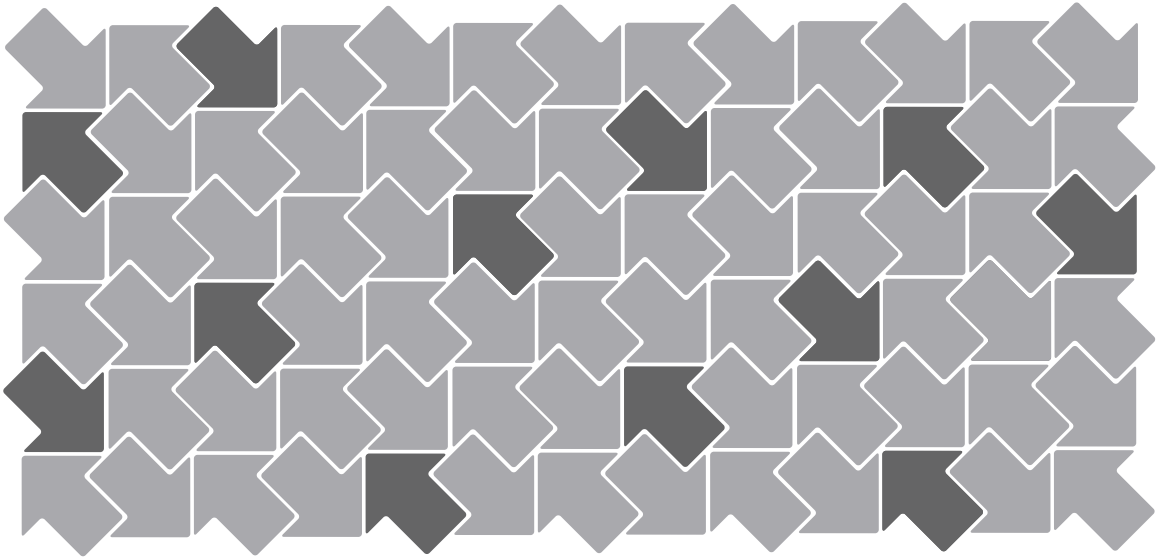
SAN Configuration Guide

ESX Server 3.0.1 and VirtualCenter 2.0.1



vmware®

SAN Configuration Guide
Revision: 20060925
Item: VI-ENG-Q206-220

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.
3145 Porter Drive
Palo Alto, CA 94304
www.vmware.com

Contents

Preface

This preface describes the contents of the *SAN Configuration Guide* and provides pointers to VMware® technical and educational resources.

This preface contains the following topics:

About This Book

This manual, the *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 main topics:

- Understanding ESX Server – Introduces ESX Server systems for SAN administrators.
- Using ESX Server with a SAN – Discusses requirements, noticeable differences in SAN setup if ESX Server is used, and how to manage and troubleshoot the two systems together.
- Enabling your ESX Server system to boot from a LUN on a SAN – Discusses requirements, limitations, and management of boot from SAN.

NOTE This manual's focus is SAN over Fibre Channel (FC). It does not discuss iSCSI or NFS storage devices. For information on iSCSI and NFS, see the *Server Configuration Guide*.

Revision History

This manual is revised with each release of the product or when necessary. A revised version can contain minor or major changes. Table P-1 provides you with the revision history of this manual.

Table P-1. Revision History

Revision	Description
20060615	ESX Server 3.0 and VirtualCenter 2.0 version of the VMware Infrastructure 3 *SAN Configuration Guide*. This is the first edition of this manual.
20060925	ESX Server 3.0.1 and VirtualCenter 2.0.1 version of the VMware Infrastructure 3 *SAN Configuration Guide*. This edition contains minor changes.

Intended Audience

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

Document Feedback

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

docfeedback@vmware.com

VMware Infrastructure Documentation

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

You can access the most current versions of this manual and other books by going to:

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

Conventions

Table P-2 illustrates the typographic conventions used in this manual.

Table P-2. Conventions Used in This Manual

Style	Elements
Blue (online only)	Cross-references and email addresses
Blue boldface (online only)	Links
Black boldface	User interface elements such as button names and menu items
`Monospace`	Commands, filenames, directories, and paths
`Monospace bold`	User input
Italic	Document titles, glossary terms, and occasional emphasis
< Name >	Variable and parameter names

Abbreviations Used in Graphics

The graphics in this manual use the abbreviations listed in Table P-3.

Table P-3. Abbreviations

Abbreviation	Description
VC	VirtualCenter
VI	Virtual Infrastructure Client
server	VirtualCenter Server
database	VirtualCenter database
host*n*	VirtualCenter managed hosts
VM#	Virtual machines on a managed host
user#	User with access permissions
dsk#	Storage disk for the managed host
datastore	Storage for the managed host

Table P-3. Abbreviations

Abbreviation	Description
SAN	Storage area network type datastore shared between managed hosts
tmplt	Template

Technical Support and Education Resources

The following sections describe the technical support resources available to you.

Self-Service Support

Use the VMware Technology Network (VMTN) for self-help tools and technical information:

- Product information – **http://www.vmware.com/products/**
- Technology information – **http://www.vmware.com/vcommunity/technology**
- Documentation – **http://www.vmware.com/support/pubs**
- VMTN Knowledge Base – **http://www.vmware.com/support/kb**
- Discussion forums – **http://www.vmware.com/community**
- User groups – **http://www.vmware.com/vcommunity/usergroups.html**

For more information about the VMware Technology Network, go to **http://www.vmtn.net**.

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**.

Overview of VMware ESX Server 1

You can use ESX Server in conjunction with a SAN (storage area network), 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 both ESX Server systems and SAN concepts. This chapter presents an overview of ESX Server concepts. It is meant for SAN administrators not yet familiar with ESX Server systems and consists of the following sections:

For in-depth information on VMware ESX Server, go to **http://www.vmware.com/products/esx/** and choose the information under **Technical Resources**. The section includes documentation, hardware compatibility lists, white papers, and more.

Introduction into ESX Server

The ESX Server architecture allows administrators to allocate hardware resources to multiple workloads in fully isolated environments called *virtual machines*.

An ESX Server system has the following key components:

- **Virtualization layer** – This layer provides the idealized hardware environment and virtualization of underlying physical resources to the virtual machines. It includes the Virtual Machine Monitor (VMM), which is responsible for virtualization, and VMkernel.

  The virtualization layer schedules both the service console running on the ESX Server host and the virtual machine operating systems. The virtualization layer manages how the operating systems access physical resources. The VMkernel needs its own drivers to provide access to the physical devices. VMkernel drivers are modified Linux drivers, even though the VMkernel is not a Linux variant.

- **Hardware interface components** – The virtual machine communicates with hardware such as CPU or disk using hardware interface components. These components include device drivers, which enable hardware-specific service delivery while hiding hardware differences from other parts of the system.

- **User interface** – Administrators can view and manage ESX Server hosts and virtual machines in several ways (see Figure 1-2).

  - A Virtual Infrastructure Client (VI Client) can connect directly to the ESX Server host. This is appropriate if your environment has only one host.
  - A VI Client can also connect to a VirtualCenter Management Server and interact with all ESX Server hosts managed by that VirtualCenter Server.
  - The VI Web Access Client allows you to perform many management tasks using a browser-based interface.
  - The service console command-line interface is used only rarely. Starting with ESX Server 3.0, the VI Client replaces the service console for most of the interactions. (Commands have changed since previous versions of ESX Server.)

  Figure 1-1 shows the VI Client with the Summary tab selected.

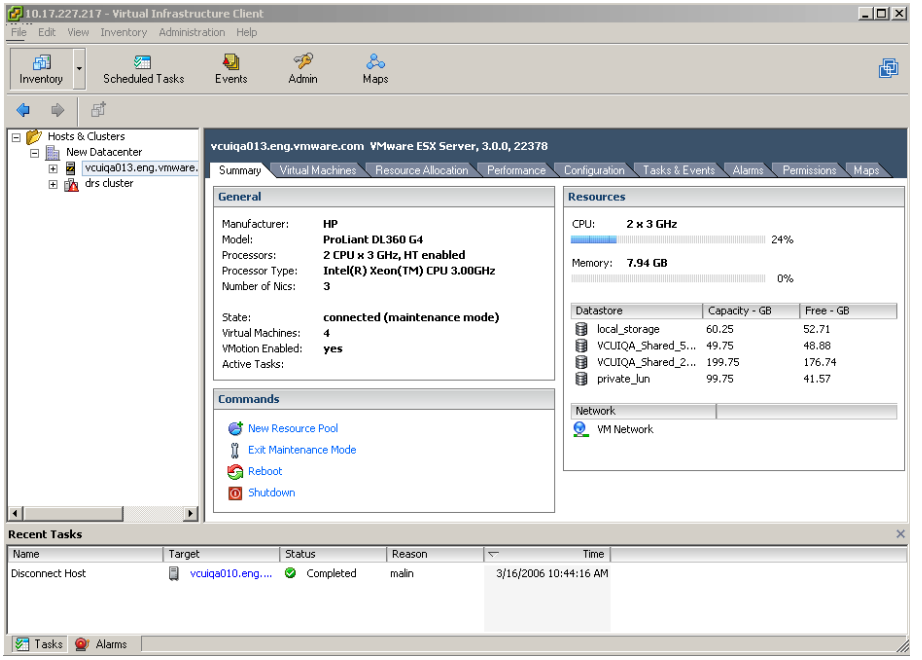


Figure 1-1. VI Client User Interface

NOTE With ESX Server 3.0 and VirtualCenter 2.0, you can perform almost all operations using the VI Client.

Figure 1-2 shows how the components interact. The ESX Server host has four different virtual machines configured. Each virtual machine runs its own guest operating system and applications. Administrators monitor the host and the virtual machines in one of two ways:

- Using a VI Client to connect to an ESX Server host directly.
- Using a VI Client to connect to a VirtualCenter Management Server. The VirtualCenter Server can manage a number of ESX Server hosts.

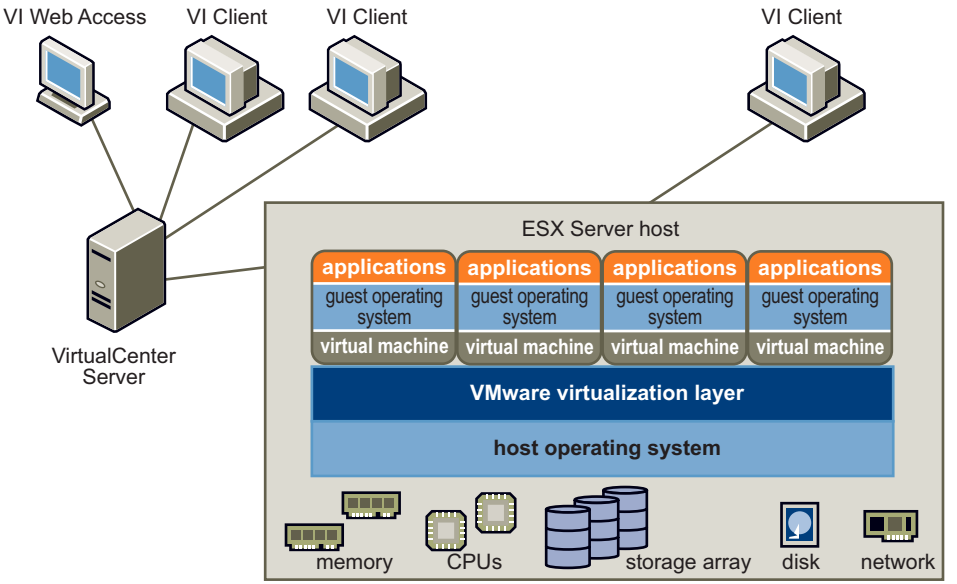


Figure 1-2. Virtual Infrastructure Environment

For additional information, see "Software and Hardware Compatibility" on page 18.

Software and Hardware Compatibility

In the VMware ESX Server architecture, the operating system of the virtual machine (the guest operating system) interacts only with the standard, x86-compatible virtual hardware presented by the virtualization layer. This allows VMware products to support any x86-compatible operating system.

In practice, VMware products support a large subset of x86-compatible operating systems that are tested throughout the product development cycle. VMware documents the installation and operation of these guest operating systems and trains its technical personnel in supporting them.

Most applications interact only with the guest operating system, not with the underlying hardware. As a result, you can run applications on the hardware of your choice as long as you install a virtual machine with the operating system the application requires.

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 provides the virtual devices that 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.

For additional information, see the following sections:

- "CPU, Memory, and Network Virtualization" on page 19
- "Virtual SCSI" on page 20

CPU, Memory, and Network Virtualization

A VMware virtual machine offers complete hardware virtualization. The guest operating system and applications running on a virtual machine can never determine directly which physical resources they are accessing (such as which physical CPU they are running on in a multiprocessor system, or which physical memory is mapped to their pages).

- **CPU Virtualization** – Each virtual machine appears to run on its own CPU (or a set of CPUs), fully isolated from other virtual machines. Registers, the translation lookaside buffer, and other control structures are maintained separately for each virtual machine.

  Most instructions are executed directly on the physical CPU, allowing resource-intensive workloads to run at near-native speed. Privileged instructions are performed safely by the virtualization layer.

  See the *Resource Management Guide* for more information.

- **Memory Virtualization** – A contiguous memory space is visible to each virtual machine. However, the allocated physical memory might not be contiguous. Instead, noncontiguous physical pages are remapped and presented to each virtual machine. With unusually memory-intensive loads, server memory becomes overcommitted. In that case, some of the physical memory of a virtual machine might be mapped to shared pages or to pages that are unmapped or swapped out.

  ESX Server performs this virtual memory management without the information the guest operating system has and without interfering with the guest operating system's memory management subsystem.

See the *Resource Management Guide* for more information.

- **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.

  See the *Server Configuration Guide* for more information.

Virtual SCSI

In an ESX Server environment, each virtual machine includes from one to four virtual SCSI HBAs (host bus adapters). These virtual adapters may appear as either Buslogic or LSI Logic SCSI controllers. They are the only types of SCSI controllers that are accessible by a virtual machine.

Each virtual disk accessible by a virtual machine through one of the virtual SCSI adapters resides in the VMFS or on a raw disk.

For more information, see the following sections:

- "Disk Configuration Options" on page 21
- "Virtual Machine File System (VMFS)" on page 22
- "Raw Device Mapping" on page 22
- "Virtual SCSI HBA" on page 23

Figure 1-3 gives an overview of storage virtualization. It illustrates both storage using VMFS and storage using raw device mapping.

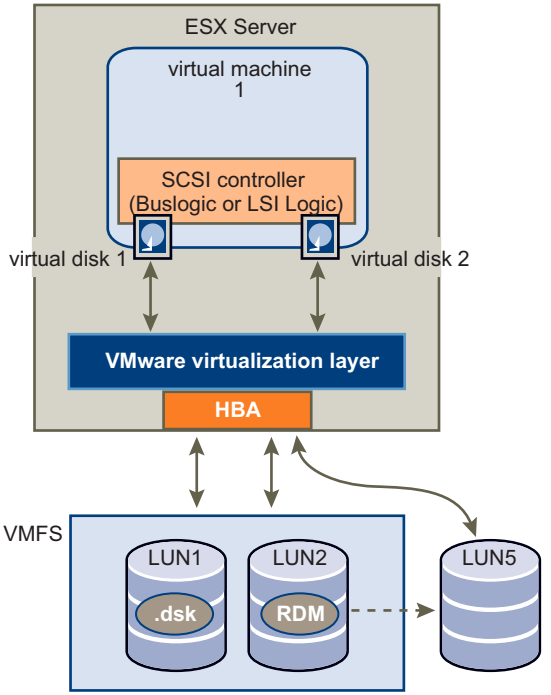


Figure 1-3. SAN Storage Virtualization

Disk Configuration Options

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

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

- Virtual machine `.vmdk` file stored on a VMFS volume. See "Virtual Machine File System (VMFS)" on page 22.
- Device mapping to a SAN LUN (logical unit number). See "Raw Device Mapping" on page 22.
- Local SCSI device passed through directly to the virtual machine (for example, a local tape drive).

