iSCSI SAN Configuration Guide

Update 2 and later for
ESX Server 3.5, ESX Server 3i version 3.5, VirtualCenter 2.5

vmware®

iSCSI SAN Configuration Guide
Revision: 20090313
Item: EN-000035-01

You can find the most up-to-date technical documentation on our Web site at:

http://www.vmware.com/support/

The VMware Web site also provides the latest product updates.

If you have comments about this documentation, submit your feedback to:

docfeedback@vmware.com

VMware, Inc.
3401 Hillview Ave.
Palo Alto, CA 94304
www.vmware.com

Contents

About This Book

This book, the *iSCSI SAN Configuration Guide*, explains how to use an ESX Server system with a storage area network (SAN). The manual discusses conceptual background, installation requirements, and management information in these topics:

- Using ESX Server with a SAN – Discusses requirements, differences in SAN setup if ESX Server is used, and how to manage the two systems together.

- Enabling your ESX Server system to boot from a LUN on a SAN – Discusses requirements, limitations, and management of booting from a SAN.

NOTE   For information on Fibre Channel (FC) or NFS storage devices, see the *Fibre Channel SAN Configuration Guide* and the *ESX Server Configuration Guide*.

The *iSCSI SAN Configuration Guide* covers both ESX Server 3.5 and ESX Server 3i version 3.5. For ease of discussion, this book uses the following product naming conventions:

- For topics specific to ESX Server 3.5, this book uses the term "ESX Server 3."

- For topics specific to ESX Server 3i version 3.5, this book uses the term "ESX Server 3i."

- For topics common to both products, this book uses the term "ESX Server."

- When the identification of a specific release is important to a discussion, this book refers to the product by its full, versioned name.

When a discussion applies to all versions of ESX Server for VMware Infrastructure 3, this book uses the term "ESX Server 3.x."

Intended Audience

The information presented in this manual is written for experienced Windows or Linux system administrators who are familiar with virtual machine technology datacenter operations.

Document Feedback

VMware welcomes your suggestions for improving our documentation. If you have comments, send your feedback to:

docfeedback@vmware.com

VMware Infrastructure Documentation

The VMware Infrastructure documentation consists of the combined VMware VirtualCenter and ESX Server documentation set.

Abbreviations Used in Figures

The figures in this book use the abbreviations listed in Table 1.

Table 1. Abbreviations

Abbreviation	Description
database	VirtualCenter database
datastore	Storage for the managed host
dsk#	Storage disk for the managed host
host*n*	VirtualCenter managed hosts
SAN	Storage area network type datastore shared between managed hosts
tmplt	Template
user#	User with access permissions
VC	VirtualCenter
VM#	Virtual machines on a managed host

Technical Support and Education Resources

The following sections describe the technical support resources available to you. To access the current versions of this book and other books, go to:

http://www.vmware.com/support/pubs

Online and Telephone Support

Use online support to submit technical support requests, view your product and contract information, and register your products. Go to:

http://www.vmware.com/support

Customers with appropriate support contracts should use telephone support for the fastest response on priority 1 issues. Go to:

http://www.vmware.com/support/phone_support.html

Support Offerings

Find out how VMware support offerings can help meet your business needs. Go to:

http://www.vmware.com/support/services

VMware Education Services

VMware courses offer extensive hands-on labs, case study examples, and course materials designed to be used as on-the-job reference tools. For more information about VMware Education Services, go to:

http://mylearn1.vmware.com/mgrreg/index.cfm

Using ESX Server with a Storage Area Network

1

You can use ESX Server in conjunction with a storage area network (SAN), a specialized high-speed network that connects computer systems to high-performance storage subsystems. Using ESX Server together with a SAN provides extra storage for consolidation, improves reliability, and helps with disaster recovery.

To use ESX Server effectively with a SAN, you must have a working knowledge of ESX Server systems and SAN concepts. Also, when you set up ESX Server hosts to use iSCSI SAN storage systems, special considerations are necessary. For more information about using ESX Server, see the *ESX Server Configuration Guide*.

This chapter discusses the following topics:

- "Understanding Virtualization" on page 12

- "Storage Area Network Concepts" on page 15

- "Overview of Using ESX Server with a SAN" on page 20

- "Basics of Using SAN Storage Systems with an ESX Server" on page 22

- "Understanding VMFS and SAN Storage Choices" on page 27

- "Understanding Data Access" on page 30

- "Path Management and Failover" on page 32

- "Choosing Virtual Machine Locations" on page 35

- "Designing for Server Failure" on page 36

Understanding Virtualization

The VMware virtualization layer is common across VMware desktop products (such as VMware Workstation) and server products (such as VMware ESX Server). This layer provides a consistent platform for development, testing, delivery, and support of application workloads and is organized as follows:

- Each virtual machine runs its own operating system (the guest operating system) and applications.

- The virtualization layer allows virtual devices to map to shares of specific physical devices. These devices include virtualized CPU, memory, I/O buses, network interfaces, storage adapters and devices, human interface devices, and BIOS.

Network Virtualization

The virtualization layer guarantees that each virtual machine is isolated from other virtual machines. Virtual machines can talk to each other only through networking mechanisms similar to those used to connect separate physical machines.

The isolation allows administrators to build internal firewalls or other network isolation environments, allowing some virtual machines to connect to the outside while others are connected only through virtual networks to other virtual machines.

For more information on networking concepts and virtual switches, see the *ESX Server Configuration Guide*.

Storage Virtualization

In an ESX Server environment, each virtual machine includes from one to four virtual SCSI initiators. These virtual adapters appear as either Buslogic or LSI Logic SCSI controllers. They are the only types of SCSI controllers that a virtual machine can access.

Each virtual disk that a virtual machine can access through one of the virtual SCSI initiators resides in the VMware Virtual Machine File System (VMFS) or on a raw disk.

Figure 1-1 gives an overview of storage virtualization. It illustrates storage using VMFS and storage using raw device mapping. It also shows how iSCSI storage is accessed through either iSCSI HBAs or by using a general-purpose NIC that uses iSCSI initiator software.

Figure 1-1. iSCSI SAN Storage Virtualization
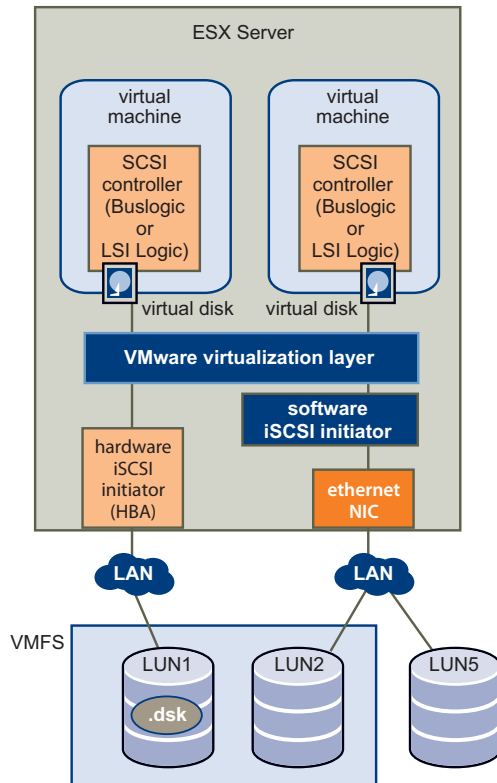
iSCSI Initiators

To access remote targets, your ESX Server host uses iSCSI initiators. Initiators transport SCSI requests and responses between the ESX Server system and the target storage device on the IP network.

ESX Server supports hardware-based and software-based iSCSI initiators:

Hardware iSCSI initiator.   A third-party host bus adapter (HBA) with iSCSI over TCP/IP capability. This specialized iSCSI adapter is responsible for all iSCSI processing and management.

Software iSCSI initiator. Code built into the VMkernel that allows an ESX Server to connect to the iSCSI storage device through standard network adapters. The software initiator handles iSCSI processing while communicating with the network adapter through the network stack. With the software initiator, you can use iSCSI technology without purchasing specialized hardware.

NOTE Guest operating systems in virtual machines cannot see iSCSI storage directly. To the guest operating systems, iSCSI storage attached to the ESX Server system appears to be available through a SCSI HBA.

The diagram in Figure 1-1 depicts two virtual machines that use different types of iSCSI initiators.

In the first example of iSCSI storage configuration, the ESX Server system uses the hardware iSCSI initiator. This specialized iSCSI initiator sends iSCSI packets to a disk over a LAN.

In the second example, the ESX Server system is configured with the software iSCSI initiator. Using the software initiator, the ESX Server system connects to a LAN through an existing NIC.

Disk Configuration Options

You can configure virtual machines with multiple virtual SCSI drives. For supported drivers, see the *Storage/SAN Compatibility Guide* at www.vmware.com/support/pubs/vi_pubs.html. The guest operating system can place limitations on the total number of SCSI drives.

Although all SCSI devices are presented as SCSI targets, physical implementation alternatives are available:

■ Virtual machine `.vmdk` file stored on a VMFS volume. See "Virtual Machine File System" on page 15.

■ Device mapping to a SAN logical unit (LUN). See "Raw Device Mapping" on page 15.

From the standpoint of the virtual machine, each virtual disk appears as if it were a SCSI drive connected to a SCSI initiator. Whether the actual physical disk device is being accessed through SCSI, iSCSI, or FC controllers is transparent to the guest operating system and to applications running on the virtual machine.

Virtual Machine File System

In a simple configuration, the virtual machines' disks are stored as files within a virtual machine file system (VMFS). When guest operating systems issue SCSI commands to their virtual disks, the virtualization layer translates these commands to VMFS file operations.

ESX Server systems use VMFS to store virtual machine files. To minimize disk I/O overhead, VMFS is optimized to run multiple virtual machines as one workload. VMFS also provides distributed locking for your virtual machine files, so that your virtual machines can operate safely in a SAN environment where multiple ESX Server hosts share a set of LUNs.

VMFS is first configured as part of the ESX Server installation. When you create a new VMFS-3 volume, it must be 1200MB or larger. See the *Installation Guide*. It can then be customized, as discussed in the *ESX Server Configuration Guide*.

A VMFS volume can be extended over 32 physical storage extents, including SAN LUNs and local storage. This allows pooling of storage and flexibility in creating the storage volume necessary for your virtual machine. With the new ESX3 Logical Volume Manager (LVM), you can extend a volume while virtual machines are running on the volume. This lets you add new space to your VMFS volumes as your virtual machine needs it.

Raw Device Mapping

A raw device mapping (RDM) is a special file in a VMFS volume that acts as a proxy for a raw device. The RDM provides some of the advantages of a virtual disk in the VMFS file system, while keeping some advantages of direct access to physical devices.

RDM might be required if you run SAN Snapshot or other layered applications in the virtual machine. RDMs better enable systems to use the hardware features inherent to SAN storage systems. See "Using Raw Device Mapping" in the *ESX Server Configuration Guide*.

Storage Area Network Concepts

If you are an ESX Server administrator planning to set up ESX Server hosts to work with SANs, you must have a working knowledge of SAN concepts. You can find information about SAN in print and on the Internet. Two web-based resources are:

■ www.searchstorage.com

■ www.snia.org

If you are new to SAN technology, read the following section to familiarize yourself with the basic terminology this document uses. To learn about basic SAN concepts, see the *SAN Conceptual and Design Basics* white paper at http://www.vmware.com/support/pubs.

This configuration guide discusses iSCSI SANs, which use Ethernet connections between computer systems, or host servers, and high-performance storage subsystems. The SAN components include host bus adapters (HBAs) or Network Interface Cards (NICs) in the host servers, switches and routers that transport the storage traffic, cables, storage processors (SPs), and storage disk systems.

To transfer traffic from host servers to shared storage, the SAN uses the iSCSI protocol that packages SCSI commands into iSCSI packets and transmits these on an Ethernet network.

Ports

In the context of this document, a port is the connection from a device into the SAN. Each node in the SAN, a host, storage device, and Ethernet switch has one or more ports that connect it to the SAN. Ports are identified in a number of ways:

IP address   Each iSCSI port has an IP address associated with it so that routing and switching equipment on your network can establish the connection between the server and storage. This is just like the IP address you assign to your computer to get access to your company's network or the Internet.

iSCSI Name   A unique name for identifying the port. The iSCSI name starts with either *iqn.* (for iSCSI qualified name) or *eui.* (for extended unique identifier). Multiple iSCSI devices can be present, with multiple iSCSI names, and can be connected through a single physical Ethernet port. iSCSI names are usually set to a unique default value, and do not need to be set by the user. An example of a VMware-generated iSCSI name is *iqn.1998-01.com.vmware:iscsitestox-68158ef2*.

iSCSI alias   A more manageable name for an iSCSI device or port, used instead of the iSCSI name. iSCSI aliases are not unique, and are intended to be just a "friendly" name to associate with a port. On an ESX Server system, the default iSCSI alias is the name of the system.

Multipathing and Path Failover

When transferring data between the host server and storage, the SAN uses a multipathing technique. Multipathing allows you to have more than one physical path from the ESX Server host to a LUN on a storage system.

If a path or any component along the path—HBA or NIC, cable, switch or switch port, or storage processor—fails, the server selects another of the available paths. The process of detecting a failed path and switching to another is called *path failover*.

Storage System Types

Storage disk systems can be of the following types:

- An active-active storage system, which allows access to the LUNs simultaneously through all the storage ports that are available without significant performance degradation. All the paths are active at all times (unless a path fails).

- An active-passive storage system, in which one port or an SP is actively providing access to a given LUN. The other ports or SPs act as backup for the LUN and can be actively providing access to other LUN I/O. I/O can be sent only to an active port. If access through the primary storage port fails, one of the secondary ports or storage processors becomes active, either automatically or through administrator intervention.

- A virtual port storage system, which allows access to all available volumes through a single virtual port. These are active-active storage devices, but hide their multiple connections though a single port. The ESX Server multipathing has no knowledge of the multiple connections to the storage. These storage systems handle port failover and connection balancing transparently. This is often referred to as "transparent failover."

Target Compared to LUN Representations

Different iSCSI storage vendors present storage to servers in different ways. Some vendors present multiple LUNs on a single target, while others present multiple targets with one LUN each (see Figure 1-2). While the way the storage is used by an ESX Server is similar, the way the information is presented through administrative tools is different.

Figure 1-2. Target compared to LUN Representations
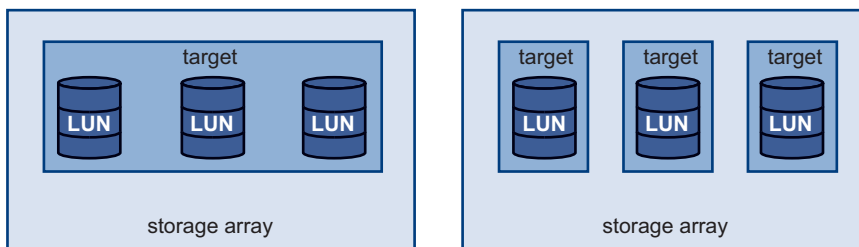
Three volumes (or LUNs, the SCSI term for logical unit, which means a portion of storage) are available in each of these configurations. In the first case, ESX Server sees one target (represented by an IQN name) but that target has three LUNs that can be used. Each of the LUNs represent individual storage volumes. In the second case, the ESX Server sees three different targets, represented by three separate IQN names. Each of these targets has one LUN.

Target numbering (though not IQN target naming) is decided by the ESX Server, so targets that are shared by different ESX Server systems might not have the same target number (vmhba1:2:3:4, where the 2 shows the target number position). Also, multiple paths to the same storage are represented by different target numbers.

LUN numbering is decided by the storage system, so LUN numbers across paths and multiple ESX Server systems always stay the same (vmhba1:2:3:4, where the 3 shows the LUN position). On storage systems with multiple targets, like the second example above, the LUN number is always zero (0).

ESX Server-based iSCSI initiators establish only one connection to each target. This means storage systems with a single target containing multiple LUNs have all LUN traffic on that one connection. With a system that has three targets with one LUN each, three connections exist between an ESX Server and the three volumes available. This information is useful when you are trying to aggregate storage traffic on multiple connections from an ESX Server with multiple iSCSI HBAs, where traffic for one target can be set to a particular HBA, while traffic for another target can use a different HBA.

This document uses the term *LUN* to mean a volume available from a storage system, even if that LUN is the only one available on an iSCSI target. This meaning is to avoid confusion with other uses of the word "volume" in this document. Although storage system vendors might not call their volumes LUNs, this is ultimately how they are represented to the ESX Server storage system.

iSCSI Naming Conventions

iSCSI uses a name to uniquely identify an iSCSI device, either target or initiator. This name is similar to the WorldWide Name associated with Fibre Channel devices, used as a way to universally identify the device.

iSCSI names are formatted two different ways. The first is by an iSCSI qualified name, commonly referred to as "an IQN name." The second, much less common method, is through an enterprise unique identifier, also referred to as "an EUI name."

iSCSI Qualified Names

iSCSI qualified names take the form `iqn.yyyy–mm.naming–authority:unique name`

yyyy-mm is the year and month when the naming authority was established.

naming-authority is usually reverse syntax of the Internet domain name of the naming authority. For example, the iscsi.vmware.com naming authority could have the iSCSI qualified name form of iqn.1998-01.com.vmware.iscsi. The vmware.com domain name was registered in January of 1998, and "iscsi" is a subdomain, maintained by vmware.com.

unique name the iscsi.vmware.com naming authority needs to ensure that any names assigned following the colon are unique, such as:

- iqn.1998-01.com.vmware.iscsi:name1

- iqn.1998-01.com.vmware.iscsi:name2

- iqn.1998-01.com.vmware.iscsi:name999

Enterprise Unique Identifiers

Enterprise unique identifiers take the form `eui.<16 hex digits>.`

For example, `eui.0123456789ABCDEF.`

The 16-hexadecimal digits are text representations of a 64-bit number of an IEEE EUI (extended unique identifier) format. The top 24 bits are a company ID that IEEE registers with a particular company. The lower 40 bits are assigned by the entity holding that company ID, and must be unique.

In many cases, the IQN format is chosen over the EUI format for readability and as a more user-friendly method of assigning names.

Overview of Using ESX Server with a SAN

Support for QLogic iSCSI HBAs and software-based iSCSI implementations allow an ESX Server system to be connected to iSCSI storage. You can then use iSCSI storage volumes to store virtual machine configuration information and application data. Using ESX Server with a SAN improves flexibility, efficiency, and reliability. It also supports centralized management as well as failover and load balancing technologies.

Benefits of Using ESX Server with a SAN

Using a SAN with ESX Server allows you to improve your environment's failure resilience:

■ You can store data redundantly and configure multiple Ethernet paths to your iSCSI storage, eliminating a single point of failure. Your enterprise is not crippled when one datacenter becomes unavailable.

■ ESX Server systems provide multipathing by default and automatically support it for every virtual machine. See "Path Management and Failover" on page 32.

■ Using a SAN with ESX Server systems extends failure resistance to the server. When you use SAN storage, all applications can instantly be restarted after host failure. See "Designing for Server Failure" on page 36.

Using ESX Server with a SAN makes high availability and automatic load balancing affordable for more applications than if dedicated hardware is used to provide standby services:

■ If virtual machines are used as standby systems for existing physical servers, shared storage is essential and a SAN is the best solution.

■ Use the VMware VMotion capabilities to migrate virtual machines seamlessly from one host to another.

■ Use VMware High Availability (HA) in conjunction with a SAN for a cold-standby solution that guarantees an immediate, automatic response.

■ Use VMware Distributed Resource Scheduler (DRS) to migrate virtual machines from one host to another for load balancing. Because storage is on a SAN storage system, applications continue running seamlessly.

■ If you use VMware DRS clusters, put an ESX Server host into maintenance mode to have the system migrate all running virtual machines to other ESX Server hosts. You can then perform upgrades or other maintenance operations.

The transportability and encapsulation of VMware virtual machines complements the shared nature of iSCSI storage. When virtual machines are located on SAN-based storage, you can shut down a virtual machine on one server and power it up on another server—or to suspend it on one server and resume operation on another server on the same network—in a matter of minutes. This allows you to migrate computing resources while maintaining consistent shared access.

Use Cases

Using ESX Server systems in conjunction with SAN is effective for the following tasks:

Maintenance with zero downtime.   When you perform maintenance, use VMware DRS or VMotion to migrate virtual machines to other servers. If shared storage is on the SAN, you can perform maintenance without interruptions to the user.

Load balancing. Use VMotion explicitly or use VMware DRS to migrate virtual machines to other hosts for load balancing. If shared storage is on a SAN, you can perform load balancing without interruption to the user.

Storage consolidation and simplification of storage layout.   If you are working with multiple hosts, and each host is running multiple virtual machines, the hosts' storage is no longer sufficient and external storage is needed. Choosing a SAN for external storage results in a simpler system architecture while giving you the other benefits listed in this section. You can start by reserving a large volume and then allocate portions to virtual machines as needed. Volume allocation and creation from the storage device needs to happen only once.

