBM Storage Host Attachment Kit installer.

Required OS packages:
+--------------------------+-----------------+
| RHEL/CentOS              | SLES            |
+--------------------------+-----------------+
| device-mapper-multipath  | multipath-tools |
| sg3_utils                | sg3_utils       |
+--------------------------+-----------------+
Attention: The installation may FAIL if any required OS package is missing.


Optional OS packages (iSCSI support):
+--------------------------+-----------------+
| RHEL/CentOS              | SLES            |
+--------------------------+-----------------+
| iscsi-initiator-utils    | open-iscsi      |
+--------------------------+-----------------+


Would you like to proceed and install the IBM Storage Host Attachment Kit? [Y/n]:
y
Please wait while the installer validates your existing configuration...
---------------------------------------------------------------
Please wait as the IBM Storage Host Attachment Kit is being installed...
---------------------------------------------------------------
Installation successful.
Please refer to the user guide for information about how to configure this host.

---------------------------------------------------------------
The IBM Storage Host Attachment Kit includes the following utilities:
xiv_attach: Interactive wizard that configures the host and verifies its
            configuration for connectivity with the IBM storage System.
xiv_devlist: Lists all storage volumes that are mapped to the host, with general
            info about non-storage volumes.
xiv_syslist: Lists all IBM storage systems that are detected by the host.
xiv_diag: Performs complete diagnostics of the host and its connectivity with
            the IBM storage System, and saves the information to a file.
xiv_fc_admin: Allows you to perform different administrative operations for
            FC-connected hosts and IBM storage systems.
xiv_iscsi_admin: Allows you to perform different administrative operations for
            iSCSI-connected hosts and IBM storage systems.
xiv_host_profiler: Collects host configuration information and performs a
            comprehensive analysis of the collected information.
---------------------------------------------------------------
```

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

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

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

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

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

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

Configuring the host for Fibre Channel using the Host Attachment Kit

Use the **xiv_attach** command to configure the Linux host. You can also create the storage system host object and host ports on the system itself. To do so, you must have a user ID and password for each storage system administrator account. Example 4-13 shows how **xiv_attach** works for Fibre Channel attachment. Your output can differ depending on your configuration.

Example 4-13 Fibre Channel host attachment configuration using the xiv_attach command

```
[root@localhost HAK_2.8.0]# xiv_attach
-------------------------------------------------------------------------------
Welcome to the IBM Storage Host Attachment wizard, version 2.8.0.
This wizard will help you attach this host to one or more IBM storage systems

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

-------------------------------------------------------------------------------
Please specify the connectivity type: [f]c / [i]scsi : f
-------------------------------------------------------------------------------
Please wait while the wizard validates your existing configuration...
Verifying multipath - multipath.conf                                         OK
Verifying multipath service(s)                                           NOT OK
-------------------------------------------------------------------------------
The wizard needs to configure this host for the IBM storage system.
Do you want to proceed? [default: yes ]:
Please wait while the host is being configured...
-------------------------------------------------------------------------------
Configuring multipath - multipath.conf                                       OK
Configuring multipath service(s)                                             OK
The host is now configured for the IBM storage system
no crontab for root

-------------------------------------------------------------------------------
Creating a host profile enables focused and better support for your host.
Would you like to add a scheduled task for host profile creation? [default: yes ]:
A scheduled task for xiv_host_profiler has been added.
-------------------------------------------------------------------------------
Please define zoning for this host and add its World Wide Port Names (WWPNs) to the
IBM storage system:
21:00:00:1b:32:8b:da:1f: [QLOGIC]: N/A
21:01:00:1b:32:ab:da:1f: [QLOGIC]: N/A
21:00:00:1b:32:8b:0c:1d: [QLOGIC]: N/A
21:01:00:1b:32:ab:0c:1d: [QLOGIC]: N/A
Press [ENTER] to proceed.

Would you like to rescan for new storage devices? [default: yes ]:
Please wait while rescanning for IBM storage devices...
-------------------------------------------------------------------------------
This host is connected to the following IBM storage arrays:
Storage Type        Serial   System Version  Host Defined  Ports Defined     Protocol
Host Name(s)
```

```
XIV                 1340010  11.6.2          No          No ports defined  FC
N/A
XIV                 1340008  11.6.2.a        No          No ports defined  FC
N/A
FlashSystem A9000   1322131  12.1.0.b        No          No ports defined  FC
N/A
FlashSystem A9000R  1320902  12.1.0.b        No          No ports defined  FC
N/A
This host is not defined on some FC-attached IBM storage systems.
Do you want to define this host on these IBM storage systems now? [default: yes ]:
Please enter a name for this host [default: localhost.localdomain ]: x3650-m2-23
Please enter a username for system 1340010 [default: admin ]:
Please enter the password of user admin for system 1340010:


Please enter a username for system 1322131 [default: admin ]:
Please enter the password of user admin for system 1322131:


Please enter a username for system 1340008 [default: admin ]:
Please enter the password of user admin for system 1340008:


Please enter a username for system 1320902 [default: admin ]:
Please enter the password of user admin for system 1320902:



Press [ENTER] to proceed.

-------------------------------------------------------------------------------
The IBM Storage Host Attachment wizard has successfully configured this host.

Press [ENTER] to exit.
```

Configuring the host for iSCSI using the Host Attachment Kit

Use the **xiv_attach** command to configure the host for iSCSI attachment of the volumes. First, make sure that the iSCSI service is running with the command **service iscsi start**.

Example 4-14 shows output when you run **xiv_attach**. Again, your output can differ depending on your configuration.

Example 4-14 iSCSI host attachment configuration using the xiv_attach command

```
# xiv_attach
-------------------------------------------------------------------------------
Welcome to the IBM Storage Host Attachment wizard, version 2.8.0.
This wizard will help you attach this host to one or more IBM storage systems

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

-------------------------------------------------------------------------------
Please specify the connectivity type: [f]c / [i]scsi : i
Would you like to enable host-side acceleration? (A scheduled task will be created)
[default: no ]:
-------------------------------------------------------------------------------
```

```
Please wait while the wizard validates your existing configuration...
Verifying multipath - multipath.conf                                         OK
Verifying multipath service(s)                                               OK
Verifying iSCSI initiator name - initiatorname.iscsi                         OK
Verifying iSCSI daemon - iscsid.conf                                         OK
Verifying iSCSI service                                                      OK
This host is already configured for the IBM storage system.
-------------------------------------------------------------------------------
Discovering new iSCSI targets ...
-------------------------------------------------------------------------------
Would you like to discover a new iSCSI target? [default: yes ]:
Please enter an IBM storage system iSCSI discovery address (iSCSI interface):
9.155.116.205
Is this host defined to use CHAP authentication with the IBM storage system? [default:
no ]:
Would you like to discover a new iSCSI target? [default: yes ]:
Please enter an IBM storage system iSCSI discovery address (iSCSI interface):
9.155.120.20
Is this host defined to use CHAP authentication with the IBM storage system? [default:
no ]:
Would you like to discover a new iSCSI target? [default: yes ]: no
Would you like to rescan for new storage devices? [default: yes ]:
-------------------------------------------------------------------------------
This host is connected to the following IBM storage arrays:
Storage Type        Serial   System Version  Host Defined  Ports Defined     Protocol
Host Name(s)
XIV                 1340010  11.6.2          No            No ports defined  FC
N/A
XIV                 1340008  11.6.2.a        No            No ports defined  FC
N/A
FlashSystem A9000   1322131  12.1.0.b        No            No ports defined  FC,iSCSI
N/A
FlashSystem A9000R  1320902  12.1.0.b        No            No ports defined  FC,iSCSI
N/A
This host is not defined on some iSCSI-attached IBM storage systems.
Do you want to define this host on these IBM storage systems now? [default: yes ]:
Please enter a name for this host [default: localhost.localdomain ]: x3650-m2-23-iSCSI

Please enter a username for system 1322131 [default: admin ]:
Please enter the password of user admin for system 1322131:

Please enter a username for system 1320902 [default: admin ]:
Please enter the password of user admin for system 1320902:

Press [ENTER] to proceed.

-------------------------------------------------------------------------------
The IBM Storage Host Attachment wizard has successfully configured this host.
```

4.2.5 Checking attached volumes

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

Example 4-15 shows use of the Host Attachment Kit to verify the volumes for an iSCSI attached volume. The **xiv_devlist** command lists all devices that are attached to a host.

Example 4-15 Verifying mapped LUNs using the Host Attachment Kit tool with iSCSI

```
# xiv_devlist
IBM storage devices
----------------------------------------------------------------------------------
------------------------------------
Device              Size (GB)  Paths  Vol Name             Vol ID  Storage ID
Storage Type       Hyper-Scale Mobility
----------------------------------------------------------------------------------
------------------------------------
/dev/mapper/mpatha  50.1       3/3    x3650-m2-23-iSCSI-1  14405   1322131
FlashSystem A9000  Idle
----------------------------------------------------------------------------------
------------------------------------


Non-IBM storage devices
...
```

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

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

Example 4-16 shows that Linux recognized three devices. By looking at the SCSI addresses in the first column, you can determine that there actually is one volume. This volume is connected through three paths. Linux creates a SCSI disk device for each of the paths.

Example 4-16 Listing attached SCSI devices

```
# lsscsi |grep /dev/sd|grep 2810XIV
[19:0:0:1]   disk    IBM      2810XIV          0000  /dev/sde
[20:0:0:1]   disk    IBM      2810XIV          0000  /dev/sdf
[21:0:0:1]   disk    IBM      2810XIV          0000  /dev/sdg
```

Linux SCSI addressing explained

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

- HBA ID: Each HBA in the system gets a host adapter instance when it is initiated. The instance is assigned regardless of whether it is parallel SCSI, Fibre Channel, or even a SCSI emulator.
- Channel ID: This field is always zero. It was formerly used as an identifier for the channel in multiplexed parallel SCSI HBAs.
- Target ID: For parallel SCSI, this is the real target ID that you set by using a jumper on the disk drive. For Fibre Channel, it represents a remote port that is connected to the HBA. This ID distinguishes between multiple paths, and between multiple storage systems.
- LUN: Logical unit numbers (LUNs) 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 4-3 shows how the SCSI addresses are generated.

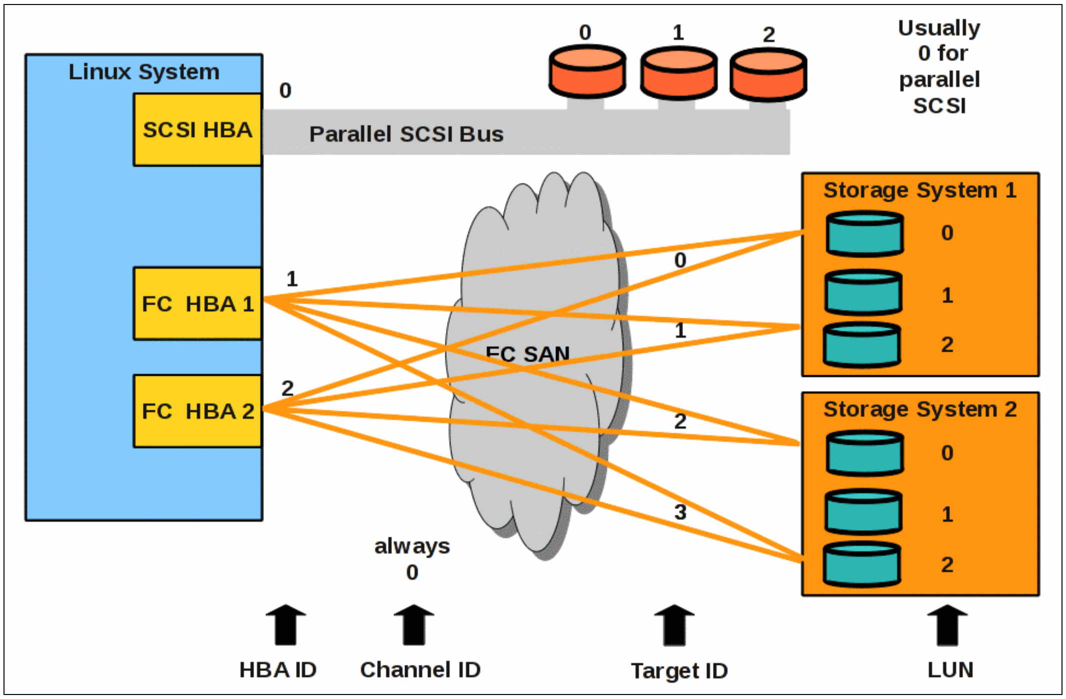


Figure 4-3 Composition of Linux internal SCSI addresses

Identifying a particular device

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

The nodes for the volumes in `/dev/disk/by-id` show a unique identifier. This identifier is composed of parts of the following numbers (see Example 4-17 on page 162):

- Worldwide node name (WWNN) of the system
- Volume serial number in hexadecimal notation

Example 4-17 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
...
```

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

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

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

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

The `/dev/disk/by-path` file contains nodes for all paths to all volumes. Here you can see the physical connection to the volumes. This connection starts with the PCI identifier of the HBAs through the remote port, represented by the WWPN, to the LUN of the volumes (see Example 4-18).

Example 4-18 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
```

Adding volumes to a Linux on z Systems system

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

Remember: Because of hardware restraints, SUSE Linux Enterprise Server 10 SP3 and IBM XIV are used for the examples. The procedures, commands, and configuration files of other distributions can differ.

In this example, Linux on z Systems has two FC HBAs assigned through z/VM. Determine the device numbers of these adapters (see Example 4-19).

Example 4-19 FCP HBA device numbers in z/VM

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

The Linux on z Systems tool to list the FC HBAs is **`lszfcp`**. It shows the enabled adapters only. Adapters that are not listed correctly can be enabled by using the **`chccwdev`** command (see Example 4-20).

Example 4-20 Listing and enabling Linux on z Systems 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 SUSE Linux Enterprise Server 10, the volume configuration files are in the `/etc/sysconfig/hardware` directory. There must be one for each HBA. Example 4-21 shows their naming scheme.

Example 4-21 HBA configuration files naming scheme example

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

Important: The configuration file that is described here is used with SUSE Linux Enterprise Server 9 and SUSE Linux Enterprise Server 10. SUSE Linux Enterprise Server 11 uses `udev` rules. These rules are automatically created by YAST when you use it to discover and configure SAN-attached volumes. They are complicated and not well documented yet, so use YAST.

The configuration files contain a remote (XIV) port and LUN pair for each path to each volume. Example 4-22 shows two XIV volumes that are defined to the HBA 0.0.0501, going through two XIV host ports.

Example 4-22 HBA configuration file example

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

The `ZFCP_LUNS="..."` statement in the file defines all the remote port to volume relations (paths) that the `zfcp` 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 id of the XIV volume. It is padded with zeros until it reaches a length of 8 bytes.

Red Hat Enterprise Linux uses the `/etc/zfcp.conf` file to configure SAN-attached volumes. It contains the same information in a different format (see Example 4-23). The three final lines in the example are comments that explain the format. They do not have to be present in the file.

Example 4-23 Format of the /etc/zfcp.conf file for Red Hat Enterprise Linux

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

4.2.6 Setting up Device Mapper Multipathing

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

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

DM-MP is able to manage path failover and failback, and load balancing for various storage architectures. Figure 4-4 shows how DM-MP is integrated into the Linux storage stack.

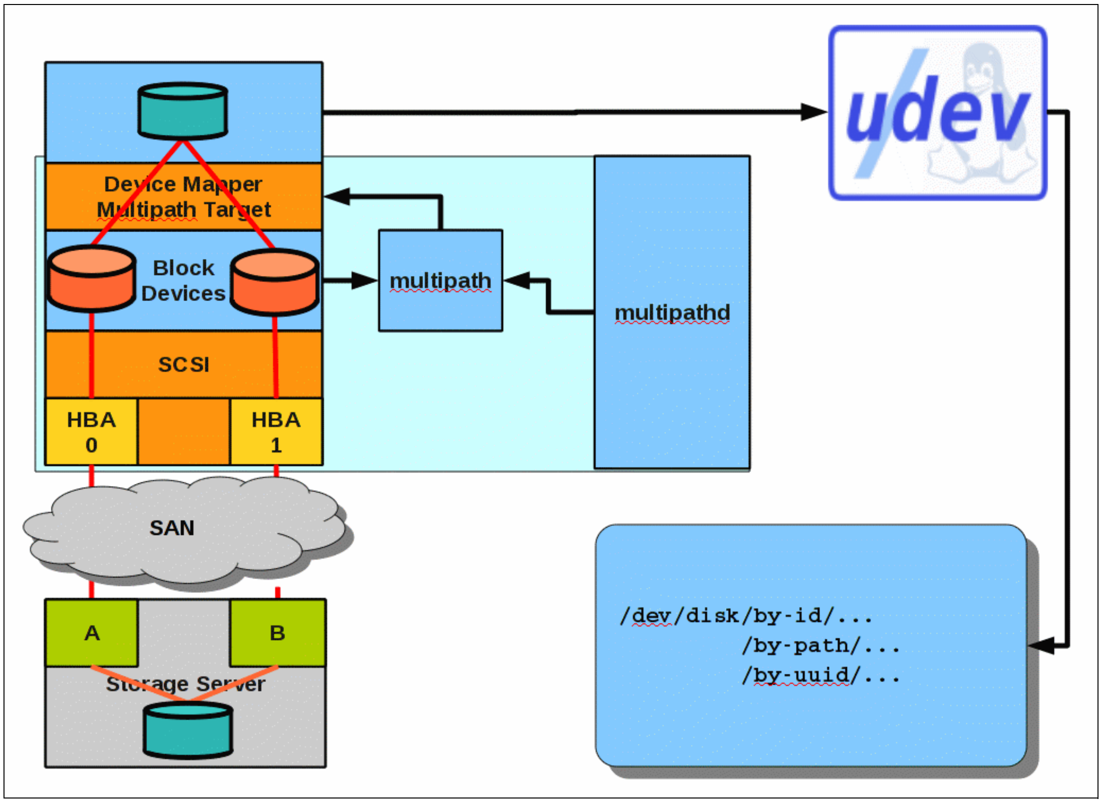


Figure 4-4 Device Mapper Multipathing in the Linux storage stack

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

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

Configuring DM-MP

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

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

The /etc/multipath.conf file is not described in detail here. For more information, see the publications in 4.1.2, “Reference material” on page 142. For more information about the settings for attachment, see 4.2.7, “Special considerations for attachment” on page 171.

One option, however, that shows up several times in the next sections needs some explanation. You can tell DM-MP to generate “user-friendly” device names by specifying this option in the /etc/multipath.conf file, as shown in Example 4-24.

Example 4-24 Specifying user-friendly names in /etc/multipath.conf

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

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

Tip: The user-friendly names differ for SUSE Linux Enterprise Server and Red Hat Enterprise Linux. They are explained in their respective sections.

Enabling multipathing for SUSE Linux Enterprise Server 11

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

You can start Device Mapper Multipathing by running two start scripts (see Example 4-25).

Example 4-25 Starting DM-MP in SUSE Linux Enterprise Server 11

```
x3650lab9:~ # /etc/init.d/boot.multipath start
Creating multipath target                                           done
x3650lab9:~ # /etc/init.d/multipathd start
Starting multipathd                                                 done
```

To have DM-MP start automatically at each system start, add these start scripts to the SUSE Linux Enterprise Server 11 system start process (see Example 4-26).

Example 4-26 Configuring automatic start of DM-MP in SUSE Linux Enterprise Server 11

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

Enabling multipathing for Red Hat Enterprise Linux 5

Red Hat Enterprise Linux includes a default `/etc/multipath.conf` file. It contains a section that blacklists all device types. You must remove or comment out these lines to make DM-MP work. A # sign in front of them will mark them as comments so they are ignored the next time DM-MP scans for devices (see Example 4-27).

Example 4-27 Disabling the blacklisting of all devices in /etc/multiparh.conf

```
...
# Blacklist all devices by default. Remove this to enable multipathing
# on the default devices.
#blacklist {
#devnode "*"
#}
...
```

Start DM-MP (Example 4-28).

Example 4-28 Starting DM-MP in Red Hat Enterprise Linux 5

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

To have DM-MP start automatically at each system start, add the start script (as shown in Example 4-29) to the Red Hat Enterprise Linux 5 system start process.

Example 4-29 Configuring automatic start of DM-MP in Red Hat Enterprise Linux 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
```

Checking and changing the DM-MP configuration

The multipath background daemon provides a user interface to print and modify the DM-MP configuration. It can be started as an interactive session with the **`multipathd -k`** command. Within this session, various options are available. Use the **`help`** command to get a list. Some of the more important options are shown in the following examples. For more information, see 4.3, “Nondisruptive SCSI reconfiguration” on page 173.