From the standpoint of the virtual machine, each virtual disk appears as if it were a SCSI drive connected to a SCSI adapter. Whether the actual physical disk device is

being accessed through SCSI, iSCSI, RAID, NFS, or Fibre Channel (FC) controllers is transparent to the guest operating system and to applications running on the virtual machine.

Virtual Machine File System (VMFS)

In a simple configuration, the virtual machines' disks are stored as files within a 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 has been 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 600MB or larger. See the *Installation and Upgrade Guide*. It can then be customized, as discussed in the *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.

For more information, see the following:

- Chapter 3, "Requirements and Installation," on page 49. The chapter lists HBAs and storage devices that are supported by ESX Server systems.
- Chapter 4, "Setting Up SAN Storage Devices with ESX Server," on page 57. The chapter provides information on configuring SAN components.

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 use Microsoft Cluster Service (MSCS) or 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 arrays. For information, see "Using Raw Device Mapping" in the *Server Configuration Guide*. See the document *Setup for Microsoft Cluster Service* for information about MSCS.

Virtual SCSI HBA

Virtual SCSI HBAs allow virtual machines access to logical SCSI devices, just as a physical HBA allows access to physical storage devices. However, in contrast to the physical HBA, the virtual SCSI HBA does not allow storage administrators (such as SAN administrators) access to the physical machine. At the same time, many virtual HBAs can be hidden behind a single (or multiple) FC HBAs. The SAN can see only the physical machine and its HBAs.

Interacting with ESX Server Systems

Generally, administrators interact with ESX Server systems in one of the following ways:

- Using a client (VI Client or VI Web Access), as discussed in the "Introduction into ESX Server" on page 16. Clients can be connected directly to the ESX Server host, or you can manage multiple ESX Server hosts simultaneously using the VirtualCenter Management Server.

  For more information on VirtualCenter Management Server, see "VMware Virtual Center" on page 23.

- Using the service console. Remember that in ESX Server 3.0 and later, use of the service console is generally not necessary and is discouraged because most administrative operations can be performed using a VI Client or VI Web Access. For scripted management, use the Virtual Infrastructure SDK.

  For more information on the service console, see "Service Console" on page 24.

VMware Virtual Center

A VirtualCenter Server can be accessed through a VI Client or VI Web Access.

- The VirtualCenter Server acts as a central administrator for VMware servers connected on a network. The server directs actions upon the virtual machines and VMware ESX Server.
- The VI Client user interface runs on Microsoft Windows. In a multi-host environment, administrators use the VI Client to make requests to the VirtualCenter server, which in turn affects its virtual machines and hosts. Connecting to an ESX Server host directly usually makes sense only in a single-server environment.
- VI Web Access allows you to connect to a VirtualCenter Server using an HTML browser.

Figure 1-4 shows the Configuration tab of a VI Client display with **Storage** selected. The selected ESX Server host connects to both SAN LUNs and to local hard disks. The difference in the display is visible only because of the names that were chosen during setup.

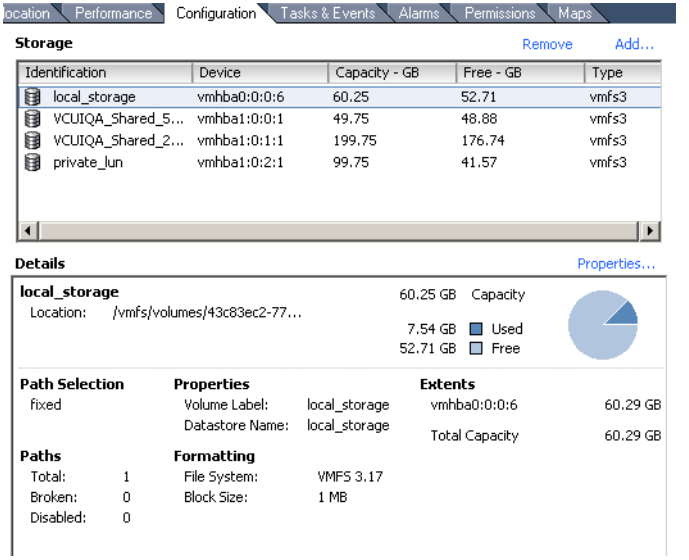


Figure 1-4. Storage Information Displayed in VI Client, Configuration Tab

Service Console

The service console is the ESX Server command-line management interface. In earlier releases, the service console was the main interface to the ESX Server host. With ESX Server 3 and later, the VI Client has priority, though some advanced administration is still performed using the service console. The service console supports ESX Server system management functions and interfaces. These include HTTP, SNMP, and API interfaces, as well as other support functions such as authentication and low-performance device access.

Because VirtualCenter functionality has been enhanced to allow almost all administrative operations, service console functionality is now limited. The service console is used only under special circumstances.

NOTE For scripted management, use the Virtual Infrastructure SDK. Service console commands might no longer be available, or might have different results, in future releases.

The service console is implemented using a modified Linux distribution. However, the service console does not correspond directly to a Linux command prompt.

For more information, see "Service Console Processes and Services" on page 25.

Service Console Processes and Services

The following ESX Server management processes and services run in the service console:

- **Host daemon (hostd)** – Performs actions in the service console on behalf of the service console and the VI Client.
- **Authentication daemon (vmauthd)** – Authenticates remote users of the VI Client and remote consoles using the user name and password database. Any other authentication store that can be accessed using the service console's Pluggable Authentication Module (PAM) capabilities can also be used. Having multiple password storage mechanisms permits the use of passwords from a Windows domain controller, LDAP or RADIUS server, or similar central authentication store in conjunction with VMware ESX Server for remote access.
- **SNMP server (net-snmpd)** – Implements the SNMP traps and data structures that an administrator can use to integrate an ESX Server system into an SNMP-based system-management tool.

In addition to these services, which are supplied by VMware, the service console can be used to run other system-wide or hardware-dependent management tools. These can include hardware-specific health monitors (such as IBM Director or HP Insight Manager), full-system backup and disaster recovery software, and clustering and high availability products.

NOTE The service console is not guaranteed to be available for general-purpose Linux hardware monitoring. It is not equivalent to a Linux shell.

Virtualization at a Glance

ESX Server virtualizes the resources of the physical system for use by the virtual machines. Multiple virtual machines share physical devices.

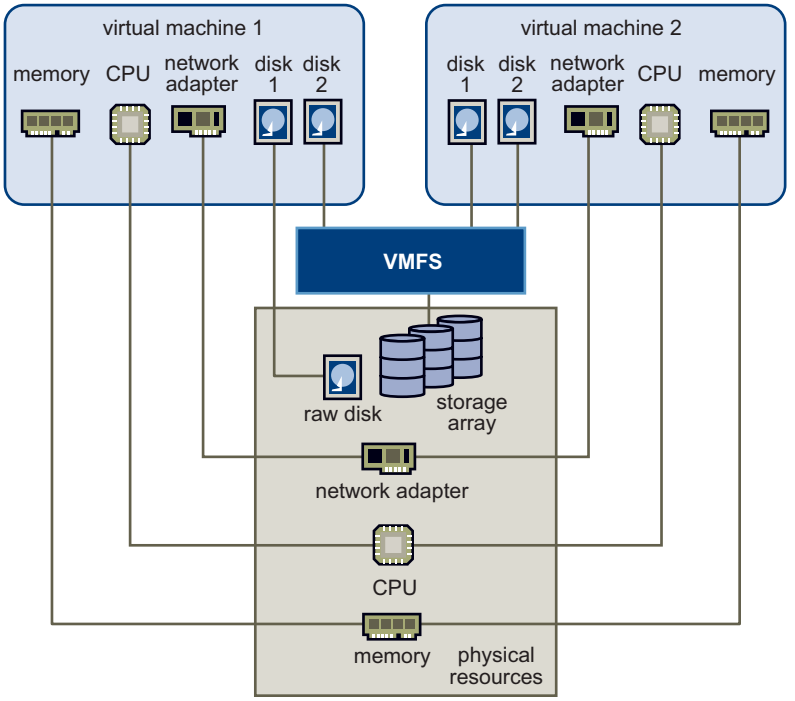


Figure 1-5. Virtual Machines Sharing Physical Resources

Figure 1-5 shows two virtual machines, each configured with:

- One CPU
- An allocation of memory and a network adapter (NIC)
- Two virtual disks

The virtual machines each use one of the CPUs on the server and access noncontiguous pages of memory, with part of the memory of one virtual machine currently swapped to disk (not shown). The two virtual network adapters are connected to two physical network adapters.

The disks are mapped as follows:

- Disk 1 of virtual machine 1 is mapped directly to a raw disk. This configuration is not usually recommended but can be advantageous under certain circumstances.
- Disk 2 of virtual machine 1 and both disks of virtual machine 2 reside on the VMFS, which is located on a SAN storage array. VMFS makes sure that appropriate locking and security is in place at all times.

Using ESX Server with SAN 2

When you set up ESX Server hosts to use FC SAN array storage, special considerations are necessary. This chapter provides introductory information on using ESX Server with a SAN array and discusses these topics:

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 recognized in the SAN industry for their wealth of information. These sites are:

- **www.searchstorage.com**
- **www.snia.org**

Because this industry changes constantly, you are encouraged to stay abreast of the latest developments by checking these resources frequently.

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

NOTE SAN administrators can skip this section and continue with the rest of this chapter.

SAN (storage area network) is a specialized high-speed network that connects computer systems, or host servers, to high performance storage subsystems. The SAN components include host bus adapters (HBAs) in the host servers, switches that help route storage traffic, cables, storage processors (SPs), and storage disk arrays.

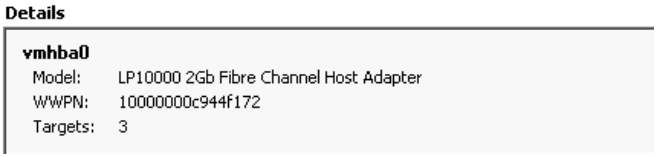
A SAN topology with at least one switch present on the network forms a **SAN fabric**.

To transfer traffic from host servers to shared storage, the SAN uses **Fibre Channel (FC)** protocol that packages SCSI commands into Fibre Channel frames.

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 fabric component, has one or more ports that connect it to the SAN. Ports can be identified in a number of ways:

- **WWPN** – World Wide Port Name. A globally unique identifier for a port which allows certain applications to access the port. The FC switches discover the WWPN of a device or host and assign a port address to the device.

To view the WWPN using a VI Client, click the host's **Configuration** tab and choose **Storage Adapters.** You can then select the storage adapter for which you want to see the WWPN.



- **Port_ID** (or port address) – Within the SAN, each port has a unique port ID that serves as the FC address for the port. This enables routing of data through the SAN to that port. The FC switches assign the port ID when the device logs into the fabric. The port ID is valid only while the device is logged on.

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 array.

If a default path or any component along the path—HBA, cable, 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 disk arrays can be of the following types:

- An **active/active disk array**, which allows access to the LUNs simultaneously through all the storage processors that are available without significant performance degradation. All the paths are active at all times (unless a path fails).
- An **active/passive disk array**, in which one SP is actively servicing a given LUN. The other SP acts as backup for the LUN and may be actively servicing other LUN I/O. I/O can be sent only to an active processor. If the primary storage processor fails, one of the secondary storage processors becomes active, either automatically or through administrator intervention.

To restrict server access to storage arrays not allocated to that server, the SAN uses **zoning**. Typically, zones are created for each group of servers that access a shared group of storage devices and LUNs. Zones define which HBAs can connect to which SPs. Devices outside a zone are not visible to the devices inside the zone.

Zoning is similar to **LUN masking**, which is commonly used for permission management. LUN masking is a process that makes a LUN available to some hosts and unavailable to other hosts. Usually, LUN masking is performed at the SP or server level.

Overview of Using ESX Server with SAN

Support for QLogic and Emulex FC HBAs allows an ESX Server system to be connected to a SAN array. You can then use SAN array LUNs 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.

See the following sections for more information:

- "Benefits of Using ESX Server with SAN" on page 30
- "Use Cases" on page 31
- "Finding Information" on page 31

Benefits of Using ESX Server with SAN

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

- You can store data redundantly and configure multiple FC fabrics 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 42.
- Using 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 45.

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

- Because shared central storage is available, building virtual machine clusters that use MSCS becomes possible. See "Server Failover and Storage Considerations" on page 46.
- If virtual machines are used as standby systems for existing physical servers, shared storage is essential and a SAN is the best solution.
- You can use the VMware VMotion capabilities to migrate virtual machines seamlessly from one host to another.
- You can use VMware HA in conjunction with a SAN for a cold-standby solution that guarantees an immediate, automatic response.

- You can use VMware DRS to automatically migrate virtual machines from one host to another for load balancing. Because storage is on a SAN array, applications continue running seamlessly.
- If you use VMware DRS clusters, you can 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 SAN storage. When virtual machines are located on SAN-based storage, it becomes possible to 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 particularly effective for the following tasks:

- Maintenance with zero downtime – When performing maintenance, you can 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 – You can 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 LUN and then allocate portions to virtual machines as needed. LUN reservation 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 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 array vendor's documentation for most setup questions. Your storage array vendor might also offer documentation on using the storage array in an ESX Server environment.
- *VMware I/O Compatibility Guide* – Lists the currently approved HBAs, HBA drivers, and driver versions.
- *VMware Storage/SAN Compatibility Guide* – Lists currently approved storage arrays.
- *VMware Release Notes* – Give information about known issues and workarounds.
- *VMware Knowledge Bases* – Have information on common issues and workarounds.

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

Specifics of Using SAN Arrays with 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:

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. See "Virtual Machine File System (VMFS)" on page 22 for some background information on VMFS and the *Server Configuration Guide* for additional information.

VMFS can:

- Coordinate access to virtual disk files – ESX Server uses file-level locks, which are managed by the VMFS distributed lock manager.
- Coordinate access to VMFS internal file system information (metadata) – ESX Server uses SCSI reservations on the entire LUN. See "Metadata Updates" on page 33.

NOTE SCSI reservations are not 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 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 due to I/O contention.

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

Figure 2-1 shows several ESX Server systems sharing the same VMFS volume.