Disaster recovery.   Having all data stored on a SAN can greatly facilitate remote storage of data backups. In addition, you can restart virtual machines on remote ESX Server hosts for recovery if one site is compromised.

Finding SAN Configuration Information

In addition to this document, a number of other resources can help you configure your ESX Server system in conjunction with a SAN:

- Use your storage vendor's documentation for most setup questions. Your storage vendor might also offer documentation on using the storage system in an ESX Server environment.

- The *Fibre Channel SAN Configuration Guide* – Discusses the use of ESX Server with Fibre Channel storage area networks.

- The *VMware I/O Compatibility Guide* – Lists the currently approved HBAs, HBA drivers, and driver versions.

- The *VMware Storage/SAN Compatibility Guide* – Lists currently approved storage systems.

- The *VMware Release Notes* – Give information about known issues and workarounds.

- The *VMware Knowledge Base* – Has information on common issues and workarounds.

For more information, see the VMware Documentation Web site at http://www.vmware.com/support/pubs.

Basics of Using SAN Storage Systems with an ESX Server

Using a SAN in conjunction with an ESX Server host differs from traditional SAN usage in a variety of ways, discussed in this section.

Network Configuration and Authentication

Before volumes using software iSCSI can be discovered, the storage network must be configured and authentication might have to be set up.

- For software iSCSI, networking for the VMkernel needs to be configured. You can verify the network configuration by using the `vmkping` utility (see Appendix B, "Utilities," on page 123). For hardware iSCSI, network parameters, such as IP address, subnet mask, and default gateway must be configured on the HBA.

- Check and change the default initiator name if necessary.

- The discovery address of the storage system must be set and should be pingable using `vmkping`.

- For CHAP authentication, enable it on the initiator and the storage system side. After authentication is enabled, it applies for all of the targets that are not yet discovered, but does not apply to targets that are already discovered. After the discovery address is set, the new volumes discovered are exposed and can be used at that point.

Sharing a VMFS Across ESX Servers

ESX Server VMFS is designed for concurrent access from multiple physical machines and enforces the appropriate access controls on virtual machine files. For background information on VMFS, see "Virtual Machine File System" on page 15. For additional information, see the *ESX Server Configuration Guide*.

VMFS can:

- Coordinate access to virtual disk files. ESX Server uses file-level locks, which the VMFS distributed lock manager manages.

- Coordinate access to VMFS internal file system information (metadata).

NOTE SCSI reservations are held during metadata updates to the VMFS volume. ESX Server uses short-lived SCSI reservations as part of its distributed locking protocol.

Because virtual machines share a common VMFS, it might be difficult to characterize peak-access periods or to optimize performance. You need to plan virtual machine storage access for peak periods, but different applications might have different peak-access periods. The more virtual machines are sharing a VMFS, the greater the potential for performance degradation because of I/O contention.

NOTE VMware recommends that you load balance virtual machines over servers, CPU, and storage. Run a mix of virtual machines on each server so that not all experience high demand in the same area at the same time.

Figure 1-3 shows several ESX Server systems sharing the same VMFS volume.

Figure 1-3. Accessing Virtual Disk Files
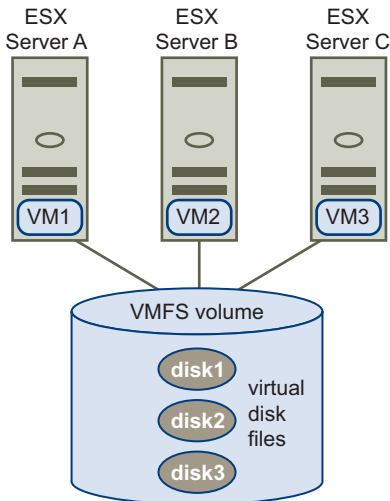
Metadata Updates

A VMFS holds files, directories, symbolic links, RDMs, and so on, along with corresponding metadata for these objects. Metadata is accessed each time the attributes of a file are accessed or modified. These operations include, but are not limited to:

- Creating, growing, or locking a file.

- Changing a file's attributes.

- Powering a virtual machine on or off.

Volume Display and Rescan

A SAN is dynamic, and which volumes are available to a certain host can change based on a number of factors including:

- New volumes created on the iSCSI storage

- Changes to volume access control

- Changes in network connectivity

The VMkernel discovers volumes when it boots and those volumes are then visible in VI Client. If changes are made to the volumes, you must rescan to see those changes.

> **CAUTION**   After you create a new VMFS volume or extend an existing VMFS volume, you must rescan the SAN storage from all ESX Server hosts that could see that particular volume (LUN). If this is not done, the shared volume might become invisible to some of those hosts.

Levels of Indirection

If you're used to working with traditional SANs, the levels of indirection can initially be confusing.

- You cannot directly access the virtual machine operating system that uses the storage. With traditional tools, you can monitor only the VMware ESX Server operating system (but not the virtual machine operating system). You use the VI Client to monitor virtual machines.

- Each virtual machine is, by default, configured with one virtual hard disk and one virtual SCSI controller during installation. You can modify the SCSI controller type and SCSI bus sharing characteristics by using the VI Client to edit the virtual machine settings, as shown in Figure 1-4. You can also add hard disks to your virtual machine. See the *ESX Server Configuration Guide*.

Figure 1-4. Setting the SCSI Controller Type
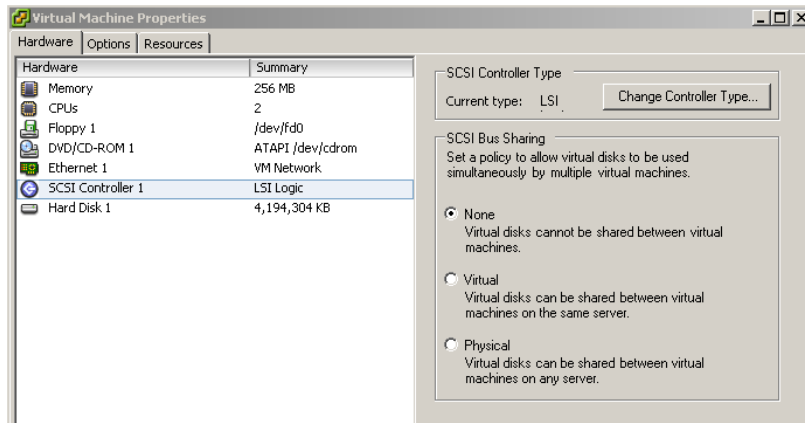
- The HBA visible to the SAN administration tools is part of the ESX Server system, not the virtual machine.

- Your ESX Server system performs multipathing for you. Multipathing software (such as PowerPath) in the virtual machine is not supported (and not required).

Data Access: VMFS or RDM

By default, a virtual disk is created in a VMFS volume during virtual machine creation. When guest operating systems issue SCSI commands to their virtual disks, the virtualization layer translates these commands to VMFS file operations. See "Virtual Machine File System" on page 15.

An alternative to VMFS is using RDMs. RDMs are special files in a VMFS volume that act as a proxy for a raw device. The RDM gives some of the advantages of a virtual disk in the VMFS while keeping some advantages of direct access to a physical device. See "Raw Device Mapping" on page 15.

Third-Party Management Applications

Most iSCSI storage hardware is packaged with storage management software. In many cases, this is a web application that can be used with any web browser connected to your network. In other cases, this software typically runs on the storage system or on a single server, independent of the servers that use the SAN for storage.

You can use this third-party management software for a number of tasks:

■ Storage system management including LUN creation, storage system cache management, LUN mapping, and LUN security.

■ Setup of replication, checkpointing, snapshotting, or mirroring.

If you decide to run the SAN management software on a virtual machine, you can run an application on a virtual machine (failover using VMotion, failover using VMware HA, and so on). Because of the additional level of indirection, however, the management software might not be able to see the SAN. This can be resolved by using an RDM. See "Layered Applications" on page 115.

> **NOTE** Whether a virtual machine can run management software successfully depends on the storage system in question.

Discovery, Authentication, and Access Control

Several mechanisms can be used to limit which volumes an ESX Server host can access on an iSCSI storage system. You must configure the ESX Server and the iSCSI storage system to support your storage access control policy.

Discovery. A discovery session is part of the iSCSI protocol, and it returns the set of targets you can access on an iSCSI storage system. The two types of discovery available on ESX Server are dynamic and static. Dynamic discovery obtains a list of accessible targets from the iSCSI storage system, while static discovery can only try to access one particular target by target name.

Authentication. iSCSI storage systems authenticate an initiator by a name and key pair. ESX Server supports the CHAP protocol, which VMware recommends for your SAN implementation. The ESX Server host and the iSCSI storage system need to have CHAP enabled, and to have common credentials. In the iSCSI login phrase, the iSCSI storage system exchanges and checks these credentials.

Access Control. A policy set up on the iSCSI storage system. Most implementations support one or more of three types of access control:

- By initiator name

- By IP address

- By the CHAP protocol

Only initiators that meet all rules attached to the iSCSI volume can access it.

Error Correction

To protect the integrity of iSCSI headers and data, the iSCSI protocol defines error correction methods known as header digests and data digests. These are disabled by default, but the user can enable them. These digests pertain to, respectively, the header and SCSI data being transferred between iSCSI initiators and targets, in both directions.

Header and data digests check the end-to-end, non-cryptographic data integrity beyond the integrity checks that other networking layers provide, such as TCP and Ethernet. They check the entire communication path, including all elements that can change the network-level traffic, such as routers, switches, and proxies.

The existence and type of the digests are negotiated when an iSCSI connection is established. When the initiator and target agree on a digest configuration, this digest must be used for all traffic between them.

Enabling header and data digests does require additional processing for both the initiator and the target and can affect throughput and CPU use performance.

Understanding VMFS and SAN Storage Choices

This section discusses the available VMFS and SAN storage choices and gives advice on how to make these choices.

Choosing Larger or Smaller LUNs

During ESX Server installation, you are prompted to create partitions for your system. Plan how to set up storage for your ESX Server systems before you perform installation. Choose one of these approaches:

- Many LUNs with one VMFS volume on each LUN

- One large LUN or many LUNs with a single VMFS volume spanning all LUNs

You can have at most one VMFS volume per LUN. You can, however, decide to use one large LUN or multiple small LUNs.

You might want fewer, larger LUNs for the following reasons:

■ More flexibility to create virtual machines without asking the SAN administrator for more space.

■ More flexibility for resizing virtual disks, doing snapshots, and so on

■ Fewer LUNs to identify and manage

You might want more, smaller LUNs for the following reasons:

■ Less contention on each VMFS because of locking and SCSI reservation issues.

■ Different applications might need different RAID characteristics.

■ More flexibility (the multipathing policy and disk shares are set per LUN).

NOTE You can divide your datacenter into servers that are best configured with fewer, larger LUNs and other servers that use more, smaller LUNs.

Making LUN Decisions

When the storage characterization for a virtual machine is not available, there is often no simple answer when you need to decide on the LUN size and number of LUNs to use. You can use one of the following approaches:

■ Predictive scheme
■ Adaptive scheme

Predictive Scheme

In the predictive scheme, you:

■ Create several LUNs with different storage characteristics.

■ Build a VMFS volume in each LUN (and label each volume according to its characteristics).

■ Locate each application in the appropriate RAID for its requirements.

■ Use disk shares to distinguish high-priority from low-priority virtual machines. Disk shares are relevant only within a given ESX Server host. The shares assigned to virtual machines on one ESX Server host have no effect on virtual machines on other ESX Server hosts.

Adaptive Scheme

In the adaptive scheme, you:

■   Create a large LUN (RAID 1+0 or RAID 5), with write caching enabled.

■   Build a VMFS in that LUN.

■   Place four or five virtual disks on the VMFS.

■   Run the applications and see whether disk performance is acceptable.

■   If performance is acceptable, you can place additional virtual disks on the VMFS. If it is not, you create a new, larger LUN, possibly with a different RAID level, and repeat the process. You can use cold migration so you do not lose virtual machines when you recreate the LUN.

Tips for Making LUN Decisions

When you make your LUN decision, keep in mind the following considerations:

■   Each LUN should have the correct RAID level and storage characteristic for applications in virtual machines that use it.

■   One LUN must contain only one VMFS volume.

■   If multiple virtual machines access the same VMFS (and therefore the same LUN), use disk shares to prioritize virtual machines.

To use disk shares to prioritize virtual machines

1   Start a VI Client and connect to a VirtualCenter Server.

2   Select the virtual machine in the inventory panel, right-click, and choose **Edit Settings** from the menu.

3   Click the **Resources** tab and click **Disk**.

4   Right-click the **Shares** column for the disk to modify and select the required value from the drop-down menu.
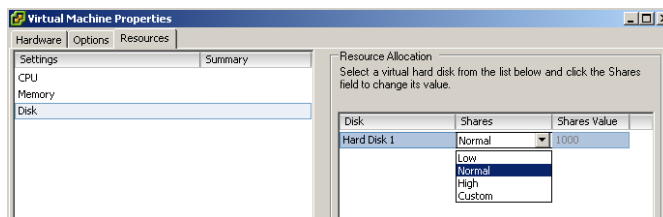
Understanding Data Access

This section discusses:

- ■ "Accessing Data on a SAN" on page 30
- ■ "How Virtual Machines Access Data" on page 30
- ■ "How Virtual Machines Access Data on a SAN" on page 31

Accessing Data on a SAN

If you are not familiar with how a physical (non-virtual) machine accesses data on a SAN storage system, see the *SAN Conceptual and Design Basics* white paper on the VMware Documentation Web site at www.vmware.com/support/pubs/.

How Virtual Machines Access Data

Virtual machines use one of the following methods to access data:

- ■ **VMFS** – In a simple configuration, the virtual machines' disks are stored as `.vmdk` files within an ESX Server VMFS. When guest operating systems issue SCSI commands to their virtual disks, the virtualization layer translates these commands to VMFS file operations.

  In a default setup, the virtual machine always goes through VMFS when it accesses a file, whether the file is on a SAN or a host's local hard drives. See "Virtual Machine File System" on page 15.

- ■ **Raw device mapping (RDM)** – An RDM is a mapping file inside the VMFS that acts as a proxy for a raw device. The RDM gives the guest operating system access to the raw device.

  VMware recommends RDM when a virtual machine must interact with a real disk on the SAN. This is the case, for example, when you make storage system snapshots or, more rarely, if you have a large amount of data that you do not want to move onto a virtual disk.

For more information on VMFS and RDMs, see the *ESX Server Configuration Guide*.

How Virtual Machines Access Data on a SAN

When a virtual machine interacts with a SAN, the following process takes place:

1   When the guest operating system in a virtual machine needs to read or write to a SCSI disk, it issues SCSI commands to the virtual disk.

2   Device drivers in the virtual machine's operating system communicate with the virtual SCSI controllers. VMware ESX Server supports two types of virtual SCSI controllers: BusLogic and LSILogic.

3   The virtual SCSI Controller forwards the command to the VMkernel.

4   The VMkernel:

■   Locates the file in the VMFS volume that corresponds to the guest virtual machine disk.

■   Maps the requests for the blocks on the virtual disk to blocks on the appropriate physical device.

■   Sends the modified I/O request from the device driver in the VMkernel to the iSCSI initiator (hardware or software).

5   If the iSCSI initiator is a hardware iSCSI initiator (iSCSI HBA), the HBA does the following:

■   Encapsulates I/O requests into iSCSI Protocol Data Units (PDUs).

■   Encapsulates iSCSI PDUs into TCP/IP packets.

■   Sends IP packets over Ethernet to the iSCSI storage system.

If the iSCSI initiator is a software iSCSI initiator, it:

■   Encapsulates I/O requests into iSCSI PDUs.

■   Sends iSCSI PDUs through TPC/IP connections.

■   The VMkernel TCP/IP stack relays TCP/IP packets to a physical NIC.

■   The physical NIC sends IP packets over Ethernet to the iSCSI storage system.

6   Depending on which port the iSCSI initiator uses to connect to the network, Ethernet switches and routers carry the request to the storage device that the host wants to access.

From the host's perspective, this storage device appears to be a specific disk, but it might be a logical device that corresponds to a physical device on the SAN.

Path Management and Failover

You can use multiple ways to manage paths and failover in an iSCSI SAN:

- SCSI storage stack failover

- NIC teaming and failover

- Array-based failover

SCSI Storage Stack Failover

ESX Server supports multipathing to maintain a constant connection between the server machine and the storage device in case of the failure of an HBA or switch. Multipathing support does not require specific failover drivers.

To support path switching, the server typically has two or more HBAs available from which the storage system can be reached using one or more switches. Alternatively, the setup might include one HBA and two storage processors so that the HBA can use a different path to reach the storage system.

Figure 1-5. Multipathing and Failover
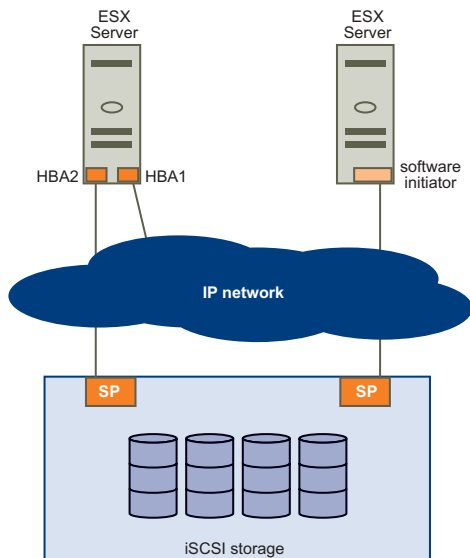
In Figure 1-5, multiple paths connect each server with the storage device. For example, if HBA1 or the link between HBA1 and the network fails, HBA2 takes over and provides the connection between the server and the network. The process of one HBA taking over for another is called *HBA failover*.

Similarly, if SP1 fails or the links between SP1 and the switches breaks, SP2 takes over and provides the connection between the switch and the storage device. This process is called *SP failover*. ESX Server supports both HBA and SP failover with its multipathing capability.

You can choose a multipathing policy for your system, either Fixed or Most Recently Used. If the policy is Fixed, you can specify a preferred path. Each LUN (disk) that is visible to the ESX Server host can have its own path policy. For information on viewing the current multipathing state and on setting the multipathing policy, see "Multipathing" on page 94.

NIC Teaming and Failover

With software iSCSI, you can connect a single virtual VMKernel iSCSI network switch to multiple physical Ethernet adapters by using the VMware Infrastructure feature called NIC teaming. NIC teaming provides network redundancy and some load balancing capabilities for iSCSI connections between ESX Server and storage systems. Similar to the SCSI multipath capabilities, NIC teaming provides failover if connections or ports on the ESX Server system fail.

NIC teaming paths do not show up as multiple paths to storage in ESX Server configurations. NIC teaming is handled entirely by the network layer and must be configured and monitored separately from the ESX Server SCSI storage multipath configuration.

Array-Based Failover

Some iSCSI storage systems manage path use of their ports automatically (transparently to ESX Server). When using one of these storage systems, ESX Server does not see multiple ports on the storage and cannot choose the storage port it connects to. These systems have a single virtual port address that ESX Server uses to initially communicate. During this initial communication, the storage system can redirect ESX Server to communicate with another port on the storage system. The iSCSI initiators in ESX Server obey this reconnection request and connect with a different port on the system. The storage system uses this technique to spread the load across available ports.

If ESX Server loses connection to one of these ports, it automatically attempts to reconnect with the virtual port of the storage system, and should be redirected to an active, usable port. This reconnection and redirection happens quickly and generally does not disrupt running virtual machines. These storage systems can also request that iSCSI initiators reconnect to the system, to change which storage port they are connected to. This allows the most effective use of the multiple ports.

Figure 1-6 shows an example of port redirection. ESX Server attempts to connect to the 10.0.0.1 virtual port. The storage system redirects this request to 10.0.0.2. ESX Server connects with 10.0.0.2 and uses this port for I/O communication.

NOTE   The storage system does not always redirect connections. The port at 10.0.0.1 could be used for traffic, also.

Figure 1-6. Port Redirection
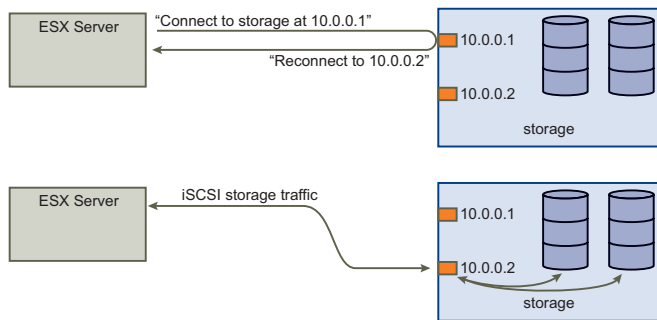
If the port on the storage system that is acting as the virtual port becomes unavailable, the storage system reassigns the address of the virtual port to another port on the system. Figure 1-7 shows an example of this type of port reassignment. In this case, the virtual port 10.0.0.1 becomes unavailable and the storage system reassigns the virtual port IP address to a different port. The second port responds to both addresses.