The **`show topology`** command (see Example 4-30) prints a detailed view of the current DM-MP configuration, including the state of all available paths.

Example 4-30 Showing 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][hwhandler=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][hwhandler=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 4-30 on page 167 shows that the paths of the multipath devices are in separate path groups. Thus, there is no load balancing between the paths. DM-MP must be configured with a special `multipath.conf` file to enable load balancing. For more information, see 4.2.7, “Special considerations for attachment” on page 171 and “Multipathing” on page 145. The Host Attachment Kit does this configuration automatically if you use it for host configuration.

You can use the **`reconfigure`** command (see Example 4-31) to tell DM-MP to update the topology after it scans the paths and configuration files. Use it to add new multipath devices after you add new volumes. For more information, see 4.3.1, “Adding and removing volumes dynamically” on page 173.

Example 4-31   Reconfigure DM-MP

```
multipathd> reconfigure
ok
```

Important: The **`multipathd -k`** command prompt of SUSE Linux Enterprise Server 11 SP1 supports the **`quit`** and **`exit`** commands to terminate. The command prompt of Red Hat Enterprise Linux 5U5 is a little older and must still be terminated by using the Ctrl+D key combination.

Although the **`multipath -l`** and **`multipath -ll`** commands can be used to print the current DM-MP configuration, use the **`multipathd -k`** interface. The **`multipath`** tool is being removed from DM-MP, and all further development and improvements will go into **`multipathd`**.

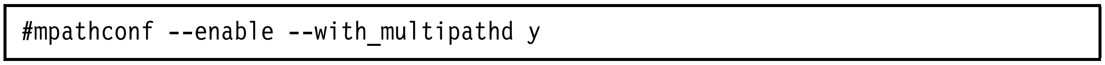
Tip: You can also issue commands in a “one-shot-mode” by enclosing them in quotation marks and typing them directly, without space, after the **`multipath -k`**. Here is an example:

```
multipathd -k“show paths”
```

Enabling multipathing for Red Hat Enterprise Linux 6 and 7

Unlike Red Hat Enterprise Linux 5, Red Hat Enterprise Linux 6 and 7 includes a new utility, **`mpathconf`**, that creates and modifies the `/etc/multipath.conf` file.

This command (see Figure 4-5) enables the multipath configuration file.



```
#mpathconf --enable --with_multipathd y
```

Figure 4-5   The mpathconf command

Be sure to start and enable the **multipathd** daemon, as shown in Figure 4-6.

```
#service multipathd start
#chkconfig multipathd on
```

Figure 4-6 Commands to ensure that multipathd is started and enabled at boot

Because the value of `user_friendly_name` in Red Hat Enterprise Linux 6 and 7 is set to `yes` in the default configuration file, the multipath devices are created as follows (where *n* is an alphabetic letter that designates the path):

`/dev/mapper/mpath`*n*

Red Hat has released numerous enhancements to the device-mapper-multipath drivers that were included starting with Red Hat Enterprise Linux 6. Make sure to install and update to the latest version, and download any bug fixes.

Accessing DM-MP devices in SUSE Linux Enterprise Server 11

The device nodes that you use to access DM-MP devices are created by `udev` in the `/dev/mapper` directory. If you do not change any settings, SUSE Linux Enterprise Server 11 uses the unique identifier of a volume as device name, as shown in Example 4-32.

Example 4-32 Multipath devices in SUSE Linux Enterprise Server 11 in /dev/mapper

```
x3650lab9:~ # ls -l /dev/mapper | cut -c 48-

20017380000cb051f
20017380000cb0520
20017380000cb2d57
20017380000cb3af9
...
```

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

SUSE Linux Enterprise Server 11 creates an extra set of device nodes for multipath devices. It overlays the former single path device nodes in `/dev/disk/by-id`. Any device mappings you did for one of these nodes before starting DM-MP are not affected. It uses the DM-MP device instead of the SCSI disk device, as shown in Example 4-33.

Example 4-33 SUSE Linux Enterprise Server 11 DM-MP device nodes in /dev/disk/by-id

```
x3650lab9:~ # 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_name` option in the `/etc/multipath.conf` file, SUSE Linux Enterprise Server 11 creates DM-MP devices with the names `mpatha`, `mpathb`, and so on, in `/dev/mapper`. The DM-MP device nodes in `/dev/disk/by-id` are not changed. They also have the unique IDs of the volumes in their names.

Accessing DM-MP devices in Red Hat Enterprise Linux

Red Hat Enterprise Linux sets the `user_friendly_name` option in its default `/etc/multipath.conf` file. The devices that it creates in `/dev/mapper` look as shown in Example 4-34.

Example 4-34 Multipath devices in Red Hat Enterprise Linux 5 in /dev/mapper

```
[root@x3650lab9 ~]# ls -l /dev/mapper/ | cut -c 45-

mpath1
mpath2
mpath3
mpath4
```

Example 4-35 shows the output from an Red Hat Enterprise Linux 6 system.

Example 4-35 Red Hat Enterprise Linux 6 device nodes in /dev/mpath

```
[root@x3650lab9 ~]# ls -l /dev/mpath/ | cut -c 39-

20017380000cb051f -> ../../dm-5
20017380000cb0520 -> ../../dm-4
20017380000cb2d57 -> ../../dm-0
20017380000cb3af9 -> ../../dm-1
```

A second set of device nodes contains the unique IDs of the volumes in their name, regardless of whether user-friendly names are specified.

In Red Hat Enterprise Linux 5, you find them in the `/dev/mpath` directory (see Example 4-36).

Example 4-36 Red Hat Enterprise Linux 6 Multipath devices

```
mpatha -> ../dm-2
 mpathap1 -> ../dm-3
 mpathap2 -> ../dm-4
 mpathap3 -> ../dm-5
 mpathc -> ../dm-6
 mpathd -> ../dm-7
```

In Red Hat Enterprise Linux 6 and 7, you find them in `/dev/mapper` (see Example 4-37).

Example 4-37 Red Hat Enterprise Linux 6 and 7 device nodes in /dev/mapper

```
# ls -l /dev/mapper/ | cut -c 43-

 mpatha -> ../dm-2
 mpathap1 -> ../dm-3
 mpathap2 -> ../dm-4
 mpathap3 -> ../dm-5
 mpathc -> ../dm-6
 mpathd ->
../dm-7
```

Using multipath devices

You can use the device nodes that are created for multipath devices just like any other block device:

- Create a file system and mount it.
- Use them with the *Logical Volume Manager (LVM)*.
- Build software RAID devices.

You can also partition a DM-MP device by using the **fdisk** command or any other partitioning tool. To make new partitions on DM-MP devices available, use the **partprobe** command. It triggers `udev` to set up new block device nodes for the partitions, as shown in Example 4-38.

Example 4-38 Using the partprobe command to register newly created partitions

```
x3650lab9:~ # fdisk /dev/mapper/20017380000cb051f
...
<all steps to create a partition and write the new partition table>
...
x3650lab9:~ # ls -l /dev/mapper/ | cut -c 48-

20017380000cb051f
20017380000cb0520
20017380000cb2d57
20017380000cb3af9
...
x3650lab9:~ # partprobe
x3650lab9:~ # ls -l /dev/mapper/ | cut -c 48-

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

Example 4-38 was created with SUSE Linux Enterprise Server 11. The method also works for Red Hat Enterprise Linux 5, but the partition names might differ.

> **Remember:** This limitation, that LVM by default does not work with DM-MP devices, does not exist in recent Linux versions.

4.2.7 Special considerations for attachment

This section addresses special considerations that apply to FlashSystem A9000, A9000R, and XIV.

Configuring multipathing

The Host Attachment Kit updates the `/etc/multipath.conf` file during installation to optimize use for FlashSystem A9000, A9000R, or XIV. If you must manually update the file, the contents of this file as it is created by the Host Attachment Kit are described next. The settings that are relevant for the storage systems are shown in Example 4-39 on page 172. Note that settings shown in the example are specific to SLES v15. For other versions, use default values.

Example 4-39   DM-MP settings for XIV for SLES v15

```
x3650lab9:~ # cat /etc/multipath.conf
devices {
device {
             vendor "IBM"
             product "2810XIV"
             path_grouping_policy group_by_prio
             path_selector "round-robin 0"
             path_checker tur
             features "0"
             no_path_retry 2
             hardware_handler "1 alua"
             prio alua
             failback 15
             rr_weight uniform
             rr_min_io 15
             rr_min_io_rq 1
             fast_io_fail_tmo 3
             dev_loss_tmo 5
             retain_attached_hw_handler no
             detect_prio no
     }
```

The `user_friendly_name` parameter is addressed in 4.2.6, “Setting up Device Mapper Multipathing” on page 164. You can add it to file or leave it out. The values for `failback`, `no_path_retry`, `path_checker`, and `polling_interval` control the behavior of DM-MP in case of path failures. Normally, do not change them. If your situation requires a modification of these parameters, see the publications that are listed in 4.1.2, “Reference material” on page 142. The `rr_min_io` setting specifies the number of I/O requests that are sent to one path before switching to the next one. The value of 15 gives good load balancing results in most cases. However, you can adjust it as necessary.

Important: Upgrading or reinstalling the Host Attachment Kit does not change the `multipath.conf` file. Ensure that your settings match the values that were shown previously.

z Systems specific multipathing settings

Testing of Linux on z Systems with multipathing has shown that for best results, set the parameters as follows:

- `dev_loss_tmo` parameter to 90 seconds
- `fast_io_fail_tmo` parameter to 5 seconds

Modify the `/etc/multipath.conf` file and add the settings that are shown in Example 4-40.

Example 4-40   z Systems specific multipathing settings

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

Make the changes effective by using the **reconfigure** command in the interactive **multipathd -k** prompt.

Disabling QLogic failover

The QLogic HBA kernel modules have limited built-in multipathing capabilities. Because multipathing is managed by DM-MP, make sure that the QLogic failover support is disabled. To check, use the **modinfo qla2xxx** command (see Example 4-41).

Example 4-41 Checking for enabled QLogic failover

```
x3650lab9:~ # modinfo qla2xxx | 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 shown in Example 4-42.

Example 4-42 Disabling QLogic failover

```
x3650lab9:~ # cat /etc/modprobe.conf
...
options qla2xxx ql2xfailover=0
...
```

After you modify this file, run the **depmod -a** command to refresh the kernel driver dependencies. Then, reload the `qla2xxx` module to make the change is effective. If you include the `qla2xxx` module in the `InitRAMFS`, you must create one.

4.3 Nondisruptive SCSI reconfiguration

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

4.3.1 Adding and removing volumes dynamically

Unloading and reloading the Fibre Channel HBA used to be the typical way to discover newly attached 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. You use a command interface that is provided by **sysfs** (see Example 4-43).

Example 4-43 Scanning for new Fibre Channel attached devices

```
x3650lab9:~ # ls /sys/class/fc_host/
host0 host1
x3650lab9:~ # echo "- - -" > /sys/class/scsi_host/host0/scan
x3650lab9:~ # echo "- - -" > /sys/class/scsi_host/host1/scan
```

First, discover which SCSI instances your FC HBAs have, then issue a **scan** command to their **sysfs** representatives. The triple dashes `("- - -")` represent the Channel-Target-LUN combination to scan. A dash causes a scan through all possible values. A number would limit the scan to the provided value. With more recent Linux versions, you can use **rescan-scsi-bus.sh** script to scan.

Tip: If the Host Attachment Kit is installed, you can use the **xiv_fc_admin -R** or **xiv_iscsi_admin -R** command to scan for new volumes.

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

Tip: For some older Linux versions, you must force the FC HBA to run a port login to recognize the newly added devices. Use the following command, which must be issued to all FC HBAs:

```
echo 1 > /sys/class/fc_host/host<ID>/issue_lip
```

If you want to remove a disk device from Linux, follow this sequence to avoid system hangs because of incomplete I/O requests:

1. Stop all applications that use the device and make sure that all updates or writes are completed.
2. Unmount the file systems that use the device.
3. If the device is part of an LVM configuration, remove it from all logical volumes and volume groups.
4. Remove all paths to the device from the system (see Example 4-44).

Example 4-44 Removing both paths to a disk device

```
x3650lab9:~ # echo 1 > /sys/class/scsi_disk/0\:0\:0\:3/device/delete
x3650lab9:~ # echo 1 > /sys/class/scsi_disk/1\:0\:0\:3/device/delete
```

The device paths (or disk devices) are represented by their Linux SCSI address. For more information, see “Linux SCSI addressing explained” on page 161. Run the **multipathd -k”show topology”** command after you remove each path to monitor the progress.

DM-MP and udev recognize the removal automatically, and delete all corresponding disk and multipath device nodes. You must remove all paths that exist to the device before you detach the device on the storage system level.

You can use **watch** to run a command periodically for monitoring purposes. This example allows you to monitor the multipath topology with a period of one second:

```
watch -n 1 'multipathd -k"show top"'
```

4.3.2 Adding and removing volumes in Linux on z Systems

The mechanisms to scan and attach new volumes do not work the same in Linux on z Systems. Commands are available that discover and show the devices that are connected to the FC HBAs. However, they do not do the logical attachment to the operating system automatically. In SUSE Linux Enterprise Server 10 SP3, use the **zfcp_san_disc** command for discovery.

Example 4-45 on page 175 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 4-45 Listing LUNs connected through a specific remote port

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

> **Remember:** In more recent distributions, **`zfcp_san_disc`** is no longer available because remote ports are automatically discovered. The attached volumes can be listed by using the **`lsluns`** script.

After you discover the connected volumes, logically attach them using **`sysfs`** interfaces. Remote ports or device paths are represented in the **`sysfs`**. There is a directory for each local-remote port combination (path). It contains a representative of each attached volume and various meta files as interfaces for action. Example 4-46 shows such a **`sysfs`** structure for a specific port.

Example 4-46 The sysfs structure for a remote port

```
lnxvm01:~ # ls -l /sys/bus/ccw/devices/0.0.0501/0x5001738000cb0191/
total 0
drwxr-xr-x 2 root root    0 2010-12-03 13:26 0x0001000000000000
...
--w------- 1 root root 4096 2010-12-03 13:26 unit_add
--w------- 1 root root 4096 2010-12-03 13:26 unit_remove
```

Add LUN `0x0003000000000000` to both available paths by using the `unit_add` metafile, as shown in Example 4-47.

Example 4-47 Adding a volume to all existing remote ports

```
lnxvm01:~ # echo 0x0003000000000000 > /sys/.../0.0.0501/0x5001738000cb0191/unit_add
lnxvm01:~ # echo 0x0003000000000000 > /sys/.../0.0.0501/0x5001738000cb0160/unit_add
```

> **Important:** You must run discovery by using **`zfcp_san_disc`** whenever new devices, remote ports, or volumes are attached. Otherwise, the system does not recognize them even if you do the logical configuration.

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

If you want to remove a volume from Linux on z Systems, complete the same steps as for the other platforms. These procedures avoid system hangs because of incomplete I/O requests:

1. Stop all applications that use the device, and make sure that all updates or writes are completed.
2. Unmount the file systems that use the device.
3. If the device is part of an LVM configuration, remove it from all logical volumes and volume groups.
4. Remove all paths to the device from the system.

Volumes can then be removed logically by using a method similar to attachment. Write the LUN of the volume into the `unit_remove` meta file for each remote port in **`sysfs`**.

Important: If you need the newly added devices to be persistent, use the methods in “Adding volumes to a Linux on z Systems system” on page 162. Create the configuration files to be used at the next system start.

4.3.3 Adding new storage system host ports to Linux on z Systems

If you connect new ports or a new system to the Linux on z Systems system, you must logically attach the new remote ports. Discover the storage system ports that are connected to your HBAs, as shown in Example 4-48.

Example 4-48 Showing connected remote ports

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

Attach the new ports logically to the HBAs. As Example 4-49 shows, a remote port is already attached to HBA `0.0.0501`. Add the second connected port to the HBA.

Example 4-49 Listing attached remote ports, attaching remote ports

```
lnxvm01:~ # ls /sys/bus/ccw/devices/0.0.0501/ | grep 0x
0x5001738000cb0191

lnxvm01:~ # echo 0x5001738000cb0170 > /sys/bus/ccw/devices/0.0.0501/port_add

lnxvm01:~ # ls /sys/bus/ccw/devices/0.0.0501/ | grep 0x
0x5001738000cb0191
0x5001738000cb0170
```

Add the second new port to the other HBA in the same way (Example 4-50).

Example 4-50 Attaching a remote port to the second HBA

```
lnxvm01:~ # echo 0x5001738000cb0181 > /sys/bus/ccw/devices/0.0.0601/port_add
lnxvm01:~ # ls   /sys/bus/ccw/devices/0.0.0601/ | grep 0x
0x5001738000cb0160
0x5001738000cb0181
```

4.3.4 Resizing volumes dynamically

You can use the additional capacity of dynamically enlarged volumes. Reducing the size is not supported. To resize volumes dynamically, complete the following steps:

1. Create a file system on one of the multipath devices and mount it. The **df** command that is shown in Example 4-51 shows the available capacity.

   Example 4-51 Checking the size and available space on a mounted file system

   ```
   x3650lab9:~ # df -h /mnt/itso_0520/
   file system           Size  Used Avail Use% Mounted on
   /dev/mapper/20017380000cb0520
                         16G  173M   15G   2% /mnt/itso_0520
   ```

2. Use the Hyper-Scale Manager (GUI) or XCLI to increase the capacity of the volume from 17 to 51 GB (decimal, as shown by the GUI). The Linux SCSI layer picks up the new capacity when you rescan each SCSI disk device (path) through **sysfs** (see Example 4-52).

   Example 4-52 Rescanning all disk devices (paths) of a 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 4-53 indicates the change in capacity.

   Example 4-53 Linux message log indicating the capacity change of a SCSI device

   ```
   x3650lab9:~ # tail /var/log/messages
   ...
   Oct 13 16:52:25 lnxvm01 kernel: [ 9927.105262] sd 0:0:0:4: [sdh] 100663296
   512-byte logical blocks: (51.54 GB/48 GiB)
   Oct 13 16:52:25 lnxvm01 kernel: [ 9927.105902] sdh: detected capacity change
   from 17179869184 to 51539607552
   ...
   ```

3. Indicate the device change to DM-MP by running the **resize_map** command of **multipathd**. The updated capacity is displayed in the output of **show topology** (see Example 4-54).

   Example 4-54 Resizing 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]
   ```

4. Resize the file system and check the new capacity, as shown in Example 4-55.

   Example 4-55 Resizing file system and checking capacity

   ```
   x3650lab9:~ # resize2fs /dev/mapper/20017380000cb0520
   resize2fs 1.41.9 (22-Aug-2009)
   file system 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 file system on /dev/mapper/20017380000cb0520 is now 12582912 blocks long.

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

> **Restrictions:** At the time of this writing, the dynamic volume increase process has the following restrictions:
>
> - It is not supported for Linux versions earlier than SUSE Linux Enterprise Server 11 SP1 and Red Hat Enterprise Linux 6.
> - The sequence works only with unpartitioned volumes.
> - The file system must be created directly on the DM-MP device.
> - Only the modern file systems can be resized while they are mounted. The ext2 file system cannot.

4.3.5 Using snapshots and remote replication targets

The snapshot and mirroring solutions create identical copies of the source volumes. The target has a unique identifier, which is made up from the WWNN of the system and volume serial number. Any metadata that is stored on the target, such as the file system identifier or LVM signature, however, is identical to that of the source. This metadata can lead to confusion and data integrity problems if you plan to use the target on the same Linux system as the source.

This section describes some methods to avoid integrity issues. It also highlights some potential traps that might lead to problems.

File system directly on a volume

The copy of a file system that is created directly on a SCSI disk device or a DM-MP device can be used on the same host as the source without modification. However, it cannot have an extra virtualization layer such as RAID or LVM. If you follow the sequence carefully and avoid the highlighted traps, you can use a copy on the same host without problems. The procedure is described on an ext3 file system on a DM-MP device that is replicated with a snapshot.

1. Mount the original file system, as shown in Example 4-56 using a device node that is bound to the unique identifier of the volume. The device node cannot be bound to any metadata that is stored on the device itself.

   Example 4-56 Mounting 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 that the data on the source volume is consistent by running the **sync** command.
3. Create the snapshot on the storage system, make it writable, and map the target volume to the Linux host. In the example, the snapshot source has the volume ID 0x0520, and the target volume has ID 0x1f93.

4. Initiate a device scan on the Linux host. For more information, see 4.3.1, “Adding and removing volumes dynamically” on page 173. DM-MP automatically integrates the snapshot target (see Example 4-57).

Example 4-57 Checking DM-MP topology for the 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. Mount the target volume to a separate mount point by using a device node that is created from the unique identifier of the volume (see Example 4-58).

Example 4-58 Mounting the target volume

```
x3650lab9:~ # mount /dev/mapper/20017380000cb1f93 /mnt/itso_fc/
x3650lab9:~ # 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 Universally Unique Identifier (UUID) or label. These IDs are stored in the data area of the volume, and are identical on both source and target. Such device nodes are ambiguous if the source and target are mapped to the host at the same time. Using them in this situation can result in data loss.