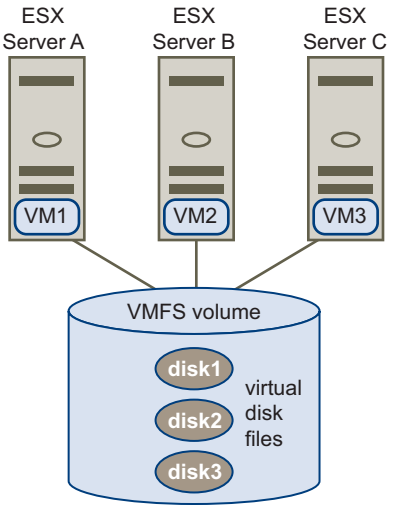


Figure 2-1. Accessing Virtual Disk Files

Metadata Updates

A VMFS holds files, directories, symbolic links, RDMs, and so on, and 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.

LUN Display and Rescan

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

- New LUNs created on the SAN storage arrays
- Changes to LUN masking
- Changes in SAN connectivity or other aspects of the SAN

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

Host Type

A LUN has a slightly different behavior depending on the type of host that is accessing it. Usually, the host type assignment deals with operating-system-specific features or issues. ESX Server systems are typically configured with a host type of Linux for LUN access.

For more information, see Chapter 6, "Managing ESX Server Systems That Use SAN Storage," on page 81 and the VMware knowledge bases.

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 editing the virtual machine settings using the VI Client, as shown in Figure 2-2. You can also add hard disks to your virtual machine. See the *Server Configuration Guide* for additional information.

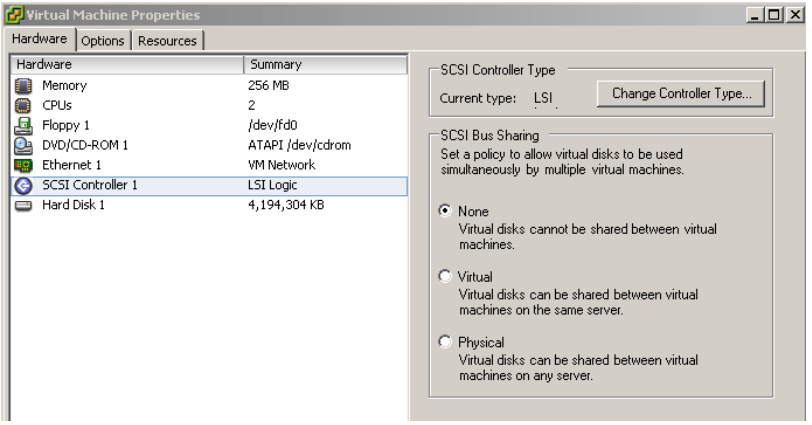


Figure 2-2. Setting the SCSI Controller Type

- 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). An exception is Dynamic Disks multipathing software in Windows virtual machines.

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 (VMFS)" on page 22.

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 22.

Third-Party Management Applications

Most SAN hardware is packaged with SAN management software. This software typically runs on the storage array 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 array management including LUN creation, array cache management, LUN mapping, and LUN security.

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

If you decide to run the SAN management software inside a virtual machine, you reap the benefits of running 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 111 for more information.

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

Zoning and ESX Server

Zoning provides access control in the SAN topology. Zoning defines which HBAs can connect to which SPs. When a SAN is configured using zoning, the devices outside a zone are not visible to the devices inside the zone.

Zoning has the following effects:

- Reduces the number of targets and LUNs presented to an ESX Server system.
- Controls and isolates paths within a fabric.
- Can prevent non-ESX Server systems from seeing a particular storage system, and from possibly destroying ESX Server VMFS data.
- Can be used to separate different environments (for example, a test from a production environment).

When you use zoning, keep in mind the following:

- ESX Server hosts that use shared storage for failover or load balancing must be in one zone.
- If you have a very large deployment, you might need to create separate zones for different areas of functionality. For example, you can separate accounting from human resources.
- It does not work well to create many small zones of, for example, two hosts with four virtual machines each.

NOTE Check with the storage array vendor for zoning best practices.

Access Control (LUN Masking) and ESX Server

Access control allows you to limit the number of ESX Server hosts (or other hosts) that can see a LUN. Access control can be useful to:

- Reduce the number of LUNs presented to an ESX Server system.
- Prevent non-ESX Server systems from seeing ESX Server LUNs and from possibly destroying VMFS volumes.

Understanding VMFS and SAN Storage Choices

This section discusses the available VMFS and SAN storage choices and gives advice on how to make them. The section covers the following topics:

- "Choosing Larger or Smaller LUNs" on page 37
- "Making LUN Decisions" on page 38
- "Tips" on page 39

Choosing Larger or Smaller LUNs

During ESX Server installation, you are prompted to create partitions for your system. You need to plan how to set up storage for your ESX Server systems before you perform installation. You can choose one of these approaches:

- Many LUNs with one VMFS volume on each LUN
- Many LUNs with a single VMFS volume spanning all LUNs

You can have at most one VMFS volume per LUN. You could, 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 going back to 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 due to locking and SCSI reservation issues.
- Different applications might need different RAID characteristics.
- More flexibility (the multipathing policy and disk shares are set per LUN).

- Use of Microsoft Cluster Service, which requires that each cluster disk resource is in its own 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 when making the decision:

- Predictive. See "Predictive Scheme" on page 38.
- Adaptive. See "Adaptive Scheme" on page 38.

Predictive Scheme

In the predictive scheme, you:

- Create several LUNs with different storage characteristics.
- Build a VMFS volume in each LUN (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. Note that 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 don't lose virtual machines when recreating the LUN.

Tips

When making your LUN decision, keep in mind the following:

- Each LUN should have the right RAID level and storage characteristic for applications in virtual machines that use it.
- One LUN must contain only one single VMFS volume.
- If multiple virtual machines access the same LUN, use disk shares to prioritize virtual machines.

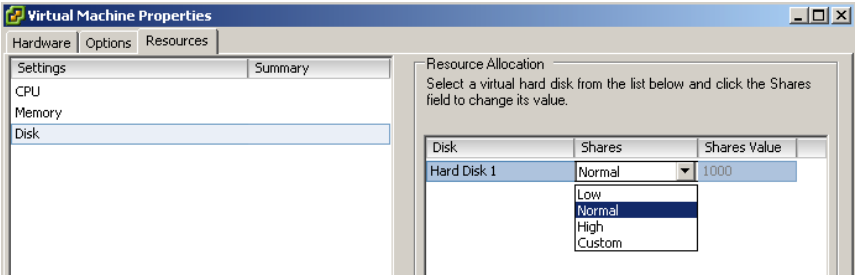
To use disk shares to prioritize virtual machines

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

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

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

4 Right-click the **Shares** column for the disk you want to modify, and select the required value from the drop-down menu.



Understanding Data Access

This section discusses:

Accessing Data on a SAN Array

If you are not familiar with how a physical (non-virtual) machine accesses data on a SAN array, 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 access data using one of the following methods:

- **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 (VMFS)" on page 22 for additional information.

- **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.

  RDM is recommended when a virtual machine must interact with a real disk on the SAN. This is the case, for example, when you make disk array snapshots or, more rarely, if you have a large amount of data that you don't want to move onto a virtual disk. It is also required for Microsoft Cluster Service setup. See the VMware document *Setup for Microsoft Cluster Service* for more information.

Figure 2-3 illustrates how virtual machines access data using VMFS or RDM.

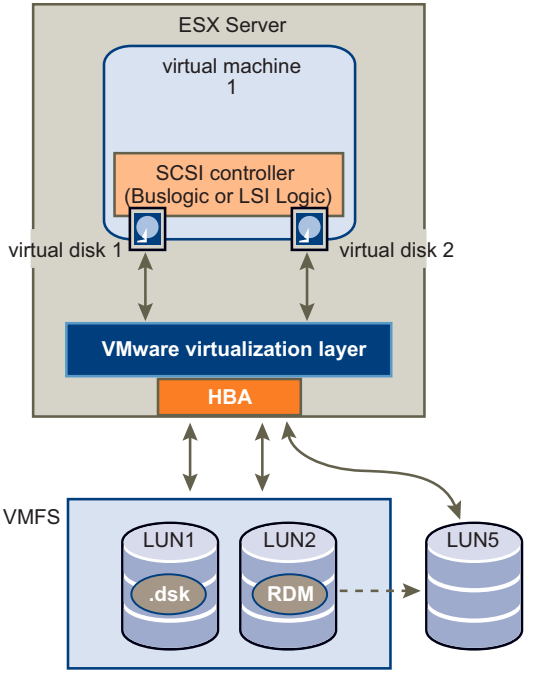


Figure 2-3. How Virtual Machines Access Data

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

How Virtual Machines Access Data on a SAN

When a virtual machine interacts with a SAN, the following steps take place:

1 When the guest operating system in a virtual machine needs to read or write to 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 physical HBA (host HBA).

5 The host HBA:

- Converts the request from its binary data form to the optical form required for transmission on the fiber optic cable.
- Packages the request according to the rules of the FC protocol.
- Transmits the request to the SAN.

6 Depending on which port the HBA uses to connect to the fabric, one of the SAN switches receives the request and routes it 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. The switch must determine which physical device has been made available to the host for its targeted logical device.

Path Management and 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, switch, SP, or FC cable. 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 array can be reached using one or more switches. Alternatively, the setup could include one HBA and two storage processors so that the HBA can use a different path to reach the disk array.

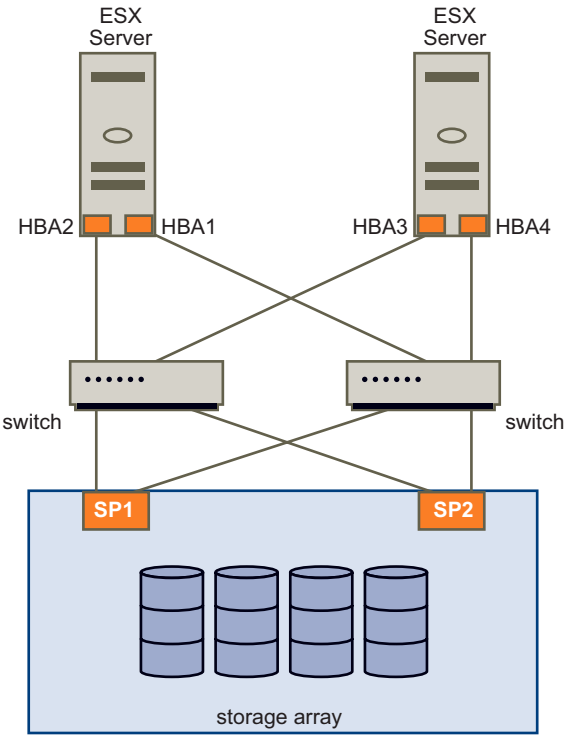


Figure 2-4. Multipathing and Failover

In Figure 2-4 multiple paths connect each server with the storage device. For example, if HBA1 or the link between HBA1 and the FC switch fails, HBA2 takes over and provides the connection between the server and the switch. 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**. VMware 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. See "Multipathing" on page 90 for information on viewing the current multipathing state and on setting the multipathing policy.

NOTE Virtual machine I/O might be delayed for at most sixty seconds while failover takes place, particularly on an active/passive array. This delay is necessary to allow the SAN fabric to stabilize its configuration after topology changes or other fabric events. In the case of an active/passive array with path policy **Fixed**, path thrashing might be a problem. See "Resolving Path Thrashing" on page 104.

Choosing Virtual Machine Locations

When you're working on optimizing performance for your virtual machines, storage location is an important factor. There is always a trade-off 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. May offer built-in snapshots to facilitate backups and Point-in-Time (PiT) restorations. Supports replication, full SP redundancy, and fibre drives. Uses high-cost spindles.
- **Mid Tier** – Offers mid-range performance, lower availability, some SP redundancy, and SCSI 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 don't 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 point-in-time (PiT) restoration requirements?
- What are its backup requirements?
- What are its replication requirements?

A virtual machine might change tiers throughout its life cycle due to 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 dual fabric, with duplication of all fabric components, protects the SAN from most fabric 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" on page 45
- "Using Cluster Services" on page 45
- "Server Failover and Storage Considerations" on page 46

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. HA requires SAN storage.

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* for detailed information.

NOTE You must be licensed to use VMware HA.

Using Cluster Services

Server clustering is a method of tying two or more servers together using a high-speed network connection so that the group of servers functions as a single, logical server. If one of the servers fails, the other servers in the cluster continue operating, picking up the operations performed by the failed server.

VMware tests Microsoft Cluster Service in conjunction with ESX Server systems, but other cluster solutions might also work. Different configuration options are available for achieving failover with clustering:

- **Cluster in a box** – Two virtual machines on one host act as failover servers for each other. When one virtual machine fails, the other takes over. (This does not protect against host failures. It is most commonly done during testing of the clustered application.)
- **Cluster across boxes** – For a virtual machine on an ESX Server host, there is a matching virtual machine on another ESX Server host.

- **Physical to virtual clustering (N+1 clustering)** – A virtual machine on an ESX Server host acts as a failover server for a physical server. Because virtual machines running on a single host can act as failover servers for numerous physical servers, this clustering method provides a cost-effective N+1 solution.

See *Setup for Microsoft Cluster Service* for more information.

Figure 2-5 shows different configuration options available for achieving failover with clustering.

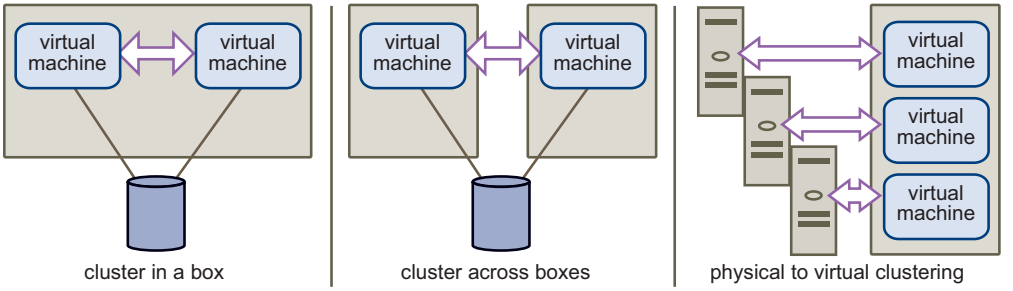


Figure 2-5. Clustering Using a Clustering Service

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 array 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 LUNs that are used by the clustered virtual machines must be seen by all ESX Server hosts. This is counterintuitive for SAN administrators, but is appropriate when using virtual machines.

  Although a LUN is accessible to a host, all virtual machines on that host do not necessarily have access to all data on that LUN. 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 using boot from SAN, each boot LUN should be seen only by the ESX Server system that is booting from that LUN. An exception is when you're trying to recover from a crash by pointing a second ESX Server system to the same LUN. In this case, the SAN LUN in question is not really a boot from SAN LUN. No ESX Server system is booting from it because it is corrupted. The SAN LUN is a regular non-boot LUN that is made visible to an ESX Server system.

Optimizing Resource Utilization

Virtual Infrastructure allows you to optimize resource allocation by migrating virtual machines from overutilized hosts to underutilized hosts. There are two options:

- Migrate virtual machines explicitly using VMotion. See "Using VMotion to Migrate Virtual Machines" on page 47.
- Migrate virtual machines automatically using VMware DRS. "Using VMware DRS to Migrate Virtual Machines" on page 48.

You can use VMotion or DRS only if the virtual disks are located on shared storage accessible to multiple servers. In most cases, SAN storage is used. For additional information on VMotion, see *Basic System Administration*. For additional information on DRS, see the *Resource Management Guide*.