Figure 1-7. Port Reassignment

NOTE   Virtual machine I/O can be delayed for up to sixty seconds while failover takes place, particularly on an active-passive array. This delay allows the SAN to stabilize its configuration after topology changes. With active-passive arrays with path policy **Fixed**, path thrashing can be a problem. See "Resolving Path Thrashing" on page 109.

Choosing Virtual Machine Locations

When you're working on optimizing performance for your virtual machines, storage location is an important factor. A trade-off always exists between expensive storage that offers high performance and high availability and storage with lower cost and lower performance. Storage can be divided into different tiers depending on a number of factors:

High Tier.   Offers high performance and high availability. Might offer built-in snapshots to facilitate backups and point-in-time (PiT) restorations. Supports replication, full SP redundancy, and SAS drives. Uses high-cost spindles.

Mid Tier.   Offers mid-range performance, lower availability, some SP redundancy, and SCSI or SAS drives. May offer snapshots. Uses medium-cost spindles.

Lower Tier.   Offers low performance, little internal storage redundancy. Uses low end SCSI drives or SATA (serial low-cost spindles).

Not all applications need to be on the highest-performance, most-available storage—at least not throughout their entire life cycle.

NOTE If you need some of the functionality of the high tier, such as snapshots, but do not want to pay for it, you might be able to achieve some of the high-performance characteristics in software. For example, you can create snapshots in software.

When you decide where to place a virtual machine, ask yourself these questions:

- How critical is the virtual machine?

- What are its performance and availability requirements?

- What are its PiT restoration requirements?

- What are its backup requirements?

- What are its replication requirements?

A virtual machine might change tiers throughout its life cycle because of changes in criticality or changes in technology that push higher-tier features to a lower tier. Criticality is relative and might change for a variety of reasons, including changes in the organization, operational processes, regulatory requirements, disaster planning, and so on.

Designing for Server Failure

The RAID architecture of SAN storage inherently protects you from failure at the physical disk level. A SAN provides multiple paths between servers and storage, which protects against network or port failures. The final step in making your whole environment failure resistant is to protect against server failure. ESX Server systems failover options are discussed in the following sections.

Using VMware HA

VMware HA allows you to organize virtual machines into failover groups. When a host fails, all its virtual machines are immediately started on different hosts. When a virtual machine is restored on a different host, it loses its memory state but its disk state is exactly as it was when the host failed (crash-consistent failover). Shared storage (such as a SAN) is required for HA. See the *Resource Management Guide*.

NOTE   You must be licensed to use VMware HA.

Server Failover and Storage Considerations

For each type of server failover, you must consider storage issues:

■   Approaches to server failover work only if each server has access to the same storage. Because multiple servers require a lot of disk space, and because failover for the storage system complements failover for the server, SANs are usually employed in conjunction with server failover.

■   When you design a SAN to work in conjunction with server failover, all volumes the clustered virtual machines use must be seen by all ESX Server hosts.

Although a volume is accessible to a host, all virtual machines on that host do not necessarily have access to all data on that volume. A virtual machine can access only the virtual disks for which it was configured. In case of a configuration error, virtual disks are locked when the virtual machine boots so no corruption occurs.

NOTE   As a rule, when you're booting from a SAN, each boot volume should be seen only by the ESX Server system that is booting from that volume. An exception is when you're trying to recover from a failure by pointing a second ESX Server system to the same volume. In this case, the SAN volume in question is not really for booting from a SAN. No ESX Server system is booting from it because it is corrupted. The SAN volume is a regular non-boot volume that is made visible to an ESX Server system.

Installation of iSCSI Initiators and Storage

2

Before ESX Server can work with a SAN, you must set up your iSCSI initiators and storage. To do this, you must first observe certain basic requirements. This chapter discusses these requirements, provides recommendations, and then details how to provide access to the SAN by installing and setting up your hardware or software iSCSI initiators.

This chapter discusses the following topics:

- "Preparing for iSCSI Storage Setup" on page 37
- "Setting Up Hardware iSCSI Initiators and Storage" on page 40
- "Setting Up Software iSCSI Initiators and Storage" on page 50

After initial installation steps are performed, you might need to modify your storage system. This is discussed in Chapter 3, "Modifying SAN Storage Systems with ESX Server," on page 65.

Preparing for iSCSI Storage Setup

To prepare for setting up your ESX Server system to use SAN storage, review the following requirements, restrictions, recommendations, and LUN allocation tips.

ESX Server SAN Requirements

The following requirements must be met before your ESX Server environment can work properly with a SAN:

- **Hardware and firmware**. Only a limited number of SAN storage hardware and firmware combinations are supported in conjunction with ESX Server systems. For an up-to-date list, see the *Storage/SAN Compatibility Guide*.

- **One VMFS volume per LUN**. Configure your system to have only one VMFS volume for each LUN. (In VMFS-3, you do not need to set accessibility.)

- Unless you're using diskless servers (booting from a SAN), do not set up the diagnostic partition on a SAN LUN. In the case of diskless servers that boot from a SAN, a shared diagnostic partition is appropriate.

- VMware recommends that you use RDMs for access to any raw disk. For more information on RDMs, see the *ESX Server Configuration Guide*.

- **Multipathing**. For multipathing to work properly, each LUN must present the same LUN number to all ESX Server hosts.

- **Queue size**. Set the BusLogic or LSILogic driver in the guest operating system to specify a large enough queue. You can set the queue depth for the physical HBA during system setup. For supported drivers, see the *Storage/SAN Compatibility Guide*.

- **SCSI Timeout**. On virtual machines running Microsoft Windows, increase the value of the SCSI `TimeoutValue` parameter to allow Windows to better tolerate delayed I/O resulting from path failover.

Restrictions

The following restrictions apply when you use ESX Server with a SAN:

- ESX Server does not support iSCSI-connected tape devices.

- You cannot use virtual-machine, multipathing software to perform I/O load balancing to a single physical LUN.

Recommendations

Consider the following recommendations when you set up your environment with ESX Server hosts and a SAN:

■ Use raw device mapping for a virtual disk of a virtual machine to use some of the hardware snapshotting functions of the storage system, or to access a disk from both a virtual machine and a physical machine in a cold-standby host configuration for data LUNs.

■ To use VMotion to move a virtual machine to a different host, the LUNs that hold the virtual disks of the virtual machines must be visible from all of the hosts.

Setting LUN Allocations

When you set LUN allocations, note the following points:

■ **Storage Provisioning.** To ensure that the ESX Server system recognizes the LUNs at startup time, provision all LUNS to the appropriate HBAs before connecting the SAN to the ESX Server system.

NOTE   VMware recommends that you provision all LUNs to all ESX Server HBAs at the same time. HBA failover works only if all HBAs see the same LUNs.

■ **VMotion and VMware DRS.** When you use VirtualCenter and VMotion or DRS, make sure that the LUNs for the virtual machines are provisioned to all ESX Server hosts. This configuration provides the greatest freedom in moving virtual machines.

■ **Active-active versus active-passive arrays**. When you use VMotion or DRS with an active-passive SAN storage device, make sure that all ESX Server systems have consistent paths to all storage processors. Not doing so can cause path thrashing when a VMotion migration occurs. See "Resolving Path Thrashing" on page 109.

For active-passive storage arrays not listed in the *Storage/SAN Compatibility Guide*, VMware does not support storage-port failover. You must connect the server to the active port on the storage system. This configuration ensures that the LUNs are presented to the ESX Server host.

Setting Up Hardware iSCSI Initiators and Storage

With hardware-based iSCSI storage, you use a specialized third-party adapter capable of accessing iSCSI storage over TCP/IP. This iSCSI initiator handles all iSCSI processing and management for your ESX Server system.

Hardware iSCSI initiators require configuration to work properly, so you must install and configure the hardware iSCSI initiators as detailed in the following sections before setting up the datastore that resides on an iSCSI storage device.

NOTE   You can configure some ESX Server systems to load balance traffic across multiple HBAs to multiple LUNs with certain active-active arrays. To do this, assign preferred paths to your LUNs so that your HBAs are being used evenly.

Installing and Viewing Hardware iSCSI Initiators
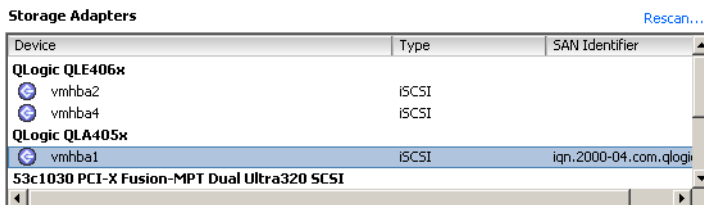
For information on which initiators are supported, see the *I/O Compatibility Guide* on the VMware Web site at www.vmware.com.

Before you begin configuring the hardware iSCSI initiator, make sure that the iSCSI HBA is successfully installed and appears on the list of initiators available for configuration. If the initiator is installed, you can view its properties.

To view the hardware iSCSI initiator properties

1   Log in to the VI Client and select a server from the inventory panel.

2   Click the **Configuration** tab and click **Storage Adapters** in the **Hardware** group.

The list of available storage adapters (initiators) appears. The iSCSI initiator appears in the list of storage adapters.

Storage Adapters		Rescan...
Device	Type	SAN Identifier
QLogic QLE406x		
vmhba2	iSCSI	
vmhba4	iSCSI	
QLogic QLA405x		
vmhba1	iSCSI	iqn.2000-04.com.qlogi
53c1030 PCI-X Fusion-MPT Dual Ultra320 SCSI		

3   Under HBA, choose the initiator to configure.

The details for the initiator appear, including the model, IP address, iSCSI name, discovery methods, iSCSI alias, and any discovered targets.

4   Click **Properties**.

The iSCSI Initiator Properties dialog box appears. The **General** tab displays additional characteristics of the initiator.
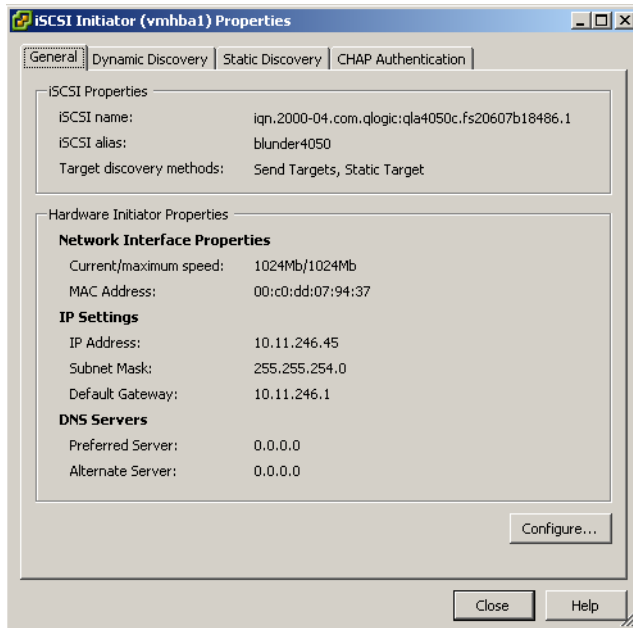
You can now configure your hardware initiator or change its default characteristics.

Configuring Hardware iSCSI Initiators

When you configure the hardware iSCSI initiator, set up your initiator's iSCSI name, IP address, and discovery addresses. VMware recommends that you set up CHAP parameters also.

After you configure your hardware iSCSI initiator, perform a rescan so that all LUNs that the initiator has access to appear on the list of storage devices available to your ESX Server host.

Setting up Naming Parameters

When you configure your hardware iSCSI initiators, make sure that their names and IP addresses are formatted properly. See "iSCSI Naming Conventions" on page 19.

To set up the iSCSI name, alias, and IP address for the hardware initiator

1   Open the **iSCSI Initiator Properties** dialog box by performing the steps listed in "To view the hardware iSCSI initiator properties" on page 40.

2   Click **Configure**.

The General Properties dialog box appears.
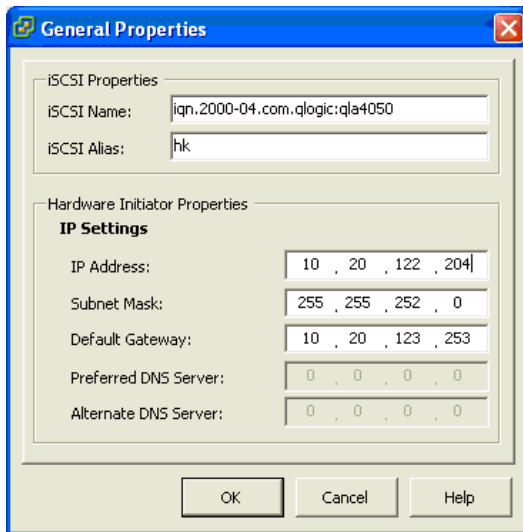
3   To change the default iSCSI name for your initiator, enter the new name.

Format the name you enter properly; otherwise, some storage devices might not recognize the hardware iSCSI initiator. You can use the default name supplied by the vendor. It does not have to be changed.

4   Enter the iSCSI alias.

The alias is a name that you use to identify the hardware iSCSI initiator.

5   If you selected **Use the following IP settings**, enter values for the following:

- **IP Address**
- **Subnet Mask**
- **Default Gateway**

6   Click **OK** to save your changes, then reboot the server for the changes to take effect.

Setting Up Discovery Addresses for the Hardware Initiator

Set up target discovery addresses so that the hardware initiator can determine which storage resource on the network is available for access. You can do this with either dynamic discovery, where all targets associated with an IP address are discovered, or with static discovery, where you must specify the IP address and the iSCSI name of the target to be seen.

To set up target discovery addresses using Dynamic Discovery

1   Open the **iSCSI Initiator Properties** dialog box by performing the steps listed in

2   In the iSCSI Initiator Properties dialog box, click the **Dynamic Discovery** tab.
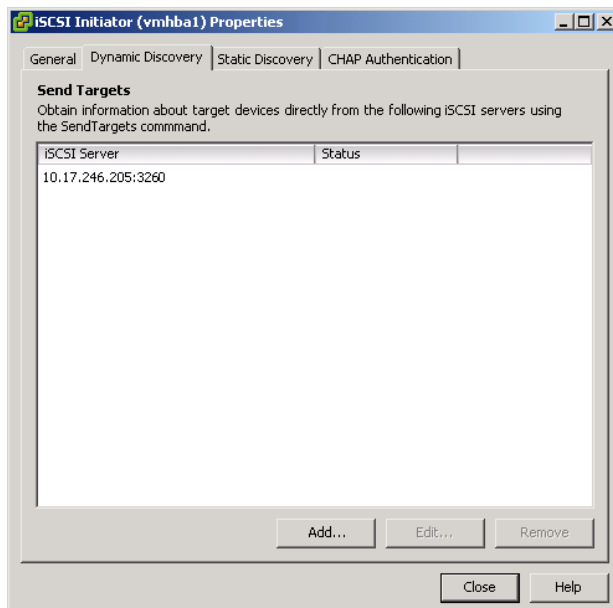
3 To add a new iSCSI target that your ESX Server host can use for a SendTargets session, click **Add**.

The **Add SendTargets Server** dialog box appears.
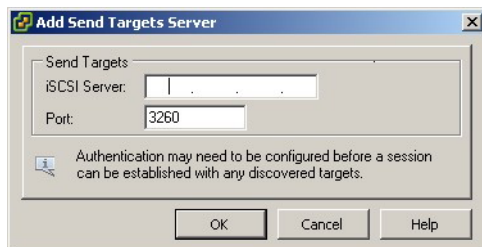
4 Enter the IP address of a storage system and click **OK**.

After your ESX Server host establishes the SendTargets session with this target device, any newly discovered targets appear in the Static Discovery list.

5 To change or delete a specific IP address, select it and click **Edit** or **Remove**.

To set up target discovery addresses by using Static Discovery

1 Open the iSCSI Initiator Properties dialog box by performing the steps listed in "To view the hardware iSCSI initiator properties" on page 40.

2 In the iSCSI Initiator Properties dialog box, click the **Static Discovery** tab.

The tab displays all dynamically discovered targets and any static targets already entered.
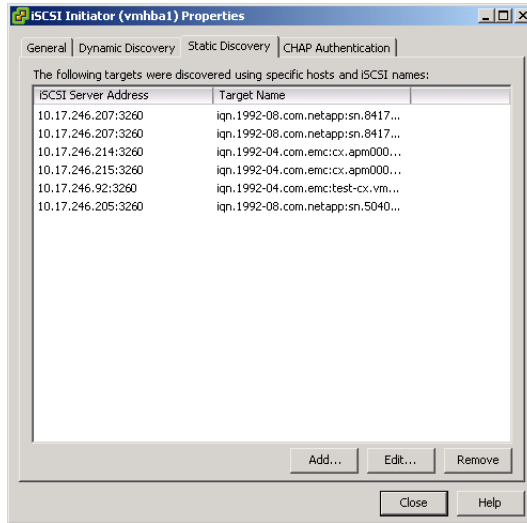
3   To add a target accessible to your ESX Server host, click **Add** and enter the target's IP address and fully qualified domain name.
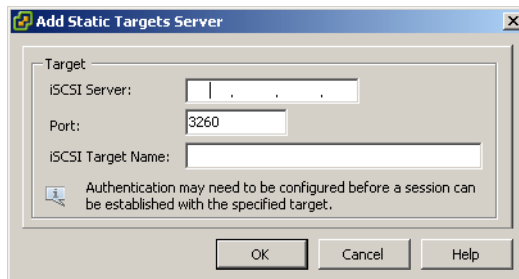
4   To change or delete a specific dynamically discovered target, select the target and click **Edit** or **Remove**.

NOTE   If you remove a dynamically discovered static target, the target can be returned to the list the next time a rescan happens, the HBA is reset, or the system is rebooted.

Setting up CHAP Parameters

When you configure your hardware iSCSI initiator, ensure that CHAP configuration matches your iSCSI storage. If CHAP is enabled on the storage, it must be enabled on the initiator. If they are enabled, you must set up the CHAP authentication credentials to match your iSCSI storage.

NOTE ESX Server hosts only support one set of CHAP credentials per initiator. You can not assign different CHAP credentials for different targets through a VI Client.

To set up CHAP parameters for the hardware initiator

1 Open the **iSCSI Initiator Properties** dialog box by performing the steps listed in

2 Click the **CHAP Authentication** tab.

The tab displays the default CHAP parameters.

3   To make any changes to the existing CHAP parameters, click **Configure**.

The CHAP Authentication dialog box opens.
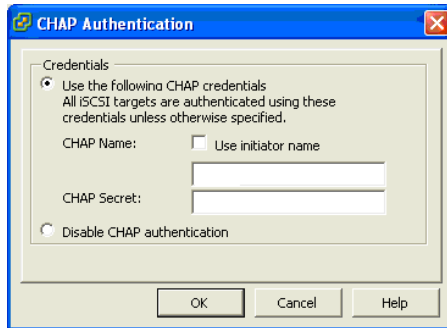
4   To keep CHAP enabled, select **Use the following CHAP credentials**.

5   Either enter a new CHAP name or select **Use initiator name**.

6   If needed, specify the **CHAP Secret**.

All new targets will use the CHAP secret to authenticate the initiator.

7   Click **OK** to save changes.

> **NOTE**   If you disable CHAP, existing sessions remain until a reboot or the storage system forces a logout, then you cannot connect to targets that require CHAP.

Adding Hardware-Initiated iSCSI Storage

When you create a datastore on a hardware-initiated iSCSI storage device, the Add Storage wizard guides you through the configuration.

To create a datastore on a hardware-initiated iSCSI device

1   Log in to the VI Client and select a server from the inventory panel.

2   Click the **Configuration** tab and click **Storage**.

3   Click **Add Storage**.

The **Select Storage Type** page appears.
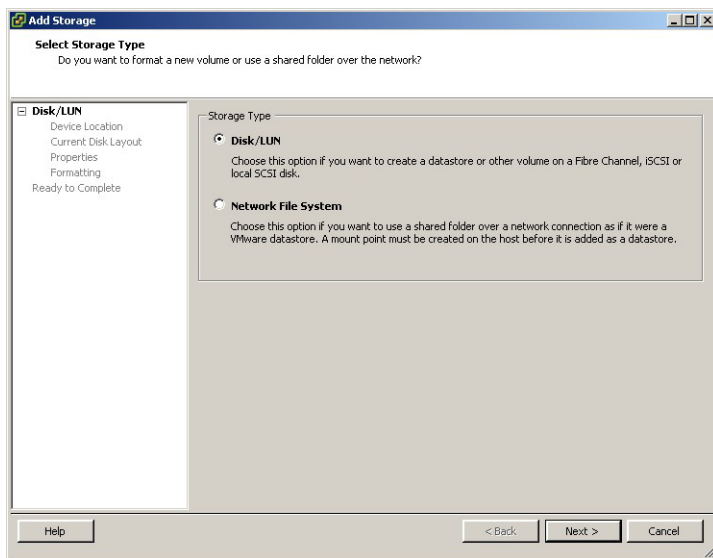
4   Select **Disk/LUN**, and click **Next**.