File system in a logical volume managed by LVM

The Linux *Logical Volume Manager* (LVM) uses metadata that is written to the data area of the disk device to identify and address its objects. If you want to access a set of replicated volumes that are under LVM control, modify this metadata so it is unique. This process ensures data integrity. Otherwise, LVM might mix volumes from the source and target sets.

A script named `vgimportclone.sh` is publicly available and automates the modification of the metadata. You can download it from this web page:

https://www.redhat.com/archives/lvm-devel/2009-May/msg00130.html

An online copy of the Linux man page for the script is available:

http://man7.org/linux/man-pages/man8/vgimportclone.8.html

Tip: The `vgimportclone` script and commands are part of the standard LVM tools for recent Red Hat Enterprise Linux and SUSE Linux Enterprise Server versions.

Complete the following steps to ensure consistent data on the target volumes and avoid mixing up the source and target. In this example, a volume group contains a logical volume that is striped over two volumes. Snapshots are used to create a point-in-time copy of both volumes. Both the original logical volume and the cloned one are then made available to the Linux system. The serial numbers of the source volumes are `1fc5` and `1fc6`, and the IDs of the target volumes are `1fe4` and `1fe5`.

1. Mount the original file system by using the LVM logical volume device (see Example 4-59).

   Example 4-59 Mounting the source volume

   ```
   x3650lab9:~ # mount /dev/vg_xiv/lv_itso /mnt/lv_itso
   x3650lab9:~ # mount
   ...
   /dev/mapper/vg_xiv-lv_itso on /mnt/lv_itso type ext3 (rw)
   ```

2. Make sure that the data on the source volume is consistent by running the **`sync`** command.
3. Create the snapshots on the storage system, unlock them, and map the target volumes `1fe4` and `1fe5` to the Linux host.
4. Initiate a device scan on the Linux host. For more information, see 4.3.1, “Adding and removing volumes dynamically” on page 173. DM-MP automatically integrates the snapshot targets, as shown in Example 4-60.

   Example 4-60 Checking 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]
   ```

   Important: To avoid data integrity issues, be sure that no LVM configuration commands are issued until step 5 is complete.

5. Run the `vgimportclone.sh` script against the target volumes, and provide a new volume group name (see Example 4-61).

   Example 4-61 Adjusting the LVM metadata of the target volumes

   ```
   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 (see Example 4-62).

Example 4-62   Activating volume group on target device and mounting the logical volume

```
x3650lab9:~ # vgchange -a y vg_itso_snap
  1 logical volume(s) in volume group "vg_itso_snap" now active
x3650lab9:~ # mount /dev/vg_itso_snap/lv_itso /mnt/lv_snap_itso/
x3650lab9:~ # mount
...
/dev/mapper/vg_xiv-lv_itso on /mnt/lv_itso type ext3 (rw)
/dev/mapper/vg_itso_snap-lv_itso on /mnt/lv_snap_itso type ext3 (rw)
```

4.4 Troubleshooting and monitoring

This section addresses topics that are related to troubleshooting and monitoring. As mentioned previously, always check that the Host Attachment Kit is installed.

After, key information can be found in the same directory as from where the installation was started, inside the `install.log` file.

4.4.1 Linux Host Attachment Kit utilities

The Host Attachment Kit includes the following utilities:

- `xiv_devlist`

  This command validates the attachment configuration. This command generates a list of multipath devices available to the operating system. The available options are listed in “The xiv_devlist command” on page 6.

- `xiv_diag`

  This command gathers diagnostic information from the operating system. The resulting compressed file can then be sent to IBM support team for review and analysis (see Example 4-63).

Example 4-63   The xiv_diag command

```
[/]# xiv_diag
Please type in a path to place the xiv_diag file in [default: /tmp]:
Creating archive xiv_diag-results_2017-10-27_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_2017-10-27_13-24-54.tar.gz to IBM-XIV for
review.
INFO: Exiting.
```

4.4.2 Multipath diagnosis

Some key diagnostic information can be found from the following `multipath` commands.

- To flush all multipath device maps:

  `multipath -F`

- To show the multipath topology (maximum information):

  `multipath -ll`

For more detailed information, use the **`multipath -ll`**, as shown in Example 4-64.

Example 4-64 Linux command multipath output that shows the correct status

```
# multipath -ll
mpatha (36001738ccce056730000000000013845) dm-3 IBM     ,2810XIV
size=47G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='service-time 0' prio=1 status=active
  |- 20:0:0:1 sdf 8:80 active ready  running
  |- 19:0:0:1 sde 8:64 active ready  running
  `- 21:0:0:1 sdg 8:96 active ready  running
```

Important: The **`multipath`** command sometimes finds errors in the `multipath.conf` file that do not exist. The following error messages can be ignored:

```
[root@b]# multipath -F
Sep 22 12:08:21 | multipath.conf line 30, invalid keyword: polling_interval
Sep 22 12:08:21 | multipath.conf line 41, invalid keyword: polling_interval
Sep 22 12:08:21 | multipath.conf line 53, invalid keyword: polling_interval
Sep 22 12:08:21 | multipath.conf line 54, invalid keyword: prio_callout
Sep 22 12:08:21 | multipath.conf line 64, invalid keyword: polling_interval
```

Another excellent command-line utility to use is the **`xiv_devlist`** command.

Important: When you are using the **`xiv_devlist`** command, note the number of paths that are indicated in the column for each device. You do not want the **`xiv_devlist`** output to show `N/A` in the paths column.

The expected output from the **`multipath`** and **`xiv_devlist`** commands is shown in Example 4-65.

Example 4-65 Shows multipath finding the devices, and updating the paths correctly

```
[root@bc-h-15-b7 ~]#multipath
create: mpathc (20017380027950251) undef IBM,2810XIV
size=48G features='0' hwhandler='0' wp=undef
`-+- policy='round-robin 0' prio=1 status=undef
  |- 8:0:0:1 sdc 8:32 undef ready running
  `- 9:0:0:1 sde 8:64 undef ready running
create: mpathd (20017380027950252) undef IBM,2810XIV
size=48G features='0' hwhandler='0' wp=undef
`-+- policy='round-robin 0' prio=1 status=undef
  |- 8:0:0:2 sdd 8:48 undef ready running
  `- 9:0:0:2 sdf 8:80 undef ready running
```

```
[root@bc-h-15-b7 ~]# xiv_devlist

XIV Devices
-------------------------------------------------------------------------------
Device         Size (GB)  Paths  Vol Name       Vol Id   XIV Id   XIV Host
-------------------------------------------------------------------------------
/dev/mapper/m  51.6       2/2    RedHat-Data_1  593      1310133  RedHat6.de.ib
pathc                                                             m.com
-------------------------------------------------------------------------------
/dev/mapper/m  51.6       2/2    RedHat-Data_2  594      1310133  RedHat6.de.ib
pathd                                                             m.com
-------------------------------------------------------------------------------


Non-XIV Devices
--------------------------
Device    Size (GB)  Paths
--------------------------
/dev/sda  50.0       N/A
--------------------------
/dev/sdb  50.0       N/A
--------------------------
```

4.4.3 Other ways to check SCSI devices

The Linux kernel maintains a list of all attached SCSI devices in the `/proc` pseudo file system, as shown in Example 4-66. The `/proc/scsi/scsi` pseudo file system contains basically the same information (apart from the device node) as the **`lsscsi`** output. It is always available, even if **`lsscsi`** is not installed.

Example 4-66 Alternate list of attached SCSI devices

```
x3650lab9:~ # cat /proc/scsi/scsi
Attached devices:
Host: scsi0 Channel: 00 Id: 00 Lun: 01
  Vendor: IBM      Model: 2810XIV          Rev: 10.2
  Type:   Direct-Access                    ANSI SCSI revision: 05
Host: scsi0 Channel: 00 Id: 00 Lun: 02
  Vendor: IBM      Model: 2810XIV          Rev: 10.2
  Type:   Direct-Access                    ANSI SCSI revision: 05
Host: scsi0 Channel: 00 Id: 00 Lun: 03
  Vendor: IBM      Model: 2810XIV          Rev: 10.2
  Type:   Direct-Access                    ANSI SCSI revision: 05
Host: scsi1 Channel: 00 Id: 00 Lun: 01
  Vendor: IBM      Model: 2810XIV          Rev: 10.2
  Type:   Direct-Access                    ANSI SCSI revision: 05
Host: scsi1 Channel: 00 Id: 00 Lun: 02
  Vendor: IBM      Model: 2810XIV          Rev: 10.2
  Type:   Direct-Access                    ANSI SCSI revision: 05
Host: scsi1 Channel: 00 Id: 00 Lun: 03
  Vendor: IBM      Model: 2810XIV          Rev: 10.2
  Type:   Direct-Access                    ANSI SCSI revision: 05
...
```

The **fdisk -l** command can be used to list all block devices, including their partition information and capacity (see Example 4-67). However, it does not include SCSI address, vendor, and model information.

Example 4-67 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

...
```

4.4.4 Performance monitoring with iostat

You can use the **iostat** command to monitor the performance of all attached disks. It is part of the `sysstat` package that is included with every major Linux distribution. However, it is not necessarily installed by default. The **iostat** command reads data that are provided by the kernel in `/proc/stats` and prints it in human readable format. For more information, see the man page of **iostat**.

4.4.5 Generic SCSI tools

For Linux, the *sg tools* allow low-level access to SCSI devices. They communicate with SCSI devices through the generic SCSI layer. This layer is represented by special device files `/dev/sg0`, `/dev/sg1`, and so on. In recent Linux versions, the *sg tools* can also access the block devices `/dev/sda`, and `/dev/sdb`. They can also access any other device node that represents a SCSI device directly.

The most useful sg tools include the following examples:

sg_inq /dev/sg*x* Prints SCSI Inquiry data, such as the volume serial number.

sg_scan Prints the SCSI host, channel, target, and LUN mapping for all SCSI devices.

sg_map Prints the `/dev/sdx` to `/dev/sgy` mapping for all SCSI devices.

sg_readcap /dev/sg*x* Prints the block size and capacity (in blocks) of the device.

sginfo /dev/sg*x* Prints SCSI inquiry and mode page data. You can also use it to manipulate the mode pages.

4.5 Boot from SAN in Linux

This section describes how you can configure a system to load the Linux kernel and operating system from a SAN-attached FlashSystem A9000, A9000R, or XIV volume. This process is illustrated with an example based on SUSE Linux Enterprise Server 11 SP1 on an x86 server with QLogic FC HBAs, booting from XIV volume. Other distributions and hardware platforms that have deviations from the example are noted. For more information about configuring the HBA BIOS to boot from SAN-attached volume, see 1.2.5, "Boot from SAN on x86 or x64 based architecture" on page 17.

4.5.1 Linux boot process

To understand how to boot a Linux system from SAN-attached volumes, you need a basic understanding of the Linux boot process. A Linux system goes through the following basic steps until it presents the login prompt:

1. OS loader.

   The system firmware provides functions for rudimentary I/O operations such as the BIOS of x86 servers. When a system is turned on, it runs the *power-on self-test (POST)* to check which hardware is available and whether everything is working. Then, it runs the operating system loader (OS loader). The OS loader uses those basic I/O routines to read a specific location on the defined system disk and starts running the code that it contains. This code is either part of the boot loader of the operating system, or it branches to the location where the boot loader is located.

   If you want to boot from a SAN-attached disk, make sure that the OS loader can access that disk. FC HBAs provide an extension to the system firmware for this purpose. In many cases, it must be explicitly activated.

   On x86 systems, this location is called the *Master Boot Record (MBR)*.

   > **Remember:** For Linux on z Systems under z/VM, the OS loader is not part of the firmware. Instead, it is part of the z/VM program `ipl`.

2. The boot loader.

   The boot loader starts the operating system kernel. It must know the physical location of the kernel image on the system disk. It then reads it in, extracts it if it is compressed, and starts it. This process is still done by using the basic I/O routines that are provided by the firmware. The boot loader can also pass configuration options and the location of the `InitRAMFS` to the kernel.

   These are most common Linux boot loaders:

   – `GRUB` (Grand Unified Bootloader) for x86 systems
   – `zipl` for z Systems
   – `yaboot` for Power Systems

3. The kernel and the InitRAMFS.

   After the kernel is extracted and running, it takes control of the system hardware. It starts and configures the following systems:

   – Memory management

   – Interrupt handling

   – The built-in hardware drivers for the hardware that is common on all systems, such as MMU and clock

It reads and extracts the InitRAMFS image, again by using the same basic I/O routines. The InitRAMFS contains more drivers and programs that are needed to set up the Linux file system tree (root file system). To be able to boot from a SAN-attached disk, the standard `InitRAMFS` must be extended with the FC HBA driver and the multipathing software. In modern Linux distributions, this process is done automatically by the tools that create the `InitRAMFS` image.

After the root file system is accessible, the kernel starts the `init()` process.

4. The `init()` process.

   The `init()` process starts the operating system itself, including networking, services, and user interfaces. The hardware is already abstracted. Therefore, `init()` is neither platform-dependent, nor are there any SAN-boot specifics.

4.5.2 Configuring the QLogic BIOS to boot from a SAN-attached volume

The first step to configure the HBA is to load a BIOS extension that provides the basic input/output capabilities for a SAN-attached disk. For more information, see 1.2.5, “Boot from SAN on x86 or x64 based architecture” on page 17.

> **Tip:** Emulex HBAs also support booting from SAN disk devices. You can enable and configure the Emulex BIOS extension by pressing Alt+E or Ctrl+E when the HBAs are initialized during server startup. For more information, see the Broadcom website:
>
> `https://www.broadcom.com/`

4.5.3 OS loader considerations for other platforms

The BIOS is the x86 specific way to start loading an operating system. This section briefly describes how this loading is done on the other supported platforms.

IBM Power Systems

When you install Linux on an IBM Power System server or LPAR, the Linux installer sets the boot device in the firmware to the drive that you are installing on. No special precautions must taken whether you install on a local disk, a SAN-attached volume, or a virtual disk provided by the VIO server.

IBM z Systems

Linux on z Systems can be loaded from traditional CKD disk devices or from Fibre Channel attached fixed block (SCSI) devices. To load from SCSI disks, the SCSI IPL feature (FC 9904) must be installed and activated on the z Systems server. The SCSI *initial program load (IPL)* is generally available on recent z Systems (IBM z10™ and later).

> **Important:** Activating the SCSI IPL feature is disruptive. It requires a power-on reset (POR) of the whole system.

Linux on z Systems can run in two configurations:

- Linux on z Systems running natively in a z Systems LPAR

  After you install Linux on z Systems, you must provide the device from which the LPAR runs the IPL in the LPAR start dialog on the z Systems *Support Element*. After it is registered there, the IPL device entry is permanent until changed.

- Linux on z Systems running under z/VM

  Within z/VM, you start an operating system with the IPL command. This command provides the z/VM device address of the device where the Linux boot loader and kernel are installed.

  When you boot from SCSI disk, you do not have a z/VM device address for the disk itself. For more information, see 4.2.1, “Platform-specific remarks” on page 146, and “z Systems” on page 148. You must provide information about which LUN the machine loader uses to start the operating system separately. z/VM provides the **`cp`** commands **`set loaddev`** and **`query loaddev`** for this purpose. Their use is shown in Example 4-68.

  Example 4-68 Setting and querying SCSI IPL device in z/VM