Using VMotion to Migrate Virtual Machines

VMotion technology enables intelligent workload management. VMotion allows administrators to manually migrate virtual machines to different hosts. Administrators can migrate a running virtual machine to a different physical server connected to the same SAN without service interruption. VMotion makes it possible to:

- Perform zero-downtime maintenance by moving virtual machines around so the underlying hardware and storage can be serviced without disrupting user sessions.
- Continuously balance workloads across the datacenter to most effectively use resources in response to changing business demands.

Figure 2-6 illustrates how you can use VMotion to migrate a virtual machine.

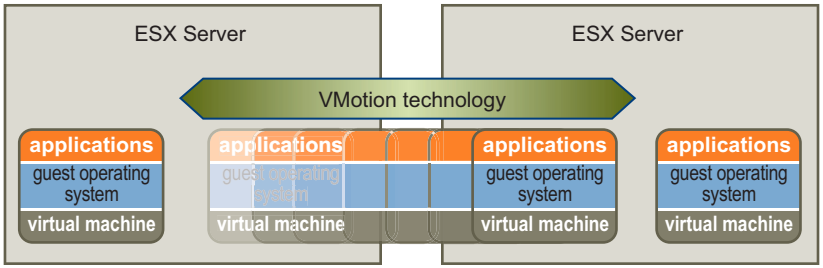


Figure 2-6. Migration with VMotion

Using VMware DRS to Migrate Virtual Machines

VMware DRS helps improve resource allocation across all hosts and resource pools. DRS collects resource usage information for all hosts and virtual machines in a VMware cluster and gives recommendations (or migrates virtual machines) in one of two situations:

- **Initial placement** – When you first power on a virtual machine in the cluster, DRS either places the virtual machine or makes a recommendation.
- **Load balancing** – DRS tries to improve resource utilization across the cluster by performing automatic migrations of virtual machines (VMotion) or by providing recommendation for virtual machine migrations.

For detailed information, see the *Resource Management Guide*.

Requirements and Installation

3

This chapter presents hardware and system requirements for using ESX Server systems with SAN storage. The chapter consists of the following sections:

This chapter lists only the most basic requirements. Before you set up your system, read Chapter 4, "Setting Up SAN Storage Devices with ESX Server," on page 57.

General ESX Server SAN Requirements

In preparation for configuring your SAN and setting up your ESX Server system to use SAN storage, review the following requirements and recommendations:

- **Hardware and firmware**. Only a limited number of SAN storage hardware and firmware combinations are supported in conjunction with ESX Server systems. See the *Storage/SAN Compatibility Guide* for an up-to-date list.
- **One VMFS volume per LUN**. Make sure that you configure your system to have only one VMFS volume per LUN. (In VMFS-3, you no longer have to set accessibility.)
- Unless you're using diskless servers, do not set up the **diagnostic partition** on a SAN LUN.

  In the case of diskless servers that boot from SAN, a shared diagnostic partition is appropriate. See "Sharing Diagnostic Partitions" on page 100 for additional information.
- Using **RDMs** is recommended for access to any raw disk from an ESX Server 2.5 or later machine. For more information on RDMs, see the *Server Configuration Guide*.
- **Multipathing**. For multipathing to work properly, each LUN must present the same LUN number to all ESX Server hosts.
- **Queue size**. Make sure the BusLogic or LSILogic driver in the guest operating system specifies a big 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, consider increasing the value of the SCSI `TimeoutValue` parameter to allows Windows to better tolerate delayed I/O resulting from path failover. See "Setting Operating System Timeout" on page 99.

For more information, see the following sections:

- "Restrictions" on page 50
- "Setting LUN Allocations" on page 51
- "Setting FC HBA" on page 51

Restrictions

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

- ESX Server does not support FC connected tape devices. These devices can be managed by the VMware Consolidated Backup proxy, which is discussed in the *Virtual Machine Backup Guide*.
- Virtual machine multipathing software cannot be used to perform I/O load balancing to a single physical LUN.
- You cannot use virtual machine logical volume manager software to mirror virtual disks. Dynamic disks in a Microsoft Windows virtual machine are an exception but require special configuration.

Setting LUN Allocations

When setting 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 Provisioning all LUNs to all ESX Server HBAs at the same time is recommended. 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 provides the greatest freedom in moving virtual machines.
- **Active/active versus active/passive arrays** – When using 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 104.

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

Setting FC HBA

During FC HBA setup, consider the following points:

- **HBA Default Settings** – FC HBAs work correctly with the default configuration settings. Follow the configuration guidelines given by your storage array vendor.

NOTE For best results, use the same model of HBA within one server. Ensure that the firmware level on each HBA is the same within one server. Having both Emulex and QLogic HBAs in the same server to the same target is not supported.

- **Static Load Balancing Across HBAs** – Some ESX Server systems can be configured 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. For example, if you have two LUNs (A and B) and two HBAs (X and Y), you can set HBA X to be the preferred path for LUN A, and HBA Y as the preferred path for LUN B. This will maximize utilization of your HBAs. Path policy must be set to **Fixed** for this case. See "To set the multipathing policy using a VI Client" on page 94.

- **Setting the Timeout for Failover** – The timeout value for detecting when a path fails is set in the HBA driver. Setting the timeout to 30 seconds ensures optimal performance and is highly recommended. To set the value, follow the instructions in "Setting the HBA Timeout for Failover" on page 98.

- **Dedicated Adapter for Tape Drives** – For best results, use a dedicated SCSI adapter for any tape drives that you are connecting to an ESX Server system. FC connected tape drives are not supported. Use the Consolidated Backup proxy, as discussed in the *Virtual Machine Backup Guide*.

  For additional information on boot from SAN HBA setup, see Chapter 5, "Using Boot from SAN with ESX Server Systems," on page 71.

Recommendations

The following provides instructions on making decisions when setting 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 disk array, or to access a disk from both a virtual machine and a physical machine in a cold standby host configuration for data LUNs.

- Use raw device mapping for the shared disks in a Microsoft Cluster Service setup. See the VMware document *Setup for Microsoft Cluster Service* for details.

- Allocate a large LUN for use by multiple virtual machines and set it up as a VMFS. You can then create or delete virtual machines dynamically without having to request additional disk space each time you add a virtual machine.

- If you want to move a virtual machine to a different host using VMotion, the LUNs that hold the virtual disks of the virtual machines must be masked so they are visible from all the hosts.

See Chapter 6, "Managing ESX Server Systems That Use SAN Storage," on page 81 for additional recommendations and troubleshooting information.

ESX Server Boot from SAN Requirements

When you have SAN storage configured with your ESX Server system, you can place the ESX Server boot image on one of the LUNs on the SAN. This configuration must meet specific criteria, discussed in this section. See "Using Boot from SAN with ESX Server Systems" on page 71 for more information.

To enable your ESX Server system to boot from SAN, you should:

- Meet the general requirements. See "General ESX Server SAN Requirements" on page 50.
- Complete tasks listed in Table 3-1.

Table 3-1. Boot from SAN Requirements

Requirement	Description
ESX Server system requirements	ESX Server 3.0 or later is recommended. When you use a 3.0 ESX Server system, RDMs are supported in conjunction with boot from SAN. For an ESX Server 2.5.x system, RDMs are not supported in conjunction with boot from SAN.
HBA requirements	The HBA BIOS for your HBA FC card must be enabled and correctly configured to access the boot LUN. See "Setting FC HBA" on page 51. The HBA should be plugged into the lowest PCI bus and slot number. This allows the drivers to detect the HBA quickly because the drivers scan the HBAs in ascending PCI bus and slot numbers, regardless of the associated virtual machine HBA number. **Note**: For precise driver and version information, see the *ESX Server I/O Compatibility Guide*.
Boot LUN considerations	■ When you boot from an active/passive storage array, the storage processor whose WWN is specified in the BIOS configuration of the HBA must be active. If that storage processor is passive, the HBA cannot support the boot process. ■ To facilitate BIOS configuration, mask each boot LUN so that it can be seen only by its own ESX Server system. Each ESX Server system should see its own boot LUN, but not the boot LUN of any other ESX Server system.

Table 3-1. Boot from SAN Requirements

Requirement	Description
SAN considerations	■ SAN connections must be through a switch fabric topology. Boot from SAN does not support direct connect (that is, connection without switches) or FC arbitrated loop connections. ■ Redundant and non-redundant configurations are supported. In the redundant case, ESX Server collapses the redundant paths so only a single path to a LUN is presented to the user.
Hardware-specific considerations	■ If you are running an IBM eServer BladeCenter and use boot from SAN, you must disable IDE drives on the blades. For additional hardware-specific considerations. check the VMware knowledge base articles and see Chapter 4, "Setting Up SAN Storage Devices with ESX Server," on page 57.

Installation and Setup Steps

Table 3-2 gives an overview of the installation and setup steps, with pointers to relevant information.

Table 3-2. Installation and Setup Steps

Step	Description	See
1	Design your SAN if it's not already configured. Most existing SANs require only minor modification to work with ESX Server.	Chapter 2, "Using ESX Server with SAN," on page 27.
2	Check that all SAN components meet requirements.	Chapter 3, "General ESX Server SAN Requirements," on page 50. *Storage/SAN Compatibility Guide*.
3	Set up the HBAs for the ESX Server hosts.	For special requirements that apply only to boot from SAN, see Chapter 3, "ESX Server Boot from SAN Requirements," on page 53. See also Chapter 5, "Using Boot from SAN with ESX Server Systems," on page 71.
4	Perform any necessary storage array modification.	For an overview, see Chapter 4, "Setting Up SAN Storage Devices with ESX Server," on page 57. Most vendors have vendor-specific documentation for setting up a SAN to work with VMware ESX Server.
5	Install ESX Server on the hosts you have connected to the SAN and for which you've set up the HBAs.	*ESX Server Installation and Upgrade Guide*.

Table 3-2. Installation and Setup Steps

Step	Description	See
6	Create virtual machines.	*Basic System Administration*.
7	Set up your system for VMware HA failover or for using Microsoft Clustering Services. This step is optional.	*VMware Resource Management Guide: ESX Server 3.0 and VirtualCenter 2.0.* *Setup for Microsoft Cluster Service.*
8	Upgrade or modify your environment as needed.	Chapter 6, "Managing ESX Server Systems That Use SAN Storage," on page 81 gives an introduction. Search the VMware knowledge base articles for machine-specific information and late-breaking news.

Setting Up SAN Storage Devices with ESX Server

4

This chapter discusses many of the storage devices supported in conjunction with VMware ESX Server. For each device, it lists the major known potential issues, points to vendor-specific information (if available), and includes information from VMware knowledge base articles.

NOTE Information in this document is updated only with each release. New information might already be available. Consult the most recent *Storage/SAN Compatibility Guide*, check with your storage array vendor, and explore the VMware knowledge base articles.

This chapter discusses the following topics:

Setup Overview

VMware ESX Server supports a variety of SAN storage arrays 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. Check *Storage/SAN Compatibility Guide* for the latest information regarding supported storage arrays.

For more details, see these sections:

- "Testing" on page 58
- "Supported Devices" on page 58

Testing

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

- **Basic Connectivity** – Tests whether ESX Server can recognize and operate with the storage array. 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.
- **Boot From SAN** – The ESX Server host boots from a LUN configured on the SAN rather than from the server itself.
- **Direct Connect** – The server connects to the array without using switches, using only an FC cable. For all other tests, a fabric connection is used. FC Arbitrated Loop (AL) is not supported.
- **Clustering** – The system is tested with Microsoft Cluster Service running in the virtual machine. See the *Setup for Microsoft Cluster Service* document included in this documentation set for information.

Supported Devices

Table 4-1 lists storage devices supported with ESX Server 3.0 and points where to find more information about using them in conjunction with ESX Server.

Table 4-1. Supported SAN Arrays

Manufacturer	Device	See
EMC	CLARiiON Storage System. Also available from FSC. Also available from Dell, Inc. as the Dell / EMC FC RAID Array family of products.	"EMC CLARiiON Storage Systems" on page 60.
	Symmetrix Storage System.	"EMC Symmetrix Storage Systems" on page 61.
IBM	IBM TotalStorage DS 4000 systems (formerly FAStT Storage system). Also available from LSI Eugenio and StorageTek.	"IBM TotalStorage DS4000 Storage Systems" on page 62.
	IBM TotalStorage Enterprise Storage Systems (previously Shark Storage systems).	"IBM TotalStorage 8000" on page 66.
	IBM SVC (SAN Volume Controller).	
Hewlett Packard	HP StorageWorks (MSA, EVA, and XP).	"HP StorageWorks Storage Systems" on page 66.
Hitachi	Hitachi Data Systems Storage. Also available from Sun and as HP XP.	"Hitachi Data Systems Storage" on page 69.
Network Appliance	Network Appliance FC SAN Storage Solutions.	"Network Appliance Storage" on page 69.

General Considerations

For all storage arrays, 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, you should consult your storage array documentation for more information.

- Unless specified for individual storage arrays discussed in this chapter, set the host type for LUNs presented to ESX Server to `Linux` or `Linux Cluster`.

- If you are using VMotion, DRS, or HA, make sure that both source and target hosts for virtual machines can see the same LUNs.

  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.

EMC CLARiiON Storage Systems

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

1 Installing and configuring your storage device.

2 Configuring zoning at the switch level.

3 Creating RAID groups.

4 Creating and binding LUNs.

5 Registering the servers connected to the SAN.

6 Creating storage groups that contain the servers and LUNs.

Use the EMC software to perform configuration. See the EMC documentation for information.

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 array. See "Resolving Path Thrashing" on page 104.

NOTE To use boot from SAN, make sure that the active storage processor is chosen for the boot LUN's target in the HBA BIOS.

For more details, see "EMC CLARiiON AX100" on page 60.

EMC CLARiiON AX100

If you work with an EMC CLARiiON storage array, read the following sections:

- "EMC CLARiiON AX100 and RDM" on page 61
- "AX100 Display Problems with Inactive Connections" on page 61

■ "Automatic Volume Resignaturing" on page 61

EMC CLARiiON AX100 and RDM

On EMC CLARiiON AX100 systems, RDMs are supported only if you use the Navisphere Management Suite for SAN administration. Navilight is not guaranteed to work 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 AX100 does not do this by default.

AX100 Display Problems with Inactive Connections

When using an AX100 FC storage device directly connected to an ESX Server systems, you must verify that all connections are operational and deregister any connections that are no longer in use. If you don't, ESX Server cannot discover new LUNs or paths.

Consider the following scenario:

1 An ESX Server system is directly connected to an AX100 storage device. The ESX Server has two FC HBAs. One of the HBAs was previously registered with the storage array and its LUNs were configured, but the connections are now inactive.

2 When you connect the second HBA in the ESX Server host to the AX100 and register it, the ESX Server host correctly shows the array as having an active connection. However, none of the LUNs that were previously configured to the ESX Server host are visible, even after repeated rescans.

To resolve this issue, remove the inactive HBA, deregister the connection to the inactive HBA, or make all inactive connections active. This causes only active HBAs to be in the storage group. After this change, rescan to add the configured LUNs.

Automatic Volume Resignaturing

Automatic volume resignaturing is not supported for AX100 storage devices. See "VMFS Volume Resignaturing" on page 112 for information on resignaturing.