The **Select Disk/LUN** page appears. This can take a few seconds depending on the number of targets that you have.

5   Select the iSCSI device to use for your datastore and click **Next**.

     The **Current Disk Layout** page appears.

6   Look over the current disk layout and click **Next**.

     The **Disk/LUN–Properties** page appears.

7   Enter a datastore name.

     The datastore name appears in the VI Client, and the label must be unique within
     the current Virtual Infrastructure instance.

8   Click **Next**.

     The **Disk/LUN–Formatting** page appears.

9   If needed, adjust the file system values and capacity you use for the datastore.
     By default, the entire free space available on the storage device is offered to you.

10   Click **Next**.

     The **Summary** page appears.

11   Review the datastore information and click **Finish**.

     This creates the datastore on the hardware-initiated iSCSI device.

Setting Additional Parameters

You can use the `esxcfg–hwiscsi` utility to configure additional parameters for your hardware iSCSI HBA. For example, some iSCSI storage systems require ARP redirection to move iSCSI traffic dynamically from one port to another. You must allow ARP redirection on your hardware iSCSI HBA.

See Appendix B, "Utilities," on page 123.

Setting Up Software iSCSI Initiators and Storage

With the software-based iSCSI implementation, you can use a standard network adapter to connect your ESX Server system to a remote iSCSI target on the IP network. The ESX Server software iSCSI initiator built into VMkernel facilitates this connection communicating with the network adapter through the network stack.

Before you configure datastores that use a software-initiated iSCSI connection to access the iSCSI storage, you must enable network connectivity and then install and configure the software iSCSI initiator.

Networking Configuration for Software iSCSI Storage

Before you can configure iSCSI storage, you must create a VMkernel port to handle iSCSI networking and (for ESX Server 3 only) a service console connection to the iSCSI network.

> **CAUTION** If you are configuring a software iSCSI implementation with ESX Server 3, you must set up both the VMkernel port and the service console connection in your networking configuration.

To create a VMkernel port for software iSCSI

1   Log in to the VI Client and select the server from the inventory panel.

    The hardware configuration page for this server appears.

2   Click the **Configuration** tab and click **Networking**.

3   Click **Add Networking**.

The **Add Network Wizard** appears.
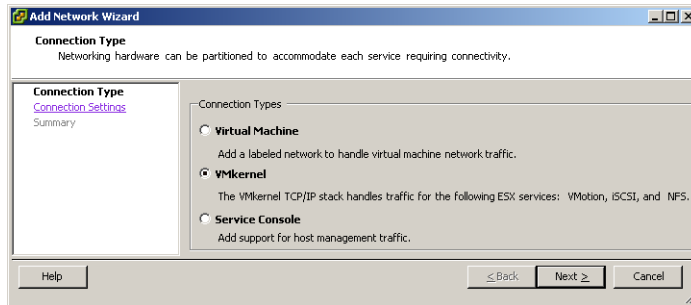
With ESX Server 3i, the option for **Service Console** does not appear in this wizard screen.

4   Select **VMkernel** and click **Next**.

This lets you connect the VMkernel, which runs services for iSCSI storage, to the physical network.

The **Network Access** page appears.

5   Select the vSwitch to use or the **Create a virtual switch** radio button.

6   Select the check boxes for the network adapters your vSwitch will use.
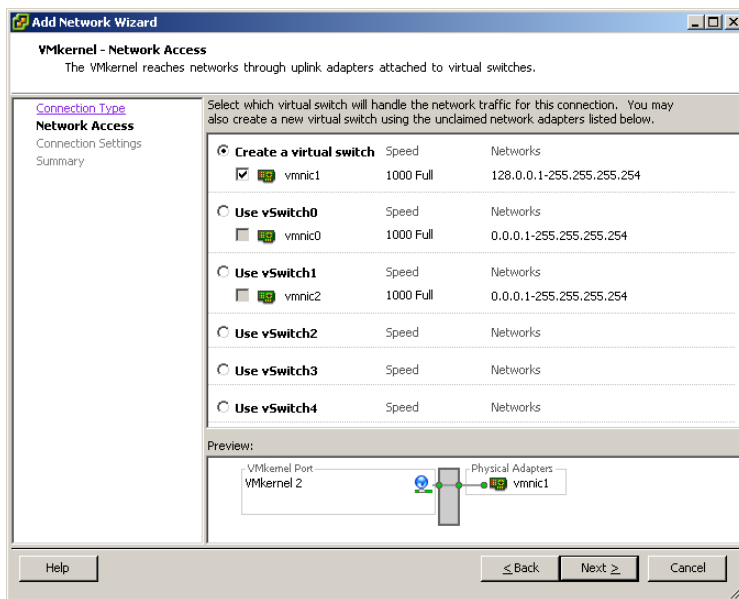
Your choices appear in the **Preview** pane.

Select adapters for each vSwitch so that virtual machines or other services that connect through the adapter can reach the correct Ethernet segment. If no adapters appear under **Create a virtual switch**, existing vSwitches are using all of the network adapters in the system.

NOTE   Do not use iSCSI on 100MB NICs.

---

7   Click **Next**.

The **Connection Settings** page appears.

8   Under **Port Group Properties**, select or enter a network label and, optionally, a VLAN ID. You can also enter or change the **IP Address** and **Subnet Mask** under **IP Settings**.

Network Label A name that identifies the port group that you are creating. This is the label that you specify when you configure a virtual adapter to be attached to this port group, when you configure iSCSI storage.

VLAN ID Identifies the VLAN that the port group's network traffic will use. VLAN IDs are not required. If you are not sure whether you need them, ask your network administrator.
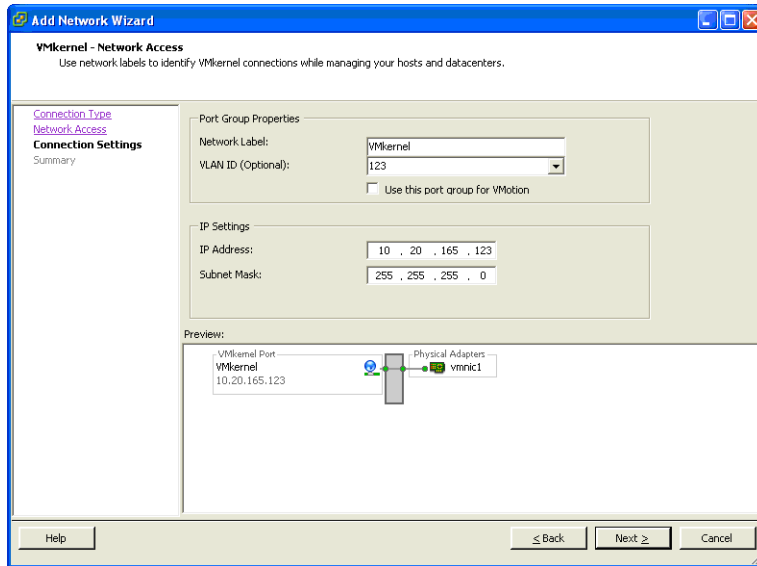
9   You receive a warning that no default gateway is set. A gateway is needed for connectivity to machines not on the same IP subnet as the service console (ESX Server 3 only) or VMkernel. To connect to multiple subnets, click **Yes** to configure one.
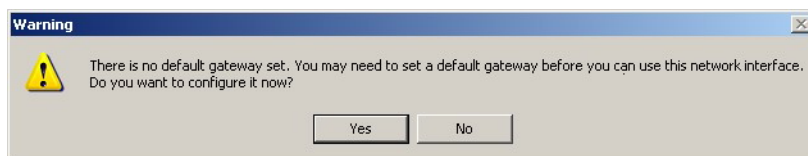
10   On the **Routing** tab in the **DNS and Routing Configuration** dialog box, the service console and the VMkernel each need their own gateway information.
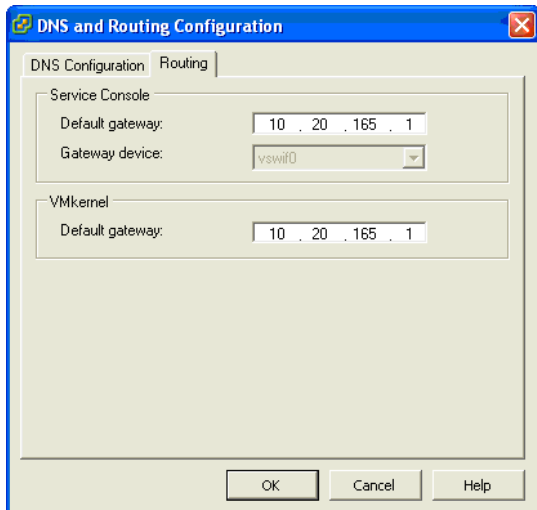
NOTE   Set a default gateway for the port that you created. You must use a valid static IP address to configure the VMkernel stack.

11   Click **OK** to save your changes and close the **DNS and Routing Configuration** dialog box.

12   Click **Next**.

13   Use the **Back** button to make any changes.

14   Review your changes on the **Ready to Complete** page and click **Finish**.

After you create a VMkernel port for iSCSI, you must create a service console connection on the same vSwitch as the VMkernel port. If you are using ESX Server 3i, this is not necessary.

To configure a service console connection for software iSCSI (ESX Server 3 only)

1   Log in to the VI Client and select the server from the inventory panel.

The hardware configuration page for this server appears.

2   Click the **Configuration** tab and click **Networking**.

3   On the right side of the screen, click **Properties** for the vSwitch associated with the VMkernel port you just created.

4 On the **Ports** tab, click **Add**.

The **Add Network Wizard** appears.
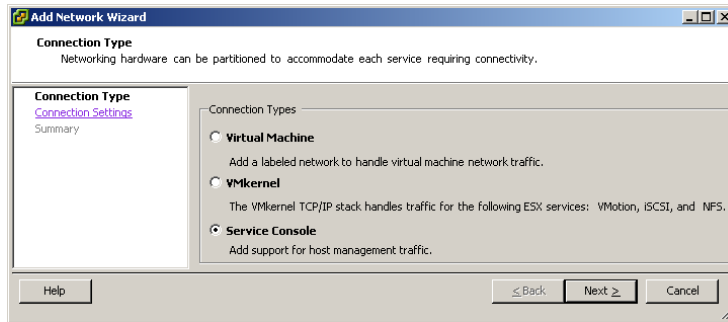
5 As a connection type, select **Service Console** and click **Next**.

The **Connection Settings** page appears.

6 In the **Port Group Properties** area, enter a network label that identifies the port group that you are creating.
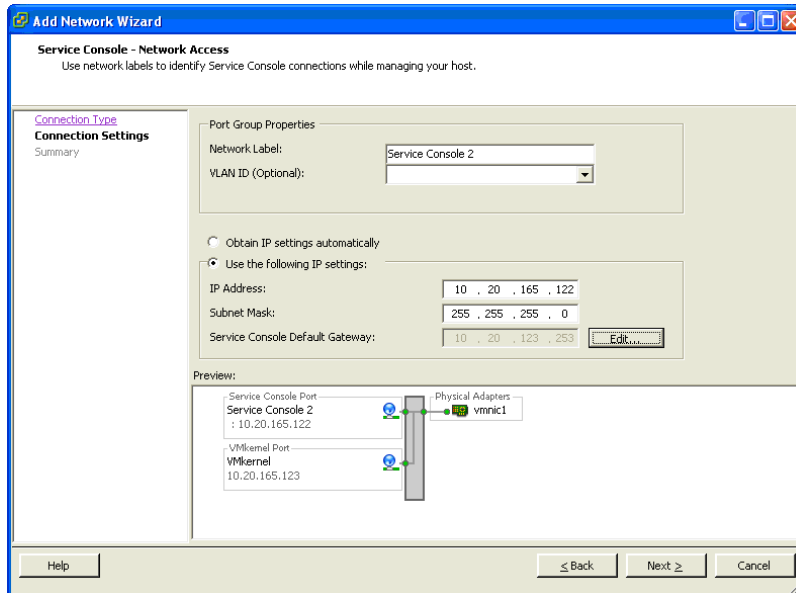
Newer ports and port groups appear at the top of the vSwitch diagram.

7   Enter the **IP Address** and **Subnet Mask,** or select the **Obtain IP setting automatically** DHCP option for the IP address and subnet mask. This must be a different IP address than the one chosen for the VMkernel.

8   Click **Edit** to set the **Service Console Default Gateway**.

9   Click **Next**.

The **Ready to Complete** page appears.
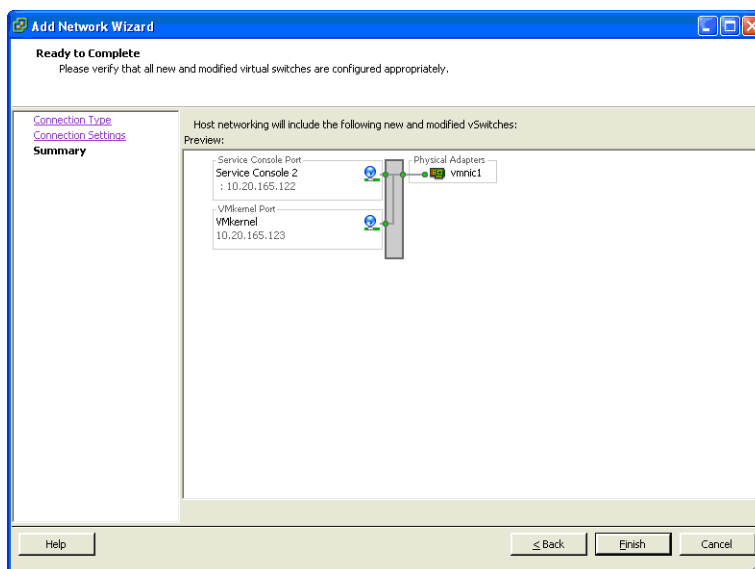
10 After you have determined that the vSwitch is configured correctly, click **Finish**.

After you create a VMkernel port and service console connection, you can enable and configure software iSCSI storage.

Configuring Software iSCSI Initiators

To configure the software iSCSI initiator, you enable it and set up its target addresses. VMware recommends that you also set up its CHAP parameters.

After you configure your software iSCSI initiator, perform a rescan, so that all LUNs that the initiator has access to appear on the list of storage devices available to your ESX Server system.

Enabling Software iSCSI Initiators

Enable your software iSCSI initiator so that ESX Server can use it.

To enable the software iSCSI initiator

1   Open the **iSCSI Initiator Properties** dialog box by performing the steps listed in "To view the software iSCSI initiator properties" on page 60.

2   Click **Configure**.

    The **General Properties** dialog box opens, displaying the initiator's status, default name, and alias.

3   To enable the initiator, select **Enabled**.
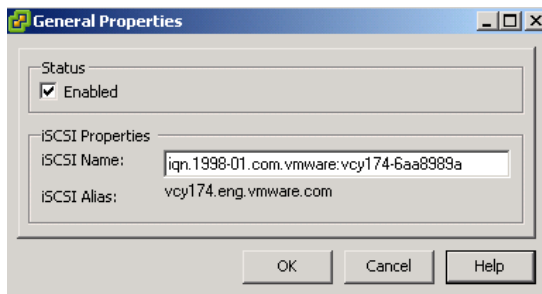
4   To change the default iSCSI name for your initiator, enter the new name. You do not need to change the default name.

    Format the name you enter properly; otherwise, some storage devices might not recognize the software iSCSI initiator.

5   Click **OK** to save your changes.

Setting up Discovery Addresses

Set up target discovery addresses so that the software initiator can determine which storage resource on the network is available for access.

To set up target discovery addresses for the software initiator

1   Open the **iSCSI Initiator Properties** dialog box by performing the steps listed in

2   Click the **Dynamic Discovery** tab.
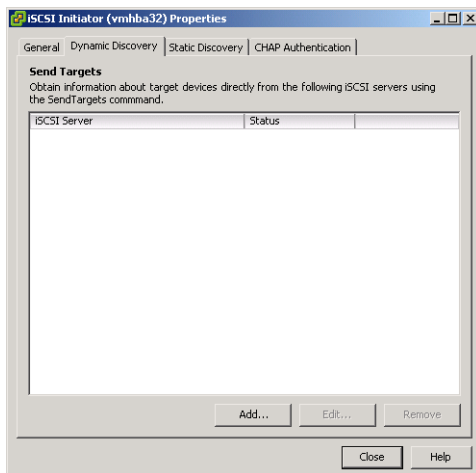
3   To add a new iSCSI target your ESX Server host can use for a SendTargets session, click **Add**.

The **Add Send Targets Server** dialog box appears.
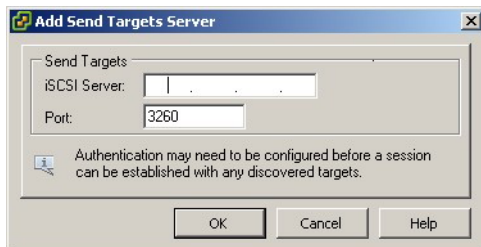
4   Enter the Send Targets server IP address and click **OK**.

5   To change or delete a Send Targets server, select the server and click **Edit** or **Remove**.

Setting up CHAP Parameters

When you configure your software iSCSI initiator, ensure that CHAP configuration matches your iSCSI storage. If CHAP is enabled on the storage, it must be enabled on the initiator. If they are enabled, you must set up the CHAP authentication credentials to match your iSCSI storage.

To set up CHAP parameters for the software initiator

1   Open the **iSCSI Initiator Properties** dialog box by performing the steps listed in "To view the software iSCSI initiator properties" on page 60.

2   Click the **CHAP Authentication** tab.
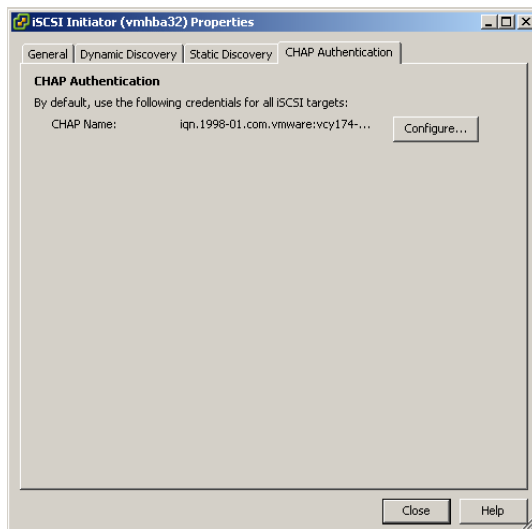
The tab displays the default CHAP parameters.

3   To make any changes to the existing CHAP parameters, click **Configure**.

The **CHAP Authentication** dialog box opens.

4   To keep CHAP enabled, select **Use the following CHAP credentials**.

5   Either enter a new CHAP name or select **Use initiator name**.

6   If needed, specify the **CHAP Secret**.

All new targets will use the CHAP secret to authenticate the initiator.
Any established sessions are not affected.

7   Click **OK** to save changes.

NOTE   If you disable CHAP, existing sessions remain until a reboot or the storage
system forces a logout, you cannot connect to targets that require CHAP.

Viewing Software iSCSI Initiators

The software iSCSI initiator that your ESX Server system uses to access a
software-initiated iSCSI storage device appears on the list of available adapters.
After configuring your software initiator, you can use the VI Client to review its
properties.

To view the software iSCSI initiator properties

1   Log in to the VI Client and select a server from the inventory panel.

2   Click the **Configuration** tab and click **Storage Adapters** under hardware.

The list of available storage adapters appears.

3   Under **iSCSI Software Adapter**, choose the available software initiator.

The details for the initiator appear, including the model, IP address, iSCSI name, discovery methods, iSCSI alias, and any discovered targets.

4   Click **Properties**.

The **iSCSI Initiator Properties** dialog box opens. The **General** tab displays additional characteristics of the software initiator.
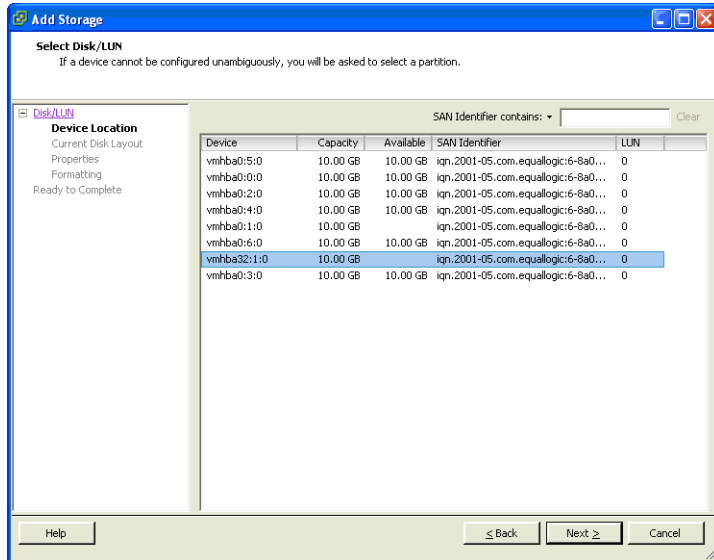
Adding Software-Initiated iSCSI Storage

When you create a datastore on a software-initiated iSCSI storage device, the Add Storage wizard guides you through the configuration.