  ```
  SET LOADDEV PORTNAME 50017380 00CB0191 LUN 00010000 00000000

  CP QUERY LOADDEV
  PORTNAME 50017380 00CB0191    LUN  00010000 00000000    BOOTPROG 0
  BR_LBA   00000000 00000000
  ```

  The port name is the 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 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.

4.5.4 Installing SUSE Linux Enterprise Server 11 SP1 on a SAN volume

With recent Linux distributions, installation on a SAN-attached volume is as easy as installation on a local disk. The process has the following extra considerations:

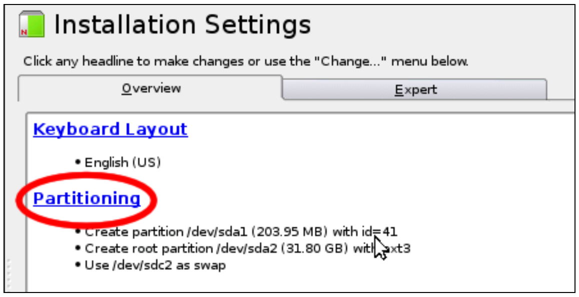
- Identifying the correct volumes to install on
- Enabling multipathing during installation

> **Tip:** After the SUSE Linux Enterprise Server 11 installation program (YAST) is running, the installation is mostly hardware independent. It works the same when it runs on an x86, IBM Power System, or z Systems server.

To install SUSE Linux Enterprise Server 11 SP1 on a volume, complete the following steps:

1. Boot from an installation DVD. Follow the installation configuration windows until you come to the Installation Settings window shown in Figure 4-7 on page 188.

   > **Remember:** The Linux on z Systems installer does not automatically list the available disks for installation. Use the Configure Disks window to discover and attach the disks that are needed to install the system by using a graphical user interface. This window is displayed before you get to the Installation Settings window. At least one disk device is required to run the installation.



Installation Settings

Click any headline to make changes or use the "Change..." menu below.

Keyboard Layout

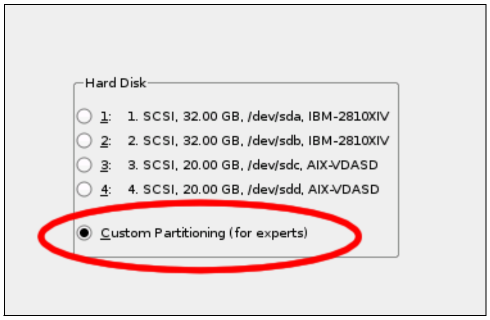
- English (US)

Partitioning

- Create partition /dev/sda1 (203.95 MB) with id=41
- Create root partition /dev/sda2 (31.80 GB) with ext3
- Use /dev/sdc2 as swap

Figure 4-7 SUSE Linux Enterprise Server 11 SP1 installation settings

2. Click **Partitioning**.
3. In the Preparing Hard Disk: Step 1 window (see Figure 4-8), make sure that **Custom Partitioning (for experts)** is selected. Which disk device is selected in the Hard Disk field does not matter. and click **Next**.



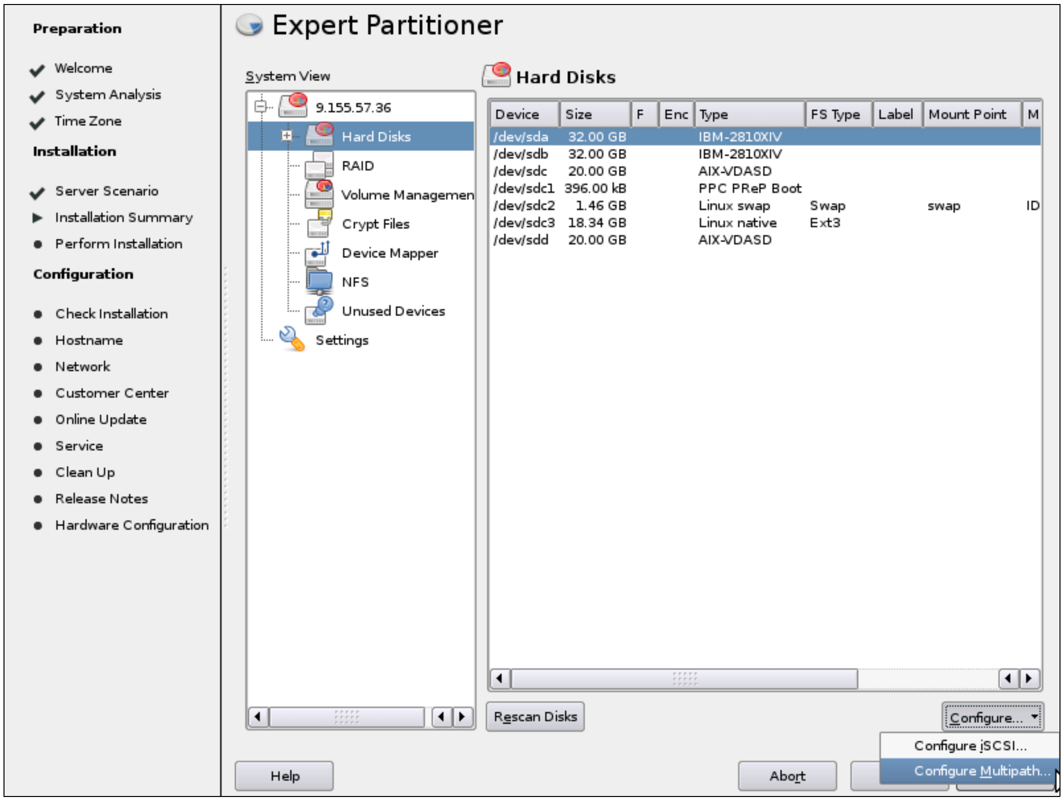
Hard Disk

○ 1: 1. SCSI, 32.00 GB, /dev/sda, IBM-2810XIV
○ 2: 2. SCSI, 32.00 GB, /dev/sdb, IBM-2810XIV
○ 3: 3. SCSI, 20.00 GB, /dev/sdc, AIX-VDASD
○ 4: 4. SCSI, 20.00 GB, /dev/sdd, AIX-VDASD

● Custom Partitioning (for experts)

Figure 4-8 Preparing Hard Disk: Step 1 window

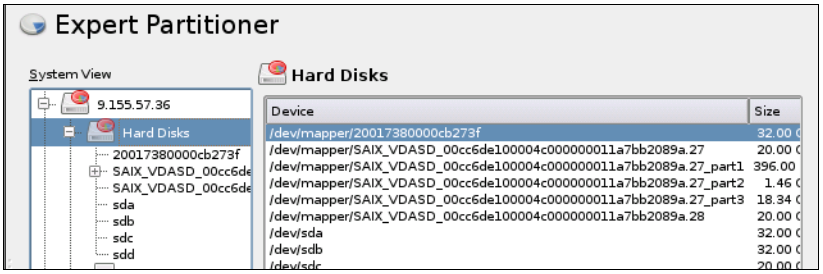
4. Enable multipathing in the Expert Partitioner window (see Figure 4-9). Select **Hard Disks** in the navigation section on the left side. Then, click **Configure** →**Configure Multipath**.



Device	Size	F	Enc	Type	FS Type	Label	Mount Point	M
/dev/sda	32.00 GB			IBM-2810XIV				
/dev/sdb	32.00 GB			IBM-2810XIV				
/dev/sdc	20.00 GB			AIX-VDASD				
/dev/sdc1	396.00 kB			PPC PReP Boot				
/dev/sdc2	1.46 GB			Linux swap	Swap		swap	ID
/dev/sdc3	18.34 GB			Linux native	Ext3			
/dev/sdd	20.00 GB			AIX-VDASD				

Figure 4-9 Enabling multipathing in the Expert Partitioner window

5. Confirm your selection, and then the tool rescans the disk devices. When finished, it presents an updated list of hard disks that also shows the multipath devices it found (see Figure 4-10).



Device	Size
/dev/mapper/20017380000cb273f	32.00
/dev/mapper/SAIX_VDASD_00cc6de100004c000000011a7bb2089a.27	20.00
/dev/mapper/SAIX_VDASD_00cc6de100004c000000011a7bb2089a.27_part1	396.00
/dev/mapper/SAIX_VDASD_00cc6de100004c000000011a7bb2089a.27_part2	1.46
/dev/mapper/SAIX_VDASD_00cc6de100004c000000011a7bb2089a.27_part3	18.34
/dev/mapper/SAIX_VDASD_00cc6de100004c000000011a7bb2089a.28	20.00
/dev/sda	32.00
/dev/sdb	32.00
/dev/sdc	20.00

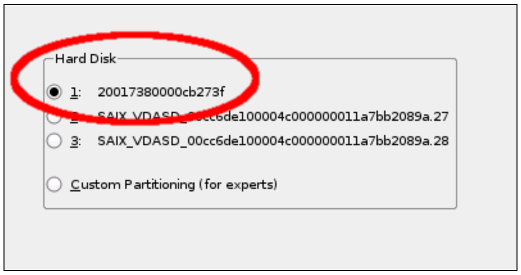
Figure 4-10 Selecting multipath device for installation

6. Select the multipath device (storage system volume) that you want to install to and click **Accept**.

7. In the Partitioner window, create and configure the required partitions for your system the same way you would on a local disk.

You can also use the automatic partitioning capabilities of YAST after the multipath devices are detected in step 5. To do so, complete the following steps:

1. Click **Back** until you see the initial partitioning window again. It now shows the multipath devices instead of the disks (see Figure 4-11).



Hard Disk

● 1: 20017380000cb273f

○ 2: SAIX_VDASD_00cc6de100004c000000011a7bb2089a.27

○ 3: SAIX_VDASD_00cc6de100004c000000011a7bb2089a.28

○ Custom Partitioning (for experts)

Figure 4-11 Preparing Hard Disk: Step 1 window with multipath devices

2. Select the multipath device that you want to install on and click **Next**.
3. Select the partitioning scheme that you want.

Important: All supported platforms can boot Linux from multipath devices. In some cases, however, the tools that install the boot loader can write only to simple disk devices. In these cases, install the boot loader with multipathing deactivated. For SUSE Linux Enterprise Server 10 and SUSE Linux Enterprise Server 11, add the **`multipath=off`** parameter to the boot command in the boot loader. The boot loader for IBM Power Systems and z Systems must be reinstalled whenever there is an update to the kernel or `InitRAMFS`. A separate entry in the boot menu allows you to switch between single and multipath mode when necessary.

For more information, see the Linux distribution documentation listed in 4.1.2, “Reference material” on page 142.

The installer does not implement any device-specific settings, such as creating the `/etc/multipath.conf` file. You must implement these settings manually after installation as explained in 4.2.7, “Special considerations for attachment” on page 171. Because DM-MP is already started during the processing of the InitRAMFS, you also must build a new `InitRAMFS` image after you change the DM-MP configuration. For more information, see “Making the FC driver available early in the boot process” on page 151.

Adding *Device Mapper* layers on top of DM-MP, such as software RAID or LVM is possible. The Linux installers support these options.

Tip: Red Hat Enterprise Linux 5.1 and later support multipathing already. Turn on multipathing by adding the **`mpath`** option to the kernel boot line of the installation system. Anaconda, the RH installer, then offers to install to multipath devices.

5

AIX connectivity

This chapter addresses the specifics for attaching IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System to host systems that are running the AIX operating system.

> **Important:** The procedures and instructions that are given here are based on a certain code, but it might not be the most recent one. For the latest support information and instructions, see IBM System Storage Interoperation Center (SSIC):
>
> `https://www.ibm.com/systems/support/storage/ssic/interoperability.wss`
>
> Host Attachment Kits and related publications can be downloaded from Fix Central:
>
> `http://www.ibm.com/support/fixcentral/`

This chapter includes the following topics:

5.1 Attaching to AIX hosts

This section provides information and procedures for attaching the storage systems to AIX on an IBM POWER® platform. Fibre Channel (FC) connectivity is addressed first, followed by the iSCSI attachment method.

FlashSystem A9000, A9000R, and XIV Storage System support different versions of the AIX operating system through FC or iSCSI connectivity.

Consider the following points that apply to all AIX releases:

- Host Attachment Kit 2.8 for AIX (current at the time of this writing) supports all AIX releases 6.1, 7.1 and 7.2.
- Dynamic LUN expansion on XIV with LVM requires XIV firmware version 10.2 or later.

5.1.1 Prerequisites

If the current AIX operating system level installed on your system is not compatible with the storage system, you must upgrade before you attach to the storage system. To determine the maintenance package or technology level that is currently installed on your system, use the **oslevel** command (see Example 5-1).

Example 5-1 Determining current AIX version and maintenance level

```
# oslevel -s
7100-01-05-1228
```

In this example, the system is running AIX 7.1.0.0 technology level 1 (7.1TL1). Use this information with the SSIC to ensure that you have an IBM supported configuration.

If AIX maintenance items are needed, consult IBM Fix Central, where you can download fixes and updates for your systems software, hardware, and operating system:

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

Before further configuring your host system or the storage system, make sure that the physical connectivity between the storage system and the POWER system is properly established. Direct attachment from the host system is not supported. If you use FC switched connections, ensure that you have functioning zoning that uses the worldwide port name (WWPN) numbers of the AIX host.

5.1.2 AIX host FC configuration

Attaching the storage system to an AIX host using FC involves the following tasks from the host side:

1. Identifying the FC host bus adapters (HBAs) and determining their WWPN values.
2. Installing the IBM Storage Host Attachment Kit for AIX.
3. Configuring multipathing.

Identifying FC adapters and attributes

To provide volumes to an AIX host, identify the FC adapters on the AIX server. Use the **lsdev** command to list all the FC adapter ports in your system (see Example 5-2). This example shows two FC ports.

Example 5-2 Listing FC adapters

```
# lsdev -Cc adapter|grep fcs
fcs0   Available 00-00 8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs1   Available 00-01 8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
```

The **lsslot** command returns the ports, and also the PCI slot where the FC adapters are in the system (see Example 5-3). This command can be used to identify in what physical slot a specific adapter is placed.

Example 5-3 Locating FC adapters

```
# lsslot -c pci | grep fcs
U5802.001.00H3722-P1-C10  PCI-E capable, Rev 1 slot with 8x lanes  fcs0 fcs1
```

To obtain the WWPN of each of the POWER system FC adapters, use the **lscfg** command (see Example 5-4).

Example 5-4 Finding FC adapter WWPN

```
# lscfg -vl fcs0
  fcs0           U5802.001.00H3722-P1-C10-T1  8Gb PCI Express Dual Port FC Adapter
(df1000f114108a03)

        Part Number.................10N9824
        Serial Number...............1A113001FB
        Manufacturer................001A
        EC Level....................D77040
        Customer Card ID Number.....577D
        FRU Number..................10N9824
        Device Specific.(ZM)........3
        Network Address.............10000000C9B7F27A
        ROS Level and ID............0278117B
        Device Specific.(Z0)........31004549
        Device Specific.(Z1)........00000000
        Device Specific.(Z2)........00000000
        Device Specific.(Z3)........09030909
        Device Specific.(Z4)........FF781116
        Device Specific.(Z5)........0278117B
        Device Specific.(Z6)........0773117B
        Device Specific.(Z7)........0B7C117B
        Device Specific.(Z8)........20000000C9B7F27A
        Device Specific.(Z9)........US1.11X11
        Device Specific.(ZA)........U2D1.11X11
        Device Specific.(ZB)........U3K1.11X11
        Device Specific.(ZC)........00000000
        Hardware Location Code......U5802.001.00H3722-P1-C10-T1
```

You can also print the WWPN of an HBA directly by issuing the following command (where `<fcs#>` is the instance of an FC HBA to query):

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

Installing the Host Attachment Kit for AIX

For AIX to correctly recognize the disks that are mapped from the storage system as MPIO 2810 XIV Disk, the IBM Storage Host Attachment Kit for AIX is required. This package also enables multipathing. At the time of this writing, Host Attachment Kit 2.8 is the most recent package that is available. You can download the file set from Fix Central:

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

> **Important:** Although AIX now natively supports FlashSystem A9000, A9000R, and XIV using Object Data Manager (ODM) changes that were back-ported to several older AIX releases, the preferred solution is to use the Host Attachment Kit. The kit automates the host definition and LUN mappings and provides valuable data collection and troubleshooting tools.

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

1. Download or copy the downloaded Host Attachment Kit to your AIX system.
2. From the AIX prompt, change to the directory where your package is located.
3. Run the **`gunzip –c IBM_Storage_Host_Attachment_Kit_<HAK_build_name>_AIX.tar.gz | tar xvf –`** command to extract the file.
4. Switch to the newly created directory and run the installation script (see Example 5-5).

Example 5-5 Installing the Host Attachment Kit for AIX

```
# ./install.sh
Welcome to the IBM Storage Host Attachment Kit installer.
Would you like to proceed and install the IBM Storage Host Attachment Kit?
[Y/n]:
y
Please wait while the installer validates your existing configuration...
---------------------------------------------------------------
Please wait as the IBM Storage Host Attachment Kit is being installed. This may
take a few minutes...
---------------------------------------------------------------
Installation successful.
Please refer to the user guide for information about how to configure this
host.

---------------------------------------------------------------
The IBM Storage Host Attachment Kit includes the following utilities:
xiv_attach: Interactive wizard that configures the host and verifies its
            configuration for connectivity with the IBM storage System.
xiv_devlist: Lists all storage volumes that are mapped to the host, with
general
            info about non-storage volumes.
xiv_syslist: Lists all IBM storage systems that are detected by the host.
xiv_diag: Performs complete diagnostics of the host and its connectivity with
            the IBM storage System, and saves the information to a file.
xiv_fc_admin: Allows you to perform different administrative operations for
            FC-connected hosts and IBM storage systems.
xiv_iscsi_admin: Allows you to perform different administrative operations for
            iSCSI-connected hosts and IBM storage systems.
xiv_host_profiler: Collects host configuration information and performs a
            comprehensive analysis of the collected information.
---------------------------------------------------------------
```

5. Zone the host to the storage system.
6. The Host Attachment Kit provides an interactive command-line utility to configure and connect the host to the storage system. At this time, only FC attachment is supported for FlashSystem A9000 and A9000R. If iSCSI attachment is needed, check the SSIC to learn which AIX versions are supported with XIV Storage system. The **xiv_attach** command starts a wizard that attaches the host to the storage system and creates the host object on the XIV. Example 5-6 shows **xiv_attach** command output.

Example 5-6 Attachment to storage system and host creation on it

```
# xiv_attach
-------------------------------------------------------------------------------
Welcome to the IBM Storage Host Attachment wizard, version 2.8.0.
This wizard will help you attach this host to one or more IBM storage systems

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

-------------------------------------------------------------------------------
Only Fiber Channel connectivity is supported on this host.
Would you like to perform Fibre Channel attachment? [default: yes ]:
-------------------------------------------------------------------------------
Please wait while the wizard validates your existing configuration...
Verifying AIX packages                                                        OK
This host is already configured for the IBM storage system.

-------------------------------------------------------------------------------
Creating a host profile enables focused and better support for your host.
Would you like to add a scheduled task for host profile creation? [default: yes ]:
A scheduled task for xiv_host_profiler has been added.
-------------------------------------------------------------------------------
Please define zoning for this host and add its World Wide Port Names (WWPNs) to the
IBM storage system:
C0:50:76:03:44:90:00:38: fcs0: [IBM]: N/A
C0:50:76:03:44:90:00:3A: fcs1: [IBM]: N/A
C0:50:76:03:44:90:00:3C: fcs2: [IBM]: N/A
C0:50:76:03:44:90:00:3E: fcs3: [IBM]: N/A
Press [ENTER] to proceed.

Would you like to rescan for new storage devices? [default: yes ]:
Please wait while rescanning for IBM storage devices...
-------------------------------------------------------------------------------
This host is connected to the following IBM storage arrays:
Storage Type        Serial   System Version  Host Defined  Ports Defined
Protocol  Host Name(s)
XIV                1340010  11.6.2         No           No ports defined  FC
N/A
XIV                1340008  11.6.2.a       No           No ports defined  FC
N/A
FlashSystem A9000   1322131  12.1.0.b       No           No ports defined  FC
N/A
FlashSystem A9000R  1320902  12.1.0.b       No           No ports defined  FC
N/A
This host is not defined on some FC-attached IBM storage systems.
Do you want to define this host on these IBM storage systems now? [default: yes ]:
Please enter a name for this host [default: p7-730-02v1.mainz.de.ibm.com ]:
p7-730-LPAR1
Please enter a username for system 1340010 [default: admin ]:
```

```
Please enter the password of user admin for system 1340010:


Please enter a username for system 1322131 [default: admin ]:
Please enter the password of user admin for system 1322131:


Please enter a username for system 1340008 [default: admin ]:
Please enter the password of user admin for system 1340008:


Please enter a username for system 1320902 [default: admin ]:
Please enter the password of user admin for system 1320902:



Press [ENTER] to proceed.

-------------------------------------------------------------------------------
The IBM Storage Host Attachment wizard has successfully configured this host.