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
- Fibrepath enabled on this port (VCM)
- SCSI 3 (SC3) set (enabled)

- Unique world wide name (UWN)
- SPC-2 (Decal) (SPC2) SPC-2 flag unset (disabled)

You configure the storage array using EMC software. Refer to your EMC documentation for information.

NOTE The ESX Server host considers any LUNs from a Symmetrix storage array 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.

IBM TotalStorage DS4000 Storage Systems

IBM TotalStorage DS4000 systems used to be called IBM FAStT. A number of storage array vendors (including LSI and StorageTek) make SAN storage arrays that are compatible with the DS4000.

See the IBM Redbook, *Implementing VMware ESX Server with IBM TotalStorage FAStT* at **http://www.redbooks.ibm.com/redbooks/pdfs/sg246434.pdf** for a detailed discussion. This section summarizes configuring your IBM TotalStorage Storage System for using SAN and Microsoft Clustering Service. See the *Setup for Microsoft Cluster Service* document for additional information.

In addition to normal configuration steps for your IBM TotalStorage storage system, you need to perform specific tasks.

For more information, see the following sections:

- "Configuring the Hardware for SAN Failover with DS4000 Storage Servers" on page 62
- "Verifying the Storage Processor Port Configuration" on page 63
- "Disabling Auto Volume Transfer" on page 64
- "Configuring Storage Processor Sense Data" on page 65
- "IBM TotalStorage DS4000 and Path Thrashing" on page 65

You must also make sure that multipathing policy is set to **Most Recently Used**. See "Viewing the Current Multipathing State" on page 91.

Configuring the Hardware for SAN Failover with DS4000 Storage Servers

To set up a highly available SAN failover configuration with DS4000 storage models equipped with two storage processors, you need the following hardware components:

- Two FC HBAs, such as QLogic or Emulex, on each ESX Server machine.

- Two FC switches connecting the HBAs to the SAN (for example, FC switch 1 and FC switch 2).
- Two SPs (for example, SP1 and SP2).

  Each SP must have at least two ports connected to the SAN.

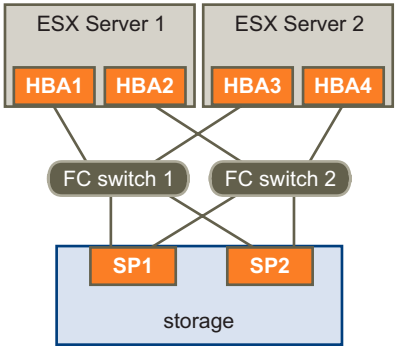


Figure 4-1. SAN Failover

Use the following connection settings for the ESX Server host as shown in Figure 4-1:

- Connect each HBA on each ESX Server machine to a separate switch. For example, connect HBA1 to FC switch 1 and HBA2 to FC switch 2.
- On FC switch 1, connect SP1 to a lower switch port number than SP2, to ensure that SP1 is listed first. For example, connect SP1 to FC switch 1 port 1 and SP2 to FC switch 1 port 2.
- On FC switch 2, connect SP1 to a lower switch port number than SP2, to ensure that SP1 is listed first. For example, connect SP1 to port 1 on FC switch 2 and SP2 to port 2 on FC switch 2.

This configuration provides two paths from each HBA, so each element of the connection can fail over to a redundant path. The order of the paths in the above configuration provides HBA and switch failover without the need to trigger SP failover. LUNs must be owned by the storage processor to which the preferred paths are connected (in the example configuration above, they should be owned by SP1).

NOTE The above examples assume that the switches are not connected through an Inter-Switch Link (ISL) in one fabric.

Verifying the Storage Processor Port Configuration

You can verify the storage processor port configuration by comparing the VI Client information with the information in the DS4000 subsystem profile.

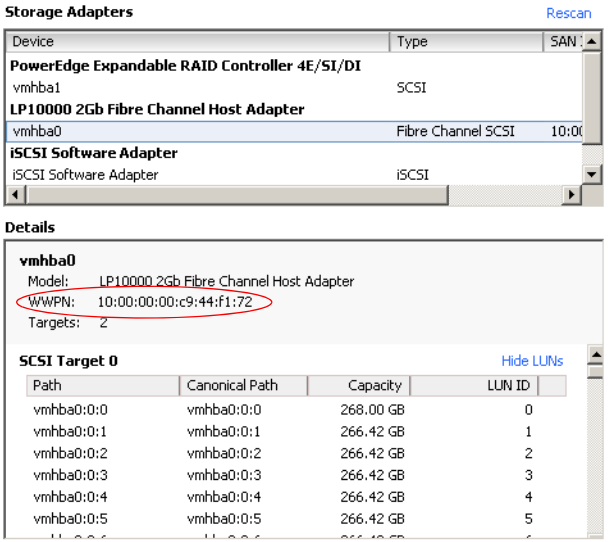
To verify storage processor port configuration

1 Connect to the ESX Server host using a VI Client.

2 Select the host and choose the **Configuration** tab.

3 Click **Storage Adapters** in the Hardware panel.

4 Select each storage adapter to see its WWPN.



5 Select **Storage** to see the available datastores.

Compare the WWPN information to the information listed in the DS4000 storage subsystem profile.

Disabling Auto Volume Transfer

To avoid the possibility of path thrashing, disable AVT (Auto Volume Transfer) on the SAN storage processors. If AVT is enabled, the two storage processors can alternately take ownership of the LUN in certain situations, resulting in performance degradation. AVT is also known as ADT (Auto Disk Transfer).

See "Resolving Path Thrashing" on page 104 for additional information.

To disable AVT, set the host type to `LNXCL` in the DS 4000 Storage Manager for each port defined in each host group containing HBAs for one or more ESX Server machines.

NOTE You must reboot the ESX Server host after you change the AVT configuration.

Configuring Storage Processor Sense Data

Storage processors can be configured to return either `Unit Attention` or `Not Ready` when quiescent. A DS4000 storage processor that is running Windows as a guest operating system should return `Not Ready` sense data when it is quiescent. Returning `Unit Attention` might cause the Windows guest to crash during a failover.

To configure the storage processors to return `Not Ready` sense data

1 Determine the index for the host type `LNXCL` by using the following commands in a shell window. Press Enter after each command.

```
SMcli.exe <ip-addr-for-SPA> show hosttopology; <Enter>
SMcli.exe <ip-addr-for-SPB> show hosttopology; <Enter>
```

The following commands assume `13` is the index corresponding to `LNXCL` in the NVSRAM host type definitions. If your storage processors have `LNXCL` at a different index, substitute that index for `13` in the following commands.

2 Execute these commands for SPA to have it return `Not Ready` sense data. Press Enter only after you have entered all commands.

```
SMcli.exe <ip-addr-for-SPA>
   set controller [a] HostNVSRAMBYTE [13,0x12]=0x01;
   set controller [a] HostNVSRAMBYTE [13,0x13]=0x00;
   reset Controller [a];
   <Enter>
```

3 Execute these commands for SPB to have it return `Not Ready` sense data. Press Enter only after you have entered all commands.

```
SMcli.exe <ip-addr-for-SPB>
   set controller [b] HostNVSRAMBYTE [13,0x12]=0x01;
   set controller [b] HostNVSRAMBYTE [13,0x13]=0x00;
   reset Controller [b];
   <Enter>
```

NOTE If you use the DS4000 Storage Manager GUI, paste the configuration commands for both storage processors into a single script and configure both storage processors at the same time. If you use `SMcli.exe`, you need to make individual connections to each storage processor.

IBM TotalStorage DS4000 and Path Thrashing

When path thrashing is detected on a DS 4000 or compatible SAN array, the following warning is logged to the vmkernel log.

```
FAStT SAN is path thrashing with another system. Check AVT setting.
```

IBM TotalStorage 8000

IBM TotalStorage 8000 systems use an active/active array that does not need special configuration in conjunction with VMware ESX Server.

To use RDMs successfully, a given LUN needs to be presented with the same LUN ID to every ESX Server host in the cluster.

In the TotalStorage Configuration Management tool, select the **Use same ID for LUN in source and target** check box.

NOTE If you are configuring the ESX Server host to use boot from SAN from a LUN on an IBM TotalStorage 8000 array, disable the internal fibre port for the corresponding blade until installation is finished.

Automatic Resignaturing

Automatic resignaturing is not supported for IBM TotalStorage 8000 systems.

HP StorageWorks Storage Systems

This section includes configuration information for the different HP StorageWorks storage systems. The section covers these topics:

- "HP StorageWorks MSA" on page 66
- "HP StorageWorks EVA" on page 68
- "HP StorageWorks XP" on page 69

For additional information, see the HP ActiveAnswers section on VMware ESX Server at **http://h71019.www7.hp.com/ActiveAnswers/cache/71086-0-0-0-121.html**.

HP StorageWorks MSA

This section lists issues of interest if you are using the active/passive version of the HP StorageWorks MSA.

Setting Profile Name to Linux

To use HP StorageWorks MSA 1000 and MSA 1500 with ESX Server systems, you must configure the FC connections between the SAN array and the ESX Server host with the Profile Name set to `Linux`.

To set the Profile Name for a connection

1 Create a static connection on the MSA 1000 using the MSA 1000 command-line interface.

See the HP StorageWorks MSA 1000 documentation for information on installing and configuring the command-line interface.

NOTE You cannot create connection settings using the HP Array Configuration utility.

2 Connect the MSA 1000 command-line interface to the MSA 1000.

3 Verify that the FC network between the MSA 1000 and the ESX Server host is working.

4 Start the command-line interface and enter the following at the prompt:

```
SHOW CONNECTIONS
```

The output displays a connection specification for each FC WWNN and WWPN attached to the MSA 1000:

```
Connection Name: <unknown>
Host WWNN = 20:02:00:a0:b8:0c:d5:56
Host WWPN = 20:03:00:a0:b8:0c:d5:57
Profile Name = Default
Unit Offset 0
Controller 1 Port 1 Status = Online
Controller 2 Port 1 Status = Online
```

5 Make sure the host's WWNN and WWPN show the correct connection for each Fiber Channel adapter on the ESX Server machine.

6 Create a static connection as follows:

```
ADD CONNECTION ESX_CONN_1 WWNN=20:02:00:a0:b8:0c:d5:56
WWPN=20:03:00:a0:b8:0c:d5:57 PROFILE=LINUX
```

7 Verify the connection as follows:

```
SHOW CONNECTIONS
```

The output displays a single connection with the WWNN and WWPN pair `20:02:00:a0:b8:0c:d5:56` and `20:03:00:a0:b8:0c:d5:57` and the Profile Name set to `Linux`:

```
Connection Name: ESX_CONN_1
Host WWNN = 20:02:00:a0:b8:0c:d5:56
Host WWPN = 20:03:00:a0:b8:0c:d5:57
Profile Name = Linux
Unit Offset = 0
Controller 1 Port 1 Status = Online
```

```
Controller 2 Port 1 Status = Online
```

NOTE Make sure WWNN = `20:02:00:a0:b8:0c:d5:56` and WWPN = `20:03:00:a0:b8:0c:d5:57` display a single connection.

There should be no connection with the Connection Name `unknown` for `WWNN= 20:02:00:a0:b8:0c:d5:56` and `WWPN = 20:03:00:a0:b8:0c:d5:57`.

8 Add static connections (with different connection name values) for each WWNN and WWPN on the ESX Server host.

Hub Controller Issues

The ESX Server system might not function correctly with the MSA hub controller. Use the 2/8 internal switch or the single port controller instead.

HP StorageWorks EVA

There are two types of HP StorageWorks EVA systems: EVA_GL, an active/passive system, and EVA_XL, an active/active system.

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

- For HP EVAgl 3000/5000 (active/passive), use the host mode type `000000002200282E`.
- For HP EVAgl firmware 4.001 (active/active firmware for GL series) and above, use the host mode type `VMware`.
- For EVA4000/6000/8000 active/active arrays with firmware below 5.031, use the host mode type `000000202200083E`.
- For EVA4000/6000/8000 active/active arrays with firmware 5.031 and above, use the host mode type `VMware`.

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 **http://h71019.www7.hp.com/ActiveAnswers/downloads/VMware3_StorageWorks_BestPractice.pdf**.

HP StorageWorks XP

For HP StorageWorks XP, set the host mode to `Windows` (not `Linux`). This system is available from Hitachi Data Systems. See "Hitachi Data Systems Storage" on page 69 for additional information.

Hitachi Data Systems Storage

This section introduces the setup for Hitachi Data Systems storage. This storage solution is also available from Sun and as HP XP storage.

- **LUN Masking** – To mask LUNs on an ESX Server host, use the HDS Storage Navigator software for best results.
- **Microcode and Configurations** – Check with your HDS representative for exact configurations and microcode levels needed for interoperability with ESX Server. If your microcode is not supported, interaction with ESX Server is usually not possible.
- **Modes** – The modes you set depend on the model you are using:
  - 9900 and 9900v uses Netware host mode.
  - 9500v series uses Hostmode1: standard and Hostmode2: SUN Cluster.

Network Appliance Storage

When configuring a Network Appliance storage device, follow instructions in these sections:

- "Setting LUN Type and Initiator Group Type" on page 69
- "Provisioning Storage" on page 69

Setting LUN Type and Initiator Group Type

You must set the appropriate LUN type and initiator group type for the storage array:

- **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.

To provision storage from a Network Appliance storage device

1 Using CLI or the FilerView GUI, 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 just 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**.

Using Boot from SAN with ESX Server Systems

5

This chapter discusses benefits of boot from SAN and describes the tasks you need to perform if you plan to have the ESX Server boot image stored on a SAN LUN.

NOTE Skip this chapter if you do not plan to have your ESX Server host boot from a SAN.

The chapter discusses the following topics:

Boot from SAN Overview

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

Use boot from SAN:

- If you don't 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 use boot from SAN:

- If you are using Microsoft Cluster Service.
- If there is a risk of I/O contention between the service console and VMkernel.

NOTE With ESX Server 2.5, you could not use boot from SAN together with RDM. With ESX Server 3.0, this restriction has been removed.

See "How Boot from SAN Works" on page 72 for more information.

How Boot from SAN Works

When you set up your system to use boot from SAN, the boot image is not stored on the ESX Server system's local disk but instead on a SAN LUN as Figure 5-1 shows.

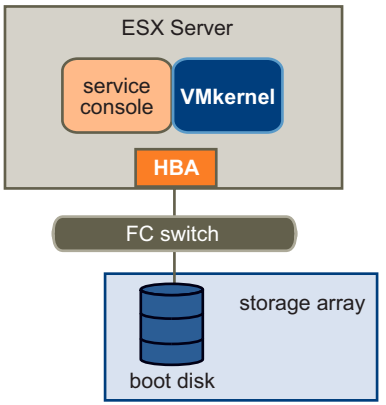


Figure 5-1. How Boot from SAN Works

On a system set up to boot from SAN:

- The HBA BIOS must designate the FC card as the boot controller. See "Setting Up the FC HBA for Boot from SAN" on page 76.

- The FC card must be configured to initiate a primitive connection to the target boot LUN.

Benefits of Boot from SAN

In a boot from SAN environment, the operating system is installed on one or more LUNs in the SAN array. The servers are informed about the boot image location. When the servers are started, they boot from the LUNs on the SAN array.

NOTE When you use boot from 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.

Getting Ready for Boot from SAN

In addition to the general ESX Server with SAN configuration tasks, you must also complete the following tasks to enable your ESX Server host to boot from SAN.