To create a datastore on a software-initiated iSCSI device

1   Log in to the VI Client and select a server from the inventory panel.

2   Click the **Configuration** tab and click **Storage**.

3   Click **Add Storage**.

The **Select Storage Type** page appears.
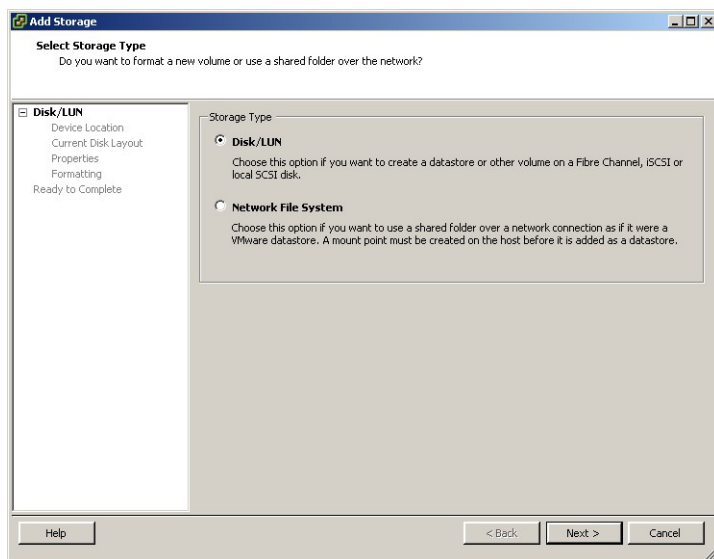
4   Select the **Disk/LUN** storage type and click **Next**.

The **Select Disk/LUN** page appears. This can take a few seconds depending on the number of targets that you have.

5   Select the iSCSI device to use for your datastore and click **Next**.

The **Current Disk Layout** page appears.

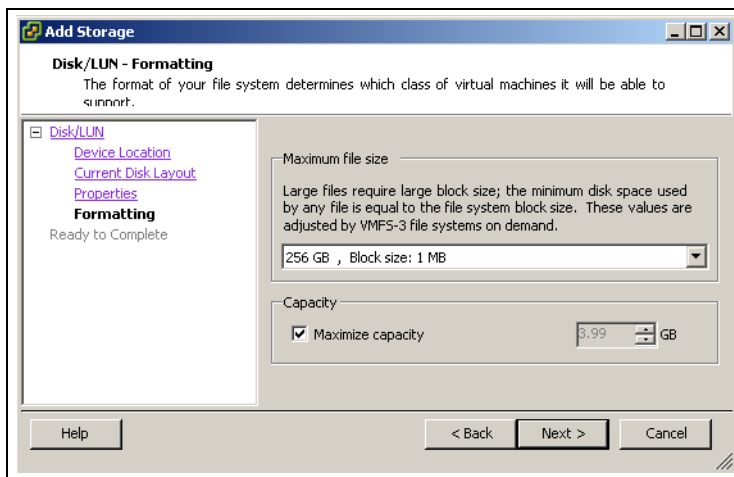6   Look over the current disk layout and click **Next**.

The **Disk/LUN–Properties** page appears.

7   Enter a datastore name.

The datastore name appears in the VI Client, and the label must be unique within the current Virtual Infrastructure instance.

8   Click **Next**.

The **Disk/LUN–Formatting** page appears.

9   If needed, adjust the file system values and capacity you use for the datastore.

By default, the entire free space available on the storage device is offered to you.

10   Click **Next**.

The **Ready to Complete** page appears.

11   Review the datastore configuration information and click **Finish**.

This creates the datastore on the software-initiated iSCSI storage device.

Modifying SAN Storage Systems with ESX Server

3

After you install your iSCSI initiators and storage, you might need to modify your storage system to ensure that it works properly with your ESX Server implementation. This chapter discusses many of the iSCSI storage systems supported in conjunction with VMware ESX Server. For each device, it lists major known potential issues, points to vendor-specific information (if available), or includes information from VMware knowledge base articles.

NOTE   Information in this document is updated only with each release. New information might already be available. Also, other iSCSI storage systems are supported but are not covered in this chapter. Consult the most recent *Storage/SAN Compatibility Guide*, check with your storage vendor, and explore the VMware knowledge base articles.

This chapter discusses the following topics:

- "Setup Overview" on page 66
- "General Considerations" on page 66
- "EMC CLARiiON Storage Systems" on page 67
- "EMC Symmetrix Storage Systems" on page 68
- "HP StorageWorks Storage Systems" on page 69
- "Network Appliance Storage Systems" on page 71
- "EqualLogic Storage Systems" on page 74
- "LeftHand Networks SAN/iQ Storage Systems" on page 75

Setup Overview

VMware ESX Server supports a variety of SAN storage systems in different configurations. Not all storage devices are certified for all features and capabilities of ESX Server, and vendors might have specific positions of support with regard to ESX Server. For the latest information regarding supported storage systems, see the *Storage/SAN Compatibility Guide*.

VMware tests ESX Server with storage systems in the following configurations:

Basic Connectivity. Tests whether ESX Server can recognize and operate with the storage system. This configuration does not allow for multipathing or any type of failover.

HBA Failover. The server is equipped with multiple HBAs connecting to one or more SAN switches. The server is robust to HBA and switch failure only.

Storage Port Failover. The server is attached to multiple storage ports and is robust to storage port failures and switch failures.

Booting from a SAN. The ESX Server host boots from a LUN configured on the SAN rather than from the server itself.

General Considerations

For all storage systems, make sure that the following requirements are met:

- LUNs must be presented to each HBA of each host with the same LUN ID number. If different numbers are used, the ESX Server hosts do not recognize different paths to the same LUN. Because instructions on how to configure identical SAN LUN IDs are vendor-specific, consult your storage documentation for more information.

- Unless specified for individual storage systems discussed in this chapter, set the host type for LUNs presented to ESX Server to Linux or Linux Cluster, if applicable to your storage system. The method ESX Server uses to access the storage system is most compatible with Linux access, however, this can vary depending on the storage system you are using.

■ If you are using VMotion, DRS, or HA, make sure that source and target hosts for virtual machines can see the same LUNs with identical LUN IDs.
SAN administrators might find it counterintuitive to have multiple hosts see the same LUNs because they might be concerned about data corruption. However, VMFS prevents multiple virtual machines from writing to the same file at the same time, so provisioning the LUNs to all required ESX Server system is appropriate.

■ If you do not have CHAP authentication set up on the LUNs that are being accessed, you must also disable CHAP on the ESX Server host. Otherwise, authentication of the storage system fails, although the LUNs have no CHAP requirement.

EMC CLARiiON Storage Systems

EMC CLARiiON storage systems work with ESX Server machines in SAN configurations. Basic configuration steps include:

1 Installing and configuring the storage device.

2 Creating RAID groups.

3 Creating and binding LUNs.

4 Registering the servers connected to the SAN.

5 Creating storage groups that contain the servers and LUNs.

Use the EMC software to perform configuration. For more information, see the EMC documentation.

NOTE This is an active-passive disk array, so related issues described elsewhere in this document apply.

To avoid the possibility of path thrashing, the default multipathing policy is Most Recently Used, not Fixed. The ESX Server system sets the default policy when it identifies the storage system. See "Resolving Path Thrashing" on page 109.

Automatic volume resignaturing is not supported for AX100i and AX150i storage devices. For information on resignaturing, see "VMFS Volume Resignaturing" on page 117.

NOTE To boot from a SAN, choose the active storage processor for the boot LUN's target in the HBA BIOS.

EMC CLARiiON AX100i and AX150i and RDM

On EMC CLARiiON AX100i and AX150i systems, RDMs are supported only if you use the Navisphere Management Suite for SAN administration. Navisphere Express is not guaranteed to configure them properly.

To use RDMs successfully, a given LUN must be presented with the same LUN ID to every ESX Server host in the cluster. The AX100i and AX150i do not do this by default.

Pushing Host Configuration Changes to the Storage System

When you use an AX100i or AX150i storage system, no host agent periodically checks the host configuration and pushes changes to the storage system. The `axnaviserverutil cli` utility is used to update the changes. This is a manual operation that you should perform as needed.

EMC Symmetrix Storage Systems

The following settings are required for ESX Server operations on the Symmetrix networked storage system:

- Common serial number (C)

- Auto negotiation (EAN) enabled

- SCSI 3 (SC3) set (enabled)

- Unique world wide name (UWN)

- SPC-2 (Decal) (SPC2) SPC-2 flag is required

You use EMC software to configure the storage system. For information, see your EMC documentation.

NOTE The ESX Server host considers any LUNs from a Symmetrix storage system that have a capacity of 50MB or less as management LUNs. These LUNs are also known as pseudo or gatekeeper LUNs. These LUNs appear in the EMC Symmetrix Management Interface and should not be used to hold data.

HP StorageWorks Storage Systems

This section includes configuration information for HP StorageWorks storage systems.

For additional information, see the section on VMware ESX Server at the HP ActiveAnswers Web site.

HP StorageWorks MSA

This section describes the setup and configuration steps needed to allow an HP StorageWorks MSA1510i storage system to communicate within an ESX Server environment.

To enable MSA1510i storage systems to communicate with ESX Server

1   Install, connect, and power up the network devices as detailed in the vendor installation document.

2   Obtain the IP address assigned to the MSA1510i controller management port.

    a   Scroll through the messages on the LCD panel until the following message appears: **603 Port MA0 IP <address>**

    b   Record the management port IP address that appears in **Basic MSA1510i information**.

3   From the server or a workstation on the MSA1510i LAN segment, open a Web browser and enter the address obtained in Step 2.

4   When prompted, enter the following default access permissions:

    User name: root

    Password: root

5   When prompted, set a unique user name and password.

6   Using the wizard, complete the following actions:

    a   Storage configuration

        i   Set the Fault Tolerant mode (RAID mode).

        ii   Assign a spare disk for appropriate RAID level.

    b   iSCSI configuration (configure an iSCSI portal)

        i   Select a data port.

        ii   Assign an IP address to the data port.

          iii   VLANs are set up on the switch and are used as one method of controlling access to the storage. If you are using VLANs, enter the VLAN ID to use (0 = not used).

          iv   The wizard suggests a default iSCSI Target Name and iSCSI Target Alias. Accept the default or enter user-defined values.

> **NOTE** To configure the remaining data ports, complete the Initial System Configuration Wizard process, and then use tasks available on the **Configure** tab.

    c   Login Settings

    d   Management Settings

7   Click **Finish** to apply the configuration settings.

> **NOTE** Wizards are available for basic configuration tasks only. Use the **Manage** and **Configure** tabs to view and change your configuration.

After initial setup, perform the following tasks to complete the configuration:

- Create an array.

- Create a logical drive.

- Create a target.

- Create a portal group.

- Associate or assign the portals created using the wizard with the portal group created.

- Map logical drives to the target.

- Add initiators (initiator IQN name and alias).

- Update the ACLs of the logical drives to provide access to initiators (select the list of initiators to access the logical drive).

See the *MSA 1510i Configuration Guide*.

HP StorageWorks EVA

The two types of HP StorageWorks EVA systems are EVA_GL, an active-passive system, and EVA_XL, an active-active system.

Set the connection type to **Custom** when you present a LUN to an ESX Server host. The value is one of the following:

■ For HP EVAgl 3000/5000 (active-passive), use the `000000002200282E` host mode type.

■ For HP EVAgl firmware 4.001 (active-active firmware for GL series) and above, use the `VMware` host mode type.

■ For EVA4000/6000/8000 active-active arrays with firmware earlier than 5.031, use the `000000202200083E` host mode type.

■ For EVA4000/6000/8000 active-active arrays with firmware 5.031 and later, use the `VMware` host mode type.

Otherwise, EVA systems do not require special configuration changes to work with an ESX Server system.

For more details, check VMware Infrastructure 3, HP StorageWorks Best Practices at the HP Web site.

Network Appliance Storage Systems

This section describes the issues and steps associated with allowing a Network Appliance storage system to communicate within an ESX Server environment.

Multipathing

When you set up multipathing between two QLogic HBAs and multiple ports on a Network Appliance storage system, give the two QLogic HBAs different dynamic or static discovery addresses to connect to the storage.

The Network Appliance storage system only permits one connection for each target and each initiator. Attempts to make additional connections cause the first connection to drop. Therefore, single QLogic HBAs should not attempt to connect to multiple IP addresses associated with the same Network Appliance target.

Setting LUN Type and Initiator Group Type

Set the appropriate LUN type and initiator group type for the storage system:

■ **LUN type** – VMware (if VMware type is not available, use Linux).

■ **Initiator group type** – VMware (if VMware type is not available, use Linux).

Provisioning Storage

You must provision storage, using either FilerView or CLI.

To provision storage by using FilerView storage management

1   Log in to Network Appliance storage system management (FilerView).

2   Create a volume.

   a   Select **Volumes** and click **Add**. Click **Next**.

   b   Select **Flexibility** (Default) or **Traditional**, then click **Next**.

   c   Enter a **Volume Name**, select a **Language**, and click **Next**.

   d   Enter values for **Containing Aggregate**, **Total Volume Size**, and **Space Guarantee** and click **Next**.

   e   Click **Commit** to create the volume.

3   Create LUNs.

   a   Select **LUNs** and click **Add**.

   b   Enter the following:

      i   **Path**: Enter a path, for example, /vol/vol1/lun1.

      ii   **LUN Protocol Type**: VMware.

      iii   **Description**: A brief description.

      iv   **Size and Unit**: Enter a size, for example, 10GB and select **Space Reserved**.

4   Create an initiator group.

   a   Select **LUNs>Initiator Group** and click **Add**.

   b   Enter the following:

      i   **Group Name**: Enter a group name

      ii   **Type**: Choose **iSCSI**.

        iii   **Operating System**: Enter **VMware**

        iv   **Initiators**: Enter fully qualified initiator names. If there is more than one initiator, each initiator has to be separated with a 'return carriage'.

   c   Click **Add**.

5   Map the LUN to the initiator group.

   a   Select **LUNs** and click **Manage**. A LUNs list appears.

   b   From this list, click the label on the **Maps** row for the specific LUNs.

   c   Click **Add Groups to Map**.

   d   Select the initiator group and click **Add**.

   e   When prompted, enter the LUN ID (any number from 0 to 255) and click **Apply**.

You can also provision the storage using a CLI.

To provision storage by using a CLI

1   Use a CLI to create an aggregate if required.

```
aggr create <vmware-aggr> <number of disks>
```

2   Create a flexible volume.

```
vol create <aggregate name> <volume size>
```

3   Create a Qtree to store each LUN.

```
qtree create <path>
```

4   Create a LUN.

```
lun create -s <size> -t vmware <path>
```

5   Create an initiator group.

```
igroup create -f -t vmware <igroup name>
```

6   Map the LUN to the initiator group you created.

```
lun map (<path>) <igroup name> <LUN ID>
```

For additional information on using Network Appliance Storage with VMware technology, see the following Network Appliance documents:

■ Network Appliance & VMware ESX Server: Instantaneous Backup & Recovery with NetApp Snapshot Technology at http://www.netapp.com/library/tr/3428.pdf.

■ Technical Case Study: Using a Network Appliance SAN with VMware to Facilitate Storage and Server Consolidation at http://www.netapp.com/library/tr/3401.pdf.

EqualLogic Storage Systems

To set up your EqualLogic storage systems to work in an ESX Server implementation, you must address the following issues:

Multipathing. No special setup is needed because EqualLogic storage systems support storage-proccesor failover that is transparent to iSCSI. Multiple iSCSI HBAs or NICs can connect to the same target or LUN on the storage side. However, normal restrictions on ESX Server configuration apply, for example NIC teaming must be correctly set up.

Creating iSCSI LUNs. From the EqualLogic web portal, right-click **Volumes**, and then select **Create Volume**.

Enable ARP redirection on hardware iSCSI HBAs. For more information on enabling ARP redirection, see "esxcfg-hwiscsi Utility" on page 126.

For more information about configuring and using EqualLogic storage systems, see the vendor's documentation.

NOTE You must add the service console and VMkernel IP addresses being used for iSCSI access to the IP address ACL on your EqualLogic storage systems. If you are using CHAP authentication or initiator name-based ACLs, however, this is not necessary.

LeftHand Networks SAN/iQ Storage Systems

SAN/iQ SANs support ESX Server iSCSI connections from a software initiator and HBAs.

Basic Configuration

Basic configuration steps include:

1   Install SAN/iQ storage nodes.

2   Create SAN/iQ management groups and clusters.

3   Create volumes.

4   Assign volumes to authentication groups and volume lists.

5   Enable ARP redirection on hardware iSCSI HBAs.
    See "esxcfg-hwiscsi Utility" on page 126.

As a best practice, configure virtual IP load balancing in SAN/iQ for all ESX Server authentication groups.

Automatic Volume Resignaturing

Enable automatic volume resignaturing for SAN/iQ storage devices to allow access to SAN/iQ snapshots and remote copies.

For more information on configuring LeftHand Networks SANs for VMware Infrastructure 3, see the *VMware Field Guide for SAN/iQ-Powered SANs*.

Booting from a SAN with ESX Server Systems

<div style="text-align: right; font-size: 3em;">4</div>

This chapter discusses the following topics:

-
-

Booting from a SAN Overview

You can boot from a SAN only with ESX Server 3 and with hardware iSCSI. When you set up your system to boot from a SAN, the boot image is not stored on the ESX Server system's local disk, but instead is stored on a SAN LUN, as Figure 4-8 shows.

Figure 4-8. How Booting from a SAN Works
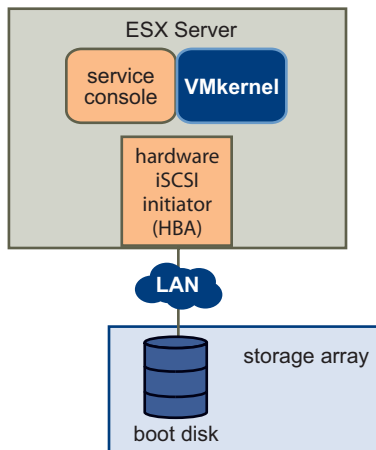
Benefits of Booting from a SAN

When booting from a SAN, the operating system is installed on one or more LUNs in the SAN storage system. The servers are informed about the boot image location. When the servers start, they boot from the LUNs on the SAN storage system.

NOTE   When you boot from a SAN in conjunction with a VMware ESX Server system, each server must have its own boot LUN.

Booting from a SAN provides numerous benefits, including:

■   **Cheaper servers** – Servers can be more dense and run cooler without internal storage.

■   **Easier server replacement** – You can replace servers and have the new server point to the old boot location.

■   **Less wasted space**.

■   **Easier backup processes** – The system boot images in the SAN can be backed up as part of the overall SAN backup procedures.

■   **Improved management** – Creating and managing the operating system image is easier and more efficient.

Deciding to Boot From a SAN

Before you consider how to set up your system for booting from a SAN, decide whether it makes sense for your environment.

Boot from a SAN:

■   If you do not want to handle maintenance of local storage.

■   If you need easy cloning of service consoles.

■   In diskless hardware configurations, such as on some blade systems.

Do not boot from a SAN if you risk I/O contention between the service console and VMkernel.

Enabling Booting from a SAN

Enabling your ESX Server host to boot from a SAN requires the completion of a number of tasks.

To enable booting from a SAN

1 Review any vendor configuration recommendations that apply to the storage system or the server booting from SAN.

2 Configure the hardware elements of your storage network, including:

   ■ SAN—see "Preparing the SAN" on page 79.

   ■ HBAs—see "Configuring iSCSI HBAs to Boot from a SAN" on page 81

3 Configure ACLs on your storage system.

   Proper access control on the storage systems is important when an ESX Server host is booting from iSCSI.

   ■ Boot LUNs should only be visible to the server using that LUN to boot. No other server or system on the SAN should be permitted to see that boot LUN.

   ■ Multiple ESX Server hosts can share a diagnostic partition. ACLs on the storage systems can allow you to do this.

4 Choose the location for the diagnostic partition.

   Diagnostic partitions can be put on the same LUN as the boot partition. Core dumps are stored in diagnostic partitions.

5 Set up your ESX Server to boot from CD-ROM first because the VMware installation CD is in the CD-ROM drive.

   To achieve this, change the system boot sequence in your system BIOS setup.

Preparing the SAN

Before you configure the iSCSI HBAs to boot from a SAN, first prepare your storage area network by checking the cabling and switch wiring and configuring the storage system.