Press [ENTER] to exit.
```

7. Create volumes on the storage system and map these volumes (LUNs) to the host system that was configured by **xiv_attach**. You can use the Hyper-Scale Manager GUI for volume creation and mapping tasks, as described in 1.4, “Logical configuration for host connectivity” on page 31. Use **cfgmgr** or **xiv_fc_admin -R** to rescan for LUNs (see Example 5-7).

Example 5-7 XIV labeled FC disks

```
# xiv_fc_admin -R
# lsdev -Cc disk
hdisk0 Available          Virtual SCSI Disk Drive
hdisk1 Available 51-T1-01 MPIO 2810 XIV Disk
hdisk2 Available 51-T1-01 MPIO 2810 XIV Disk
```

8. Use the **xiv_devlist** command to get more information about the mapped LUNs (see Example 5-8).

Example 5-8 The xiv_devlist command

```
# xiv_devlist -x
IBM storage devices
-------------------------------------------------------------------------------
---------------------
Device       Size (GB)  Paths  Vol Name      Vol ID  Storage ID  Storage Type
Hyper-Scale Mobility
-------------------------------------------------------------------------------
---------------------
/dev/hdisk1  50.1       12/12  ITSO_AIX_002  14407   1322131     FlashSystem A9000
Idle
-------------------------------------------------------------------------------
---------------------
/dev/hdisk2  50.1       12/12  ITSO_AIX_001  14406   1322131     FlashSystem A9000
Idle
-------------------------------------------------------------------------------
--------------------------
```

To add disks to the system, complete the following steps:

1. Use the GUI or XCLI to map the new LUNs to the AIX server.
2. On the AIX system, run **`xiv_fc_admin -R`** to rescan for the new LUNs.
3. Use **`xiv_devlist`** to confirm that the new LUNs are present to the system.

Although other AIX commands, such as **`cfgmgr`**, also can be used, these commands are built within the Host Attachment Kit commands.

Portable Host Attachment Kit Install and usage

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

The **`xiv_fc_admin`** command can be used to confirm that the AIX server is running a supported configuration and ready to attach to the storage. Use the **`xiv_fc_admin -V`** command to verify the configuration and be notified if any OS component is missing. The **`xiv_attach`** command must be run the first time that the server is attached to the array. It is used to scan for new storage system LUNs and configure the server to work with XIV.

Do not run the **`xiv_attach`** command more than once. If LUNs are added in the future, use the **`xiv_fc_admin -R`** command to scan for the new LUNs. For more information about these commands and others in the portable Host Attachment Kit, see 5.1.5, “Host Attachment Kit utilities” on page 210.

Using a network drive

Complete the following steps to use the portable Host Attachment Kit package from a network drive:

1. Extract the files from the following file into a shared folder on a network drive:

   ```
   IBM_Storage_Host_Attachment_Kit_<HAK_build_name>_Portable.tar.gz
   ```

2. Mount the shared folder to each host computer you intend to use the Host Attachment Kit on. The folder must be recognized and accessible as a network drive.

You can now use the Host Attachment Kit on any host to which the network drive is mounted.

To run commands from the portable Host Attachment Kit location, use a period and forward slash character combination (`./`) before every command.

> **Tip:** Whenever a newer Host Attachment Kit version is installed on the network drive, all hosts to which that network drive was mounted have access to that version.

Using a portable USB flash drive

Complete the following steps to use the portable Host Attachment Kit package from a USB flash drive:

1. Extract the files from the `tar.gz` file into a folder on the USB flash drive.
2. Plug the USB flash drive into any host on which you want to use the Host Attachment Kit.
3. Run any Host Attachment Kit utility from the drive.

For more information about setting up hosts that use the portable Host Attachment Kit, see “AIX MPIO” on page 199.

Removing the Host Attachment Kit software

In some situations, you must remove the Host Attachment Kit. In most cases, when you are upgrading to a new version, the Host Attachment Kit can be installed without uninstalling the older version first. Check the release notes and instructions to determine the best procedure.

If the Host Attachment Kit is locally installed on the host, you can uninstall it without detaching the host from the storage system.

The portable Host Attachment Kit packages do not require the uninstallation procedure. You can delete the portable Host Attachment Kit directory on the network drive or the USB flash drive to uninstall it. For more information about the portable Host Attachment Kit, see "Portable Host Attachment Kit Install and usage" on page 197.

The regular uninstallation removes the locally installed Host Attachment Kit software without detaching the host. This process preserves all multipathing connections to the storage system.

Use the following command to uninstall the Host Attachment Kit software:

```
# /opt/xiv/host_attach/bin/uninstall
```

The **`uninstall`** command removes the following components:

- IBM Storage Solutions External Runtime Components
- IBM Storage Host Attachment Kit tools

If you get the message the following message, use the package management service to remove the Host Attachment Kit:

```
Please use the O/S package management services to remove the package
```

The package name is `xiv.hostattachment.tools`. To remove the package, use the **`installp -u hostattachment.tools`** command (see Example 5-9).

Example 5-9 Uninstalling the Host Attachment Kit

```
# installp -u hostattachment.tools
+-----------------------------------------------------------------------------+
                    Pre-deinstall Verification...
+-----------------------------------------------------------------------------+
Verifying selections...done
Verifying requisites...done
Results...

SUCCESSES
---------
  Filesets listed in this section passed pre-deinstall verification
  and will be removed.

  Selected Filesets
  -----------------
  hostattachment.tools 2.8.0.0                # Support tools for Storage co...

  << End of Success Section >>

FILESET STATISTICS
------------------
    1  Selected to be deinstalled, of which:
        1  Passed pre-deinstall verification
  ----
    1  Total to be deinstalled

+------------------------------------------------------------------------------+
                           Deinstalling Software...
+------------------------------------------------------------------------------+

installp: DEINSTALLING software for:
        hostattachment.tools 2.8.0.0

Removing dynamically created files from the system
Finished processing all filesets.  (Total time:  59 secs).

+------------------------------------------------------------------------------+
                                Summaries:
+------------------------------------------------------------------------------+

Installation Summary
--------------------
Name                        Level           Part        Event       Result
--------------------------------------------------------------------------------
hostattachment.tools        2.8.0.0         USR         DEINSTALL   SUCCESS
```

AIX MPIO

AIX Multipath I/O (MPIO) is an enhancement to the base OS environment that provides native support for multipath FC storage attachment. MPIO automatically discovers, configures, and makes available every storage device path. The storage device paths provide high availability and load balancing for storage I/O. MPIO is part of the base AIX kernel, and is available with the current supported AIX levels.

The MPIO base functionality is limited. It provides an interface for vendor-specific path control modules (PCMs) that allow for implementation of advanced algorithms.

For more information, see IBM Knowledge Center:

https://ibm.biz/BdsbTd

Configuring devices as MPIO or non-MPIO devices

Configuring storage system devices as MPIO provides the best solution. However, if you are using a third-party multipathing solution, you might want to manage the XIV 2810 device with the same solution. Using a solution that is not from IBM usually requires the devices to be configured as non-MPIO devices.

The AIX **`manage_disk_drivers`** command can switch a device between MPIO and non-MPIO. This command can be used to change how the device is configured. All disks are converted.

> **Restriction:** Converting one storage system disk to MPIO and another storage system disk to non-MPIO is not possible.

After you run either of the following **manage_disk_drivers** commands to switch a device, reboot the system for the configuration change to take effect:

- To switch XIV 2810 devices from MPIO to non-MPIO, run the following command and reboot:

  ```
  manage_disk_drivers -o AIX_non_MPIO -d 2810XIV
  ```

- To switch XIV 2810 devices from non-MPIO to MPIO, run the following command and reboot:

  ```
  manage_disk_drivers -o AIX_AAPCM -d 2810XIV
  ```

To display the present settings, run the following command:

```
manage_disk_drivers -l
```

Disk behavior algorithms and queue depth settings

By using the storage systems in a multipath environment, you can change the disk behavior algorithm between `round_robin` and `fail_over` mode. The default disk behavior mode is `round_robin`, with a queue depth setting of 40.

Check the disk behavior algorithm and queue depth settings, as shown in Example 5-10.

Example 5-10 Viewing disk behavior and queue depth

```
# lsattr -El hdisk2 | grep -e algorithm -e queue_depth
algorithm       round_robin                                 Algorithm True
queue_depth     40                                          Queue DEPTH True
```

If the application is I/O intensive and uses large block I/O, the queue_depth and the max transfer size might need to be adjusted. Such an environment typically needs a `queue_depth` of 64 - 256, and *`max_tranfer`*`=0x100000`. Typical values are 40 - 64 as the queue depth per LUN, and 512 - 2048 per HBA in AIX.

Performance tuning

This section provides some performance considerations to help you adjust your AIX system to best fit your environment. If you boot from a SAN-attached LUN, create a mksysb image or a crash-consistent snapshot of the boot LUN before you change the HBA settings.

AIX includes the following performance considerations:

- Use multiple threads and asynchronous I/O to maximize performance on the storage system.
- Check with **`iostat`** on a per path basis for the LUNs to ensure that the load is balanced across all paths.
- Verify the HBA queue depth and per LUN queue depth for the host are sufficient to prevent queue waits. However, ensure that they are not so large that they overrun the storage system queues.

  For XIV, queue limit is 1400 per XIV port and 256 per LUN per WWPN (host) per port. Do not submit more I/O per XIV port than the 1400 maximum it can handle.

  For FlashSystem A9000 and A9000R, queue limit is 2048 per storage system port and 256 per LUN per WWPN (host) per port. Do not submit more I/O per storage system port than the 2048 maximum it can handle.

  The limit for the number of queued I/O for an HBA on AIX systems with 8-Gb HBAs is 4096. This limit is controlled by the `num_cmd_elems` attribute for the HBA, which is the maximum number of commands that AIX queues.

If necessary, increase it to the maximum value, which is 4096. The exception is if you have HBAs of 1 Gbps, 2 Gbps, or 4 Gbps, in which cases the maximum is lower.

- The other setting to consider is the `max_xfer_size`. This setting controls the maximum I/O size the adapter can handle. The default is 0x100000. If necessary, increase it to 0x200000 for large IOs, such as backups.

Check these values by using **`lsattr -El fcsX`** for each HBA, as shown in Example 5-11.

Example 5-11 The lsattr command

```
# lsattr -El fcs0
DIF_enabled   no         DIF (T10 protection) enabled                       True
bus_intr_lvl             Bus interrupt level                                False
bus_io_addr   0xff800    Bus I/O address                                    False
bus_mem_addr  0xffe76000 Bus memory address                                 False
bus_mem_addr2 0xffe78000 Bus memory address                                 False
init_link     auto       INIT Link flags                                    False
intr_msi_1    135616     Bus interrupt level                                False
intr_priority 3          Interrupt priority                                 False
lg_term_dma   0x800000   Long term DMA                                      True
max_xfer_size 0x100000   Maximum Transfer Size                              True
num_cmd_elems 500        Maximum number of COMMANDS to queue to the adapter True
pref_alpa     0x1        Preferred AL_PA                                    True
sw_fc_class   2          FC Class for Fabric                                True
tme           no         Target Mode Enabled                                True

# lsattr -El fcs1
DIF_enabled   no         DIF (T10 protection) enabled                       True
bus_intr_lvl             Bus interrupt level                                False
bus_io_addr   0xffc00    Bus I/O address                                    False
bus_mem_addr  0xffe77000 Bus memory address                                 False
bus_mem_addr2 0xffe7c000 Bus memory address                                 False
init_link     auto       INIT Link flags                                    False
intr_msi_1    135617     Bus interrupt level                                False
intr_priority 3          Interrupt priority                                 False
lg_term_dma   0x800000   Long term DMA                                      True
max_xfer_size 0x100000   Maximum Transfer Size                              True
num_cmd_elems 500        Maximum number of COMMANDS to queue to the adapter True
pref_alpa     0x1        Preferred AL_PA                                    True
sw_fc_class   2          FC Class for Fabric                                True
tme           no         Target Mode Enabled                                True
```

The maximum number of commands AIX queues to the adapter and the transfer size can be changed with the **`chdev`** command. Example 5-12 shows how to change these settings. The system must be rebooted for the changes to take effect.

Example 5-12 The chdev command

```
# chdev -a 'num_cmd_elems=4096 max_xfer_size=0X200000' -l fcs0 -P
fcs0 changed

# chdev -a 'num_cmd_elems=4096 max_xfer_size=0X200000' -l fcs1 -P
fcs1 changed
```

The changes can be confirmed by running the **lsattr** command again (see Example 5-13).

Example 5-13 The lsattr command confirmation

```
#  lsattr -El fcs0
...
max_xfer_size 0X200000   Maximum Transfer Size                              True
num_cmd_elems 4096       Maximum number of COMMANDS to queue to the adapter True
...
# lsattr -El fcs1
...
max_xfer_size 0X200000   Maximum Transfer Size                              True
num_cmd_elems 4096       Maximum number of COMMANDS to queue to the adapter True
...
```

To check the disk queue depth, periodically run **iostat -D 5**. If the `avgwqsz` (average wait queue size) or `sqfull` is consistently greater than zero, increase the disk queue depth. The maximum disk queue depth is 256. However, do not start at 256 and work down because you might flood the storage system with commands and waste memory on the AIX server. For most environments, 64 is a good number. For more information about AIX disk queue depth tuning, see the following paper in the Techdocs Library:

http://www-01.ibm.com/support/docview.wss?uid=tss1td105745

The default disk behavior algorithm is `round_robin` with a queue depth of 40.

Example 5-14 shows how to adjust the disk behavior algorithm and queue depth setting. In the command, `<hdisk#>` stands for an instance of an hdisk.

Example 5-14 Changing disk behavior algorithm and queue depth command

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

If you want the `fail_over` disk behavior algorithm, load-balance the I/O across the FC adapters and paths. Set the path priority attribute for each LUN so that $1/n^{th}$ of the LUNs are assigned to each of the *n* FC paths.

Useful MPIO commands

The following commands are used to change priority attributes for paths that can specify a preference for the path that is used for I/O. The effect of the priority attribute depends on whether the disk behavior algorithm attribute is set to `fail_over` or `round_robin`:

- For **algorithm=fail_over**, the path with the higher priority value handles all the I/O. If a path failure occurs, the other path is used. After a path failure and recovery, I/O is redirected down the path with the highest priority if you have IY79741 installed.

  If you want the I/O to go down the primary path, use **chpath** to disable and then re-enable the secondary path. If the priority attribute is the same for all paths, the first path that is listed with **lspath -Hl *<hdisk>*** is the primary path. Set the primary path to priority value `1`, the next path's priority (in case of path failure) to 2, and so on.

- For **algorithm=round_robin**, I/O goes down each path equally if the priority attributes are the same. If you set `pathA` priority to 1 and `pathB` to 255, 255 IOs are sent down `pathB` for every I/O going down `pathA`.

To change the path priority of an MPIO device, use the **chpath** command. An example of this process is shown in Example 5-17 on page 203.

Initially, use the **lspath** command to display the operational status for the paths to the devices, as shown in Example 5-15.

Example 5-15 The lspath command shows the paths for hdisk2

```
# lspath -l hdisk2 -F status:name:parent:path_id:connection
Enabled:hdisk2:fscsi0:0:5001738027820170,1000000000000
Enabled:hdisk2:fscsi0:1:5001738027820160,1000000000000
Enabled:hdisk2:fscsi0:2:5001738027820150,1000000000000
Enabled:hdisk2:fscsi0:3:5001738027820140,1000000000000
Enabled:hdisk2:fscsi0:4:5001738027820180,1000000000000
Enabled:hdisk2:fscsi0:5:5001738027820190,1000000000000
Enabled:hdisk2:fscsi1:6:5001738027820162,1000000000000
Enabled:hdisk2:fscsi1:7:5001738027820152,1000000000000
Enabled:hdisk2:fscsi1:8:5001738027820142,1000000000000
Enabled:hdisk2:fscsi1:9:5001738027820172,1000000000000
Enabled:hdisk2:fscsi1:10:5001738027820182,1000000000000
Enabled:hdisk2:fscsi1:11:5001738027820192,1000000000000
```

The **lspath** command can also be used to read the attributes of a path to an MPIO-capable device (see Example 5-16). The *<connection>* information is either "*<SCSI ID>*, *<LUN ID>*" for SCSI (for example "5, 0") or "*<WWN>*, *<LUN ID>*" for FC devices (as in Example 5-16).

Example 5-16 The lspath command reads attributes of the 0 path for hdisk2

```
# lspath -AHE -l hdisk2 -p fscsi0 -w "5001738027820170,1000000000000"
attribute value              description  user_settable

scsi_id   0x20ac00           SCSI ID      False
node_name 0x5001738027820000 FC Node Name False
priority  1                  Priority     True
```

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

Example 5-17 shows the use of the **chpath** command with the storage system. The command sets the primary path to `fscsi0` using the first path listed. There are two paths from the switch to the storage for this adapter. For the next disk, set the priorities to 4, 1, 2, and 3. In failover mode, assume the workload is relatively balanced across the hdisks. This setting balances the workload evenly across the paths.

Example 5-17 The chpath command

```
# chpath -l hdisk2 -p fscsi0 -w "5001738027820160,1000000000000" -a priority=2
path Changed
# chpath -l hdisk2 -p fscsi1 -w "5001738027820162,1000000000000" -a priority=3
path Changed
# chpath -l hdisk2 -p fscsi1 -w "5001738027820152,1000000000000" -a priority=4
path Changed
```

The **rmpath** command unconfigures or undefines, or both, one or more paths to a target device. You cannot unconfigure (undefine) the last path to a target device by using the **rmpath** command. The only way to unconfigure (undefine) the last path to a target device is to unconfigure the device itself. Use the **rmdev** command to do so.

5.1.3 AIX host iSCSI configuration

To ensure that your AIX version is supported for iSCSI attachment (for iSCSI hardware or software initiator), check IBM SSIC:

https://www.ibm.com/systems/support/storage/ssic/interoperability.wss

At the time of this writing, AIX iSCSI attachment to FlashSystem A9000 and A9000R is *not* supported.

For iSCSI, no Host Attachment Kit is required. Make sure that your system is equipped with the required file sets by running the **lslpp** command (see Example 5-18).

Example 5-18 Verifying installed iSCSI file sets in AIX

```
# lslpp -la "*.iscsi*"
  Fileset                      Level  State      Description
  ----------------------------------------------------------------------------
Path: /usr/lib/objrepos
  devices.common.IBM.iscsi.rte
                             7.1.3.0  COMMITTED  Common iSCSI Files
                            7.1.3.15  COMMITTED  Common iSCSI Files
  devices.iscsi.disk.rte    7.1.0.15  COMMITTED  iSCSI Disk Software
  devices.iscsi.tape.rte     7.1.0.0  COMMITTED  iSCSI Tape Software
  devices.iscsi_sw.rte       7.1.3.0  COMMITTED  iSCSI Software Device Driver
                            7.1.3.15  COMMITTED  iSCSI Software Device Driver
                             7.1.4.0  COMMITTED  iSCSI Software Device Driver

Path: /etc/objrepos
  devices.common.IBM.iscsi.rte
                             7.1.3.0  COMMITTED  Common iSCSI Files
                            7.1.3.15  COMMITTED  Common iSCSI Files
  devices.iscsi_sw.rte       7.1.3.0  COMMITTED  iSCSI Software Device Driver
                            7.1.3.15  COMMITTED  iSCSI Software Device Driver
                             7.1.4.0  COMMITTED  iSCSI Software Device Driver
```

Current limitations when using iSCSI

The code available at the time of writing has the following limitations when you are using the iSCSI software initiator in AIX:

- iSCSI is supported through a single path. No MPIO support is provided.
- The **xiv_iscsi_admin** command does not discover new targets on AIX. You must manually add new targets.
- The **xiv_attach** wizard does not support iSCSI.

Volume Groups

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

- Configure Volume Groups that are created using iSCSI devices to be in an inactive state after reboot. After the iSCSI devices are configured, manually activate the iSCSI-backed Volume Groups. Then, mount any associated file systems.

  > **Restriction:** Volume Groups are activated during a different boot phase than the iSCSI software. For this reason, you cannot activate iSCSI Volume Groups during the boot process.

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

I/O failures

To avoid I/O failures, consider these guidelines:

- If connectivity to iSCSI target devices is lost, I/O failures occur. Before you do anything that causes longterm loss of connectivity to the active iSCSI targets, stop all I/O activity and unmount iSCSI-backed file systems.
- If a loss of connectivity occurs while applications are attempting I/O activities with iSCSI devices, I/O errors eventually occur. You might not be able to unmount iSCSI-backed file systems because the underlying iSCSI device remains busy.
- File system maintenance must be performed if I/O failures occur because of loss of connectivity to active iSCSI targets. To do file system maintenance, run the **fsck** command against the affected file systems.

Configuring the iSCSI software initiator and the server

To connect AIX to the storage system through iSCSI, complete the following steps:

1. Get the *iSCSI qualified name* (IQN) on the AIX server, and set the maximum number of targets by using the *System Management Interface Tool* (SMIT):

   a. Select **Devices**.

   b. Select **iSCSI**.

   c. Select **iSCSI Protocol Device**.

   d. Select **Change / Show Characteristics of an iSCSI Protocol Device**.

   e. Select the device and verify the iSCSI Initiator Name value. The Initiator Name value is used by the iSCSI Target during login.

   > **Tip:** A default initiator name is assigned when the software is installed. This initiator name can be changed to match local network naming conventions.

   You can also issue the **lsattr** command to verify the initiator_name parameter, as shown in Example 5-19.

   Example 5-19 Checking initiator name