To enable boot from SAN

1 Ensure that the configuration settings meet the basic boot from SAN requirements.

See "ESX Server Boot from SAN Requirements" on page 53.

2 Prepare the hardware elements.

This includes your HBA, network devices, and storage system. Refer to the product documentation for each device. See also "Setting Up the FC HBA for Boot from SAN" on page 76.

3 Configure LUN masking on your SAN.

This ensures that each ESX Server host has a dedicated LUN for the boot partitions. The boot LUN must be dedicated to a single server.

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. See "Sharing Diagnostic Partitions" on page 100.

The rest of this section lists the tasks you need to complete before you can successfully boot your ESX Server machine from SAN. It covers the following topics:

Before You Begin

You need to review the following:

1 Review the recommendations or sample setups for the type of setup you want:

- Single or redundant paths to the boot LUN.
- FC switch fabric.
- Any specific recommendations that apply to the type of storage array you have.

2 Review restrictions and requirements including:

- Boot-from-SAN restrictions.
- The vendor's recommendation for the storage array to be used for booting from SAN.
- The vendor's recommendation for the server booting from SAN.

3 Find the WWN for the boot path HBA using one of the following methods:

- Go into the FC HBA BIOS upon boot.
- Find the WWN on the physical card. It is similar to a MAC address.

LUN Masking in Boot from SAN Mode

Proper LUN masking is critical in boot from SAN mode.

- Each server may see only its own boot LUN, not the boot LUNs of other servers.

- Multiple servers can share a diagnostic partition. You can use LUN masking to achieve this. See "Sharing Diagnostic Partitions" on page 100.

Preparing the SAN

This section lists the steps for preparing the SAN storage array for boot from SAN. Steps 3-7 are specific to boot from SAN, while steps 1 and 2 are not.

To prepare the SAN

1. Connect the FC and Ethernet cables, referring to any cabling guide that applies to your setup.

   Check the FC switch wiring, if there is any.

2. Configure the storage array.

   a. From the SAN storage array, make the ESX Server host visible to the SAN. This is often referred to as creating an object.

   b. From the SAN storage array, set up the ESX Server host to have the WWPNs of the host's FC adapters as port names or node names.

   c. Create LUNs.

   d. Assign LUNs.

   e. Record the IP addresses of the FC switches and storage arrays.

   f. Record the WWPN for each storage processor and host adapter involved.

NOTE If you use scripted installation to install ESX Server in boot from SAN mode, you need to take special steps to avoid unintended data loss. For more information, see VMware knowledge base article 1540 at **http://www.vmware.com/support/kb/enduser/std_adp.php?p_faqid=1540**.

3. Configure the HBA BIOS for boot from SAN, as discussed in the following sections:

   - "Setting Up the QLogic FC HBA for Boot from SAN" on page 76.
   - "Setting Up the Emulex FC HBA for Boot from SAN" on page 78

4. Boot your ESX Server system from the ESX Server installation CD.

   See the *Installation and Upgrade Guide* for additional information.

The QLogic BIOS uses a search list of paths (`wwpn:lun`) to locate a boot image. If one of the `wwpn:lun` paths is associated with a passive path (as could be the case with

CLARiiON or IBM TotalStorage DS 4000 systems), the BIOS stays with the passive path and does not locate an active path. If you are booting your ESX Server system from a SAN LUN, the boot fails while the host tries to access the passive path.

Minimizing the Number of Initiators

Be sure the zone contains the minimum number of host and storage ports possible. The Emulex and QLogic BIOS can become unresponsive if there are several other initiators in the same zone and you try to select a boot LUN.

For example, if there are fifteen initiators and four Symmetrix ports in one zone, you might not be able to select a boot device from either the Emulex or QLogic BIOS because it becomes unresponsive. If you zone the two host ports to see only the four storage ports, selecting a boot LUN becomes possible.

Setting Up the FC HBA for Boot from SAN

This section discusses how to set up the HBAs:

- "Setting Up the QLogic FC HBA for Boot from SAN" on page 76
- "Setting Up the Emulex FC HBA for Boot from SAN" on page 78

Setting Up the QLogic FC HBA for Boot from SAN

Configuring the QLogic HBA BIOS to boot ESX Server from SAN includes the following tasks:

- "Enabling the QLogic HBA BIOS" on page 76
- "Enabling the Selectable Boot" on page 77
- "Selecting the Boot LUN" on page 77
- "Setting Up Your System to Boot from CD-ROM First" on page 78

NOTE If you are using an IBM BladeCenter, disconnect all your local disk drives from the server.

Enabling the QLogic HBA BIOS

When configuring the QLogic HBA BIOS to boot ESX Server from SAN, start with enabling the QLogic HBA BIOS.

To enable the QLogic HBA BIOS

1 Enter the BIOS Fast!UTIL configuration utility:

   a Boot the server.

   b While booting the server, press `Ctrl-Q`.

2 Perform the appropriate action depending on the number of HBAs.

   - If you have only one host bus adapter (HBA), the Fast!UTIL Options page appears. Skip to step 3.
   - If you have more than one HBA, select the HBA manually:

     i In the Select Host Adapter page, use the arrow keys to position the cursor on the appropriate HBA.

     ii Press Enter.

3 In the Fast!UTIL Options page, select **Configuration Settings** and press Enter.

4 In the Configuration Settings page, select **Host Adapter Settings** and press Enter.

5 Set the BIOS to search for SCSI devices:

   a In the Host Adapter Settings page, select **Host Adapter BIOS**.

   b Press Enter to toggle the value to **Enabled**.

   c Press Esc to exit.

Enabling the Selectable Boot

You need to enable the selectable boot.

To enable the selectable boot

1 Choose **Selectable Boot Settings** and press Enter.

2 In the Selectable Boot Settings page, choose **Selectable Boot**.

3 Press Enter to toggle the value to **Enabled**.

Selecting the Boot LUN

If you are using an active/passive storage array, the selected SP must be on the preferred (active) path to the boot LUN. If you are not sure which SP is on the active path, use your storage array management software to find out. The target IDs are created by the BIOS and might change upon each reboot.

To select the boot LUN

1 Use the cursor keys to select the first entry in the list of storage processors, and press Enter to open the Select Fibre Channel Device page.

2 Use the cursor keys to select the chosen storage processor (SP), and press Enter.

- **If the SP has only one LUN attached**, it is automatically selected as the boot LUN, and you can skip to step 3.
- **If the SP has more than one LUN attached**, the Select LUN page opens. Use the arrow keys to position to the chosen LUN, and press Enter.

  If any remaining storage processors show in the list, position to those entries and press C to clear the data.

3 Press Esc twice to exit, and press Enter to save the setting.

Setting Up Your System to Boot from CD-ROM First

Because the VMware installation CD is in the CD-ROM drive, you must set up your system to boot from CD-ROM first. To achieve this, you need to change the system boot sequence in your system BIOS setup.

For example, on the IBM X-Series 345 server, do the following:

1 During your system power up, enter the system BIOS Configuration/Setup Utility.

2 Select **Startup Options** and press Enter.

3 Select **Startup Sequence Options** and press Enter.

4 Change **the First Startup Device** to `[CD-ROM]`.

You can now install the ESX Server system, as discussed in the *Installation and Upgrade Guide*.

Setting Up the Emulex FC HBA for Boot from SAN

Configuring the Emulex HBA BIOS to boot ESX Server from SAN includes the following tasks:

- Enabling the Boot BIOS prompt (see "To enable the BootBIOS prompt")
- Enabling the Boot BIOS (see "To enable the BIOS")

To enable the BootBIOS prompt

1 From the ESX Server service console or a Linux command prompt, run `lputil`.

NOTE Consider booting the ESX Server host from a Linux Administration CD that loads the Emulex driver, then run `lputil` from there.

2 Select <3> **Firmware Maintenance.**

3 Select an adapter.

4 Select <6> **Boot BIOS Maintenance.**

5 Select <1> **Enable Boot BIOS.**

To enable the BIOS

1 Reboot the ESX Server machine.

2 Press <ALT+E> at the Emulex prompt.

   a Select an adapter (with BIOS support).

   b Select <2> **Configure Adapter's Parameters.**

   c Select <1> **Enable or Disable BIOS.**

   d Select <1> to enable BIOS.

   e Select <x> to exit and <N> to return to the main menu.

3 From the Emulex main menu:

   a Select the same adapter.

   b Select <1> **Configure Boot Devices.**

   c Select the location for the Boot Entry.

   d Enter the two-digit boot device

   e Enter the two-digit (HEX) starting LUN (for example, `08`).

   f Select the boot LUN

   g Select <1> **WWPN.** (Boot this device using WWPN, not DID).

   h Select <x> to exit and <Y> to reboot.

4 Boot into the system BIOS and move Emulex first in the boot controller sequence.

5 Reboot and install on a SAN LUN.

Managing ESX Server Systems That Use SAN Storage

6

This chapter can help you with managing your ESX Server system, with using SAN storage effectively, and with troubleshooting. It discusses these topics:

Issues and Solutions

Table 6-1 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 6-1. Issues and Solutions

Issue	Solution
A LUN is not visible in the VI Client.	See "Resolving Display Issues" on page 85.
A shared LUN and a VMFS filesystem formatted on it is not visible to all ESX Server hosts that access the LUN.	See "Issues with Displaying the Same LUN IDs Across Hosts" on page 113.
You want to 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 90.
You want to 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 90.
You need to increase the Windows disk timeout to avoid disruption during failover.	See "Setting Operating System Timeout" on page 99.
You need to customize driver options for the QLogic or Emulex HBA.	See "Setting Device Driver Options for SCSI Controllers" on page 99.
The server is unable to access a LUN, or access is slow.	Path thrashing might be the problem. See "Resolving Path Thrashing" on page 104.
Access is slow.	If you have a lot of LUNs/VMFS volumes, and all of them are VMFS-3, unload the VMFS-2 driver by typing at a command-line prompt: `vmkload_mod –u vmfs2` You will see a significant increase in the speed of management operations such as refreshing datastores and rescanning storage adapters.
You have added a new LUN or a new path to storage and want to see it in the VI Client.	You have to rescan. See "Using Rescan" on page 87.

Potential Problems

Follow these guidelines to avoid potential problems:

- Place only one VMFS volume onto each LUN. Multiple VMFS volumes on one LUN are *not* recommended.

- Do not change the path policy the system sets for you. 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 85.

This section covers the following topics:

Viewing HBA Information

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

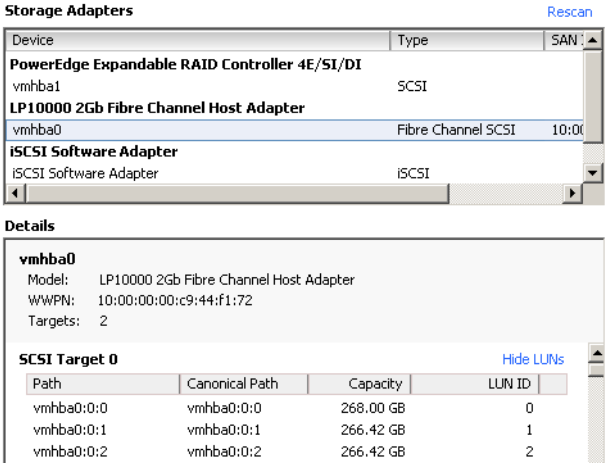
To see a list of HBA types

1 Select the host for which you want to see the HBAs 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 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 for which you want to see the storage devices and click the **Configuration** tab.

2 In the Hardware panel, choose **Storage (SCSI, SAN, and NFS).**

The list of datastores (volumes) appears in the Storage panel.

The display shows the whole VMFS for the selected host. Only storage that has been formatted with VMFS is included in the display. See Figure 6-1 for an example.

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 0:0:14 is a LUN, and vmhba 0:0:14:1 is a partition. One VMFS volume can have multiple extents.

NOTE The abbreviation vmhba refers to the physical HBA (QLogic or Emulex) on the ESX Server system, not to the SCSI controller used by the virtual machines.

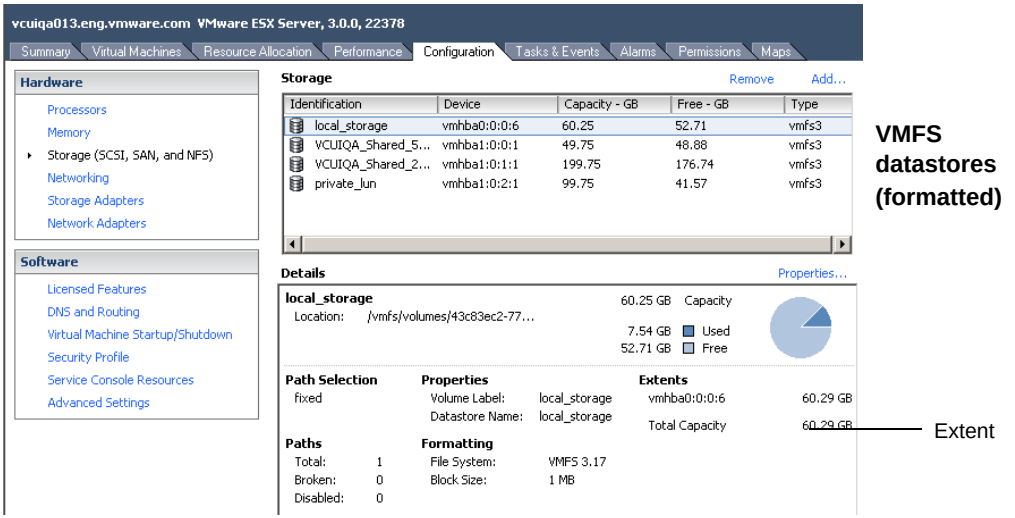


Figure 6-1. Viewing Storage Information

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

See "To view the current multipathing state" on page 91 for multipathing information.

Resolving Display Issues

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

- "Understanding LUN Naming in the Display" on page 86
- "Resolving Issues with LUNs That Are Not Visible" on page 86
- "Using Rescan" on page 87
- "Removing Datastores" on page 88

If you are using an AX100 storage array, inactive connections can cause display problems. See "AX100 Display Problems with Inactive Connections" on page 61.

Understanding LUN Naming in the Display

In the VI Client, a LUN is displayed 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 may 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 could change to `vmhba1:1:3`. The numbers have the following meaning:

- The first number, the SCSI HBA, changes if there is an FC or iSCSI network outage 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 there is a change in the mappings in the FC or 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 view LUNs using the VI Client.

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

- **Cable connectivity** – If you don't see a port, the problem could be cable connectivity or zoning. Check the cables first.
- **Zoning** – Limits access to specific storage devices, increases security, and decreases traffic over the network. Some storage vendors allow only single-initiator zones. In that case, an HBA can be in multiple zones to only one target. Other vendors allow multiple-initiator zones. See your storage vendor's documentation for zoning requirements. Use the SAN switch software to configure and manage zoning.
- **LUN masking** – If an ESX Server host sees a particular storage device but not the expected LUNs on that device, it might be that LUN masking has not been set up properly.

  For boot from 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 disk array software to make sure 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 Using

Disk.MaxLUN" on page 89 and "Masking LUNs Using Disk.MaskLUNs" on page 90.

- **Storage processor** – If a disk array 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 disk arrays, only one SP is active and the other SP is passive until there is a failure. 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.