To prepare the SAN

1   Connect network cables, referring to any cabling guide that applies to your setup.

2   Ensure IP connectivity between your storage system and server.
    This includes proper configuration of any routers or switches on your storage network. Storage systems must be able to ping the iSCSI HBAs in your ESX Server hosts.

3   Configure the storage system:

    a   Create a volume (or LUN) on the storage system for ESX Server to boot from.

    b   Configure the storage system so that the ESX Server system has access to the assigned LUN. This could involve updating ACLs with the IP addresses, iSCSI IQN names, and the CHAP authentication parameter you use on the ESX Server system. On some storage systems, in addition to providing access information for the ESX Server host, you must also explicitly associate the assigned LUN with the host.

    c   Ensure that the LUN is presented to the ESX Server system as LUN 0. (On storage systems that present volumes as multiple targets rather than multiple LUNS, the volumes are always presented as LUN 0).

    d   Ensure that no other system has access to the configured LUN.

    e   Record the iSCSI (IQN) name and IP addresses of the targets assigned to the ESX Server host. You need this information to configure your iSCSI HBA.

CAUTION   If you use scripted installation to install ESX Server when booting from a SAN, you need to take special steps to avoid unintended data loss. See VMware knowledge base article 1540 at
http://www.vmware.com/support/kb/enduser/std_adp.php?p_faqid=1540.

Configuring iSCSI HBAs to Boot from a SAN

This section discusses how to configure a QLogic iSCSI HBA for booting from a SAN.

On a system set up to boot from a SAN:

- The system BIOS must designate the iSCSI card as the boot controller.

- The BIOS must be enabled on the iSCSI HBA to locate the target boot LUN.

To enable the QLogic iSCSI HBA to Boot from a SAN

1   During server POST, press Crtl+q to enter the QLogic iSCSI HBA configuration menu.

2   Select the I/O port to configure.
    By default, the Adapter Boot mode is set to Disable.

3   Configure the HBA.

   a   From the **Fast!UTIL Options** menu, select **Configuration Settings**>**Host Adapter Settings**.

   b   Configure the following settings for your host adapter: initiator IP address, subnet mask, gateway, initiator iSCSI name, and CHAP (if required).

4   Configure iSCSI Boot Settings.

a   From the **Fast!UTIL Options** menu, select **Configuration Settings** > **iSCSI Boot Settings**.

b   Before you can set SendTargets, set Adapter Boot mode to **Manual**.

```
                        QLogic Fast!UTIL
                    ═══Selected Adapter═══
   Adapter  Boot Mode I/O Address Slot Bus Device Function MAC Address
   QLA405x  Disable   E800        02   0F  03     1        00-C0-DD-07-66-6A



                   ═iSCSI Boot Settings═
   Boot Device     Lun    Target   │ Boot Mode │ Name
   Primary:        0      0.0.0.0  │           │
   Alternate:      0      0.0.0.0  │ Disable   │
                                   │ Manual    │
   Adapter Boot Mode: Disable      │ DHCP      │
   Primary Boot Device Settings
   Alternate Boot Device Settings
   DHCP Boot Settings

              Press "C" to clear selected boot device or <F1>
              to display complete iSCSI name of boot device
```

c   Select **Primary Boot Device Settings**.

   i   Enter the discovery **Target IP** and **Target Port**.

   ii   You can leave the **Boot LUN** and **iSCSI Name** fields blank if only one iSCSI target and one LUN are at the specified address to boot from. Otherwise, you must specify these fields to ensure that you do not boot from a volume for some other system. After the target storage system is reached, these fields will be populated after a rescan.

   iii Save changes.
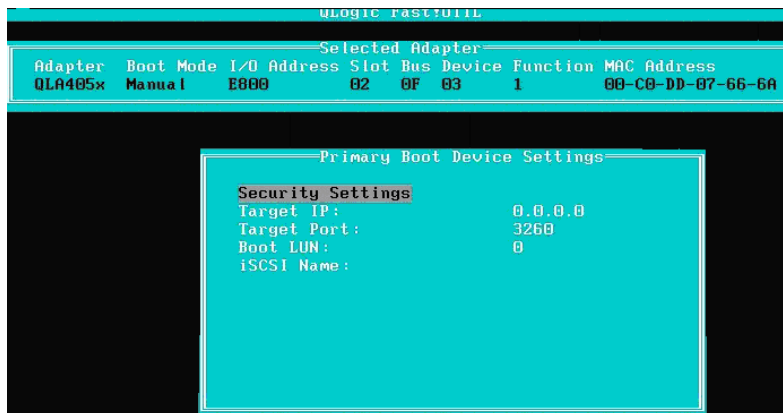
d   From the **iSCSI Boot Settings** menu, select the primary boot device. An auto rescan of the HBA is made to find new target LUNS.

e   Select the iSCSI target.

NOTE   If more then one LUN exists within the target, you can choose a specific LUN ID by pressing **Enter** after you locate the iSCSI device.
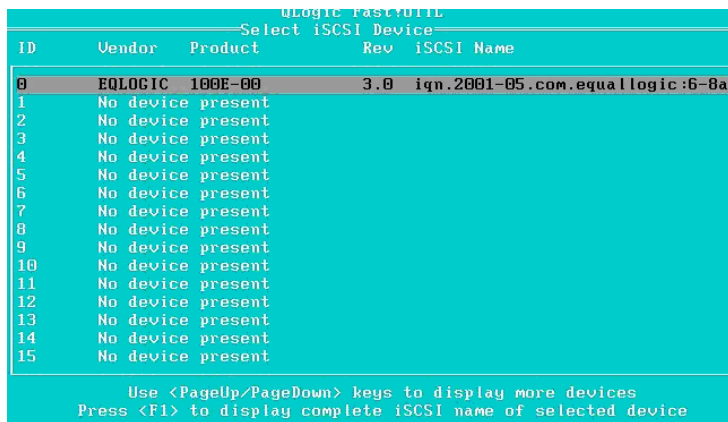
f Return to the **Primary Boot Device Setting** menu. After the rescan, the **Boot LUN** and **iSCSI Name** fields are populated. Change the value of **Boot LUN** to the desired LUN ID.

5 Save your changes and restart the system.

For more information and more up-to-date details about QLogic host adapter configuration settings, see the QLogic host adapter read me file at the QLogic web site.

Managing ESX Server Systems That Use SAN Storage

5

This chapter can help you manage your ESX Server system, use SAN storage effectively, and perform troubleshooting.

This chapter discusses the following topics:

Issues and Solutions

Table 5-1 lists the issues that are most frequently encountered and either explains how to resolve them or points to the location where the issue is discussed.

Table 5-1. Issues and Solutions

Issue	Solution
A LUN is not visible in the VI Client.	See "Resolving Display Issues" on page 89.
Understand how path failover is performed or change how path failover is performed.	The VI Client allows you to perform these actions. See "Multipathing" on page 94.
View or change the current multipathing policy or preferred path, or disable or enable a path.	The VI Client allows you to perform these actions. See "Multipathing" on page 94.
Increase the Windows disk timeout to avoid disruption during failover.	See "Setting Guest Operating System Timeout" on page 101.
The server cannot access a LUN, or access is slow.	Path thrashing might be the problem. See "Resolving Path Thrashing" on page 109.
You added a new LUN or a new path to storage and want to see it in the VI Client.	Rescan. See "Using Rescan" on page 90.

Guidelines for Avoiding SAN Problems

Follow these guidelines to avoid potential problems with your SAN configuration:

- Place only one VMFS volume on each LUN. Multiple VMFS volumes on one LUN is not recommended.

- Do not change the path policy the system sets for you unless you understand the implications of making such a change. In particular, working with an active-passive array and setting the path policy to **Fixed** can lead to path thrashing.

Getting Information

This section explains how to find information about HBAs, status, multipathing, and so on. If you experience problems when performing these tasks, see "Resolving Display Issues" on page 89.

Viewing HBA Information

Use the VI Client to display all available storage adapters and their information.

To see a list of HBA types

1   Select the host to see the HBAs of and click the **Configuration** tab.

You can view a list of all storage devices from the **Summary** tab. However, you cannot see details or manage the device from there.

2   In the **Hardware** panel, choose **Storage Adapters**.

The list of storage adapters appears. You can select each adapter for additional information.

Viewing Datastore Information

Use the VI Client to display all formatted datastores and review details about a specific datastore.

To view all storage devices and details about them

1 Select the host to see the storage devices of and click the **Configuration** tab.

NOTE   The **Service Console Resources** link under the Software heading appears for ESX Server 3 only. ESX Server 3i does not provide a service console.

2 In the **Hardware** panel, choose **Storage**.

The list of datastores (volumes) appears in the **Storage** panel. The display shows the whole VMFS for the selected host. Only storage that was formatted with VMFS is included in the display.

3 To view details about any datastore, select it.

The **Details** panel displays additional information. This includes the location and capacity, number of paths, path policy, and properties. It also includes extent information. An extent is a VMFS-formatted partition (a piece of a LUN). For example, vmhba 0:6:0 is a LUN, and vmhba 0:6:0:1 is a partition. One VMFS volume can have multiple extents.

NOTE   The abbreviation vmhba refers to the HBA on the ESX Server system, not to the SCSI controller that the virtual machines use.

4 Click **Properties** to view and change properties.

Resolving Display Issues

This section discusses how to troubleshoot common status and visibility issues.

If you are using an AX100i or AX150i storage system, inactive connections can cause display problems.

Understanding LUN Naming in the Display

In the VI Client, a LUN appears as a sequence of three or four numbers, separated by colons:

`<SCSI HBA>:<SCSI target>:<SCSI LUN>:<disk partition>`

If the last number is 0 or not displayed, the name refers to the entire LUN.

The first three numbers in an ESX device name might change, but still refer to the same physical device. For example, `vmhba1:2:3` represents SCSI LUN3, attached to SCSI target 2, on SCSI HBA 1. When the ESX Server system is rebooted, the device name for LUN 3 might change to `vmhba1:1:3`. The numbers have the following meaning:

■ The first number, the SCSI HBA, changes if an iSCSI network outage occurs at the time the system is booted or rescanned and ESX is required to access the physical device over a different SCSI HBA.

■ The second number, the SCSI target, changes if a change occurs in the mappings in the iSCSI targets visible to the ESX Server host.

■ The third number, the SCSI LUN, never changes.

Resolving Issues with LUNs That Are Not Visible

You can use the VI Client to view LUNs.

If the display (or output) differs from what you expect, check the following:

Cable connectivity. If you do not see a port, the problem could be cable connectivity or routing. Check the cables first. Ensure that cables are connected to the ports and a link light indicates that the connection is good. If each end of the cable does not show a good link light, replace the cable.

Routing . Controls connectivity between different subnets on your Ethernet configuration. If your ESX Server system and iSCSI storage are not on the same subnet, ensure that appropriate routing exists between the subnets. Also, ensure that the subnet mask and gateway address are set correctly on the iSCSI storage and the iSCSI initiator in the ESX Server host.

Access Control. If the expected LUNs do not appear after rescan, access control might not be configured correctly on the storage system side:

■ If CHAP is configured, ensure that it is enabled on the ESX Server host and matches the storage system setup.

■ If IP-based filtering is used, ensure that the iSCSI HBA or the VMkernel port group IP address and service console IP address are allowed.

■ If you are using initiator name-based filtering, ensure that the name is a qualified iSCSI name and matches the storage system setup.

For booting from a SAN, ensure that each ESX Server host sees only required LUNs. Do not allow any ESX Server host to see any boot LUN other than its own. Use storage system software to make sure that the ESX Server host can see only the LUNs that it is supposed to see.

Ensure that the **Disk.MaxLUN** and **Disk.MaskLUNs** settings allow you to view the LUN you expect to see. See "Changing the Number of LUNs Scanned by Using Disk.MaxLUN" on page 92 and "Masking LUNs by Using Disk.MaskLUNs" on page 93.

Storage processor. If a storage system has more than one storage processor, make sure that the SAN switch has a connection to the SP that owns the LUNs you want to access. On some storage systems, only one SP is active and the other SP is passive until a failure occurs. If you are connected to the wrong SP (the one with the passive path) you might not see the expected LUNs, or you might see the LUNs but get errors when trying to access them.

Software iSCSI Network Configuration. The software iSCSI initiator in ESX Server requires that a service console network port and a VMkernel network port have access to the iSCSI storage. The software initiator uses the service console for iSCSI discovery and error handling. It uses the VMkernel for data transfer between the ESX Server system and the iSCSI storage. See "Networking Configuration for Software iSCSI Storage" on page 50.

Using Rescan

Perform a rescan each time you:

■ Create new LUNs on a SAN.

■ Change the LUN masking on an ESX Server host storage system.

■ Reconnect a cable.

- Make a change to a host in a cluster.

- Change CHAP settings or add new discovery addresses.

NOTE Do not rescan when a path is unavailable. If one path fails, the other takes over and your system continues to be fully functional. If, however, you rescan at a time when a path is not available, the ESX Server host might remove the path from its list of paths to the device. The path cannot be used by the ESX Server host until the next time a rescan is performed while the path is active.

To perform a rescan

1   In the VI Client, select a host and click the **Configuration** tab.

2   In the **Hardware** panel, choose **Storage Adapters** and click **Rescan** above the **Storage Adapters** panel.

    You can also right-click an individual adapter and click **Rescan** to rescan just that adapter.

Removing Datastores

Using the VI Client, you can remove a datastore from being used as storage for virtual machines. You cannot remove targets in software iSCSI, except by rebooting.

To remove a datastore

1   Turn off all virtual machines that use the datastore being removed.

2   Select and remove each virtual machine from the inventory by right-clicking the virtual machine and clicking **Remove from Inventory**.

3   Click the **Configuration** tab and click **Storage** to display all storage devices.

4   Select the datastore to remove and click **Remove**.

At this point the datastore is removed and the inventory should be refreshed automatically.

NOTE After you remove a datastore from an ESX Server host, mask or remove the LUN from the storage system and rescan with the VI Client to prevent ESX Server from discovering the LUN.

Advanced LUN Display Configuration

This section discusses a number of advanced configuration options.

Changing the Number of LUNs Scanned by Using Disk.MaxLUN

By default, the VMkernel scans for LUN 0 to LUN 255 for every target (a total of 256 LUNs). You can change the Disk.MaxLun parameter to change this number. This change might improve LUN discovery speed.

NOTE You cannot discover LUNs with a LUN ID number that is greater than 255.

Reducing the value can shorten rescan time and boot time. The time to rescan LUNs depends on several factors, including the type of storage system and whether sparse LUN support is enabled. See "Changing Sparse LUN Support by Using DiskSupportSparseLUN" on page 94.

To change the value of Disk.MaxLUN

1   In the VI Client inventory panel, select the host, click the **Configuration** tab and click **Advanced Settings**.

2   Select **Disk**.

3   Scroll down to **Disk.MaxLUN**, change the existing value to the value of your choice, and click **OK**.
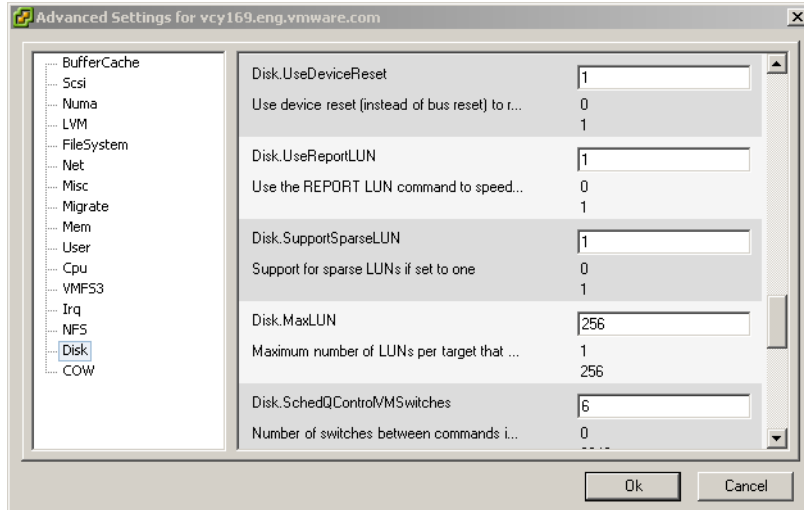
Masking LUNs by Using Disk.MaskLUNs

The **Disk.MaskLUNs** parameter allows you to mask specific LUNs on specific HBAs. Masked LUNs are not touched or accessible by the VMkernel, even during initial scanning.

Use this option to prevent the ESX Server system from accessing some iSCSI LUNs without the storage system's masking mechanisms.

To change the value of Disk.MaskLUNs

1   In the VI Client inventory panel, select the host, click the **Configuration** tab and click **Advanced Settings**.

2   Select **Disk**.

3   Scroll down to **Disk.MaskLUNs,** change the existing value to the value of your choice, and click **OK**.

> ⚠ **CAUTION** If a target, LUN, or vmhba number changes because of a server or SAN reconfiguration, the incorrect LUN might be masked or exposed.

Changing Sparse LUN Support by Using DiskSupportSparseLUN

By default, the VMkernel is configured to support sparse LUNs, that is, a case where some LUNs in the range 0 to N-1 are not present, but LUN N is present. For example, if a storage system presents LUNs numbering 0, 6, and 23, but no LUNs with any of the numbers between these.

If all LUNs are sequential, you can change the **Disk.SupportSparseLUN** parameter. This change decreases the time needed to scan for LUNs.

To change the value of Disk.SupportSparseLUN

1   In the VI Client inventory panel, select the host, click the **Configuration** tab, and click **Advanced Settings**.

2   In the Advanced Settings dialog box, select **Disk**.

3   Scroll down to **Disk.SupportSparseLUN**, change the value to **0**, and click **OK**.

Multipathing

SAN implementations with a high number of LUNs and paths to those LUNs can cause ESX Server to run out of resources before all of the paths are enumerated. This scenario prevents ESX Server from seeing all of the paths to the storage. To avoid this situation, reduce the path count to the LUNs.

For an introduction to multipathing concepts, see "Path Management and Failover" on page 32.

Viewing the Current Multipathing State

You can use the VI Client to view the current multipathing state.

To view the current multipathing state

1   In the VI Client inventory panel, select a host and click the **Configuration** tab.

2   In the **Storage** panel, select one of the datastores.

Information about that datastore appears in the **Details** panel.

3   To view additional information, or to change the multipathing policy, select **Properties** above the **Details** panel.

4   In the **Extents** panel, select the extent to view or change information.

The **Extent Device** panel displays information about the extent, the path selection algorithm, the available paths, and the active path.

The display includes information on the status of each path to the device extent. The following path information appears:

■   **Active** – The path is working and is the current path being used to transfer data.

■   **Disabled** – The path is disabled and no data can be transferred.

- **Standby** – The path is working but is not currently used for data transfer.

- **Dead** – The software cannot connect to the disk through this path.

5 If you are using path policy **Fixed** and want to see which path is the **Preferred** path, click **Manage Paths**.

The preferred path is marked with an asterisk (*) in the fourth column.
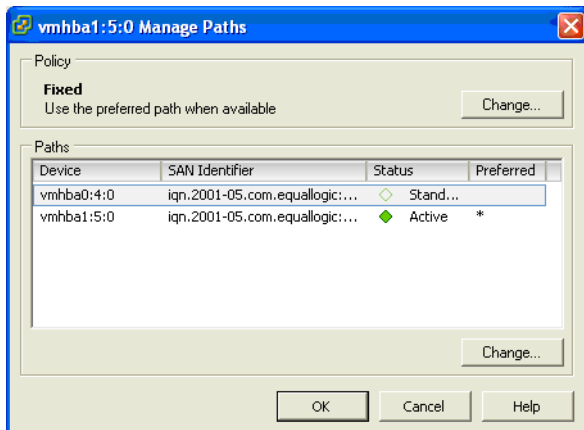
Table 5-2 summarizes how the behavior of an ESX Server system changes, depending on the type of storage system and the failover policy.

Table 5-2. Path Policy Effects

Policy/Controller	Active-Active	Active-Passive
Most Recently Used	Administrator action is required to fail back after path failure.	Administrator action is required to fail back after path failure.
Fixed	VMkernel resumes using the preferred path when connectivity is restored.	VMkernel attempts to resume using the preferred path. This can cause path thrashing or failure because another SP now owns the LUN. See "Resolving Path Thrashing" on page 109.

NOTE   iSCSI storage systems that provide transparent failover behave as active-active.

Active Paths

ESX Server does not typically perform I/O load balancing across paths for a given LUN. At any one time, only a single path is used to issue I/O to a given LUN. This path is known as the active path.

■   If the path policy of a LUN is set to Fixed, ESX Server selects the path marked as Preferred as the active path.

 If the preferred path is disabled or unavailable, the ESX Server system uses an alternative working path as the active path.

■   If the path policy of a LUN is set to Most Recently Used, the ESX Server host selects an active path to the LUN that prevents path thrashing. The preferred path designation is not considered.

NOTE   In some SAN terminology, the term *active* means any path that is available for issuing I/O to a LUN. From the ESX Server host's point of view, the term active means the one and only path that the ESX Server host is using to issue I/O to a LUN.