```
# lsattr -El iscsi0
disc_filename  /etc/iscsi/targets
Configuration file                          False
disc_policy    file                                           Discovery
Policy                                 True
```

```
initiator_name iqn.com.ibm.de.mainz.p7-730-02v1.hostid.099b7683 iSCSI
Initiator Name                         True
isns_srvnames  auto                                          iSNS Servers
IP Addresses                    True
isns_srvports                                                iSNS Servers
Port Numbers                    True
max_targets    16                                            Maximum
Targets Allowed                      True
num_cmd_elems  200                                           Maximum
number of commands to queue to driver True
```

f. The Maximum Targets Allowed field corresponds to the maximum number of iSCSI targets that can be configured. If you reduce this number, you also reduce the amount of network memory pre-allocated for the iSCSI protocol during configuration.

2. Define the AIX server on storage system as described in “Defining a host” on page 35.
3. Create the LUNs in storage system and map them to the AIX iSCSI server as illustrated in “Mapping LUNs to a host” on page 36, starting from LUN 0. If you are not using LUN 0, you might see the warning that is shown in Example 5-20.

Example 5-20 Warning by not using LUN 0

```
# cfgmgr -l iscsi0
cfgmgr: 0514-621 WARNING: The following device packages are required for
        device support but are not currently installed.
devices.iscsi.array
```

4. Determine the iSCSI IP addresses in the storage system as described in 1.3.5, “Identifying iSCSI ports” on page 29.
5. Find the IQN of the storage system as described in “Getting the XIV iSCSI Qualified Name” on page 25.
6. Return to the AIX system and add the storage system iSCSI IP address, port name, and IQN to the `/etc/iscsi/targets` file. This file must include the iSCSI targets for the device configuration.

> **Tip:** The iSCSI `targets` file defines the name and location of the iSCSI targets that the iSCSI software initiator attempts to access. This file is read every time that the iSCSI software initiator is loaded.

Each uncommented line in the file represents an iSCSI target. iSCSI device configuration requires that the iSCSI targets can be reached through a properly configured network interface. Although the iSCSI software initiator can work using a 10/100 Ethernet LAN, it is designed for use with at least a separate gigabit Ethernet network.

Include your specific connection information in the `targets` file, as shown in Example 5-21.

Example 5-21 Inserting connection information into the /etc/iscsi/targets file in AIX

```
# cat /etc/iscsi/targets
...
9.155.120.20 3260 iqn.2005-10.com.xivstorage:01322131
```

7. Enter the following command at the AIX prompt:

   ```
   cfgmgr -l iscsi0
   ```

   This command runs the following actions:

   – Reconfigures the software initiator.

   – Causes the driver to attempt to communicate with the targets listed in the `/etc/iscsi/targets` file.

   – Defines a new hdisk for each LUN found on the targets.

8. Run the **`lsdev -Cc disk`** command to view the new iSCSI devices. Example 5-22 shows two iSCSI disks.

   Example 5-22 iSCSI confirmation

   ```
   # lsdev -Cc disk
   hdisk0 Available          Virtual SCSI Disk Drive
   hdisk1 Available 51-T1-01 MPIO 2810 XIV Disk
   hdisk2 Available 50-T1-01 MPIO 2810 XIV Disk
   hdisk3 Available          Other iSCSI Disk Drive
   hdisk4 Available          Other iSCSI Disk Drive
   ```

   > **Exception:** If the appropriate disks are not defined, review the configuration of the initiator, the target, and any iSCSI gateways to ensure correctness. Then, rerun the **`cfgmgr`** command.

iSCSI performance considerations

To ensure the best performance, enable the following features of the AIX Gigabit Ethernet Adapter and the iSCSI Target interface:

- TCP Large Send
- TCP send and receive flow control
- Jumbo frame

The first step is to confirm that the network adapter supports jumbo frames. *Jumbo frames* are Ethernet frames that support more than 1500 bytes. Jumbo frames can carry up to 9000 bytes of payload, but some care must be taken when using the term. Many different Gigabit Ethernet switches and Gigabit Ethernet network cards can support jumbo frames. Check the network card specification or the vendor's support website to confirm that the network card supports this function.

You can use **`lsattr`** to list some of the current adapter device driver settings. Enter **`lsattr -E -l`** ***`ent0`***, where ***`ent0`*** is the adapter name. Make sure that you are checking and modifying the correct adapter. A typical output is shown in Example 5-23.

Example 5-23 The lsattr output that displays adapter settings

```
# lsattr -E -l ent0
alt_addr       0x000000000000   Alternate ethernet address                 True
busintr        167              Bus interrupt level                        False
busmem         0xe8120000       Bus memory address                         False
chksum_offload yes              Enable hardware transmit and receive checksum  True
compat_mode    no               Gigabit Backward compatability             True
copy_bytes     2048             Copy packet if this many or less bytes     True
delay_open     no               Enable delay of open until link state is known True
failback       yes              Enable auto failback to primary            True
failback_delay 15               Failback to primary delay timer            True
```

```
failover       disable          Enable failover mode                          True
flow_ctrl      yes              Enable Transmit and Receive Flow Control      True
intr_priority  3                Interrupt priority                            False
intr_rate      10000            Max rate of interrupts generated by adapter   True
jumbo_frames   no               Transmit jumbo frames                         True
large_send     yes              Enable hardware TX TCP resegmentation         True
media_speed    Auto_Negotiation Media speed                                   True
rom_mem        0xe80c0000       ROM memory address                            False
rx_hog         1000             Max rcv buffers processed per rcv interrupt   True
rxbuf_pool_sz  2048             Rcv buffer pool, make 2X rxdesc_que_sz        True
rxdesc_que_sz  1024             Rcv descriptor queue size                     True
slih_hog       10               Max Interrupt events processed per interrupt  True
tx_que_sz      8192             Software transmit queue size                  True
txdesc_que_sz  512              TX descriptor queue size                      True
use_alt_addr   no               Enable alternate ethernet address             True
```

In the example, `jumbo_frames` are `off`. When this setting is not enabled, you cannot increase the network speed. Set up the `tcp_sendspace`, `tcp_recvspace`, `sb_max`, and `mtu_size` network adapter and network interface options to optimal values.

To see the current settings, use **`lsattr`** to list the settings for `tcp_sendspace`, `tcp_recvspace`, and `mtu_size` (see Example 5-24).

Example 5-24 The lsattr output that displays interface settings

```
lsattr -E -l en0
alias4                        IPv4 Alias including Subnet Mask          True
alias6                        IPv6 Alias including Prefix Length        True
arp           on              Address Resolution Protocol (ARP)         True
authority                     Authorized Users                          True
broadcast                     Broadcast Address                         True
mtu           1500            Maximum IP Packet Size for This Device    True
netaddr       9.155.87.120    Internet Address                          True
netaddr6                      IPv6 Internet Address                     True
netmask       255.255.255.0   Subnet Mask                               True
prefixlen                     Prefix Length for IPv6 Internet Address   True
remmtu        576             Maximum IP Packet Size for REMOTE Networks True
rfc1323                       Enable/Disable TCP RFC 1323 Window Scaling True
security      none            Security Level                            True
state         up              Current Interface Status                  True
tcp_mssdflt                   Set TCP Maximum Segment Size              True
tcp_nodelay                   Enable/Disable TCP_NODELAY Option         True
tcp_recvspace                 Set Socket Buffer Space for Receiving     True
tcp_sendspace                 Set Socket Buffer Space for Sending       True
```

Example 5-24 shows that all values are true, and that `mtu` is set to `1500`.

To change the `mtu` setting, enable `jumbo_frames` on the adapter. Issue the following command:

```
chdev -l ent0 -a jumbo_frames=yes -P
```

Reboot the server by entering **shutdown -Fr**. Check the interface and adapter settings and confirm the changes (see Example 5-25).

Example 5-25 The adapter settings after you make the changes

```
# lsattr -E -l ent0
...
jumbo_frames   yes              Transmit jumbo frames                          True
...
```

Example 5-26 shows that the `mtu` value was changed to 9000.

Example 5-26 The mtu value changed to 9000

```
# lsattr -E -l en0
...
mtu           9000          Maximum IP Packet Size for This Device     True
...
```

Use the **/usr/sbin/no -a** command to show the `sb_max`, `tcp_recvspace`, and `tcp_sendspace` values (see Example 5-27).

Example 5-27 Checking values by using the /usr/sbin/no -a command

```
# /usr/sbin/no -a
...
                    sb_max = 1048576
...
             tcp_recvspace = 16384
             tcp_sendspace = 16384
...
```

Check the following settings:

tcp_sendspace Specifies how much data the sending application can buffer in the kernel before the application is blocked on a send call.

tcp_recvspace Specifies how many bytes of data the receiving system can buffer in the kernel on the receiving sockets queue.

sb_max Sets an upper limit on the number of socket buffers queued to an individual socket. It therefore controls how much buffer space is used by buffers that are queued to a sender socket or receiver socket.

Use the following values for these settings:

- For `tcp_sendspace`, `tcp_recvspace`, and `sb_max`: The maximum transfer size of the iSCSI software initiator is 256 KB. Assuming that the system maximums for `tcp_sendspace` and `tcp_recvspace` are set to 262144 bytes, use the **ifconfig** command to configure a gigabit Ethernet interface by using the following command:

  ```
  ifconfig en0 9.155.87.120 tcp_sendspace 262144 tcp_recvspace 262144
  ```

- For `sb_max`: Set this network option to at least `524288`, and preferably `1048576`. The `sb_max` sets an upper limit on the number of socket buffers queued. Set this limit with the command **/usr/sbin/no -o sb_max=1048576**.

5.1.4 Management volume LUN 0

According to the SCSI standard, the storage system maps itself in every map to LUN 0 for inband FC management. This LUN serves as the “well known LUN” for that map. The host can then issue SCSI commands to that LUN that are not related to any specific volume. This device is displayed as a normal hdisk in the AIX operating system.

You might want to eliminate this management LUN on your system, or need to assign the LUN 0 number to a specific volume.

5.1.5 Host Attachment Kit utilities

The Host Attachment Kit includes the following useful utilities, as described in 1.1.3, “Host Attachment Kit” on page 5:

- `xiv_devlist`
- `xiv_diag`
- `xiv_attach`
- `xiv_fc_admin`
- `xiv_iscsi_admin` (`xiv_iscsi_admin` is *not* supported on AIX)
- `xiv_detach` (applicable to Windows Server only)

5.2 Boot from SAN in AIX

This section describes a SAN boot implementation for the IBM POWER System (formerly IBM System p) in an AIX v6.1 environment. Similar steps can be followed for other AIX environments.

When you use AIX SAN boot with FlashSystem A9000, A9000R, or XIV, the default MPIO is used. During the boot sequence, AIX uses the bootlist to find valid paths to a LUN or hdisk that contains a valid boot logical volume (hd5). However, a maximum of five paths can be defined in the bootlist, while the storage system multipathing setup results in more than five paths to a hdisk.

A fully redundant configuration establishes 12 paths (see Figure 1-3 on page 10). When FlashSystem A9000, A9000R HyperSwap, or Hyper-Scale Mobility are used, ensure that the adapter bios targets (WWPNs) from both storage systems are defined.

For example, consider two hdisks (hdisk0 and hdisk1) that contain a valid boot logical volume, both having 12 paths to the storage system. To set the bootlist for hdisk0 and hdisk1, issue the following command:

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

The use of the **`bootlist`** command displays the list of boot devices, as shown in Example 5-28.

Example 5-28 Displaying the bootlist

```
# bootlist -m normal -o
hdisk0 blv=hd5 pathid=0
```

Example 5-28 shows that hdisk1 is not present in the bootlist. Therefore, the system cannot boot from hdisk1 if the paths to hdisk0 are lost.

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

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

Implement SAN boot with AIX by using one of the following methods:

- For a system with an already installed AIX operating system, mirror the `rootvg` volume to the SAN disk.
- For a new system, start the AIX installation from a bootable AIX CD installation package or use Network Installation Management (NIM).

The *mirroring* method is simpler to implement than using the NIM.

5.2.1 Creating a SAN boot disk by mirroring

The mirroring method requires that you have access to an AIX system that is up and running. Locate an available system where you can install AIX on an internal SCSI disk.

To create a boot disk on the system, complete the following steps:

1. Select a logical drive that is the same size or larger than the size of `rootvg` currently on the internal SCSI disk. Verify that your AIX system can see the new disk with the **lspv -L** command, as shown in Example 5-29.

Example 5-29 The lspv command

```
# lspv -L
hdisk0           00cc6de1b1d84ec9                      rootvg           active
hdisk1           none                                  None
hdisk2           none                                  None
hdisk3           none                                  None
hdisk4           none                                  None
hdisk5           00cc6de1cfb8ea41                      None
```

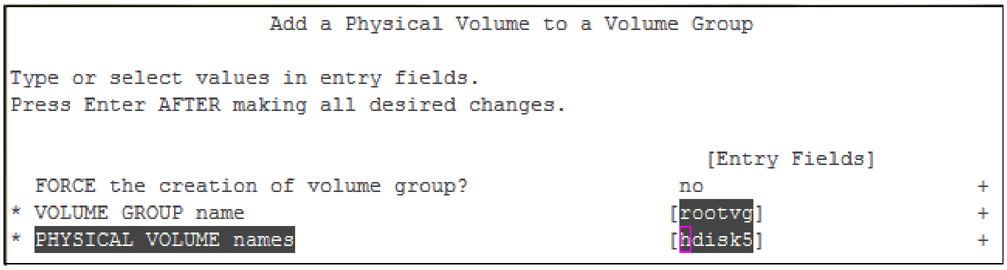
2. Verify the size with the **xiv_devlist** command to make sure that you are using an (external) disk. Example 5-30 shows that hdisk0 is 32 GB, hdisks 1 - 5 are attached, and they are external LUNs. Notice that hdisk1 is only 17 GB, so it is not large enough to create a mirror.

Example 5-30 The xiv_devlist command

```
# ./xiv_devlist
                                                                                XIV
Devices
-----------------------------------------------------------------------------
Device       Size (GB)  Paths  Vol Name       Vol Id   XIV Id   XIV Host
-----------------------------------------------------------------------------
/dev/hdisk1  17.2       2/2    ITSO_Anthony_  1018     1310114  AIX_P570_2_lp
                               Blade1_Iomete                    ar2
                               r
-----------------------------------------------------------------------------
/dev/hdisk2  1032.5     2/2    CUS_Jake       230      1310114  AIX_P570_2_lp
                                                                ar2
-----------------------------------------------------------------------------
/dev/hdisk3  34.4       2/2    CUS_Lisa_143   232      1310114  AIX_P570_2_lp
                                                                ar2
-----------------------------------------------------------------------------
/dev/hdisk4  1032.5     2/2    CUS_Zach       231      1310114  AIX_P570_2_lp
                                                                ar2
```

```
------------------------------------------------------------------------------
/dev/hdisk5  32.2       2/2      LPAR2_boot_mi  7378     1310114  AIX_P570_2_lp
                                 rror                              ar2
------------------------------------------------------------------------------


Non-XIV Devices
-----------------------------
Device       Size (GB)  Paths
-----------------------------
/dev/hdisk0  32.2       1/1
-----------------------------
```

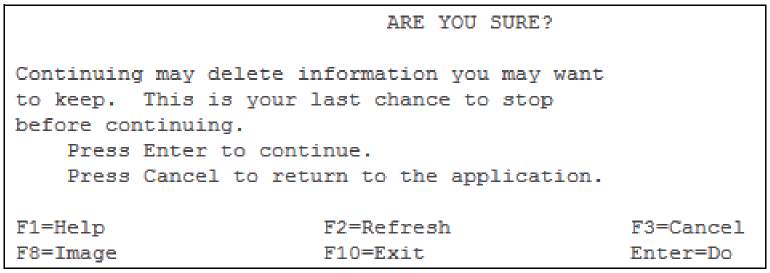
3. Add the new disk to the rootvg volume group by clicking **smitty vg →Set Characteristics of a Volume Group →Add a Physical Volume to a Volume Group**.
4. Keep the `Force the creation of volume group` value set to `no`.
5. Enter the Volume Group name (in this example, `rootvg`) and Physical Volume name that you want to add to the volume group (see Figure 5-1).



```
                      Add a Physical Volume to a Volume Group

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

                                                       [Entry Fields]
  FORCE the creation of volume group?                  no                  +
* VOLUME GROUP name                                   [rootvg]             +
* PHYSICAL VOLUME names                               [hdisk5]             +
```

Figure 5-1 Adding the disk to the rootvg

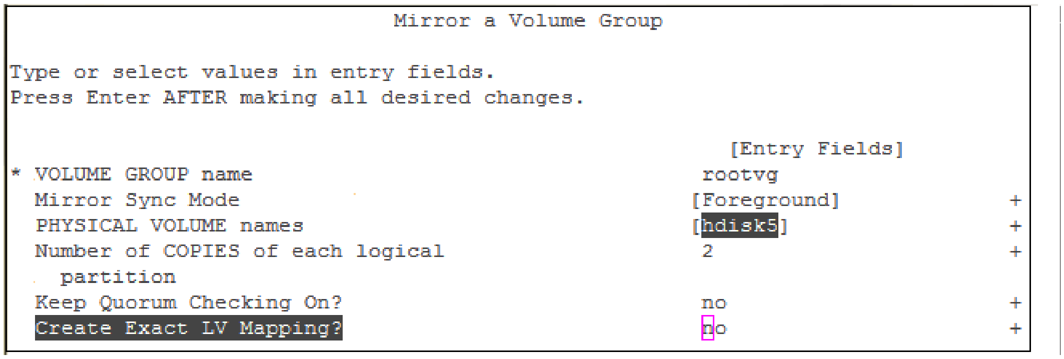
Figure 5-2 shows the settings confirmation.



```
                              ARE YOU SURE?

Continuing may delete information you may want
to keep.  This is your last chance to stop
before continuing.
    Press Enter to continue.
    Press Cancel to return to the application.

F1=Help                  F2=Refresh               F3=Cancel
F8=Image                 F10=Exit                 Enter=Do
```

Figure 5-2 Adding disk confirmation

6. Create the mirror of `rootvg`. If the `rootvg` is already mirrored, create a third copy on the new disk by clicking **smitty vg** →**Mirror a Volume Group** (see Figure 5-3).