Using Rescan

Perform a rescan each time you:

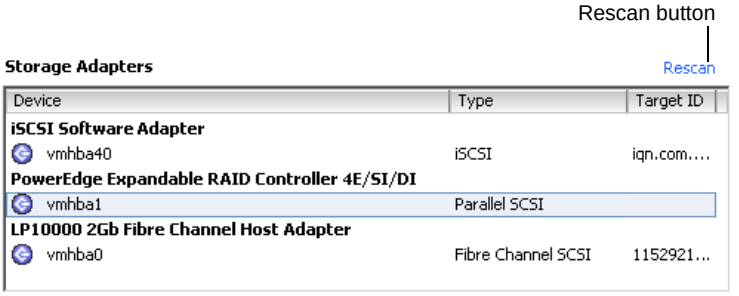
- Zone a new disk array on the SAN to an ESX Server host.
- Create new LUNs on a SAN disk array.
- Change the LUN masking on an ESX Server host disk array.
- Reseat a cable.
- Make a change to a host in a cluster.

NOTE Do not rescan when a path is down. 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 removes 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.

NOTE You can also select 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.

NOTE To remove the datastore from the ESX Server host, you must mask or remove the LUN from the SAN array, and rescan from the VI Client.

To remove a datastore

1. Power down all virtual machines that use the datastore you want to remove.
2. In the Inventory panel, select the host.
3. Use VMotion to migrate all virtual machines you want to keep.
4. Click the **Configuration** tab and click **Storage** to display all storage devices.
5. Select the datastore you want to remove and click **Remove**.
6. Click **Rescan** to update the view of available storage options.

Advanced LUN Display Configuration

This section discusses a number of advanced configuration options:

- "Changing the Number of LUNs Scanned Using Disk.MaxLUN" on page 89
- "Masking LUNs Using Disk.MaskLUNs" on page 90
- "Changing Sparse LUN Support Using DiskSupportSparseLUN" on page 90

Changing the Number of LUNs Scanned 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 can't discover LUNs with a LUN ID number that is higher than 255.

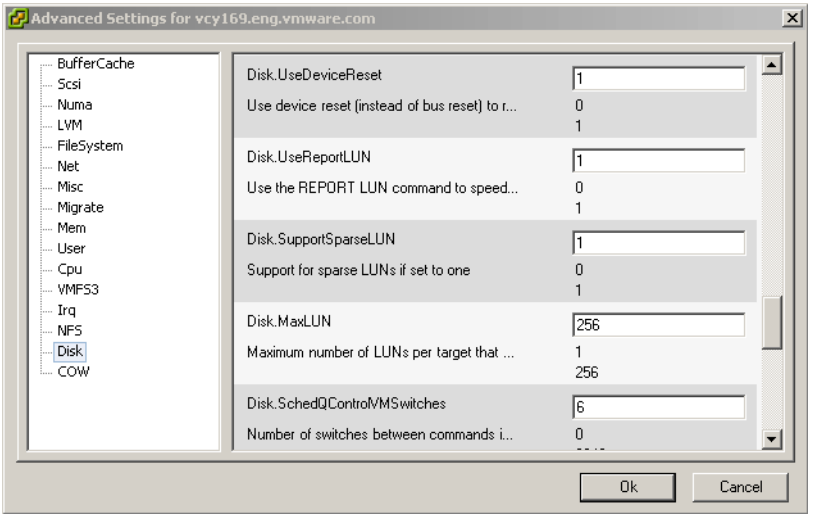
Reducing the value can shorten both rescan time and boot time. The time to rescan LUNs depends on several factors, including the type of storage array and whether sparse LUN support is enabled. See "Changing Sparse LUN Support Using DiskSupportSparseLUN" on page 90.

To change the value of Disk.MaxLUN

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

2 In the dialog box that appears, select **Disk**.

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



Masking LUNs 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 when you want to prevent the ESX Server system from accessing some FC LUNs, but do not want to use the FC switch or FC device LUN masking mechanims.

To change the value of Disk.MaskLUNs

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

2 In the dialog box that appears, 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 may be masked or exposed.

Changing Sparse LUN Support 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.

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's inventory panel, select the host, click the **Configuration** tab, and click **Advanced Settings.**

2 In the dialog box that appears, select **Disk**.

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

Multipathing

This section discusses the following multipathing topics:

- "Path Management and Manual Load Balancing" on page 96

For an introduction to multipathing concepts, see "Path Management and Failover" on page 42.

Viewing the Current Multipathing State

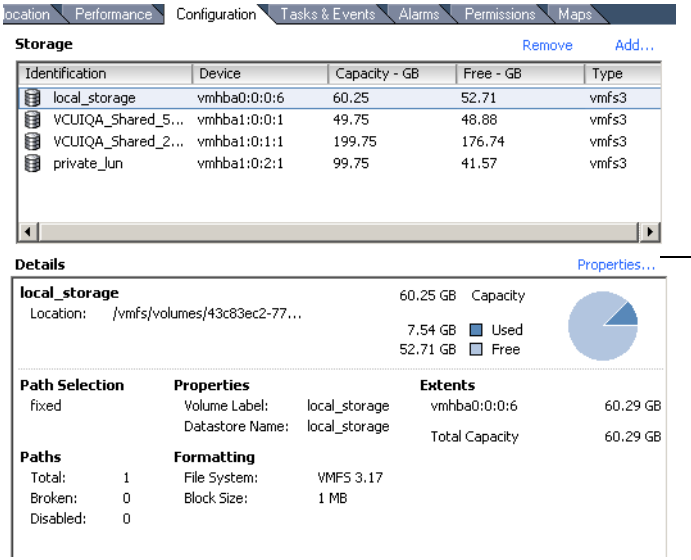
You can use the VI Client to use the current multipathing state.

To view the current multipathing state

1 In the VI Client's 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 for which you want 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 for transferring data.
- **Disabled** – The path has been 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.

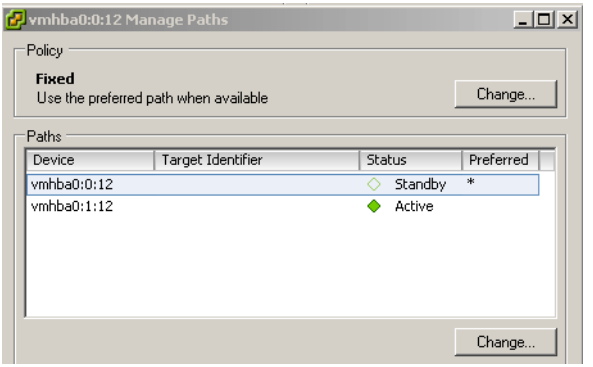


Table 6-2 summarizes how the behavior of an ESX Server system changes, depending on the type of array and the failover policy.

Table 6-2. 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 104.

See "Active Paths" on page 93 for more information.

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 alternate 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.

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's 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 alternate 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.

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

NOTE Manually changing **Most Recently Used** to **Fixed** is not recommended. The system sets this policy for those arrays that require it.

To set the multipathing policy using a VI Client

1. In the VI Client's inventory panel, select the host and click the **Configuration** tab.
2. In the Hardware panel, select **Storage**.
3. Select the datastore for which you want to change the multipathing policy, and click **Properties** in the Details panel.
4. In the Extent panel, select the device for which you want to make the change, and click **Manage Paths** in the Extent Device panel on the right.
5. Select the multipathing policy in the dialog box that appears and click **Done**.

NOTE For active/passive storage devices, **Most Recently Used** is highly recommended.

Disabling and Enabling Paths

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

To disable a path

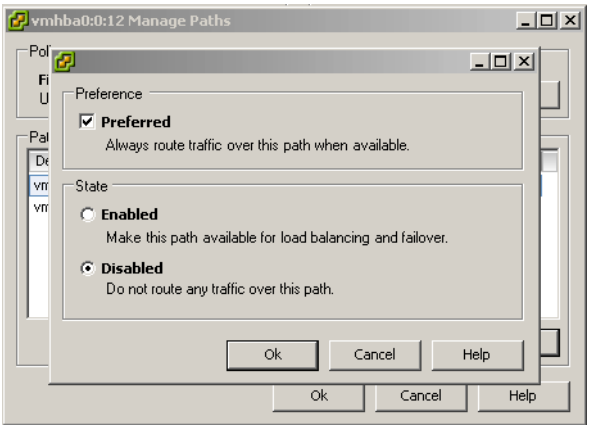
1 In the VI Client's inventory panel, select the host and click the **Configuration** tab.

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

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

4 In the **Extent** panel, select the device for which you want to make the change, 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 select the **Disabled** radio button to disable the path.



To enable a path

If you have disabled a path (for example, for maintenance), you can enable it by following the steps for disabling a path, but clicking the **Enabled** radio button.

Setting the Preferred Path (Fixed Path Policy Only)

If you are using path policy **Fixed**, 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 for which you want to disable a path, and click **Properties** in the Details panel.

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

5 Select the path you want to make the preferred path 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 set up your system to use different paths to different LUNs by changing the preferred path for the different HBAs. This is possible only for active/active SPs, and requires that you have path policy set to **Fixed**. See "Example Using Manual Load Balancing" on page 96 for an example.

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

Example Using Manual Load Balancing

When using an active/active array, you can set up your environment for load balancing. Assume the following setup, shown in Figure 6-2:

- Active/Active SPs
- An ESX Server system
- Four Fibre Channel HBAs in each server
- Director class software

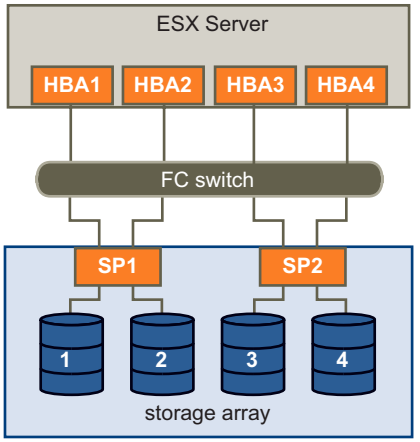


Figure 6-2. Manual Load Balancing

For load balancing, set the preferred paths as follows.

- LUN 1: vmhba1:1:1
- LUN 2: vmhba2:1:2
- LUN 3: vmhba3:2:3
- LUN 4: vmhba4:2:4

See "Setting the Preferred Path (Fixed Path Policy Only)" on page 95 for information.

NOTE Load balancing can be performed with as few as two HBAs, although this example uses four.

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 array. You can determine the active path—the path currently used by the server—by looking at the LUN's properties.

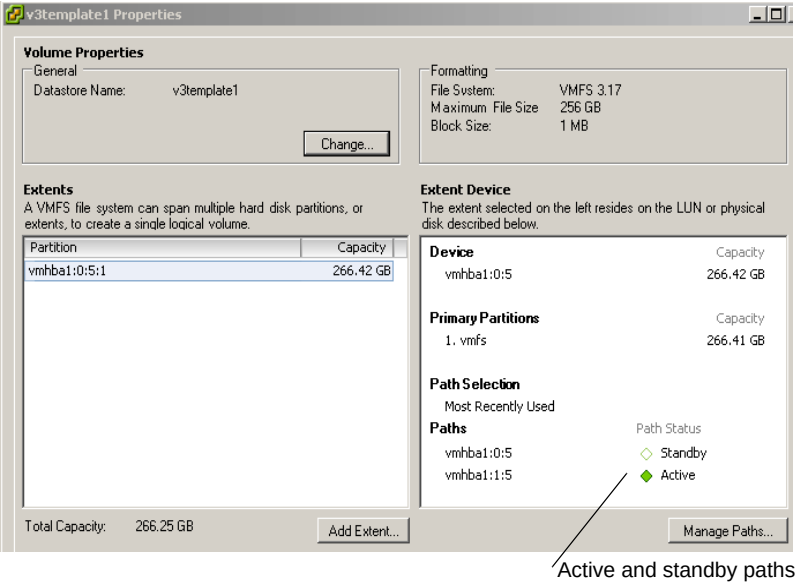


Figure 6-3. Active and Standby Paths

When an FC cable is pulled, I/O might pause for 30-60 seconds until the FC driver determines that the link is down and until failover has occurred. 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 hang. After failover is complete, I/O resumes normally.

In case of disastrous events that include 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.

For more information, see the following sections:

- "Setting the HBA Timeout for Failover" on page 98
- "Setting Device Driver Options for SCSI Controllers" on page 99
- "Setting Operating System Timeout" on page 99

Setting the HBA Timeout for Failover

The timeout value for I/O retry operations is usually set in the HBA BIOS driver. (You might also want to change operating system timeout, as discussed in "Setting Operating System Timeout" on page 99.)

VMware recommends that you set the timeout value to 30 seconds.

- For QLogic HBAs, the timeout value is `2*n + 5` seconds, where `n` is the value of the `PortDownRetryCount` parameter of the BIOS of the QLogic card. You can change the path failure detection time by changing the value of the module parameter `qlport_down_retry` (whose default value comes from the BIOS setting). The recommended setting for this parameter is 14.
- For Emulex HBAs, you can modify the path failure detection time by changing the value of the module parameters `lpfc_linkdown_tmo` (the default is 30) and `lpfc_nodedev_tmo` (the default is `30`). The sum of the values of these two parameters determines the path failure detection time. The recommended setting for each is the default.

To change these parameters, you must pass an extra option to the driver, such as `qlport_down_retry` or `lpfc_linkdown_tmo`. The following section explains how you can pass these options to the driver.

Setting Device Driver Options for SCSI Controllers

This section sets device driver options for QLogic, Emulex, or other SCSI card drivers.

To set device driver options for QLogic, Emulex, or other SCSI card drivers

1 Back up the file `/etc/vmware/esx.conf`, and open it for editing.

The file includes a section for each SCSI device, as in the following example:

```
/device/002:02.0/class = "0c0400"
/device/002:02.0/devID = "2312"
/device/002:02.0/irq = "19"
/device/002:02.0/name = "QLogic Corp QLA231x/2340 (rev 02)"
/device/002:02.0/options = ""
/device/002:02.0/owner = "vmkernel"
/device/002:02.0/subsysDevID = "027d"
/device/002:02.0/subsysVendor = "1014"
/device/002:02.0/vendor = "1077"
/device/002:02.0/vmkname =
```

2 Find the `options` line right under the `name` line and modify it as appropriate.

3 Repeat for every SCSI adapter that is controlled by the same driver if needed.

Setting 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 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 first **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've made 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 `/` and `/boot` partitions of the service console. The different components are discussed in more detail in the *Installation and Upgrade Guide*.

For more information, see "Sharing Diagnostic Partitions" on page 100.

Sharing Diagnostic Partitions

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

However, for diskless servers that boot from 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 zoned 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 there is only one diagnostic slot 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 Problems

This section gives some tips for avoiding and resolving problems:

- Document everything. Include information about zoning, access control, storage, switch, server and FC HBA configuration, software and firmware versions, and storage cable plan.
- Plan for failure:
  - Take your topology maps and make several copies. 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 didn't miss a critical failure point in your design.
- Disconnect the Fibre Channel HBAs during local installation when you install an ESX Server host on a production system.

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

- Ensure that the Fibre Channel 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, FC switch statistics, and storage performance statistics.