Setting a LUN Multipathing Policy

The following multipathing policies are currently supported:

■   **Fixed** – The ESX Server host always uses the preferred path to the disk when that path is available. If it cannot access the disk through the preferred path, it tries the alternative paths. Fixed is the default policy for active-active storage devices.

■   **Most Recently Used** – The ESX Server host uses the most recent path to the disk until this path becomes unavailable. That is, the ESX Server host does not automatically revert back to the preferred path. Most Recently Used is the default policy for active-passive storage devices and is required for those devices.

■   **Round Robin** – The ESX Server host uses an automatic path selection rotating through all available paths. In addition to path failover, round robin supports load balancing across the paths.

NOTE   Round robin load balancing is experimental and not supported for production use. For more information, see the *Round-Robin Load Balancing* white paper.

The ESX Server host sets the multipathing policy according to the make and model of the storage system it detects. If the detected storage system is not supported, it is treated as active-active. For a list of supported storage systems, see the *Storage/SAN Compatibility Guide*.

NOTE   VMware recommends that you not change Most Recently Used to Fixed. The system sets this policy for the storage systems that require it.

To set the multipathing policy by using a VI Client

1   In the VI Client inventory panel, select the host and click the **Configuration** tab.

2   In the **Hardware** panel, select **Storage**.

3   Select the datastore to change the multipathing policy for, and click **Properties** in the **Details** panel.

4   In the **Extent** panel, select the device to make the change for, and click **Manage Paths** in the **Extent Device** panel on the right.

5   In the Manage Paths dialog box, click **Change**.

6   Select the multipathing policy in the dialog box that appears and click **Done**.

NOTE   For active-passive storage devices, VMware recommends Most Recently Used.

Disabling and Enabling Paths

You can temporarily disable paths for maintenance or other reasons. You can do so using the VI Client.

To disable or enable a path

1   In the VI Client inventory panel, select the host and click the **Configuration** tab.

2   In the **Hardware** panel, select **Storage**.

3   Select the device to disable a path for, and click **Properties** in the **Details** panel.

4   In the **Extent** panel, select the device to make the change for, and click **Manage Paths** in the **Extent Device** panel on the right.

5   Click **Change** in the **Paths** panel of the **Manage Paths** dialog box and click **Disabled** to disable the path or **Enabled** to enable the path.
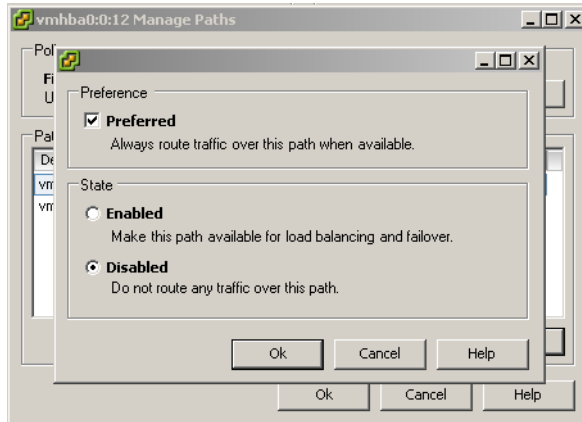
Setting the Preferred Path (Fixed Path Policy Only)

If you are using a Fixed path policy, the server always uses the preferred path when available.

To set the preferred path

1   In the VI Client inventory pane, select the host and click the **Configuration** tab.

2   In the Hardware panel, select **Storage**.

3   Select the device to set a preferred path for and click **Properties** in the **Details** panel.

4   In the **Extent** panel, select the device to make the change to, and click **Manage Paths** in the **Extent Device** panel on the right.

5   Select the path to make the preferred path for and click **Change**.

6   In the **Preference** pane, click **Preferred**.

If **Preferred** is not an option, make sure that the path policy is **Fixed**.

7   Click **OK** and click **OK** again to exit the dialog boxes.

Path Management and Manual Load Balancing

Balancing loads among available paths improves performance. You can configure your system to use different paths to different LUNs by changing the preferred path for the HBAs. You can do this only for active-active SPs, and requires that you have the path policy set to **Fixed**.

If a path fails, the surviving paths carry all the traffic. Path failover might take a minute or more, because the network might converge with a new topology to try to restore service. This delay is necessary to allow the SAN to stabilize its configuration after topology changes or other network events.

The following example demonstrates how manual load balancing is performed.

When you use an active-active array, you can set up your system for load balancing. Assume the following setup, shown in Figure 5-9:

- Active-Active SPs
- An ESX Server system
- Two iSCSI HBAs

Figure 5-9. Manual Load Balancing
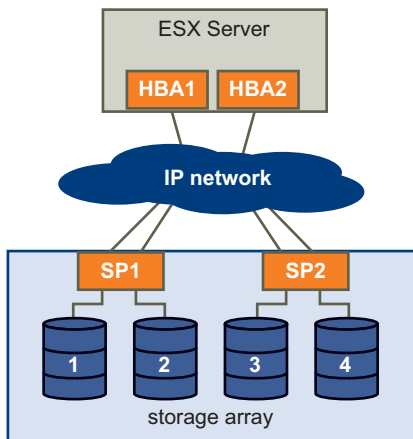
For load balancing, set the preferred paths as follows.

- LUN 1: vmhba1:1:1
- LUN 2: vmhba2:1:2
- LUN 3: vmhba1:2:3
- LUN 4: vmhba2:2:4

Path Failover

Path failover refers to situations when the active path to a LUN is changed from one path to another, usually because of some SAN component failure along the current path. A server usually has one or two HBAs and each HBA sees one or two storage processors on a given SAN storage system. You can determine the active path—the path that the server currently uses—by looking at the LUN's properties.

Figure 5-10. Active and Standby Paths
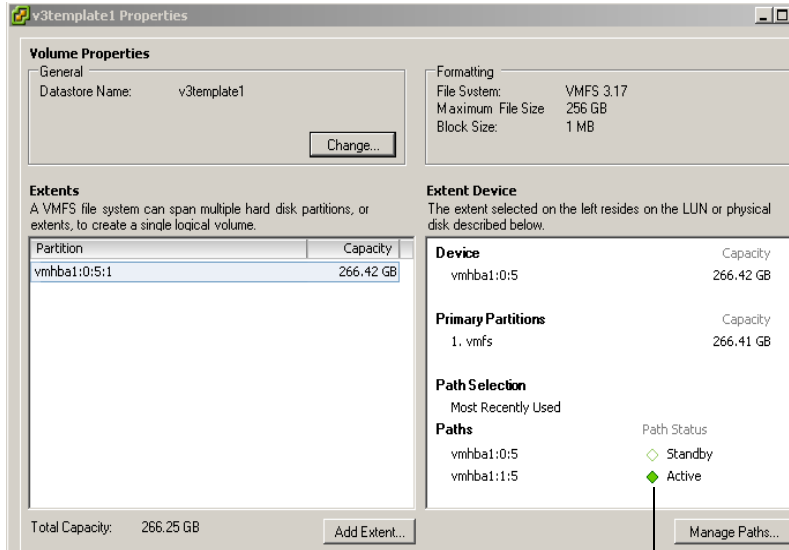
Active and standby paths

When a network cable is pulled, I/O might pause for 30 to 60 seconds until the iSCSI driver determines that the link is unavailable and until failover occurs. As a result, the virtual machines (with their virtual disks installed on SAN storage) can appear unresponsive. If you attempt to display the host, its storage devices, or its adapter, the operation might appear to stall. After failover is complete, I/O resumes normally.

In case of multiple breakages, all connections to SAN storage devices might be lost. If none of the connections to the storage device is working, some virtual machines might encounter I/O errors on their virtual SCSI disks.

Setting Guest Operating System Timeout You might want to increase the standard disk timeout value so that a Windows guest operating system is not extensively disrupted during failover.

For Windows 2000 and Windows Server 2003 guest operating systems, you can set operating system timeout by fusing the registry.

To set operating system timeout for Windows servers

1   Back up your Windows registry.

2   Select **Start**>**Run**, type **regedit.exe** and click **OK**.

3   In the left-panel hierarchy view, double-click **HKEY_LOCAL_MACHINE**, then **System**, then **CurrentControlSet**, then **Services**, and then **Disk**.

4   Select the **TimeOutValue** and set the data value to **x03c** (hexadecimal) or **60** (decimal).

    After you make this change, Windows waits at least 60 seconds for delayed disk operations to complete before it generates errors.

5   Click **OK** to exit the **Registry Editor**.

VMkernel Configuration

When you install your ESX Server system, decide where to place different storage elements such as the root (`/`) and `/boot` partitions of the service console (ESX Server 3 only).

Sharing Diagnostic Partitions

When you use a hardware iSCSI initiator, if your ESX Server host has a local disk, that disk is most appropriately used for the diagnostic partition. One reason is that if an issue with remote storage causes a core dump, the core dump is lost and resolving the issue becomes more difficult.

However, for diskless servers that boot from a SAN, multiple ESX Server systems can share one diagnostic partition on a SAN LUN. If more than one ESX Server system is using a LUN as a diagnostic partition, that LUN must be configured so that all the servers can access it.

Each server needs 100MB of space, so the size of the LUN determines how many servers can share it. Each ESX Server system is mapped to a diagnostic slot. VMware recommends at least 16 slots (1600MB) of disk space if servers share a diagnostic partition.

If only one diagnostic slot is on the device, all ESX Server systems sharing that device map to the same slot. This can easily create problems. If two ESX Server systems perform a core dump at the same time, the core dumps are overwritten on the last slot on the diagnostic partition.

If you allocate enough memory for 16 slots, it is unlikely that core dumps are mapped to the same location on the diagnostic partition, even if two ESX Server systems perform a core dump at the same time.

Avoiding and Resolving SAN Problems

This section gives some tips for avoiding and resolving problems with your SAN configuration:

- Document everything. Include information about configuration, access control, storage, switch, server and iSCSI HBA configuration, software and firmware versions, and storage cable plan.

- Plan for failure:

  - Make several copies of your topology maps. For each element, consider what happens to your SAN if the element fails.

  - Cross off different links, switches, HBAs and other elements to ensure you did not miss a critical failure point in your design.

⚠ **CAUTION**   The installer lets you erase any accessible disks, including SAN LUNs in use by other servers.

- Ensure that the iSCSI HBAs are installed in the correct slots in the ESX Server host, based on slot and bus speed. Balance PCI bus load among the available busses in the server.

- Become familiar with the various monitor points in your storage network, at all visibility points, including ESX Server performance charts, Ethernet switch statistics, and storage performance statistics.

Optimizing SAN Storage Performance

The major factors for optimizing a typical iSCSI environment are storage system performance, server performance and network performance. If the network environment is properly configured, the iSCSI components should provide adequate throughput and low enough latency for iSCSI initiators and targets. If the network is congested and links, switches or routers are saturated, iSCSI performance suffers and might not be adequate for ESX Server environments.

Storage System Performance

If issues occur with storage system performance, consult your storage system vendor's documentation for any relevant information.

When you assign LUNs, remember that you can access each LUN through a number of ESX Server hosts, and that a number of virtual machines can run on each host. One LUN used by an ESX Server host can service I/O from many different applications running on different operating systems. Because of this diverse workload, the RAID group that contains the ESX Server LUNs should not include LUNs that other hosts use that are not running ESX Server for I/O intensive applications.

Enable read caching and write caching.

Load balancing is the process of spreading server I/O requests across all available SPs and their associated host server paths. The goal is to optimize performance in terms of throughput (I/O per second, megabytes per second, or response times).

SAN storage systems require continual redesign and tuning to ensure that I/O is load balanced across all storage system paths. To meet this requirement, distribute the paths to the LUNs among all the SPs to provide optimal load balancing. Close monitoring indicates when it is necessary to manually rebalance the LUN distribution. See "Path Management and Manual Load Balancing" on page 100 for an example.

Tuning statically balanced storage systems is a matter of monitoring the specific performance statistics (such as I/O operations per second, blocks per second, and response time) and distributing the LUN workload to spread the workload across all the SPs.

NOTE Dynamic load balancing is not currently supported with ESX Server.

Server Performance

Ensuring optimal server performance requires looking at a number of factors. Each server application must have access to its designated storage with the following items:

- High I/O rate (number of I/O operations per second)

- High throughput (megabytes per second)

- Minimal latency (response times)

Because each application has different requirements, you can meet these goals by choosing an appropriate RAID group on the storage system. To achieve performance goals, perform the following tasks:

■　Place each LUN on a RAID group that provides the necessary performance levels. Pay attention to the activities and resource use of other LUNS in the assigned RAID group. A high-performance RAID group that has too many applications doing I/O to it might not meet the performance goals that an application running on the ESX Server host requires.

■　Provide each server with a sufficient number of NICs or HBAs to allow maximum throughput for all the applications hosted on the server for the peak period. I/O spread across multiple ports provides higher throughput and less latency for each application.

■　To provide redundancy for software iSCSI, in the initiator bind multiple NICs to the vswitch used for iSCSI connectivity.

■　Allocate adequate performance capacity when you allocate LUNs or RAID groups for ESX Server systems, multiple operating systems will use and share that resource. As a result, the performance required from each LUN in the storage subsystem can be much higher if you are working with ESX Server systems than if you are using physical machines. For example, if you expect to run four I/O-intensive applications, allocate four times the performance capacity for the ESX Server LUNs.

■　When you use multiple ESX Server systems in conjunction with a VirtualCenter Server, the performance needed from the storage subsystem increases correspondingly.

Match the number of outstanding I/Os that applications running on an ESX Server system with the number of I/Os the SAN can handle.

Network Performance

A typical SAN consists of a collection of computers connected to a collection of storage systems through a network of switches. Often, numerous computers are accessing the same storage. Figure 5-11 shows several computer systems connected to a storage system through an Ethernet switch. In this configuration, each system is connected through a single Ethernet link to the switch, which is also connected to the storage system through a single Ethernet link. In most configurations, with modern switches and typical traffic, this is not a problem.

Figure 5-11. Single Ethernet Link Connection to Storage
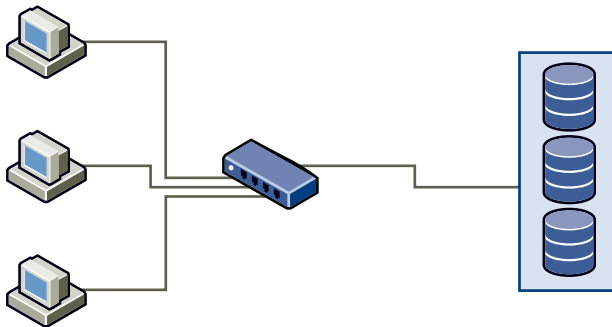
When systems read data from storage, the maximum response from the storage is to send enough data to fill the link between the storage systems and the Ethernet switch. It is unlikely that any single system or virtual machine will get full use of the network speed, but this is the expected situation when many systems share one storage device.

When writing data to storage, multiple systems or virtual machines might attempt to fill their links. As Figure 5-12 shows, when this happens, the switch between the systems and the storage system has to drop data. This happens because, while it has a single connection to the storage device, it has more traffic to send to the storage system than a single link can carry. In this case, the switch drops network packets because the amount of data it can transmit is limited by the speed of the link between it and the storage system.

Figure 5-12. Dropped Packets
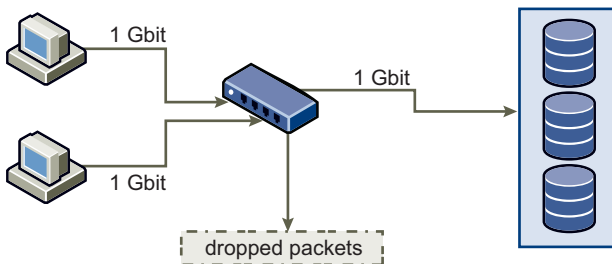
Recovering from dropped network packets results in large performance degradation. In addition to time spent determining that data was dropped, the retransmission uses network bandwidth that could otherwise be used for current transactions.

iSCSI traffic is carried on the network by the Transmission Control Protocol (TCP). TCP is a reliable transmission protocol that ensures that dropped packets are retried and eventually reach their destination. TCP is designed to recover from dropped packets and retransmits them quickly and seamlessly. However, when the switch discards packets with any regularity, network throughput suffers significantly. The network becomes congested with requests to resend data and with the resent packets, and less data is actually transferred than in a network without congestion.

Most Ethernet switches can buffer, or store, data and give every device attempting to send data an equal chance to get to the destination. This ability to buffer some transmissions, combined with many systems limiting the number of outstanding commands, allows small bursts from several systems to be sent to a storage system in turn.

If the transactions are large and multiple servers are trying to send data through a single switch port, a switch's ability to buffer one request while another is transmitted can be exceeded. In this case, the switch drops the data it cannot send, and the storage system must request retransmission of the dropped packet. For example, if an Ethernet switch can buffer 32KB on an input port, but the server connected to it thinks it can send 256KB to the storage device, some of the data is dropped.

Most managed switches provide information on dropped packets, similar to the following:

```
*: interface is up
IHQ: pkts in input hold queue      IQD: pkts dropped from input queue
OHQ: pkts in output hold queue     OQD: pkts dropped from output queue
RXBS: rx rate (bits/sec)           RXPS: rx rate (pkts/sec)
TXBS: tx rate (bits/sec)           TXPS: tx rate (pkts/sec)
TRTL: throttle count
```

Interface	IHQ	IQD	OHQ	OQD	RXBS	RXPS	TXBS	TXPS	TRTL
* GigabitEthernet0/1	3	9922	0	0	476303000	62273	477840000	63677	0

In this example from a Cisco switch, the bandwidth used is 476303000 bits/second, which is less than half of wire speed. In spite of this, the port is buffering incoming packets and has dropped quite a few packets. The final line of this interface summary indicates that this port has already dropped almost 10,000 inbound packets in the IQD column.

Configuration changes to avoid this problem involve making sure several input Ethernet links are not funneled into one output link, resulting in an oversubscribed link. When a number of links transmitting near capacity are switched to a smaller number of links, oversubscription is a possibility.

Generally, applications or systems that write a lot of data to storage, such as data acquisition or transaction logging systems, should not share Ethernet links to a storage device. These types of applications perform best with multiple connections to storage devices.

Figure 5-13, shows multiple connections from the switch to the storage.

Figure 5-13. Multiple Connections from Switch to Storage
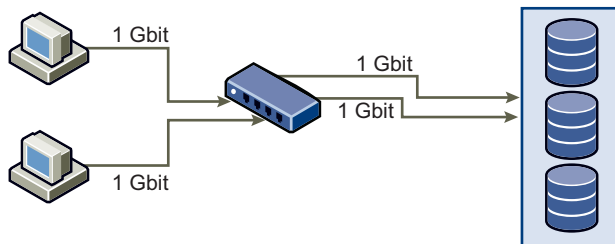
Using VLANs or VPNs does not provide a suitable solution to the problem of link oversubscription in shared configurations. VLANs and other virtual partitioning of a network provide a way of logically designing a network, but do not change the physical capabilities of links and trunks between switches. When storage traffic and other network traffic end up sharing physical connections, as they would with a VPN, the possibility for oversubscription and lost packets exists. The same is true of VLANs that share interswitch trunks. Performance design for a SANs must take into account the physical limitations of the network, not logical allocations.

Resolving Performance Issues

This section discusses performance monitoring and possible ways of resolving performance issues. For best performance, place each virtual machine on the appropriate tier of storage. See "Choosing Virtual Machine Locations" on page 35.

Monitoring Performance

The VI Client offers extensive facilities for collecting performance information. The information is then graphically displayed in the VI Client. The VI Client updates its display periodically. For information, see the *Virtual Infrastructure User's Guide*.

With ESX Server 3, you can also use the `esxtop` utility, available from the service console. For information about `esxtop`, see the *Resource Management Guide*, or look at the man page from the service console. You can use `esxtop` to monitor performance in real time. If you are using ESX Server 3i, the `resxtop` utility provides similar functionality.

Checking Ethernet Switch Statistics

Many Ethernet switches provide methods for monitoring switch health. Switches that have ports operating near maximum throughput much of the time do not provide optimum performance. If you have ports in your iSCSI SAN running near the maximum, reduce the load. If the port is connected to an ESX Server system or iSCSI storage, you can reduce the load by using manual load balancing as described in "Path Management and Manual Load Balancing" on page 100.

If the port is connected between multiple switches or routers, consider installing additional links between these components to handle more load. Ethernet switches also commonly provide information about transmission errors, queued packets, and dropped Ethernet packets. If the switch regularly reports any of these conditions on ports being used for iSCSI traffic, performance of the iSCSI SAN will be poor. For additional information on properly configuring your iSCSI network, see "Network Performance" on page 105.

Resolving Path Thrashing

If your server cannot access a LUN, or access is very slow, you might have a problem with path thrashing (also called LUN thrashing). Path thrashing might occur when two hosts access the LUN through different SPs and, as a result, the LUN is never actually available.