```
                              Mirror a Volume Group

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

                                                          [Entry Fields]
* VOLUME GROUP name                                     rootvg
  Mirror Sync Mode                                     [Foreground]            +
  PHYSICAL VOLUME names                                [hdisk5]                +
  Number of COPIES of each logical                      2                      +
    partition
  Keep Quorum Checking On?                              no                     +
  Create Exact LV Mapping?                              no                     +
```

Figure 5-3 Creating a rootvg mirror

Enter the volume group name that you want to mirror (`rootvg`, in this example).

7. Select one of the following mirror sync modes:

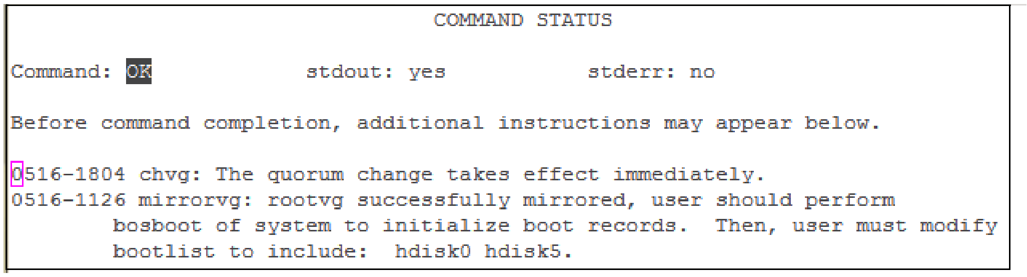
   – **Foreground**: This option causes the command to run until the mirror copy synchronization completes. The synchronization can take a long time. The amount of time depends mainly on the speed of your network and how much data you have.

   – **Background**: This option causes the command to complete immediately, and mirror copy synchronization occurs in the background. With this option, it is not obvious when the mirrors complete their synchronization.

   – **No Sync**: This option causes the command to complete immediately without running any type of mirror synchronization. If this option is used, the new remote mirror copy exists but is marked as stale until it is synchronized with the **`syncvg`** command.

8. Select the Physical Volume name. You added this drive to your disk group (see Figure 5-1 on page 212). The number of copies of each logical volume is the number of physical partitions that are allocated for each logical partition. The value can be one to three. A value of two or three indicates a mirrored logical volume. Leave the Keep Quorum Checking on and Create Exact LV Mapping settings at `no`.

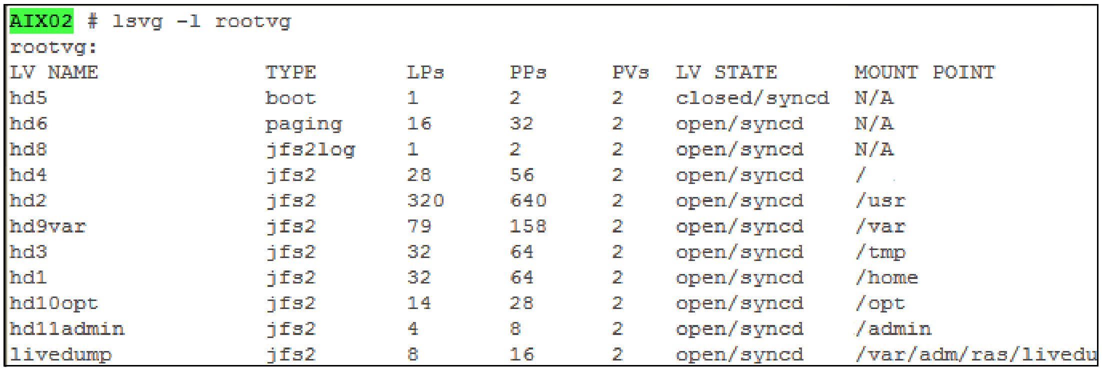
   After the volume is mirrored, you see confirmation that the mirror was successful, as shown in Figure 5-4.



```
                                 COMMAND STATUS

Command: OK            stdout: yes           stderr: no

Before command completion, additional instructions may appear below.

0516-1804 chvg: The quorum change takes effect immediately.
0516-1126 mirrorvg: rootvg successfully mirrored, user should perform
        bosboot of system to initialize boot records.  Then, user must modify
        bootlist to include:  hdisk0 hdisk5.
```

Figure 5-4 Mirror completed

9. Verify that all partitions are mirrored with **`lsvg -l rootvg`** (see Figure 5-5). The physical volume (PVs) column displays as two or three, depending on the number you chose when you created the mirror.



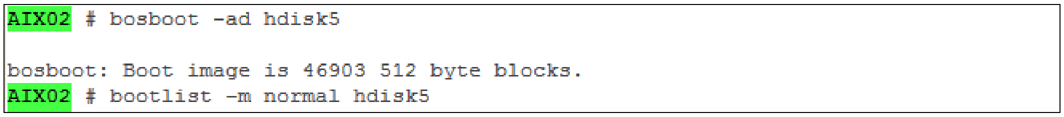
```
AIX02 # lsvg -l rootvg
rootvg:
LV NAME             TYPE       LPs     PPs     PVs  LV STATE      MOUNT POINT
hd5                 boot       1       2       2    closed/syncd  N/A
hd6                 paging     16      32      2    open/syncd    N/A
hd8                 jfs2log    1       2       2    open/syncd    N/A
hd4                 jfs2       28      56      2    open/syncd    /
hd2                 jfs2       320     640     2    open/syncd    /usr
hd9var              jfs2       79      158     2    open/syncd    /var
hd3                 jfs2       32      64      2    open/syncd    /tmp
hd1                 jfs2       32      64      2    open/syncd    /home
hd10opt             jfs2       14      28      2    open/syncd    /opt
hd11admin           jfs2       4       8       2    open/syncd    /admin
livedump            jfs2       8       16      2    open/syncd    /var/adm/ras/livedu
```

Figure 5-5 Verifying that all partitions are mirrored

10. Re-create the boot logical drive, and change the normal boot list with the following commands:

    ```
    bosboot -ad hdiskx
    bootlist -m normal hdiskx
    ```

    Figure 5-6 shows the output after you run the commands.



```
AIX02 # bosboot -ad hdisk5

bosboot: Boot image is 46903 512 byte blocks.
AIX02 # bootlist -m normal hdisk5
```

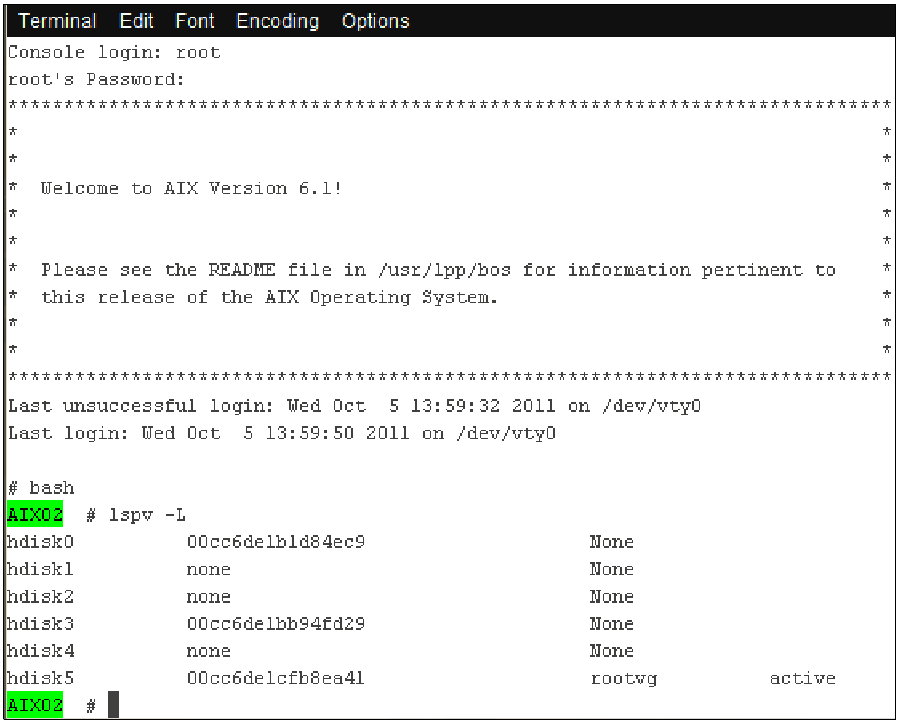
Figure 5-6 Relocating boot volume

11. Select the `rootvg` volume group and the original hdisk that you want to remove, then click **smitty vg →Unmirror a Volume Group**.
12. Select **rootvg** for the volume group name ROOTVG and the internal SCSI disk you want to remove.
13. Click **smitty vg →Set Characteristics of a Volume Group →Remove a Physical Volume from a Volume Group**.
14. Run the following commands again:

    ```
    bosboot -ad hdiskx
    bootlist -m normal hdiskx
    ```

At this stage, the creation of a bootable disk on the storage system is completed. Restarting the system makes it boot from the SAN (XIV) disk.

After the system reboots, use the **lspv -L** command to confirm that the server is booting from the hdisk (see Figure 5-7).



```
Console login: root
root's Password:
*******************************************************************************
*                                                                             *
*                                                                             *
*  Welcome to AIX Version 6.1!                                                *
*                                                                             *
*                                                                             *
*  Please see the README file in /usr/lpp/bos for information pertinent to    *
*  this release of the AIX Operating System.                                  *
*                                                                             *
*                                                                             *
*******************************************************************************
Last unsuccessful login: Wed Oct  5 13:59:32 2011 on /dev/vty0
Last login: Wed Oct  5 13:59:50 2011 on /dev/vty0

# bash
AIX02  # lspv -L
hdisk0          00cc6de1b1d84ec9                    None
hdisk1          none                                None
hdisk2          none                                None
hdisk3          00cc6de1bb94fd29                    None
hdisk4          none                                None
hdisk5          00cc6de1cfb8ea41                    rootvg          active
AIX02  #
```

Figure 5-7 SAN boot disk confirmation

5.2.2 Installation on external storage from bootable AIX CD-ROM

To install AIX on storage system disks, complete the following preparations:

1. Update the FC adapter (HBA) microcode to the latest supported level.
2. Make sure that you have an appropriate SAN configuration and the host is properly connected to the SAN.
3. Make sure that the zoning configuration is updated and at least one LUN is mapped to the host.

> **Tip:** If the system cannot see the SAN fabric at login, configure the HBAs at the server open firmware prompt.

Because a SAN allows access to many devices, identifying the hdisk to install to can be difficult. Use the following method to facilitate the discovery of the lun_id to hdisk correlation:

1. If possible, zone the switch or disk array such that the system being installed can discover only the disks to be installed to. After the installation completes, you can reopen the zoning so the system can discover all necessary devices.
2. If more than one disk is assigned to the host, make sure that you are using one of the following methods:
   - Assign Physical Volume Identifiers (PVIDs) to all disks from an installed AIX system that can access the disks. Assign PVIDS by using the following command (where X is the appropriate disk number):

     ```
     chdev -a pv=yes -l hdiskX
     ```

Create a table mapping PVIDs to physical disks. Make the PVIDs visible in the installation menus by selecting option **77 display more disk info**. You can also use the PVIDs to do an unprompted NIM installation.

- Another way to ensure that the selection of the correct disk is to use Object Data Manager (ODM) commands:

  i. Boot from the AIX installation CD-ROM.

  ii. From the main installation menu, click **Start Maintenance Mode for System Recovery →Access Advanced Maintenance Functions →Enter the Limited Function Maintenance Shell**.

  iii. At the prompt, issue one of the following commands:

  ```
  odmget -q "attribute=lun_id AND value=OxNN..N" CuAt
  odmget -q "attribute=lun_id" CuAt (list every stanza with lun_id
  attribute)
  ```

  In the command, `OxNN..N` is the `lun_id` that you are looking for. This command prints the ODM stanzas for the hdisks that have that `lun_id`.

  iv. Enter `Exit` to return to the installation menus.

The Open Firmware implementation can boot from only lun_ids 0 - 7. The firmware on the FC adapter (HBA) promotes this `lun_id` to an 8-byte FC LUN ID. The firmware does this promotion by adding a byte of zeros to the front and 6 bytes of zeros to the end. For example, lun_id 2 becomes 0x0002000000000000. The lun_id is normally displayed without the leading zeros. Be careful when you are installing because the procedure allows installation to lun_ids outside of this range.

To install on external storage, complete the following steps:

1. Insert an AIX CD that has a bootable image into the CD-ROM drive.
2. Select **CD-ROM** as the installation device to make the system boots from the CD. The way to change the bootlist varies by model. In most System p models, you use the system management services (SMS) menu. For more information, see the user's guide for your model.
3. Allow the system to boot from the AIX CD image after you leave the SMS menu.
4. After a few minutes, the console displays a window that directs you to press a key to use the device as the system console.
5. A window prompts you to select an installation language.
6. The Welcome to the Base Operating System Installation and Maintenance window is displayed. Change the installation and system settings for this system to select a FC-attached disk as a target disk. Enter `2` to continue.
7. On the Installation and Settings window, enter `1` to change the system settings and select the **New and Complete Overwrite** option.
8. On the Change (the destination) Disk window, select the FC disks that are mapped to your system. To see more information, enter `77` to display the detailed information window that includes the PVID. Enter `77` again to show WWPN and LUN ID information. Type the number, but do not press Enter, for each disk that you choose. Typing the number of a selected disk clears the device of any existing data. Be sure to include an storage system disk.
9. After you select the FC-attached disks, the Installation and Settings window is displayed with the selected disks. Verify the installation settings, and then enter `0` to begin the installation process.

> **Important:** Verify that you made the correct selection for root volume group. The existing data in the destination root volume group is deleted during Base Operating System (BOS) installation.

When the system reboots, a window displays the address of the device from which the system is reading the boot image.

5.2.3 AIX SAN installation with NIM

NIM is a client/server infrastructure and service that allows remote installation of the operating system. It manages software updates, and can be configured to install and update third-party applications. The NIM server and client file sets are part of the operating system. A separate NIM server must be configured to keep the configuration data and the installable product file sets.

Deploy the NIM environment, and ensure that the following configurations are completed:

- The NIM server is properly configured as the NIM master and the basic NIM resources are defined.
- The FC adapters are already installed on the system onto which AIX is to be installed.
- The FC adapters are connected to a SAN, and on the storage system have at least one logical volume (LUN) mapped to the host.
- The target system (NIM client) currently has no operating system installed, and is configured to boot from the NIM server.

For more information about how to configure a NIM server, see the *NIM Setup Guide*:

`http://www.ibm.com/support/docview.wss?uid=isg3T1010383`

Before the installation, modify the `bosinst.data` file, where the installation control is stored. Insert your appropriate values at the following stanza:

```
SAN_DISKID
```

This stanza specifies the worldwide port name and a logical unit ID for FC-attached disks. The worldwide port name and logical unit ID are in the format that is returned by the **`lsattr`** command (that is, 0x followed by 1–16 hexadecimal digits). The `ww_name` and `lun_id` are separated by two slashes (//):

```
SAN_DISKID = <worldwide_portname//lun_id>
```

Here is an example:

```
SAN_DISKID = 0x0123456789FEDCBA//0x2000000000000
```

Or, you can specify PVID (this example is with internal disk):

```
target_disk_data:
PVID = 000c224a004a07fa
SAN_DISKID =
CONNECTION = scsi0//10,0
LOCATION = 10-60-00-10,0
SIZE_MB = 34715
HDISKNAME = hdisk0
```

To install AIX SAN with NIM, complete the following steps:

1. Enter the `#` **`smit nim_bosinst`** command.

2. Select the **lpp_source** resource for the BOS installation.
3. Select the **SPOT** resource for the BOS installation.
4. Select the **BOSINST_DATA to use during installation** option, and select a `bosinst_data` resource that can run a non-prompted BOS installation.
5. Select the **RESOLV_CONF to use for network configuration** option, and select a `resolv_conf` resource.
6. Click the **Accept New License Agreements** option, and select **Yes**. Accept the default values for the remaining menu options.
7. Press Enter to confirm and begin the NIM client installation.
8. To check the status of the NIM client installation, enter the following command:

```
# lsnim -l va09
```

6

Clients connecting through VIOS

This chapter explains IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System connectivity through Virtual I/O Server (VIOS) to AIX. VIOS is a component of PowerVM that provides the ability for logical partitions (LPARs) that are VIOS clients to share resources.

> **Important:** The procedures and instructions that are given here are based on code that was available at the time of writing. For the latest support information and instructions, see IBM System Storage Interoperation Center (SSIC):
>
> `https://www.ibm.com/systems/support/storage/ssic/interoperability.wss`
>
> Host Attachment Kits can be downloaded from Fix Central:
>
> `http://www.ibm.com/support/fixcentral/`

This chapter includes the following topics:

- 6.1, “IBM PowerVM overview” on page 220
- 6.2, “Virtual I/O Server” on page 221
- 6.3, “Node Port ID Virtualization” on page 222
- 6.4, “General guidelines” on page 223

6.1 IBM PowerVM overview

Virtualization on IBM Power Systems servers provides a rapid and cost-effective response to many business needs. Virtualization capabilities have become an important element in planning for IT floor space and servers. Growing commercial and environmental concerns create pressure to reduce the power footprint of servers. With the escalating cost of powering and cooling servers, consolidation and efficient utilization of the servers is becoming critical.

Virtualization on Power Systems servers allows an efficient utilization of servers by reducing the following needs:

- Server management and administration costs because there are fewer physical servers
- Power and cooling costs with increased utilization of existing servers
- Time to market because virtual resources can be deployed immediately

IBM PowerVM is a virtualization technology for AIX, IBM i, and Linux environments on IBM POWER® processor-based systems. It is a special software appliance that is tied to IBM Power Systems, which are the converged IBM i and IBM p server platforms. It is licensed on a POWER processor basis.

PowerVM offers a secure virtualization environment with the following features and benefits:

- Consolidates diverse sets of applications that are built for multiple operating systems (AIX, IBM i, and Linux) on a single server.
- Virtualizes processor, memory, and I/O resources to increase asset utilization and reduce infrastructure costs.
- Dynamically adjusts server capability to meet changing workload demands.
- Moves running workloads between servers to maximize availability and avoid planned downtime.

Virtualization technology is offered in three editions on Power Systems:

- PowerVM Express Edition
- PowerVM Standard Edition
- PowerVM Enterprise Edition

PowerVM provides logical partitioning technology by using the following features:

- Either the Hardware Management Console (HMC) or the Integrated Virtualization Manager (IVM)
- Dynamic logical partition (LPAR) operations
- IBM Micro-Partitioning® and VIOS capabilities
- N_Port ID Virtualization (NPIV)

6.1.1 PowerVM Express Edition

PowerVM Express Edition is available only on some IBM Power Systems servers and on all IBM BladeCenter blade servers that have Power Architecture technology. It is designed for clients who are looking for an introduction to more advanced virtualization features.

With PowerVM Express Edition, you can create up to three partitions on a server (two client partitions and one for the VIOS and IVM) that use virtual Small Computer System Interface (SCSI), NPIV, and shared processors, also called Micro-Partitioning technology.

To create more than two client logical partitions that use shared processors, virtual SCSI, or NPIV, you must purchase either the Standard Edition or the Enterprise Edition and enter the activation code.

All virtualization features can be managed by using the IVM, including the following examples:

- Micro-Partitioning
- Shared processor pool
- VIOS
- PowerVM LX86
- Shared dedicated capacity
- NPIV
- Virtual tape

6.1.2 PowerVM Standard Edition

For clients who are ready to gain the full value from their server, IBM offers the PowerVM Standard Edition. This edition provides the most complete virtualization functionality for UNIX and Linux in the industry, and is available for all IBM Power Systems servers.

With PowerVM Standard Edition, you can create up to 254 partitions on a server. You can use virtualized disk and optical devices, and try out the shared processor pool. All virtualization features can be managed by using an HMC or the IVM. These features include Micro-Partitioning, shared processor pool, Virtual I/O Server, PowerVM Lx86, shared dedicated capacity, NPIV, and virtual tape.

6.1.3 PowerVM Enterprise Edition

PowerVM Enterprise Edition is offered starting with IBM POWER6® servers. It includes all the features of the PowerVM Standard Edition, plus the PowerVM Live Partition Mobility and the IBM Active Memory™ Sharing capability.

With PowerVM Live Partition Mobility, you can move a running partition from one IBM POWER 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.

The PowerVM Active Memory Sharing technology enables selected logical partitions to share memory from a single pool of physical memory. A new level of abstraction managed by the hypervisor supports the Active Memory Sharing technology.

6.2 Virtual I/O Server

Virtual I/O Server (VIOS) is virtualization software that runs in a separate partition of the POWER system. VIOS provides virtual storage and networking resources to one or more client partitions.

VIOS owns physical I/O resources such as Ethernet and SCSI/FC adapters. It virtualizes those resources for its client LPARs to share them remotely using the built-in hypervisor services. These client LPARs can be created quickly, and typically own only real memory and shares of processors without any physical disks or physical Ethernet adapters.

With Virtual SCSI support, VIOS client partitions can share disk storage that is physically assigned to the VIOS LPAR. VIOS owns the physical adapters, such as the Fibre Channel storage adapters, that are connected to the storage system. The logical unit numbers (LUNs) of the physical storage devices that are detected by VIOS are mapped to VIOS virtual SCSI (VSCSI) server adapters. The VSCSI adapters are created as part of its partition profile.

The client partition connects to the VIOS VSCSI server adapters by using the hypervisor. The corresponding VSCSI client is adapters that are defined in its partition profile. VIOS runs SCSI emulation and acts as the SCSI target for the operating system.

Figure 6-1 shows an example of the VIOS owning the physical disk devices, and their virtual SCSI connections to two client partitions.

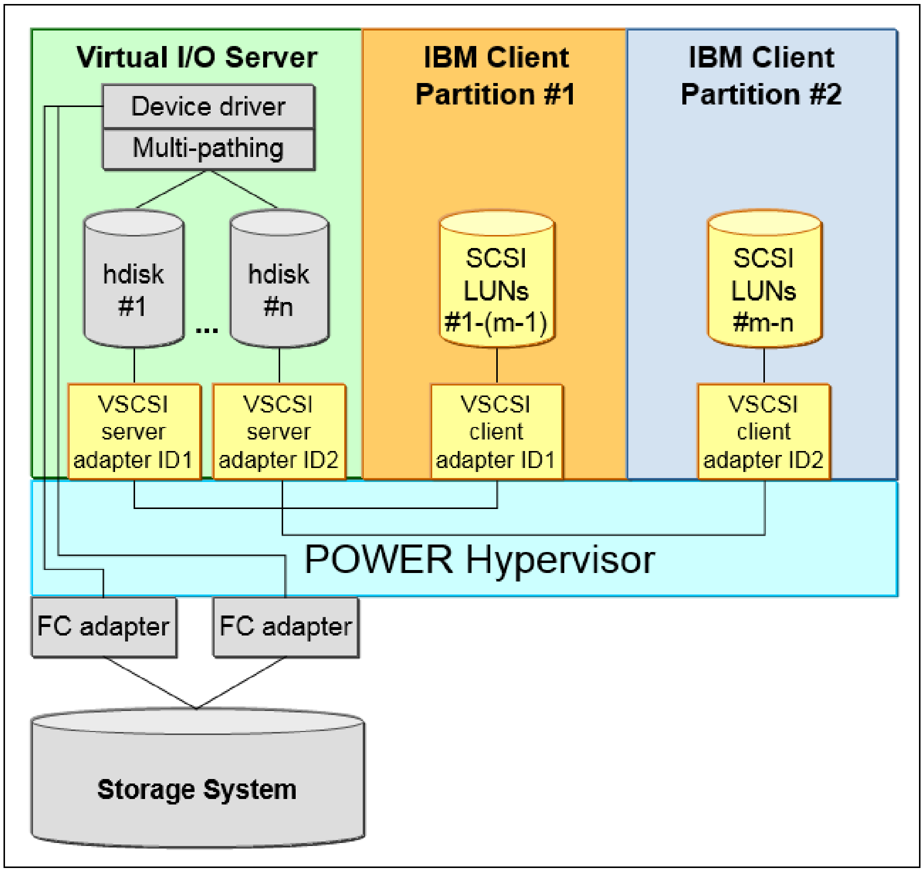


Figure 6-1 VIOS virtual SCSI support

6.3 Node Port ID Virtualization

The VIOS technology has been enhanced to boost the flexibility of IBM Power Systems servers with support for NPIV. NPIV simplifies the management and improves performance of Fibre Channel SAN environments. It does so by standardizing a method for Fibre Channel ports to virtualize a physical node port ID into multiple virtual node port IDs. The VIOS takes advantage of this feature, and can export the virtual node port IDs to multiple virtual clients. The virtual clients see this node port ID and can discover devices as though the physical port was attached to the virtual client.

The VIOS does not do any device discovery on ports that use NPIV. Therefore, no devices are shown in the VIOS connected to NPIV adapters. The discovery is left for the virtual client, and all the devices that are found during discovery are detected only by the virtual client. This way, the virtual client can use FC SAN storage-specific multipathing software on the client to discover and manage devices.

For more information about PowerVM virtualization management, see *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590:

http://www.redbooks.ibm.com/abstracts/sg247590.html

6.4 General guidelines

This section presents general guidelines for the storage system that is connected to a host. With the grid architecture and massive parallelism inherent to FlashSystem A9000, A9000R, and XIV Storage System, the general approach is to always maximize the use of all storage system resources.

6.4.1 Physical Fibre Channel adapters and virtual SCSI adapters

You can connect up to 4,095 LUNs per target, and up to 510 targets per port on a VIOS physical FC adapter. You can assign up to 16 LUNs to one VSCSI adapter. Therefore, you can use the number of LUNs to determine the number of virtual adapters that you need.

6.4.2 Multipath with two Virtual I/O Servers

The storage system is connected to an IBM client partition through the VIOS. For redundancy, connect the storage system to an IBM client with at least two VIOS partitions, independent from using vSCSI or NPIV.

6.4.3 Distributing connectivity

The goal for host connectivity is to create a balance of the resources in the storage system. Balance is achieved by distributing the physical connections across the interface modules or grid controllers. A host usually manages multiple physical connections to the storage device for redundancy purposes by using at least one SAN connected switch. The ideal is to distribute these connections across each of the interface modules. This way, the host uses the full resources of each module to which it connects for maximum performance.

You do not need to connect each host instance to each interface module. However, when the host has more than one physical connection, have the connections (cabling) spread across separate interface modules.

Similarly, if multiple hosts have multiple connections, you must distribute the connections evenly across the interface modules.

6.4.4 Zoning SAN switches

To maximize balancing and distribution of host connections to the storage system, create zones for the SAN switches. In these zones, have each host adapter connected to each interface module or grid controller and through redundant SAN switches. For more information, see for XIV 1.2.2, “Fibre Channel configurations” on page 9 and 1.2.3, “Zoning” on page 13, and for FlashSystem A9000 and A9000R see 2.2.2, “Fibre Channel configurations” on page 51 and 2.2.3, “Zoning” on page 56.

7

VMware connectivity

IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System are excellent choices for your VMware storage requirements. They achieve consistent high performance by balancing the workload across physical resources. This chapter addresses operating system-specific general connectivity considerations for host connectivity of VMware ESX and ESXi servers.

This chapter includes the following topics:

Note: For more information, refer to the IBM Redbooks publication *Using the IBM Spectrum Accelerate Family in VMware Environments: IBM XIV, IBM FlashSystem A9000 and IBM FlashSystem A9000R, and IBM Spectrum Accelerate*, REDP-5425.

7.1 Integration concepts and implementation guidelines

This section is for IT decision makers, storage administrators, and VMware administrators. It offers an overview of FlashSystem A9000, A9000R, and XIV Storage System and VMware integration concepts, and general implementation guidelines.

Note: FlashSystem A9000 and A9000R support VMware ESXi servers version 5.0 or later.

At a fundamental level, the goal of both the storage system and VMware's storage features is to significantly reduce the complexity of deploying and managing storage resources. With the storage systems, storage administrators can provide consistent tier-1 storage performance and quick change-request cycles. This support is possible because they need perform little planning and maintenance to keep performance levels high and storage optimally provisioned.

The following underlying strategies are built into the vSphere storage framework to insulate administrators from complex storage management tasks, and non-optimal performance and capacity resource utilization:

- Make storage objects much larger and more scalable, reducing the number that must be managed by the administrator.
- Extend specific storage resource-awareness by attaching features and profiling attributes to the storage objects.
- Help administrators make the correct storage provisioning decision for each virtual machine or even fully automate the intelligent deployment of virtual machine storage.
- Remove many time-consuming and repetitive storage-related tasks, including the need for repetitive physical capacity provisioning.

Clearly, vCenter relies upon the storage subsystem to fully support several key integration features to effectively implement these strategies. Appropriately compatible storage is essential.

7.2 vSphere traditional storage architectural overview

The vSphere storage architecture, including physical and logical storage elements, is shown in Figure 7-1 on page 227.

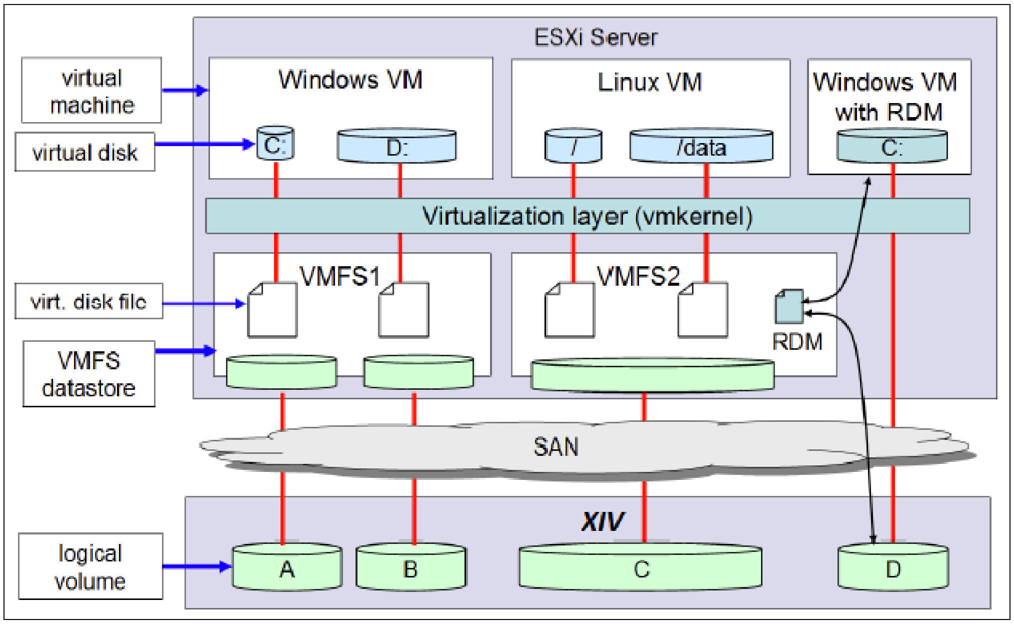


Figure 7-1 ESX/ESXi basic storage elements in the vSphere infrastructure

Although not intended to thoroughly explore vSphere storage concepts and terminology here, the essential components and their relationships provide the foundational framework necessary to understand upcoming integration principles.

The VMware file system (VMFS) is the central abstraction layer that acts as a medium between the storage and the hypervisor layers. The current generation of VMFS includes the following distinguishing attributes, among others:

- Clustered file system: Purpose-built, high performance clustered file system for storing virtual machine files on shared storage (Fibre Channel and iSCSI). The primary goal of the design of VMFS is as an abstraction layer between the VMs and the storage to efficiently pool and manage storage as a unified, multi-tenant resource.
- Shared data file system: Enable multiple vSphere hosts to read and write from the same data store concurrently.
- Online insertion or deletion of nodes: Add or remove vSphere hosts from VMFS volume with no impact to adjacent hosts or VMs.
- On-disk file locking: Ensure that the same virtual machine is not accessed by multiple vSphere hosts concurrently.

7.3 VMware vSphere Virtual Volumes (VVols)

Before the availability of vSphere Virtual Volumes, a virtual machine (VM) in a VMware environment would be presented a disk in the form of a file called a VMware disk (VMDK). This file represented a physical disk to the VM and then can be accessed by the operating system that is installed on the VM in the same way as a physical volume on a regular server.

The VMDK file was then placed onto the VMware file system (VMFS) hosted by a standard volume (LUN), for example implemented on external storage system such as FlashSystem A9000, A9000R, or XIV.

Although this design has the advantage of simplicity, it also imposes constraints and limitations on the management of the VM data. Indeed, the storage administrator and the VMware Administrator need to agree about the size and placement of volumes in the storage array before the deployment of VMs. This approach presents scalability and granularity issues, and cannot respond to the needs of businesses in a dynamic fashion. It also inhibits using advanced storage system functions such as instant snapshots and replication, and complicates backup solutions.

With the availability of the vSphere Virtual Volumes technology, each VM disk can now be mapped to an external storage volume.

> **Note:** With VVols, the storage system becomes aware of individual VMDK files, and data operations such as snapshot and replication can be performed directly by the storage, at the VMDK level rather than the entire VMFS data store. At the time of writing, Virtual Volumes are not supported with FlashSystem A9000 and A9000R.

Figure 7-2 shows how VVols changes the landscape of storage in a virtualized environment.

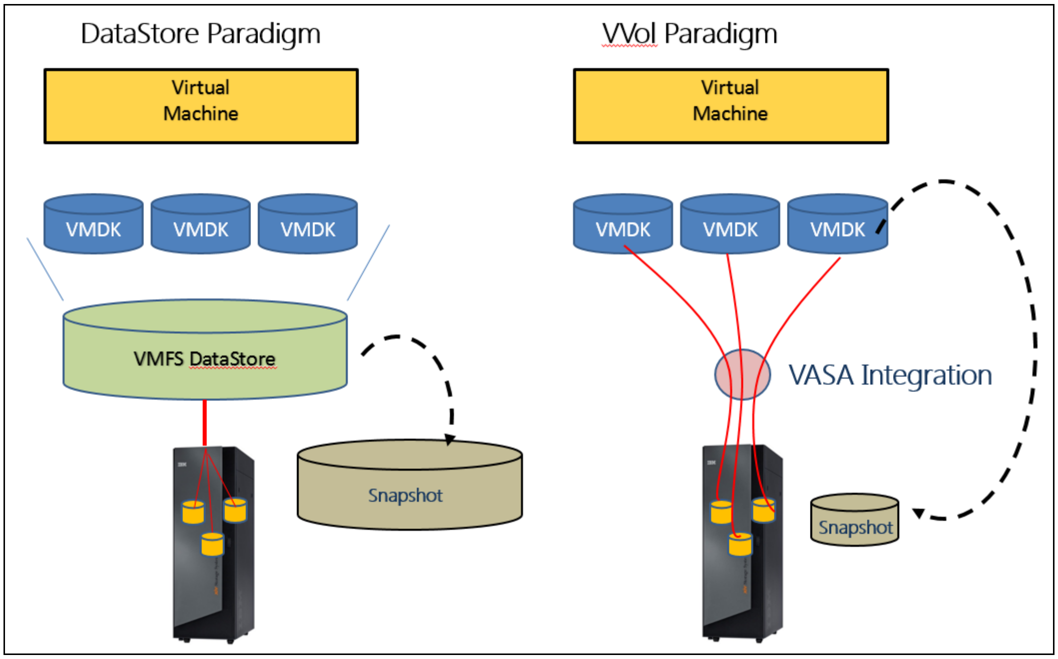


Figure 7-2 VMFS data store and VVols paradigms

7.4 VMware general connectivity guidelines

When you implement Fibre Channel connectivity for the storage system in a vSphere environment, adhere to the following practices:

- Use host cluster groups on storage system for LUN assignment for vSphere HA.
- Configure single initiator zones.

- At the time of this writing, VMware specifies that there can be a maximum of 1024 paths and 256 LUNs per ESX/ESXi host. The following conditions must be simultaneously satisfied to achieve the optimal storage configuration:
  - Effectively balance paths across the following objects:
    - Host HBA ports
    - XIV interface modules, FlashSystem A9000 or A9000R grid controllers
  - Ensure that the wanted minimum number of host paths per LUN and the wanted minimum number of LUNs per host can be simultaneously met by using the optimum number of paths per volume, as described in 1.2.2, “Fibre Channel configurations” on page 9 and 1.2.2, “Fibre Channel configurations” on page 9.
- Configure the Path Selection Plug-in (PSP) multipathing based on the vSphere version:
  - Use Round Robin policy if the vSphere version is vSphere 4.0 or later.
  - Use Fixed Path policy if the vSphere version is earlier than vSphere 4.0.
  - Do *not* use the Most Recently Used (MRU) policy.

When you implement iSCSI connectivity for the storage systems in a vSphere environment, adhere to the following practices:

- One VMkernel port group per physical network interface card (NIC):
  - VMkernel port is bound to physical NIC port in vSwitch, which creates a path.
  - Creates a 1-to-1 path for VMware NMP.
  - Uses the same PSP as for FC connectivity.
- Enable jumbo frames for throughput-intensive workloads (must be done at all layers).
- Use Round Robin PSP to enable load balancing across all modules. Each initiator should see a target port on each module.
- Queue depth can also be changed on the iSCSI software initiator. If more bandwidth is needed, the LUN queue depth can be modified.

Table 7-1 lists the maximum values for storage elements in VSphere 5.0, 5.1, 5.5, 6.0, and 6.5.

Table 7-1 Notable storage maximums in vSphere 5.0, 5.1, 5.5, 6.0, and 6.5.

Storage element limit	Maximum
Virtual disk size	2 TB minus 512 bytes
Virtual disks per host	2048
LUNs per Host	256 or 512[a]
Total number of paths per host	1024 or 2048[a]
Total number of paths to a LUN	32
LUN size	64 TB
Concurrent storage vMotions per datastore	8
Concurrent storage vMotions per host	2

[a] vSphere 6.5 only

Table 7-2 lists maximum values for virtual volumes in vSphere 6.0.

Table 7-2 Notable virtual volume maximums in vSphere 6.0

Storage element limit	Maximum
Data virtual volume size	62 TB
Number of virtual volumes bound to a host	64,000
Number of protocol endpoints (PEs) per host	256
Storage container size	2^64
Storage container per host	256
Configured VVol managed storage arrays per host	64

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 publications

The following Redbooks publications provide more information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *FlashSystem A9000 and A9000R, XIV and Spectrum Accelerate with IBM SAN Volume Controller Best Practices*, REDP-5408.
- *IBM FlashSystem A9000 and IBM FlashSystem A9000R Architecture and Implementation,* SG24-8345
- *Hyper-Scale Manager for IBM Spectrum Accelerate Family: XIV, FlashSystem A9000 and A9000R, IBM Spectrum Accelerate*, SG24-8376
- *IBM FlashSystem A9000 and A9000R Business Continuity Solutions*, REDP-5401
- *IBM HyperSwap for IBM FlashSystem A9000 and A9000R*, REDP-5434
- *IBM XIV Storage System Architecture and Implementation*, SG24-7659
- *IBM XIV Storage System Business Continuity Functions, SG24-7759*
- *Using IBM Spectrum Accelerate Family in VMware Environments*, REDP-5425

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System: Host Attachment and Interoperability

(0.2"spine)
0.17"<->0.473"
90<->249 pages

SG24-8368-01

ISBN 0738457892

Printed in U.S.A.