Optimizing SAN Storage Performance

The two major factors for optimizing a typical SAN environment are storage array performance and server performance. If the environment is properly configured, the SAN fabric components (particularly the SAN switches) are only *minor* contributors because of their low latencies relative to servers and storage arrays. Ensure that the paths through the switch fabric are not saturated, that is, the switch fabric is running at the highest throughput (for example, 2GB/s).

For more information, see these sections:

- "Storage Array Performance" on page 102

- "Server Performance" on page 102

Storage Array Performance

If there are issues with storage array performance, be sure to consult your storage array vendor's documentation for any relevant information.

When assigning LUNs, remember that each LUN is accessed by 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 containing the ESX Server LUNs should not include LUNs used by other hosts that are not running ESX Server for I/O intensive applications.

Make sure read/write caching is enabled.

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 arrays require continual redesign and tuning to ensure that I/O is load balanced across all storage array 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 96 for an example.

Tuning statically balanced storage arrays 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:

- 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 array. To achieve performance goals:

- Place each LUN on a RAID group that provides the necessary performance levels. Pay particular attention to the activities and resource utilization 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 performance goals required by an application running on the ESX Server host.
- Make sure that each server has a sufficient number of HBAs to allow maximum throughput for all the applications hosted on the server for the peak period. I/O spread across multiple HBAs provide higher throughput and less latency for each application.
- To provide redundancy in the event of HBA failure, make sure the server is connected to a dual redundant fabric.
- When allocating LUNs or RAID groups for ESX Server systems, keep in mind that 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 using multiple ESX Server systems in conjunction with a VirtualCenter Server, the performance needed from the storage subsystem increases correspondingly.
- The number of outstanding I/Os needed by applications running on an ESX Server system should match the number of I/Os the HBA and storage array can handle.

Resolving Performance Issues

This section discusses performance monitoring and possible ways of resolving performance issues:

NOTE For best performance, place each virtual machine on the appropriate tier of storage. See "Choosing Virtual Machine Locations" on page 44 for information.

Monitoring Performance

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

You can also use the `esxtop` tool, 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.

Resolving Path Thrashing

If your server is unable to 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 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.

This problem can also occur on a direct connect array (such as AX100) with HBA failover on one or more nodes.

Path thrashing is a problem that you typically won't 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 is issuing I/Os at a time. Path thrashing does not become a problem.

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

To resolve path thrashing

- Ensure all hosts sharing the same set of LUNs on those active/passive arrays 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.
- Make sure the path policy is set to **Most Recently Used** (the default).

See "Understanding Path Thrashing" on page 105 for more information.

Understanding Path Thrashing

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

- For active/passive arrays, all the sectors on the storage that make up a given LUN can be accessed by only one LUN at a time. The ownership is passed around between the storage processors. The reason is that storage arrays use caches and SP A must not write something to disk that invalidates SP B's cache. Because the SP has to flush the cache when it's done with its operation, it takes a little time to move the ownership. During that time, no I/O's to the LUN can be processed by either SP.
- For active/active arrays, the algorithms allow more fine-grained access to the storage and synchronize caches. Access can happen concurrently through any SP without extra time required.

Arrays with AVT are active/passive arrays that attempt to look like active/active arrays by passing the ownership of the LUN to the various SPs as I/O arrives. This approach is fine in a clustering setup, but if many ESX Server systems access the same LUN concurrently through different SPs, the result is LUN thrashing.

Consider how path selection works:

- On an active/active array the system starts sending I/O down 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 down that path.
  - If the ESX Server host cannot find such 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 this 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 continuously cause the ownership of the LUN to move between the two SP's, effectively ping-ponging the ownership of the LUN. Because the system moves the ownership quickly, the storage array 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 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 bandwidth is limited by the queue depth of the storage adapter.

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 to the number of your choice and click **OK**.

5 Reboot the server.

This change can impact disk bandwidth scheduling, but experiments have shown improvements for disk-intensive workloads.

If you adjust this value in the VMkernel, you might also want to adjust the queue depth in your storage adapter. See "Setting Maximum Queue Depth for HBAs" on page 107.

Removing VMFS-2 Drivers

If you have a lot of LUNs/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 from the service console requiring opening files or doing metadata updates. (See "Metadata Updates" on page 33.)

There can be performance degradation if such operations are happening frequently on multiple servers accessing the same VMFS. For instance, it is not recommended to run many virtual machines from multiple servers that are using virtual disk snapshots on the same VMFS. Limit the number of VMFS file operations that are executed from the service console when many virtual machines are running on the VMFS.

Setting Maximum Queue Depth for HBAs

If you notice unsatisfactory performance of your HBAs, you can change their maximum queue depth. See the following for more information:

- "Adjusting Queue Depth for a QLogic HBA" on page 107
- "Adjusting Queue Depth for an Emulex HBA" on page 108

Adjusting Queue Depth for a QLogic HBA

You can adjust the maximum queue depth for a QLogic qla2x00 series adapter with the following procedure.

To set maximum queue depth for a QLogic HBA

1 Log on to the service console as the root user.

2 Back up the file `/etc/vmware/esx.conf` and open it for editing.

The file looks similar to the following example:

```
/device/002:02.0/class = "0c0400"
/device/002:02.0/devID = "2312"
/device/002:02.0/irq = "19"
/device/002:02.0/name = "QLogic Corp QLA231x/2340 (rev 02)"
```

```
/device/002:02.0/options = ""
/device/002:02.0/owner = "vmkernel"
/device/002:02.0/subsysDevID = "027d"
/device/002:02.0/subsysVendor = "1014"
/device/002:02.0/vendor = "1077"
/device/002:02.0/vmkname =
```

3 Find the `options` line right under the `name` line and modify it to specify the maximum queue depth, as follows (where `nn` is the queue depth maximum):

```
/device/001:02.0/options = "ql2xmaxqdepth=nn"
```

NOTE The second character in `ql2xmaxqdepth` is a lowercase "L".

4 Save your changes and reboot the server.

Adjusting Queue Depth for an Emulex HBA

You can adjust the maximum queue depth for an Emulex HBA with the following procedure.

To change the queue depth of an Emulex HBA

1 Log on to the service console as root.

2 Verify which Emulex HBA module is currently loaded:

```
vmkload_mod –l | grep lpfcdd
```

Depending on the model of the HBA, the module can be one of the following:

```
lpfcdd_7xx
lpfcdd_732
```

3 For a single instance of an Emulex HBA on the system, run the following commands. The example shows the `lpfcdd_7xx` module. Use the appropriate module based on the outcome of Step 2.

a `esxcfg–module –s lpfc0_lun_queue_depth=16 lpfcdd_7xx`

b `esxcfg–boot –b`

In this case, the HBA represented by `lpfc0` will have its LUN queue depth set to 16.

4 For multiple instances of an Emulex HBA being present on the system, run the following commands:

a
```
esxcfg–module –s "lpfc0_lun_queue_depth=16
lpfc1_lun_queue_depth=16" lpfcdd_7xx
```

```
b   esxcfg-boot -b
```

In this case, both HBAs `lpfc0` and `lpfc1` will have their LUN queue depths set to 16.

5 Reboot.

SAN Storage Backup Considerations

This section discusses backups in the SAN environment:

Backups in a SAN Environment

Within 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 within 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, the scheduled backup should be re-examined. It might be better to 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, then 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* for more information.

Using third-party software has the advantage of a uniform environment. However, you have to consider that 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 both VMFS and RDMs. If both are supported, you can make either a snapshot of the whole virtual machine file system for a host, or snapshots for the individual virtual machines (one per 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 for additional information.

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 choosing 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, you should investigate a third-party backup solution to see whether it better suits your needs.

Layered Applications

SAN administrators are customarily using specialized array-based software for backup, disaster recovery, data mining, forensics, and configuration test. ESX Server administrators might be used to working with tools included with the ESX Server host for performing the same operations. When you use an ESX Server system in conjunction with a SAN, you need to decide whether array-based or host-based tools are more suitable for your particular situation.

See the following sections for more information:

- "Array-Based (Third-Party) Solution" on page 111
- "File-Based (VMware) Solution" on page 112

Array-Based (Third-Party) Solution

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

- If you use the array-based solution, RDM (not VMFS) is usually the appropriate choice. If you do not intend to use RDM, you must check the vendor's documentation to see if operations on virtual disks are supported.
- Array-based solutions usually result in more comprehensive statistics. With RDM, data always go along the same path, which results in easier performance management.
- Security tends to be more manageable when you use RDM and an array-based solution because with RDM, virtual machines more closely resemble physical machines.

File-Based (VMware) Solution

When considering a file-based solution using VMware tools and VMFS, be aware of the 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 and replication 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 ESX Server tools is easier.
- ESX Server administrators who use the file-based solution are more independent from the SAN administrator.

VMFS Volume Resignaturing

VMFS volume resignaturing allows you to make a hardware snapshot of a VMFS volume and access that snapshot from an ESX Server system. It involves resignaturing the volume UUID and creating a new volume label. You can control resignaturing as follows:

- Use the **LVM.EnableResignature** option to turn auto-resignaturing on or off (default is off).
- As a rule, a LUN should appear with the same LUN ID to all hosts that access the LUN. Under certain circumstances, you need to change this behavior using the **LVM.DisallowSnapshotLUN** option.

See these sections for more information:

- "Mounting Original and Snapshot VMFS Volumes" on page 113

- "Issues with Displaying the Same LUN IDs Across Hosts" on page 113
- "Understanding Resignaturing Options" on page 114

Mounting Original and Snapshot VMFS Volumes

You can mount both, the original and snapshot VMFS volumes on the same ESX Server host.

To mount original and snapshot VMFS volumes

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

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

3 Perform the following tasks repeatedly, as needed:

   a Make the snapshot.

   b Add the snapshot to the storage array.

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

4 Rescan the volume.

   After rescan, the LUN appears as `/vmfs/volumes/snap-<DIGIT>-<old-label>`.

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

   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.

5 If necessary, set the **LVM.EnableResignature** option to **0** after resignaturing is complete.

Issues with Displaying the Same LUN IDs Across Hosts

Normally, a LUN appears with the same LUN ID to all hosts that access the LUN. On some arrays, however, it may not be possible to display the LUN with the same LUN ID across hosts. As a result, the ESX Server system incorrectly detects the LUN as a snapshot and places it offline.

Examples of storage arrays for which the same LUN ID may not be visible for a given LUN across hosts are Clariion AX100 and few IBM TotalStorage Enterprise Storage Systems (previously Shark Storage systems).

NOTE If you use Clariion AX100 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, you need to recreate these LUNS.

To resolve issues with invisible LUNs on certain arrays

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

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

3 Select **LVM** in the left panel and set **LVM.DisallowSnapshotLUN** to **0** in the right panel.

4 Rescan all VMFS volumes.

After the rescan, all VMFS volumes are available.

CAUTION With **LVM.DisallowSnapshotLUN** set to **0**, snapshots *cannot* be exposed to the ESX Server host. For more information, see "State 3 - EnableResignature=no, DisallowSnapshotLUN=no" on page 115.

Understanding Resignaturing Options

This section discusses how the **EnableResignature** and **DisallowSnapshotLUN** options interact and explains the following three states resulting from changing these options:

- "State 1 - EnableResignature=no, DisallowSnapshotLUN=yes (default)" on page 114
- "State 2 - EnableResignature=yes, (DisallowSnapshotLUN is irrelevant)" on page 115
- "State 3 - EnableResignature=no, DisallowSnapshotLUN=no" on page 115

State 1 - EnableResignature=no, DisallowSnapshotLUN=yes (default)

In this state:

- You cannot bring snapshots of VMFS volumes by the array into the ESX Server host.
- LUNs formatted with VMFS must have the same ID for each ESX Server host.

State 1 is the safest state but

- It can cause issues with VMFS on some arrays (like IBM TotalStorage 8000 arrays and the EMC AX100) which do not always present the same LUNs with the same ID everywhere unless you take special care.
- You lose the ability to snapshot a VMFS volume and bring it into an ESX Server system. To do that, change the **LMV.EnableResignature** setting to **1**.

State 2 - EnableResignature=yes, (DisallowSnapshotLUN is irrelevant)

In this state:

- You can bring snapshots of VMFS volumes into the same or different servers
- VMFS volumes containing LUNs from IBM TotalStorage 8000 or AX100 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 his situation at all costs.

State 3 - EnableResignature=no, DisallowSnapshotLUN=no

With these settings, snapshots *cannot* be exposed to the ESX Server host. This is similar to ESX Server 2.x behavior.

If you have an IBM TotalStorage 8000 or AX100 that can not be configured to present the same LUN numbers to all servers for some reason, you need this setting to allow all ESX Server system to use the same LUNs for features like VMotion, VMware DRS, and VMware HA.

When changing these settings, consider the following potential issues:

- 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 non-deterministic behavior is dangerous.
- 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 is can corrupt your filesystem.

Multipathing Checklist

A

This appendix provides a checklist of multipathing setup requirements for different storage arrays.

Component	Comments
All storage arrays	Write cache must be disabled if not battery backed.
Topology	No single failure should cause both HBA and SP failover, especially with active-passive storage arrays.
IBM TotalStorage DS 4000 (formerly FastT)	Default host type must be `LNXCL`. Host type must be `LNXCL`. AVT (Auto Volume Transfer) is disabled in this host mode.
HDS 99xx and 95xxV family	HDS 9500V family (Thunder) requires two host modes: ■ Host Mode 1: `Standard`. ■ Host Mode 2: `Sun Cluster` HDS 99xx family (Lightning) and HDS Tabma (USP) require host mode set to `Netware`.
EMC Symmetrix	Enable the SC3 setting. Disable the SPC2 setting.
EMC Clariion	All Initiator records must have: ■ Failover Mode = 1 ■ Initiator Type = “Clariion Open” ■ Array CommPath = “Enabled” or 1
HP MSA	Host type must be `Linux.` Set the connection type for each HBA port to `Linux`.

Component	Comments
HP EVA	For EVA3000/5000 firmware 4.001 and above, and EVA4000/6000/8000 firmware 5.031 and above, 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
HP XP	For XP 128/1024/10000/12000, the host mode should be set to `0C (Windows)` — that is, zeroC (Windows).
NetApp	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. All FC HBAs must be of the same model.

Utilities

B

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 tool provides a fine-grained look at ESX Server resource utilization 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.

storageMonitor Utility

The storageMonitor utility monitors SCSI sense errors experienced by storage devices attached to VMware ESX Server. The utility gathers sense error information by periodically polling the storageMonitor running inside the VMkernel, and sends error information to standard output, 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 as per SCSI-3 specification.

If no configuration file is specified, storageMonitor parses the default configuration file /etc/vmware/storageMonitor.conf to filter certain errors and allow other errors to be displayed. You can run storageMonitor in interactive mode or daemon mode using the –d option.

See the following for more information:

- "Options" on page 120
- "Examples" on page 121

Options

You can invoke 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 specified by the user. If the –s option is also specified, output is written to standard out as well.
-h	Displays help information.

Table B-1. storageMonitor Command-Line Options (Continued)

Option	Description
-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 standard out. 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 standard out (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 using the configuration file `myconf.conf`. Writes output to syslog. By default, `syslog` is located at `/var/log/storageMonitor`.

```
storageMonitor –d –l mylog.log –s
```

Runs `storageMonitor` in daemon mode using the default configuration file. Sends output to `mylog.log` instead of `syslog`. Also writes output to standard out because the `–s` option is specified.

Index

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Z