Usually, only specific SAN configurations in conjunction with the following conditions can cause the path thrashing:

- You are working with an active-passive array. Path thrashing only occurs on active-passive arrays. For active-active arrays or arrays that provide transparent failover, path thrashing does not occur.

- Path policy is set to **Fixed**.

- Two hosts access the LUN using opposite path order. For example, Host A is set up to access the lower-numbered LUN through SP A. Host B is set up to access the lower-numbered LUN through SP B.

- Path thrashing can also occur if Host A lost a certain path and can use only paths to SP A while host B lost other paths and can use only paths to SP B.

Path thrashing can also occur on a direct-connect storage system (such as AX100i or AX150i) with HBA failover on one or more nodes.

Path thrashing is a problem that you typically will not experience with other operating systems:

■ No other common operating system uses shared LUNs for more than two servers (that setup is typically reserved for clustering).

■ For clustering, only one server issues I/Os at a time and path thrashing does not become a problem.

In contrast, multiple ESX Server systems might be issuing I/O to the same LUN concurrently.

To resolve path thrashing

■ Ensure that all hosts sharing the same set of LUNs on those active-passive arrays that access the same storage processor simultaneously.

■ Correct any cabling inconsistencies between different ESX Server hosts and SAN targets so that all HBAs see the same targets in the same order.

■ Set the path policy to Most Recently Used (the default).

Understanding Path Thrashing

In all storage systems, the SPs are like independent computers that have access to some shared storage. Algorithms determine how concurrent access is handled.

■ For active-passive arrays, only one LUN at a time can access all the sectors on the storage that make up a given LUN. The ownership is passed between the storage processors. Storage systems use caches and SP A must not write something to disk that invalidates the SP B cache. Because the SP has to flush the cache when it's finished with its operation, it takes a little time to move the ownership. During that time, neither SP can process I/O to the LUN.

■ For active-active arrays, the algorithms allow more detailed access to the storage and synchronize caches. Access can happen concurrently through any SP without extra time required.

Consider how path selection works:

■ On an active-active array the system starts sending I/O on the new path.

■ For active-passive arrays, the ESX Server system checks all standby paths. The SP at the end of the path that is currently under consideration sends information to the system on whether it currently owns the LUN.

■   If the ESX Server system finds an SP that owns the LUN, that path is selected and I/O is sent on that path.

■   If the ESX Server host cannot find such a path, the ESX Server host picks one of the paths and sends the SP (at the other end of the path) a command to move the LUN ownership to this SP.

Path thrashing can occur as a result of the following path choice: If server A can reach a LUN only through one SP, and server B can reach the same LUN only through a different SP, they both continually cause the ownership of the LUN to move between the two SPs. Because the system moves the ownership quickly, the storage system cannot process any I/O (or can process only very little). As a result, any servers that depend on the LUN start timing out I/O.

Equalizing Disk Access Between Virtual Machines

You can adjust the maximum number of outstanding disk requests with the **Disk.SchedNumReqOutstanding** parameter in the VI Client. When two or more virtual machines are accessing the same LUN, this parameter controls the number of outstanding requests that each virtual machine can issue to the LUN. Adjusting the limit can help equalize disk access between virtual machines.

This limit is inapplicable when only one virtual machine is active on a LUN. In that case, the queue depth of the storage adapter and the network's capacity to pass traffic limits the bandwidth.

To set the number of outstanding disk requests

1   In the VI Client, select the host in the inventory panel.

2   Click the **Configuration** tab and click **Advanced Settings**.

3   Click **Disk** in the left panel and scroll down to **Disk.SchedNumReqOutstanding**.

4   Change the parameter value and click **OK**.

5   Reboot the server.

This change can affect disk bandwidth scheduling, but might also lead to improvements for disk-intensive workloads.

If you adjust this parameter value in the VMkernel, you might also want to adjust the queue depth in your storage adapter. See "Setting Maximum Queue Depth for Software iSCSI" on page 112.

Removing VMFS-2 Drivers

If you have a lot of LUNs and VMFS volumes, and all of them are VMFS-3, you can potentially improve performance by unloading the VMFS-2 driver. At a command-line prompt, type:

```
vmkload_mod -u vmfs2
```

A significant increase in the speed of certain management operations like refreshing datastores and rescanning storage adapters should result.

Reducing SCSI Reservations

Operations that require getting a file lock or a metadata lock in VMFS result in short-lived SCSI reservations. SCSI reservations lock an entire LUN. Excessive SCSI reservations by a server can cause performance degradation on other servers accessing the same VMFS.

Examples of operations that require getting file locks or metadata locks include:

- Virtual machine power on

- VMotion

- Virtual machines running with virtual disk snapshots

- File operations that require opening files or doing metadata updates (See "Metadata Updates" on page 24.)

Performance degradation can occur if such operations are happening frequently on multiple servers accessing the same VMFS. For instance, VMware recommends that you do not run many virtual machines from multiple servers that are using virtual disk snapshots on the same VMFS. Limit the number of VMFS file operations when many virtual machines are running on the VMFS.

Setting Maximum Queue Depth for Software iSCSI

If you notice unsatisfactory performance for your software iSCSI LUNs, you can change their maximum queue depth by using the `esxcfg-module` command as follows:

```
esxcfg-module  -s  iscsi_max_lun_queue=value  iscsi_mod
```

After you issue this command, reboot your system.

The **iscsi_max_lun_queue** parameter is used to set the maximum outstanding commands, or queue depth, for each LUN accessed through the software iSCSI adapter. The default is 32, and the valid range is 1 to 255.

CAUTION Setting the queue depth higher than the default can decrease the total number of LUNs supported.

SAN Storage Backup Considerations

In the SAN environment, backups have two goals. The first goal is to archive online data to offline media. This process is repeated periodically for all online data on a time schedule. The second goal is to provide access to offline data for recovery from a problem. For example, database recovery often requires retrieval of archived log files that are not currently online.

Scheduling a backup depends on a number of factors:

- Identification of critical applications that require more frequent backup cycles within a given period of time.

- Recovery point and recovery time goals. Consider how precise your recovery point needs to be, and how long you are willing to wait for it.

- The rate of change (RoC) associated with the data. For example, if you are using synchronous-asynchronous replication, the RoC affects the amount of bandwidth required between the primary and secondary storage devices.

- Overall impact on SAN environment, storage performance (while backing up), and other applications.

- Identification of peak traffic periods on the SAN (backups scheduled during those peak periods can slow the applications and the backup process).

- Time to schedule all backups in the datacenter.

- Time it takes to back up an individual application.

- Resource availability for archiving data; usually offline media access (tape).

Include a recovery-time objective for each application when you design your backup strategy. That is, consider the time and resources necessary to reprovision the data. For example, if a scheduled backup stores so much data that recovery requires a considerable amount of time, examine the scheduled backup. Perform the backup more frequently, so that less data is backed up at a time and the recovery time decreases.

If a particular application requires recovery within a certain time frame, the backup process needs to provide a time schedule and specific data processing to meet this requirement. Fast recovery can require the use of recovery volumes that reside on online storage to minimize or eliminate the need to access slow offline media for missing data components.

Snapshot Software

Snapshot software allows an administrator to make an instantaneous copy of any single virtual disk defined within the disk subsystem. Snapshot software is available at different levels:

- ESX Server hosts allow you to create snapshots of virtual machines. This software is included in the basic ESX Server package.

- Third-party backup software might allow for more comprehensive backup procedures and might contain more sophisticated configuration options.

Administrators make snapshots for a variety of reasons, including:

- Backup

- Disaster recovery

- Availability of multiple configurations, versions, or both

- Forensics (looking at a snapshot to find the cause of problems while your system is running)

- Data mining (looking at a copy of your data to reduce load on production systems)

Using a Third-Party Backup Package

If you are using third-party backup software, make sure that the software is supported with ESX Server hosts. See the *Backup Software Compatibility Guide*.

Using third-party software has the advantage of a uniform environment. However, the additional cost of the third-party snapshotting software can become higher as your SAN grows.

If you use snapshots to back up your data, consider the following points:

- Some vendors support snapshots for VMFS and RDMs. If both are supported, you can make either a snapshot of the whole VMFS for a host, or snapshots for the individual virtual machines (one for each disk).

- Some vendors support snapshots only for a setup using RDM. If only RDM is supported, you can make snapshots of individual virtual machines.

See your storage vendor's documentation.

NOTE ESX Server systems also include a Consolidated Backup component, which is discussed in detail in the *Virtual Machine Backup Guide*.

Choosing Your Backup Solution

When you choose your backup solution, consider that a backup can be one or all of these:

- Crash consistent
- File-system consistent
- Application consistent

VMware offers a file-system-consistent backup. In most cases, a file-system-consistent backup allows you to completely recover from failure. However, if your applications require synchronization across file systems or with a database, the VMware solution might not provide enough consistency. In these cases, investigate a third-party backup solution to see whether it better suits your needs.

Layered Applications

SAN administrators customarily use specialized array-based software for backup, disaster recovery, data mining, forensics, and configuration testing.

Storage providers typically supply two types of advanced services for their LUNs—snapshotting and replication.

- Snapshotting creates space with efficient copies of LUNs that share common blocks of data. In general, snapshotting is used locally on the same storage system as the primary LUN for quick backups, application testing, forensics, or data mining.

- Replication creates full copies of LUNs. Replicas are usually made to separate storage systems, possibly separate sites to protect against major outages that incapacitate or destroy an entire storage system or site.

When you use an ESX Server system in conjunction with a SAN, decide whether array-based or host-based tools are more suitable for your particular situation.

Array-Based (Third-Party) Solution

When you consider an array-based solution, consider the following points:

NOTE   ESX Server systems also include a consolidated backup component, which is discussed in detail in the *Virtual Machine Backup Guide*.

- Array-based solutions usually result in more comprehensive statistics. With RDM, data always takes the same path, which results in easier performance management.

- Security is more transparent to the storage administrator when you use RDM and an array-based solution because with RDM, virtual machines more closely resemble physical machines.

- If you use an array-based solution, physical compatibility RDMs are often used for the storage of virtual machines. If you do not intend to use RDM, check the storage vendor documentation to see if operations on LUNs with VMFS volumes are supported. Furthermore, if you use array operations on VMFS LUNs, carefully read the section on resignaturing.

File-Based (VMFS) Solution

When you consider a file-based solution that uses VMware tools and VMFS (instead of the array tools), be aware of the following points:

- Using VMware tools and VMFS is better for provisioning: one large LUN is allocated and multiple VMDK files can be placed on that LUN. With RDM, a new LUN is required for each virtual machine.

- Snapshotting is included with your ESX Server host at no extra cost. The file-based solution is therefore more cost-effective than the array-based solution.

- For ESX Server administrators, using VMFS is easier.

- ESX Server administrators who use the file-based solution are more independent from the SAN administrator.

VMFS Volume Resignaturing

ESX servers need to be able to differentiate between their VMFS volumes and use a volume signature to do so. When a VMFS volume is replicated or a snapshot is taken, the resulting LUN copy has the same signature as the source. When an ESX Server sees two LUNs with the same signature, the ESX Server must handle the condition to prevent downtime caused by confusion over which LUN it should be using to access the registered virtual machines. Resignaturing is a feature introduced in ESX Server 3.0 to solve this problem.

NOTE   When a LUN needs to be resignatured, an alert appears in the vmkernel log. If you encounter such an alert, set your resignaturing options appropriately, as described in the following sections.

Mounting Original, Snapshot, or Replica VMFS Volumes

You can mount original, snapshot, or replica VMFS volumes on the same ESX Server host.

To mount original, snapshot, or replica VMFS volumes

1   Perform the required storage tasks:

a   Make the storage system snapshot or replica.

b   Configure access control to allow ESX Server to access the snapshot or replica.

2   In the VI Client, select the host in the inventory panel.

3   Click the **Configuration** tab and click **Advanced Settings**.

4   Select **LVM** in the left panel, then set the **LVM.EnableResignature** option to **1**.

5   Rescan for any new LUNs or VMFS volumes.

After the rescan, the copied VMFS volume appears as
`/vmfs/volumes/snap–<DIGIT>–<old–label>`.

If the VMX file for any of the virtual machines or the VMSD file for virtual machine snapshots contains `/vmfs/volumes/<label or UUID>/` paths, you must change these items to reflect the resignatured volume path.

6   Set the **LVM.EnableResignature** option to **0** after resignaturing is complete.

NOTE   Any virtual machines on this new snapshot volume are not auto-discovered. You must manually register the virtual machines.

Understanding Resignaturing Options

This section discusses how the EnableResignature and DisallowSnapshotLUN options interact and explains the three states that result from changing these options:

- State 1: EnableResignature=0, DisallowSnapshotLUN=1 (the ESX Server 3.x default)

- State 2: EnableResignature=1 (DisallowSnapshotLUN is not relevant)

- State 3: EnableResignature=0, DisallowSnapshotLUN=0 (ESX Server 2.x behavior)

State 1: EnableResignature=0, DisallowSnapshotLUN=1 (default)

In this state:

- You cannot bring snapshots or replicas of VMFS volumes made by the storage system into the ESX Server host regardless of whether the ESX Server has access to the original LUN.

- LUNs formatted with VMFS must have the same ID for each ESX Server host.

State 1 is the safest state but:

- If you use Clariion AX100i or AX150i with Navisphere Express, you cannot configure the same LUN ID across storage groups. You must instead use a version of Navisphere software that has more comprehensive management capabilities.

- For IBM TotalStorage 8000, if you have LUNs that are not configured to present the same LUN ID to all servers, you need to either use the settings in state 3 or recreate the LUNs from scratch.

State 2: EnableResignature=1, (DisallowSnapshotLUN is not relevant)

In this state:

- You can safely bring snapshots or replicas of VMFS volumes into the same servers as the original and they are automatically resignatured.

- VMFS volumes containing LUNs from AX100i or AX150i that are not presented with the same LUN numbers to all servers effectively lose the ability to use the virtual machines stored on that VMFS volume. Avoid this situation.

State 3: EnableResignature=0, DisallowSnapshotLUN=0

This is similar to ESX Server 2.x behavior. In this state, the ESX Server assumes that it sees only one replica or snapshot of a given LUN and never tries to resignature. This is ideal in a DR scenario where you are bringing a replica of a LUN to a new cluster of ESX Servers, possibly on another site that does not have access to the source LUN. In such a case, the ESX Server uses the replica as if it is the original.

If you have an AX100i or AX150i that cannot be configured to present the same LUN numbers to all servers for some reason, you need this setting to allow all ESX Server systems to use the same LUNs for features like VMotion, VMware DRS, and VMware HA.

Do not use this setting if you are bringing snapshots or replicas of a LUN into a server with access to the original LUN. This can have destructive results including:

- If you create snapshots of a VMFS volume one or more times and dynamically bring one or more of those snapshots into an ESX Server, only the first copy is usable. The usable copy is most likely the primary copy. After reboot, it is impossible to determine which volume (the source or one of the snapshots) is usable. This nondeterministic behavior should be avoided.

- If you create a snapshot of a spanned VMFS volume, an ESX Server host might reassemble the volume from fragments that belong to different snapshots. This can corrupt your file system.

Multipathing Checklist

<div style="text-align: right">**A**</div>

This appendix provides a checklist of multipathing setup requirements for different storage systems.

Table A-1. Multipathing Setup Requirements

Component	Comments
All storage systems	Write cache must be disabled if not battery backed.
Topology	No single failure should cause HBA and SP failover, especially with active-passive storage arrays.
EMC Symmetrix	Enable the SPC2 and SC3 settings. Contact EMC for the latest settings.
EMC Clariion	Set the Advanced Setting for the ESX Server host: All Initiator records must have: ■ Failover Mode = 1 ■ Initiator Type = Clariion Open ■ Array CommPath = "Enabled" or 1
HP MSA	No specific requirements
HP EVA	For EVA3000/5000 firmware 4.001 and later, and EVA4000/6000/8000 firmware 5.031 and later, set the host type to `VMware`. Otherwise, set the host mode type to `Custom`. The value is: ■ EVA3000/5000 firmware 3.x: 000000002200282E ■ EVA4000/6000/8000: 000000202200083E
NetApp	No specific requirements
EqualLogic	No specific requirements

Table A-1. Multipathing Setup Requirements (Continued)

Component	Comments
LeftHand	No specific requirements
ESX Server Configuration	Set the following Advanced Settings for the ESX Server host: ■ Set **Disk.UseLunReset** to **1** ■ Set **Disk.UseDeviceReset** to **0** A multipathing policy of Most Recently Used must be set for all LUNs hosting clustered disks for active-passive arrays. A multipathing policy of Most Recently Used or Fixed may be set for LUNs on active-active arrays. Allow ARP redirection if the storage system supports transparent failover. See "esxcfg-hwiscsi Utility" on page 126.

Utilities

<div style="text-align: right; font-size: 3em; font-weight: bold; color: gray;">B</div>

In most cases, the VI Client is well-suited for monitoring an ESX Server host connected to SAN storage. Advanced users might, at times, want to use some command-line utilities for additional details.

This appendix provides information on the following utilities:

-
-
-
-
-

esxtop Utility

The esxtop command-line utility provides a detailed look at ESX Server 3 resource use in real time. It runs on the ESX Server host's service console. For detailed information about esxtop, see the *Resource Management Guide* or type **man esxtop** at the command-line prompt.

NOTE   If you are using ESX Server 3i, the resxtop utility provides similar functionality. See *Remote Command-Line Interface Installation and Reference Guide*.

storageMonitor Utility

The storageMonitor utility monitors SCSI sense errors that storage devices attached to VMware ESX Server experience. The utility gathers sense-error information by periodically polling the storageMonitor running inside the VMkernel, and sends error information to a standard output file, a file, or the system log. It formats error information before sending it to output. For example, it converts sense-error codes to corresponding text according to SCSI-3 specification.

If no configuration file is specified, storageMonitor parses the /etc/vmware/storageMonitor.conf default configuration file to filter certain errors and allow other errors to be displayed. You can use the -d option to run storageMonitor in interactive mode or daemon mode.

Options

You can start storageMonitor from the ESX Server command line using one of the following options.

Table B-1. storageMonitor Command-Line Options

Option	Description
<config-file>	Allows you to specify a configuration file. If this option is left unspecified, the default is used.
	The configuration file specifies which type of errors storageMonitor should allow and which ones it should filter before displaying them.
	The default configuration file illustrates the format of the entries.
-d	Specifies that storageMonitor should be run in daemon mode. When this option is specified, all output goes either to syslog or to a log file that the user specifies. If the -s option is also specified, output is written to a standard output file as well.
-h	Displays help information.
-l <log_file>	When this option is specified, output from the program is written to <log_file>. This option is valid only if the -d option is also specified.
-p <poll_interval>	Allows you to specify the interval (in seconds) used for polling kernel resident storage and for retrieving the status or errors of the storage devices. If this option is not specified, the default polling interval of 10 seconds is used.
-s	Specifies that storageMonitor should send output to a standard output file. This option is only valid if you start storageMonitor in daemon mode (-d option is specified).

Examples

storageMonitor -p 60

Sets the polling interval to 60 seconds. Sends output to a standard output file (because storageMonitor is not running in daemon mode). Uses the filters specified in the default configuration file before sending the output.

storageMonitor -d -c myconf.conf

Runs storageMonitor in daemon mode by using the myconf.conf configuration file. Writes output to syslog. By default, syslog is located at /var/log/storageMonitor.

storageMonitor -d -l mylog.log -s

Runs storageMonitor in daemon mode by using the default configuration file. Sends output to mylog.log instead of syslog. Also writes output to a standard output file because the -s option is specified.

esxcfg-swiscsi Utility

The esxcfg-swiscsi utility allows you to enable or disable software iSCSI on ESX hosts.

Usage example:

esxcfg-swiscsi [-e][-d][-h][-q][-s] <vmkernel SCSI adapter name>

Table B-2. esxcfg-swiscsi Command-Line Options

Option	Description
-e	Enables software iSCSI.
-d	Disables software iSCSI. Using this option can lead to problems if you are using iSCSI volumes.
-q	Checks if software iSCSI is on or off.
-s	Scans for disks available through the software iSCSI interface.
-h	Displays help information.

esxcfg-hwiscsi Utility

The esxcfg-hwiscsi utility allows you to configure supported parameters for hardware iSCSI.

Usage example:
/sbin/esxcfg-hwiscsi [-l] [-a allow|deny] [-h] <vmkernel SCSI adapter name>

Table B-3. esxcfg-hwiscsi Command-Line Options

Option	Description
-l	Lists current configuration (overrides settings options).
-a	Allows or denies ARP redirection on adapter.
-h	Displays help information.

vmkping Utility

The vmkping utility allows you to verify the VMkernel networking configuration.

Usage example:
vmkping [options] [host|IP address]

Table B-4. vmkping Command-Line Options

Option	Description
-D	VMkernel TCP stack debug mode.
-c <count>	Sets packet count.
-i <interval>	Sets interval.
-s <size>	Sets send size.

Index